MiningVis: Visual Analytics of the Bitcoin Mining Economy

Natkamon Tovanich, Nicolas Soulié, Nicolas Heulot, and Petra Isenberg

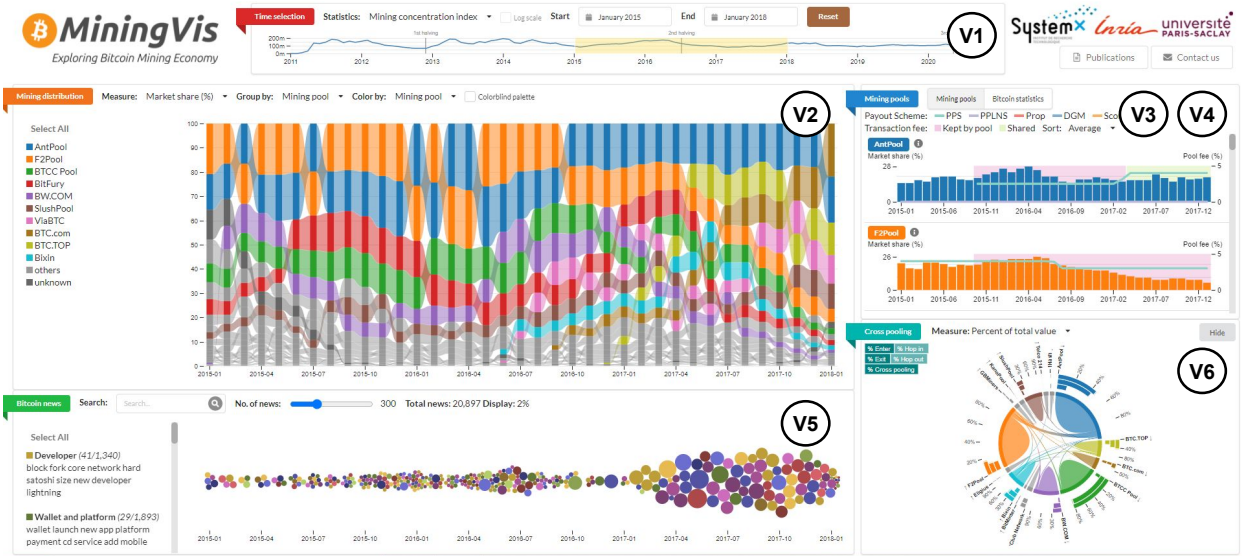


Fig. 1. A screenshot of the MiningVis tool. (V1) The time filter view allows analysts to select a time interval of interest. (V2) The mining distribution view shows the evolution of the mining pools as a ribbon chart. (V3) The mining pool details view shows mining power and pool characteristics for each mining pool in a compound chart. (V4) The Bitcoin statistics view shows Bitcoin network statistics as a list of area charts (see Fig. 4). (V5) The Bitcoin news view displays a swarm plot of the news distribution over time. (V6) The cross pooling view represents the total reward of cross pooling miners in mining pools at the time interval on a chord diagram.

Abstract—We present a visual analytics tool, MiningVis, to explore the long-term historical evolution and dynamics of the Bitcoin mining ecosystem. Bitcoin is a cryptocurrency that attracts much attention but remains difficult to understand. Particularly important to the success, stability, and security of Bitcoin is a component of the system called "mining." Miners are responsible for validating transactions and are incentivized to participate by the promise of a monetary reward. Mining pools have emerged as collectives of miners that ensure a more stable and predictable income. MiningVis aims to help analysts understand the evolution and dynamics of the Bitcoin mining ecosystem, including mining market statistics, multi-measure mining pool rankings, and pool hopping behavior. Each of these features can be compared to external data concerning pool characteristics and Bitcoin news. In order to assess the value of MiningVis, we conducted online interviews and insight-based user studies with Bitcoin miners. We describe research questions tackled and insights made by our participants and illustrate practical implications for visual analytics systems for Bitcoin mining.

Index Terms—Visual analytics, Bitcoin, Bitcoin mining, mining pools, pool hopping

1 Introduction

Cryptocurrencies have gained increased attention from the public in recent years as a possible alternative to standard fiat currencies (e.g., U.S. Dollar, Euro) [11]. Given the still new and emerging phenomenon of cryptocurrencies, researchers, analysts, and policy makers still have only very limited models on which to base decisions and recommendations regarding cryptocurrencies and require exploratory tools to understand behaviors, trends, and patterns related to the phenomenon. To this aim, we introduce MiningVis, an exploratory visual analytics tool that focuses on Bitcoin mining activities and actors. Not only mining is crucial to the stability and future of most cryptocurrencies and all the people who heavily invested in those currencies, but it also has implications on society, economy, and the environment [2, 21, 31]. As such, mining is an important aspect of cryptocurrencies to study.

Bitcoin is the first and the highest valued cryptocurrency. To transfer Bitcoin users of the system generate transactions. These transactions, if valid, are stored in an open public blockchain—a type of database in which a certain number of transactions are grouped in blocks. The blockchain is maintained in a decentralized way by a network of peers. In contrast to fiat currencies, Bitcoin, and generally cryptocurrencies, are not controlled by any financial institution, government, or organization. Without central control, an alternative mechanism is required to make sure that transactions are valid; that people spend Bitcoin they actually own and that they spend it only once [42].

In Bitcoin and most other cryptocurrencies, the process of validating transactions is called "mining". Bitcoin miners collect incoming transactions and store them into a block of transactions with a given size. The first miner to solve a computational puzzle (with *proof-of-*

- *Natkamon Tovanich is with IRT SystemX, Paris-Saclay, 91120, Palaiseau, France and also with Université Paris-Saclay, CNRS, Inria, LISN, 91405, Orsay, France. E-mail: natkamon.tovanich@irt-systemx.fr.*
- *Nicolas Soulié is with Université Paris-Saclay, Univ Evry, IMT-BS, LITEM, 91025, Evry-Courcouronnes, France. E-mail: nicolas.soulie@imt-bs.eu.*
- *Nicolas Heulot is with IRT SystemX, Paris-Saclay, 91120, Palaiseau, France. E-mail: nicolas.heulot@irt-systemx.fr.*
- *Petra Isenberg is with Université Paris-Saclay, CNRS, Inria, LISN, 91405, Orsay, France. E-mail: petra.isenberg@inria.fr.*

©2021 IEEE. This is the author's version of the article that will be published in IEEE Transactions on Visualization and Computer Graphics. The final version of this article is available at: xx.xxxx/TVCG.2021.xxxxxxx/

work) wins the right to propose its block as the next block of the blockchain [1]. The winning miner receives a financial reward in the form of new Bitcoin value (called *block reward*) plus *transaction fees*. The block is then proposed to the entire network of miners who verify the validity of all its transactions and only then the financial reward is given out. Clear economic incentives are, thus, involved in mining.

Our work followed a data-first design methodology [44], motivated by the presence of an interesting real-world dataset from which we later derived specific stakeholder analysis questions. Our work began from a survey on Bitcoin visualization and analyses [59] during which we remarked a lack of tools and systems that help to analyze Bitcoin mining in detail. Due to the primary economic incentives of Bitcoin mining we partnered with an economist working on Bitcoin as our main stakeholder. To study this kind of activity, the economist needed to analyze multiple factors—both internal and external to the activity—that affect the evolution and dynamics of Bitcoin mining. In-depth analysis of such a complex socio-technical system required exploring historical data to relate all relevant factors in a holistic view. Besides, little is still known about the behavior of miners and whether behaviors have remained and do remain stable and predictable. Therefore, a multi-facet visual analytics tool is necessary to help researchers and miners observe this activity's evolution over the long term. Such a tool can impact studying not only Bitcoin mining, but also the activity more broadly across related cryptocurrencies.

Motivated by the potential impact and necessity for understanding Bitcoin mining, we iteratively developed MiningVis through a two-year design cycle. We ultimately designed MiningVis as a visual analytics tool for long-term historic analysis that brings together a large number of complementary types of data for in-depth analysis. The tool consists of multiple-linked views that allow analysts to filter to a time interval of interest, analyze network-internal mining activity in detail, and relate this activity to contextual information for economic analysis. Complementing the tool contribution, we discuss the results of two-week online user study with eight Bitcoin miners aimed to understand research questions of interest to them and insights they were able to make with our tool. We collected positive results on overall usability and visualization value but also learned about differences between the types of questions relevant to miners and economic researchers. We also report on our first experience with the micro-entries methodology as a potential to study "in-the-wild" insight capture.

2 Background on Bitcoin Mining and Mining Pools

MiningVis is a visual analytics tool for the analysis of Bitcoin mining. It focuses, in particular, on the analysis of mining pools that emerged in response to several specific economic problems of Bitcoin mining.

2.1 What are Mining Pools?

Bitcoin mining requires a large amount of effort and some luck. Miners compete to solve a computational-intensive puzzle while validating transactions. The ability to win is largely dependent on the power of the mining hardware used (called *hash rate*). However, buying powerful hardware once is not enough. The higher the mining power in the entire network, the more difficult it becomes for miners to mine a new block because the Bitcoin protocol adjusts the *mining difficulty*. As a consequence, income from mining is unpredictable and based not only on one's hardware but also the activity of all other miners in the network. In order to have a more predictable income, miners have started to form teams called "mining pools" [12]. The current mining practice is an interaction of three main actors: (1) miners who own mining hardware, (2) mining pools, and (3) the Bitcoin protocol that regulates the mining activity.

Mining pools take on a significant role between its individual mining members and the Bitcoin blockchain. Pools join mining power to maximize the likelihood of successful mining and share the reward among all members. In order to be competitive, mining pools race to attract miners to improve their market share by providing better reward systems. Miners typically join a mining pool to increase their chance of a reward [36, 52]. To maximize their rewards, some miners switch to a more attractive pool (called *pool hopping*) [3] or participate in multiple mining pools at a time (called *cross-pooling*) [49]. Their behaviors, in turn, directly affect the mining power, and as a result, the competitiveness of mining pools.

2.2 The Dangers of Mining Pools

In the Bitcoin mining ecosystem three major risks to its security and sustainability exist: a) the risk that one pool becomes dominant, b) the risk that mining stops carrying economic incentives and miners stop participating, and c) the carbon footprint of the Bitcoin mining.

As mining pools represent larger groups of miners, they play a major but also potentially dangerous role in the network. To guarantee the security and trustworthiness of the Bitcoin network in the long term, none of the mining pools (or the combination of a few pools) should dominate the market. When one or more collaborating pools gain the majority of the total mining power, they can perform a 51% majority attack and decide to validate invalid transactions [13]. As large and specific investments are needed to mine successfully, only a few large and persistent mining pools emerge [49]. Hence, the danger of a 51% attack is real and persistent in the Bitcoin network. The potential risk raises questions on which factors lead to mining pool concentration and how miners collectively react when concern about an attack emerges.

The second issue relates to financial incentive and mining reward. On what is called a *"halving days"*, the Bitcoin protocol halves the *block reward*. Halving days happen every 210,000 blocks (≈4 years) to control the Bitcoins supply. There will be 33 halving days in total after which *transaction fees* will remain the only compensation for miners. Transaction fees are expected to increase to compensate for the decrease of block rewards [42]. A theoretical study previously cast doubt on the real compensation effect of these freely fixed transaction fees [25]. In fact, early evidence showed that transaction fees are very low and seem to be driven by social norm rather than economic reason [41]. Analyzing miners' behavior around halving days and the evolution of transaction fees allows observing the consequence of changed mining rewards on the behavior of mining pools and individual miners.

The third issue concerns the carbon footprint of the Bitcoin mining activity. Tremendous amounts of electricity are needed to run mining hardware [4, 35, 57, 64]. Bitcoin price increases have incentivized miners to compete in a prisoner's dilemma to upgrade their hardware continuously [66]. This violates the *one-CPU-one-vote* policy initially envisioned by Nakamoto [42] and causes a negative externality on climate change [19, 39, 40]. Alternative solutions, such as *proof-of-stake*, have been recently promoted to reduce the excessive energy consumption of cryptocurrencies [43]. Analyzing miners' revenue function and the environmental impact would help inform policy makers to develop sustainable mining regulation and policy [21, 29, 63, 64].

3 Analysis Factors and Motivation

Understanding mining and mining pool history, as well as the characteristics of major players, is important to different kinds of people. Miners are interested in understanding which pools to join; mining pool owners want to assess the competition; researchers want to study the economic phenomena of Bitcoin mining and model behaviors; or regulators may want to define guidelines for this type of activity. Previous work in this domain theoretically analyzed miners' motivation behind pool selection and miners' migration. A comprehensive review analyzed this work based on game theory [38] and most other work similarly took on specific viewpoints considering a limited number of mining pools [3, 41, 49, 66, 67]. In collaboration with our economist co-author, we focused on a broad view of the phenomenon across Bitcoin's history and all pools. We target the following analysis factors (AF1–AF3):

AF1: Bitcoin regulation and the evolution of mining pools. An excessive concentration of pool market shares is dangerous to Bitcoin's operations and raises questions about potential mining regulations. The economist wanted to analyze factors that pose a risk to pool concentration and detect critical periods in Bitcoin mining. Periods characterized by significant variations of bitcoin values (e.g., halving days, peak in bitcoin value) impact mining pools and should be analyzed. Furthermore, external data such as news about Bitcoin might give insights to better understand the impact of regulatory changes in some countries.

AF2: Pool managers' behavior and the competition to attract miners. Mining pools have different *payout schemes* [50] that define which and how much miners are paid. Payouts are reduced by *pool fees* [5] a pool may keep to pay for operating costs. We would expect that pools with lower fees pay more constant income to miners for a given payout scheme. These pools may therefore attract miners and exhibit increasing mining power. Pool competition might be visible by convergence toward a limited number of (best) payout schemes and a decrease in pool fees. Such information is crucial for economic debates around pool viability as these fees correspond to a significant part of their income. An analysis of the data might provide additional evidence of the importance of payout schemes and pool fees in attracting miners.

AF3: Miners' decision to join or leave a pool and its impact on mining pool market shares. Miners are economic agents who consider the cost and benefit of mining. In this respect, Bitcoin value, payout schemes, and pool fees are major determinants of miners' expected income. Miners' migration data (i.e., *pool hopping* and *cross-pooling*) helps to test assumptions and better understand the drivers of miners' pool choice. Such analyses are critical to understand a pool's growth or decline and to which extent miners behave as rational economic agents.

From an economic viewpoint, these three analysis directions are associated with the sustainability of Bitcoin mining. in the long term. To make empirical analyses of Bitcoin mining and mining pools more accessible, we developed a web-based visual analytics tool called MiningVis. Our tool targets two types of users: 1) researchers who look for exploratory analysis tools to generate hypotheses and models of mining pools dynamics, and 2) Bitcoin miners and users who would like to see economic data and make decisions related to their personal goals. In contrast to most closely related previous work [69], we offer longitudinal data over the entire Bitcoin history, combine blockchain internal with external data, and contribute a user study of the tool.

4 Data Acquisition

The first step in our development was to obtain the relevant data to analyze. Typically, economic analysis aims to explain a phenomenon and generalize it as a mathematical model. Therefore, the analysis our economist collaborator wanted to conduct required data acquisition from multiple sources with variables for which potential relationships could be established. As a consequence, we obtained five datasets related to our main analysis directions. The datasets are publicly available at [61], except for the data obtained from sources that are already public and which we cannot reshare (D3 and D4).

D1: The mining pool distribution dataset gives information about the evolution of mining pools over time. It was derived from identifying mining pools that received the reward for each individual block in the Bitcoin blockchain. We extended the initial dataset compiled by Romiti et al. [49] (data up to: 2018–12) to extract the mining pool that mined each block. Blocks that could not be attributed to any known mining pool were marked as "unknown". Our final dataset at the time of writing includes data until 2020-09 and contains the mining power of each mining pool for each month. We computed multiple quantitative measures related to mining power: hash rate, market share (normalized hash rate), and total reward received from mining (in BTC and USD). Additionally, we calculated the estimated electricity consumption of each mining pool based on the CBECI index [4] to allow to judge the externality of mining on the environment.

D2: The mining pool characteristics dataset includes external data about mining pools that we obtained from public sources [6, 7] and manually cleaned. The attributes include the primary location of the pool (nominal), payout scheme (nominal), pool fee (quantitative), and whether the pool kept transaction fees or shared it to miners (binary). We also extracted the wiki history to track attribute changes as mining pools adapt their payout policies to compete with other pools.

D3: The Bitcoin network statistics dataset contains multiple quantitative measurements of Bitcoin network's state over time. We collected daily statistics from Blockchain.info [10] and averaged those measures to a monthly time window. Examples of network statistics for each month include the market price, total hash rate, total block reward, total transaction fees, mining difficulty, the number of transactions, electricity consumption [4], and the global energy price index [27].

D4: The Bitcoin news dataset lists headlines from the Press forum in Bitcointalk.org [8], where users posted links to news articles related to Bitcoin. Each news contains the date when the news was published by its source, the news headline, and the number of replies and views. We chose to use the number of views and replies as an attractiveness indicator of the news from the Bitcointalk's members. We derived an importance score of each news items as $views \times (replies + 1)$.

The news we collected covered various Bitcoin-related topics. To group related articles, we used the Word Network Topic Model method [71] that proved to be simple and effective for short texts. We extracted 15 topics and the top-10 keywords for each topic.

D5: The miners' migration dataset describes miners' migration between mining pools over months. We detect miners who participate in more than one pool (*cross-pooling*) within a month and calculate the number of miners' addresses and the total reward those miners received from each pool. Besides, we also detect the flow of miners that join a mining pool for the first time (*enter*), move from a pool to another (*pool hopping*), and quit the mining pool (*exit*) between months. The process of obtaining this dataset is complex and covered in one of our previous articles focused on pool hopping [62].

5 Design Process and Task Analysis

We started our design process by generating prototypes of basic charts built with Altair and Jupyter notebooks. The economist used the prototypes to examine the empirical evidence, generate hypotheses to validate, and develop new research questions that started subsequent design iterations. Our collaborator was already able to make scientific discoveries about Bitcoin mining processes while we were developing and discussing these prototypes; showing the potential success of these types of close collaborations for Visual Analytics researchers.

From our discussion about prototypes, we published his findings on historic changes in mining pool evolution related to market concentration, reward rules, and pool locations in a workshop paper [60]. This publication also shows examples of the early independent simple charts we built. It became apparent, that some questions could not be solved with simple charts and simple datasets. For example, his questions regarding pool hopping required the development of a dedicated algorithm that we published with his findings and images of additional Jupyter notebook visualization prototypes at a Blockchain-focused conference [62]. The work discussed in these two previous papers built the foundation for this present paper but only focused on specific narrow elements of Bitcoin mining. Here we go significantly beyond this prior work and discuss our final and encompassing web-based Visual Analytics MiningVis tool and the results from our user study.

For the final design of MiningVis, we derived visualization analysis tasks by observing which actions the economist repeatedly performed on our prototypes. In Table 1, we summarize the relationship of our derived analysis factors (AF1–AF3), the data we collected (D1–D5), and the visualization tasks (T1.1–T3.2). The tasks are:

1. **Temporal overview of mining pool rankings.** Analyzing the Bitcoin mining activity requires gaining an overview of the emergence and evolution of mining pools (T1.1) as well as assessing the competition among them by comparing market share, mining power and rank of pools over time (T1.2).
2. **Temporal zoom and filter** to explore relevant time periods to investigate in more detail. The economist frequently searched for time periods of importance to Bitcoin mining (T1.3). He looked in particular at periods where one or two mining pools were close to dominating the market share ("concentration index") as well as outliers in Bitcoin statistics (e.g., market price, transaction fees).
3. **Comparison of mining pool characteristics.** The evolution of mining pools may be associated with some characteristics that affect miners' pool choice. To explain pool distributions, the economist compared the evolving characteristics of competing pools to identify characteristics that could affect the mining distribution (T2.1, T2.4).
4. **Relating mining migration behavior with blockchain-internal data.** Comparison of internal pool characteristics with pool distribution (T2.2) and detection of miners' migration among pools (T3.1)

Table 1. Mapping design levels of MiningVis tool from analysis direction, data, analysis tasks, to visualization (view).

Analysis Factor	Data	Analysis Task	Visualization (View)
AF1	D1	T1.1: Explore the evolution of mining pool distribution and rankings over time	V2
	D1	T1.2: Compare the evolution of rankings and mining power of mining pools over time	V2
	D1, D2, D3	T1.3: Identify outlier time periods in mining pool concentration and Bitcoin statistics	V1, V2
	D2	T1.4: Browse Bitcoin statistics measures that correlate with the mining activity	V4
	D4	T1.5: Browse Bitcoin news that explain the mining activity evolution in that time	V5
AF2	D1, D3	T2.1: Compare mining pool characteristics with the mining pool distribution over time	V2, V3
	D1, D3	T2.2: Lookup the detail mining power and characteristics of the mining pool of interests	V3
	D2, D3	T2.3: Identify characteristics of mining pools that dominates the market share over time	V2, V3, V4
	D3	T2.4: Compare mining power and pool characteristics across multiple pools	V3
AF3	D5	T3.1: Identify the significant miner's migration flow between mining pools.	V6
	D5, D2	T3.2: Lookup the mining pool characteristics of the pools that have the high miner's migrations	V6, V3

at a focused time period. The economist looked at pool distributions, characteristics, and miners' migration to identify characteristics that dominate the market (T2.3) and explain miners' behavior selecting or switching pools (T3.2).

5. **Relating external contextual information.** The economists browsed external factors including Bitcoin statistic measures (T1.4) or specific news events (T1.5) to find additional explanations for observations related to the evolution of mining distribution.

Based on the analysis tasks, we designed MiningVis' coordinated views. Table 1's last column lists the views we developed for each task and data. The final version of MiningVis was implemented with D3.js.

6 Related Work

Our work is most related to prior work on blockchain visualization. We were also inspired by work on similar analysis tasks: on rank comparison, temporal text visualizations, and abstract flow visualizations.

6.1 Blockchain data visualization

We previously published a systematic review of visualizations of blockchain data [59]. Out of the sources we surveyed, 79% focused on Bitcoin as the most prominent blockchain. Most visualizations came from online sources and used basic representations of blockchain network statistics and individual transaction details. In contrast, dedicated visualization tools for in-depth analyses are still rare.

Several other Bitcoin visualizations relate to our work in their focus on showing Bitcoin user activities. The most comprehensive visual analytics tools of these are: BitVis [58] which displays multiple graph visualizations to analyze transaction networks partially aggregated by user; BitConduite [30] that uses a clustering algorithm to group Bitcoin users with the same characteristics to investigate further. With this tool we share our use of a clustering algorithm to group transaction addresses that might belong to the same user; and BitExTract [70] which is a dedicated visualization to analyze transaction activities of Bitcoin exchange services. Similar to BitExTract, we focused on a particular activity in Bitcoin rather than proposing a general-purpose blockchain exploration tool focused on transaction.

Even though mining is an essential activity in the Bitcoin blockchain, most existing visualizations simply show mining pool market shares as pie and area charts [10, 16]. We found only one prior work in the visualization literature with a similar goal as ours. SuPoolVisor [69] is a visual analytics system that tracks mining pools' daily computational power, and reward distribution network. The tool focuses on showing the payout network between two pools and top miners' rewards over a short period. In contrast to this tool, MiningVis focuses on mining activity as a long-term macroeconomic process and considers both internal and external incentives that affect mining pool dynamics and miners' migration. The competition among mining pools is the focus of our tool rather than the individual miners of previous work. Moreover, we contribute a user study to understand the usability and value of our tool for Bitcoin miners and economic researchers.

6.2 Rank visualization

Tasks T1.1 and T1.2 required the visualization of mining pool rankings over time. We considered several previous solutions: The Rank chart stacks ranked objects in multiple columns (e .g., per year) and connects the same items with a line, similar to parallel coordinates with discrete rank steps [14, Fig. 63]. LineUp similarly ranks items in a table-like structure and focuses on exposing the comparison of multiple attributes in columns connected by ribbons. Rank charts recently inspired Gap Charts that use ribbons to encode both a rank and a quantitative difference between ranked items (y-axis) over time (x-axis) [46]. For rank visualizations in large datasets, Xia et al. proposed another table-like visualization to show changing ranks (y-axis) of most viewed pages on Wikipedia. In each column (x-axis), they added a time-series glyph to observe trends in a particular month [68]. In contrast to these approaches, we wanted each column to represent a stacked bar chart to show the total value of each measure per time step. More similar to our needs is RankExplorer which applies a stream chart variant in which quantities are stacked per ranked item [56]. In our tool, we decided to use a ribbon chart (Fig. 1 (V2)) to encode mining pools with stacked bars for each time point and ribbons connecting the same ranked items. This choice allowed us to show different quantitative measures as bars and their changes in rank and quantity by ribbons.

6.3 Stock and flow visualization

For Task T3.1, we model miners' migration data as a flow relationship between mining pools with cross-pooling miners flowing between mining pools. Common visualization types that can encode this kind of data are Sankey and Chord Dependency diagrams.

Sankey diagrams [48, 55] show the flow of values from one stock to another as a nodes and links where their width encodes a flow value. A famous example is Minard's map showing the number of French soldiers during Napoleon's invasion of Russia in 1812 [47]. It encodes the flow of men in both space and time. Telly's Stock and Flow Visualization [17] divides students into groups and tracks their high school advancement. Rosvall and Bergstrom [51] use an Alluvial diagram to show the scientific network community changing over time.

Chord dependency diagrams encode the inter-relationship among stocks in a circular form. The arcs' length shows the node's value, while the ribbon's size shows the flow's value relating two stocks. Circos [33] adopts this representation to show the similarity between genomic intervals. The work adds an outer circle in which detailed information in various forms is encoded, for example as, bar charts, time series, or customized visualizations. Among Bitcoin empirical research, Parino et al. [45] use the Circos diagram to show the total amount of Bitcoin values (flow) transfer between countries (stock). In our analysis tool, we represent flow of miners between pools (cross-pooling). The stock equivalent for us is the size of the pool. We chose the chord dependency diagram (Fig. 1 (V6)) because it effectively shows the crossing amount among many pools (and also self-loops) with high data-to-ink ratio.

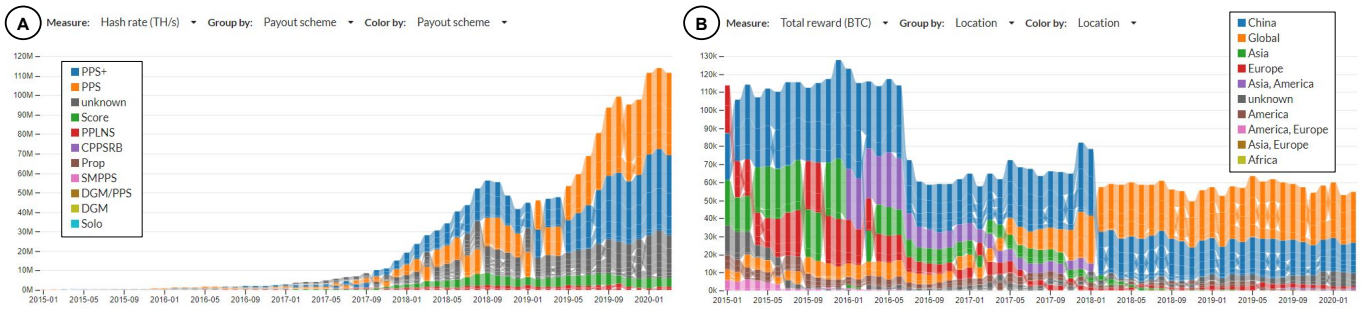


Fig. 2. The mining pool distribution view (V2) with different configurations. (A) *Measure: Hash rate (TH/s), Grouped by: Payout scheme, and Color by: Payout scheme.* The economist detected that the hash rate is growing while mining pools with the PPS payout scheme dominated in the market after Apr 2015. (B) *Measure: Total reward, Grouped by: Location, Color by: Location.* The economist detected the halving in rewards in July 2016 when Chinese pools dominated the market. After the local peak between Dec 2017–Jan 2018, global pools overtake Chinese pools in the market.

6.4 Text visualization

MiningVis includes a view to visualize news headlines over time in relation to mining pool rankings. The Text Visualization Browser [34] surveys the text visualization domain and provides a taxonomy based on multiple aspects (e.g., analysis task, visualization task, data domain, visual encoding). Particularly relevant to our task, T1.5 works on temporal text data. The stream graph, an extension of ThemeRiver [23], is one method to encode evolving topics in large-scale documents. CiteRivers [24] is an example of past work that applied a stream graph to show the citation popularity of multiple topics over time as a stream chart. Word clouds are overlayed on the top of the stream chart to show top keywords for each topic at each time. CloudLines [32] proposed different visual representation that shows the intensity of separated topics (x-axis) as a dots stream (y-axis). The article also proposed a smoothing function to determine the dot size based on text counts over time series. In contrast to this work, we focused on the visualization of individual news items rather than topics or themes. As such, our work is more closely related to past work on the exploration of documents over time. TimeLineCurator [20] is a visual analytics system that shows a timeline of event distributions extracted from text documents. The tool provides coordinated views that allow users to browse and curate events by editing or adding the document. BrandSediment [37] adopted a visual sedimentation [26] metaphor to show brand perception (document) distribution in each trait (topic) as a facet of beeswarm charts. We adopted the same concept in our Bitcoin news view (Fig. 1 (V5)) due to its ability to encode data with two additional visual channels (hue, size) per news item.

7 Visualization Design

Given the variety of information we wanted to make available for exploratory analysis, we designed MiningVis to include six coordinated views; each showing multiple different metrics. The display centers around a larger ranking visualization that allows analysts to gain an overview of the Bitcoin mining pool competition and then offers supporting information about factors that may impact mining activities.

7.1 V1: Time selection view

At the top of the tool, a time series chart serves as a historic overview of mining statistics and a filter for the remaining views (Fig. 1 (V1)).

Visual Encoding: The time series shows a selected network statistic measure for the entire history of Bitcoin mining. By default, the timeline shows the mining pool concentration index as an indication to which extent the mining pool distribution risks being dominated by just a few pools [60]. Analysts can select other Bitcoin statistics measures (e.g., total reward, market price, mining difficulty) from a dropdown menu and switch to a log scale. The time series highlights halving days as important events related to mining rewards.

Interaction: Analysts can filter all remaining views (V2–V6) to a specific time interval by brushing on the time axis or specifying a range with the calendar inputs. The selection will then trigger the other views to filter the information to the specified time interval.

7.2 V2: Mining distribution view

The mining distribution view (Fig. 1 (V2)) allows analysts to detect dominating mining pool, the rise and decline of mining pools, and to find characteristics that possibly lead to changes in ranking.

Visual Encoding: We chose a temporal ribbon chart that allows analysts to see both the changing rank and a quantitative metric related to each mining pool. The ribbon chart is filtered to the time period selected in V1 and users can choose among multiple statistical measures and how to group and color the pools. The chart's multiple stacked bars sorted by rank show data aggregated and displayed by months. By default mining pools are sorted from the highest value at the top of the stack to the lowest one at the bottom per month. The same mining pool is connected across months with a ribbon to highlight its rank changing. We encode the top-10 mining pools for the selected time in distinct colors, while the remaining pools are colored in grey.

Fig. 2 shows examples of different configurations of the ribbon chart. Analysts can select the mining power measure, e .g., market share, hash rate, total reward, and transaction fees. They can also chose mining pool characteristics to display in different color hue scales, i .e., the name of the pool, its payout scheme, and location. Using the coloring mechanism, analysts can see patterns for the characteristics of the top mining pool. Furthermore, analysts can group those mining pools by the same characteristics to see if any characteristics correlate with the mining growth and domination in the market. Within each group, mining pools are sorted by the selected measure.

Interaction: Besides the measures, group-by, and color-by selectors, we provide two ways to highlight mining pools or their characteristics. First, analysts can click on the left side labels which will increase transparency of the unselected pools and consequently highlight those that fall into the selection. The selection also affects and filters the mining pools displayed on the mining pool details view (V3). Second, analysts can draw a brush on the ribbon chart to filter both mining pools and highlight a specific time range. The highlighted timeframe is also represented in views V3–V6. Finally, like in all other views, detail-on-demand is available on hover via tooltips that show the mining pool name and exact value of the selected measure for this pool.

Design Alternatives: In the early prototypes, we considered and prototyped bubble chart to show the mining power over time ([60, Fig. 1]). The economist was able to observe each mining pool's mining power but found it difficult to detect rank changes and the extent to which multiple top pools dominated the market. We also considered using a gap chart design, but found it difficult to estimate the total hash rate of mining pools over time. In the end, we opted for a ribbon chart as quantitative values were easier to see with bar height than circle size. It also allowed us to display total mining power in the same view, which required an additional chart in our prototype.

7.3 V3: Mining pool details view

We designed the mining pool details view (Fig. 1 (V3)) to encode details for each mining pool in small multiple charts. The view allows analysts to look at the evolution of each mining pool in detail while also correlating it with additional pool characteristics not present in V2.

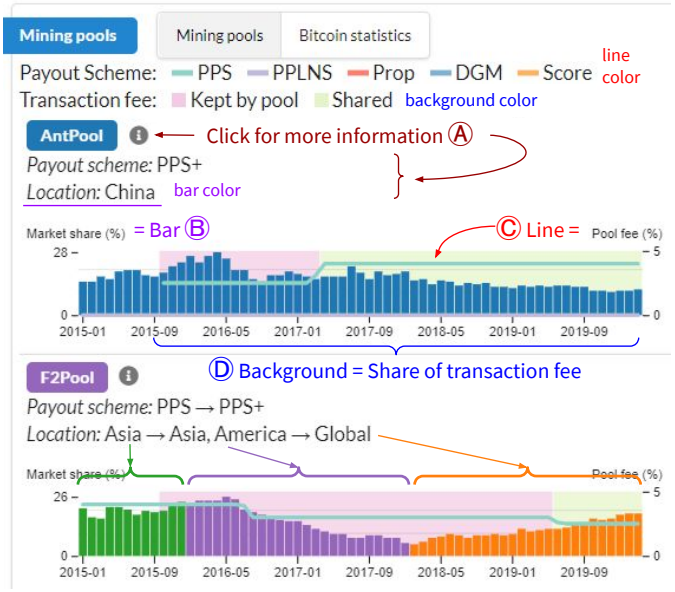


Fig. 3. The mining pool details view (V3) shows the detail of each mining pool as a dual-axis chart. The measure on the left y-axis and the bar color are selected from V2: *Measure: Market share (%)* and *Color by: Location*, respectively. The economists found that the AntPool market share started to decrease after its pool fee increased. F2Pool has changed the primary location three times, and its market share started to increase again after it changed to the global pool (orange).

Visual Encoding: We use a temporal bar chart to encode aggregated (per month) mining pool measures (Fig. 3 (B)). Each bar chart is normalized to the maximum measure of each pool instead of a maximum across all pools. This helps to see the mining power of small pools more clearly. The color of each bar corresponds to the one in V2 to help cross-comparison of these two views.

In order to address T2.3, we added two additional visual encodings to the bar chart: the pool fees kept by the pool (quantitative), and whether the mining pool shares the transaction fee to its miners (nominal). Due to limited screen space we chose a dual-axis encoding instead of an additional chart and show the pool fee as a line chart (Fig. 3 (C)). As the pool fee may vary according to the payout scheme, we use different line colors for different schemes. This dual-axis allows analysts to estimate a possible correlation between pool measures and pool fees. The information about the share of the transaction fee (binary) is encoded as the background color (Fig. 3 (D)). It helps to see when a mining pool changed its policy.

Interaction: The charts in V3 are mostly controlled by selections made in V2 as they are meant as accompanying detail. Analysts can click the info icon for a text description of the pool's characteristics (Fig. 3 (A)). Additional interactions are tooltips for detail-in-demand.

Design Alternatives: As V3 and V2 are closely related, we selected bar charts as the main encoding in V3. However, we discussed design alternatives for the dual-axis encoding extensively as these types of charts had been critiqued, and careful design has been advised [18, 22, 28, 54]. We considered providing an additional bar or line chart above each pool's bar chart. After some tests, we decided against it due to screen-space issues and because these additional charts took away to focus on the display of the selected measure. Our collaborating economist preferred the dual-axis chart as he was familiar with this chart type from the economic literature. To address some problems, we opted for overlaying a line chart on the bar chart to emphasize the two different types of data and provide summary data in text form. Nevertheless, we acknowledge that this chart is relatively complex but allows us to correlate mining power and pool fee directly.



Fig. 4. The Bitcoin statistics view (V4) shows each Bitcoin statistics measure as the area chart. Analysts can browse the list of statistics measures and correlate with measures shown in V2.

7.4 V4: Bitcoin statistics view

Apart from measures directly related to mining pools, broader Bitcoin statistics can influence people's behavior related to mining pools and can help to explain competition. We provide two views, V4 and V5, to help analysts find contextual information for observations made from V2 and V3: Bitcoin statistics (T1.4) and Bitcoin-related news (T1.5).

Visual Encoding: The Bitcoin statistics view (Fig. 4) is located in the same position as V3 reachable via a toggle bar. We encode 17 different factors each as a gray temporal area chart including Bitcoin-internal statistics such as the total number of blocks mined, the total hash rate, the median confirmation time for a block, total rewards paid out, or the total transaction fees. In addition, we calculated statistics with external data such as fees converted to USD according to the current market price or the trade volume in USD. As such, the list of Bitcoin statistics provides information about the status of the Bitcoin network. For example, the number of transactions implies the demand of users; the amount of Bitcoins in transactions means the supply of currency circulating in the market; and the average waiting time indicates the network capacity to verify transactions.

Interaction: The view offers details-on-demand via a tooltip.

Design Alternatives: We considered coloring the area chart instead of using gray but opted against it in order not to confuse the view with V2 and V3 that include an explicit color encoding. Representing the data as a line instead of area charts or bar charts would also be possible without compromising the readability of the data. Again, we opted for a slightly different chart type in order to make it visually obvious which view was currently shown.

7.5 V5: Bitcoin news view

News can be useful to give context for understanding historic mining behavior. Individual and groups of news articles might explain trends or patterns in the Bitcoin measures directly or might themselves have influenced miners' and mining pools' behavior. The Bitcoin news view (Fig. 1 (V5)) allows analysts to browse news that are potentially relevant to Bitcoin mining activities. To structure and group the relevant news in our large corpus of documents (D4), we used topic modeling and designed a swarm plot with search functionality. The dataset we collected included >30,000 postings that cannot all be displayed in one chart. We, therefore, sort the news by its importance score and show only by default the top 300 news on the plot. To preserve news distribution over time, we choose the top news in proportion to the number of news for each month.

Visual Encoding: We use a swarm plot to display each individual news item in a compact fashion across a timeline. It trades off accurate position across the timeline for an overlap-free layout. Each circle represents one news article and its size corresponds to the calculated importance score. The larger the size, the more frequently the posting was read or commented on. We use color to indicate topic membership. The news topics are encoded in different colors, and the list of keywords

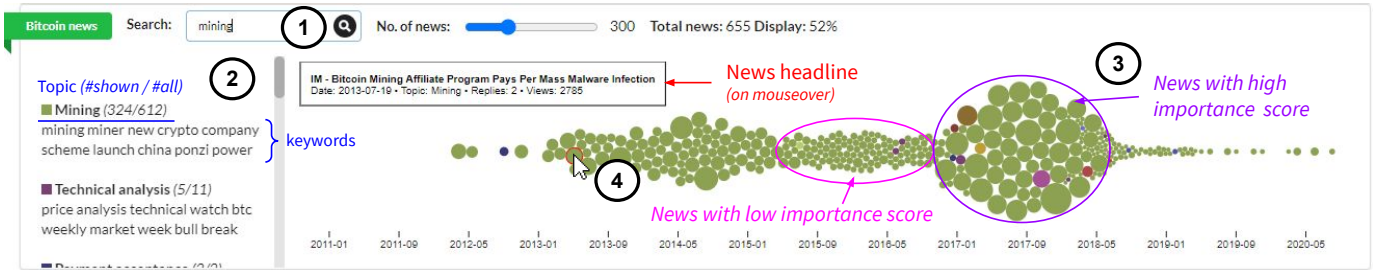


Fig. 5. An example of the Bitcoin news view. (1) The economist filter news for the term "mining". (2) The left panel shows the list of topics and keywords. (3) The swarm plot shows the distribution of news over time. (4) The news headline is displayed when we mouse over on a news circle.

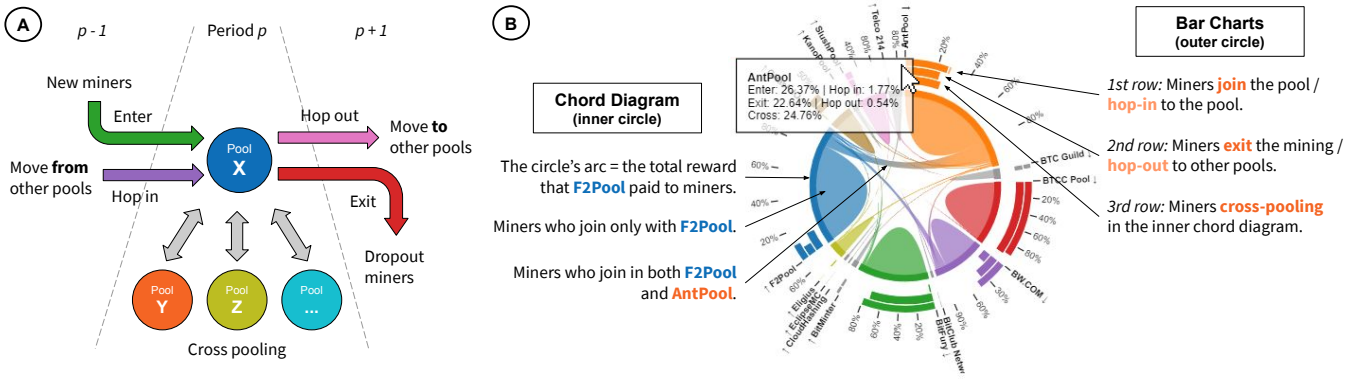


Fig. 6. The cross pooling view. (A) The stock and flow of miners' migration in a mining pool. This is a simplified version of our previous miners' migration flow model [62]. (B) Our visual representation of the miners' migration flows as a chord diagram and bar charts. The economist observed the high cross-pooling between F2Pool and AntPool and detect both pools have the higher incoming miners than outgoing ones (in the bar chart).

listed on the left. The numbers behind the topic label indicate the number of news shown in the swam plot versus the total news.

Interaction: We demonstrate an example use case for the view in Fig. 5. Analysts can browse the news by hovering circles to see tooltips with news headlines on the top-left of the chart. When the analyst clicks on the circle, it will open a new tab to the news source. On the top panel, the list of news can also be narrowed down by keyword search. Analysts can use a slider to specify the number of news showing in the chart. The panel also shows the number of total news and the percentage of news that the chart currently displays. Finally, analysts can use the topics panel on the left to filter news by topic.

Design Alternatives: Our first design used a stream chart to focus on news topic distribution over time. However, the economist wanted to browse individual news items to find detailed information that would explain mining distributions he observed. We then considered using a simple news list sorted by importance coupled with a keyword search. However, with this design analysts would not gain an overview of the news distribution over time. We finally, chose a beeswarm plot because it uses a compact design to show the distribution of news while still allowing us to encode topic membership and importance.

7.6 V6: Cross pooling view

Miners' flows are complicated to visualize because they involve several processes: new, hop-in, cross-pooling, hop-out, and dropout miners within and across the time frame (Fig. 6 (A)). We designed the cross pooling view to represent and summarize migration patterns.

Visual Encoding: The cross-pooling miners during a selected time interval can be considered as a flow of miners between mining pools. Fig. 5 (B) shows the visual encoding of the view. We used a chord diagram to display a metric related to miners crossing between pools; the total amount of miners' rewards (default) or the total number of miner addresses. The diagram shows the metric encoded as the outside arcs' length. The flow between mining pools represents the total amount of the metric for those miners who cross-pooled. We represent the average percentage of the miner's migration statistics per month (i.e., new, exit, hopping in, hopping out, cross pooling) as stacked bar charts around the outer circle. For each pool, three rows represent the percentage of miners incoming (new and hopping in), outgoing (dropout and hopping out), and cross-pooling with the mining pool, respectively.

Interaction: Analysts can hover over the flow or stacked bars to see the exact value. They can also change the metric from the total reward to the number of miner addresses. The total reward is a weighted average that considers the impact of large players in the pool. The measure is more robust than just the number of miners.

Design Alternatives: Cross-pooling is represented as a flow to highlight miners that receive a reward from multiple pools. Previous work had used glyph charts [69] to show numbers over time but does not highlight the association between pools. Due to the fluctuations of miner counts in the data [62], we chose to show aggregated data to reduce outliers. A heatmap matrix is an alternative to display the percentage of cross-pooling miners ([62, Fig. 6]) but the color coding turned out to be problematic due to the different largely varying scale of mining pools. In contrast, the chord diagram encodes the scale of mining pools as the arcs that allow analysts to see how large the mining pool is and detect significant miners' flow between pools.

8 User Study

Our tool was designed with an economist and to focus on economic analysis questions. While we developed the tool, we identified additional potential user groups with economic interests in mining. Our goal was to understand whether the tool intended for an economic analysis would also be applicable to day-to-day Bitcoin users, in particular miners as one of the largest potential user groups. In the user study, we therefore conducted a multi-week online user study with eight Bitcoin miners to study the usability and relevance of the tool for them.[1] We report the full background and survey responses of every participant in our OSF repository `osf.io/ud2c9/`.

[1]The research ethics board of the Université Paris-Saclay has approved this study, CER-Paris-Saclay-2020-062.

Table 2. List of tasks for participants in the first interview study and the usage patterns we have observed.

Task	Question	Usage Pattern	Description
1	Which mining pool received the most BTC in August 2019?	V1 → V2 (8)	8 quickly used V1 to zoom close to the month and then select the "Total Reward (BTC)" measure in V2.
2	Can you explain why GHash.IO dominated the Bitcoin mining activity and then stopped its operation?	V2 → V1 → V2/V5/V3 (4) V2 → V5 (3) V1 → V2 → V3 (1)	7 detected the GHash.IO pool from V2 quickly, then selected the pool in the legend, and looked at the market share domination and decline. 4 then filtered the time to see patterns more clearly in V1. 7 searched for the news with "GHash.IO" that could explain the decline.
3	Can you identify when top mining pools nearly dominated the market and posed a risk to the 51% attack?	V2 → V3 (5) V2 → V1/V5 → V3 (2) V3 (1)	7 searched for pools with outstanding market shares in V2 and looked for detail in V3 later. 1 looked at V3 directly 6 were satisfied with naming the pools and stopped looking for mining pool details. 1 exceptionally searched for the news about 51% attack.
4	Can you find a historical anomaly in the transaction fees? If yes, can you explain what else happened in the same period?	V2 → V1 → V4 (4) V2 → V4 → V4/V5 (2) V2 → V5 → V4 (2)	8 changed the measure to "Transaction Fee" and spotted the anomaly in V2. 4 proceeded to use V1 to zoom-in on the time with extreme transaction fees. 4 browsed the news with the "Transaction Fee" keyword before finding the Bitcoin statistics in V4 that correlated with this event.

8.1 Participant recruitment

We asked potential participants to sign up for our user study in the popular Bitcointalk.org forum as well as Reddit channels dedicated to Bitcoin (*r/Bitcoin* and *r/BitcoinMining*). As compensation, we offered participants free access to our tool. Prospective participants filled a consent form, contact e-mail, country of residence, and short questions about their Bitcoin and mining experience. We needed country information to check the legality of Bitcoin mining in their country. We received 11 complete responses and were able to recruit 8 participants.

Participant Background: All participants were male and had experience with Bitcoin mining. Six mined Bitcoin as part of a mining pool, one mined Bitcoin alone and one worked with Bitcoin as part of his job. Participants had 4.38 years of experience on average in Bitcoin mining before the user study (sd: 2, median: 4.5).

8.2 Procedure

The study consisted of : 1) a 1h introductory session & interview, 2) a 2-week free exploration phase, and 3) a follow-up interview and survey.

8.2.1 Introductory Session

The study began with a 1h online interview with each participant. Before its start, we asked participants to fill a consent form and answer questions about their data visualization experience. All participants reported to be very familiar and comfortable with reading data visualizations. In the first 15 minutes, we talked about participants' interests or any specific analysis questions they may already have. Then, we gave a 15-minute presentation about the features of MiningVis. Next, we asked participants to think-aloud while exploring our tool for around 30 minutes with a set of short exploratory tasks we had prepared. The tasks listed in Table 2 aimed to get participants familiar with various aspects of the tool and to see how participants would intuitively go about answering them with MiningVis. At the end of the session, we asked participants for their first impression about the tool's data displays and usability and explained the subsequent steps of the user study.

8.2.2 Free-Exploration Phase

We gave participants free access (login and password) to our tool and asked them to use it at their own pace for two weeks. We wanted to learn which questions people were trying to answer and which findings and insights they were able to make. To make insight reports as lightweight as possible we designed mini-logs inspired by the micro-entries [9] methodology. Through a mini-form reachable from the MiningVis interface we asked participants to report a) any interesting pattern they found using the tools, b) the views in the tool they used, and c) explain what they observed. We encouraged them to make at least five submissions and reminded them once throughout the two week phase.

8.2.3 Final Interview and Survey

After two weeks, we contacted all eight participants for a follow-up interview and survey. In addition to what information they found useful or missing we asked them about their thoughts on the micro-entries form as the method was just recently proposed for insight evaluation. The survey consisted of a System Usability Scale (SUS) [15] and a variant of the ICE-T questionnaire [65]. Four participants agreed to be interviewed and five filled the survey.

8.3 Results

Next, we report our most important findings related to tool and methods.

8.3.1 Usage Patterns and Usability

During the think-aloud session in the first interview, we observed participants solve four tasks. When participants got stuck, we explained views they had not considered yet to increase familiarity with the tool. We tracked the sequence of views participants interacted with for each task and summarized the patterns in Table 2.

We observed that participants were comfortable and easily interacted with the three central MiningVis views: V1 (the overall timeline), V2 (the ranking view), as well as V5 (the news view). The most frequently used view was V2, which aligns with our expectation (see Table 1). They used V5 frequently to look for explanations in the news. We found that participants often overlooked the detail provided by V3 and V4, and occasionally reminded them about those views and visual encoding details. After the session, they reported being slightly overwhelmed by the amount of information in the tool. Participants never used V6 to investigate miners' mobility even though it might have provided information for Task 2. We found out that pool hopping was not a familiar metric that related to participants' own work or Bitcoin usage.

After completing the tasks, participants shared their first impressions on the tool. In general, they complimented our tool and mentioned that it was the most complete tool about mining data they were familiar with. They also pointed out interactions that mismatched their first intuition but that they quickly learned how to use the tool. After the two-week free exploration we asked participants to fill the System Usability Scale. The responses showed an average score of 73.5/100 (sd: 13.3, median: 72.5) which is an above average (68) SUS score [53] but still shows room for improvement. No individual question scores were below average (on a negative to positive response scale); with the highest scores given for confidence of use and learnability. The lowest (but still positive) scores were given for ease of use. From feedback outside the survey itself, we found that in particular the unavailability of real-time mining data kept people from considering extended future use of the tool. One participant from the first interview even declined to participate in the follow-up due to a lack of real-time data.

Usability Improvements: Based on feedback and observations, we improved the tool's usability: 1) Several participants expected the highlight brush in V2 to zoom in on a specific time interval instead of

only highlighting pools and time intervals. We added a highlight brush and a "Focus" button on V1 to allow zooming-in on the highlighted time period. 2) Participants wanted to sort mining pools in V3. We added a dropdown that allows analysts to sort mining pools based on average, min, or max of the displayed pool measure. 3) We added a search panel in V4 to allow users to find specific Bitcoin measures.

8.3.2 Insights and Findings

We received 12 micro-entries from four participants. We removed four entries from our analysis: three were feature suggestions and one was a general tool review. Participants reported to log in 5.2 times on average (sd: 4.38, median: 3) within the two weeks. All but one participant mentioned to only have spent a limited amount of time with the tool during the free-exploration phase and therefore did not make deep insights. We coded each entry according to analysis factors from Sect. 3. 6 entries referred to AF1 (evolution of pools), 4 to AF2 (pool payments), and 1 to AF3 (miner decisions). Participants related V2 (6 entries) to their entries, followed by V1 (3), V5 (3), V3 (2), and V6 (1).

Overall, the eight insight-based micro-entries showed a focus on higher-level mining activity evolution rather than on a specific mining pool. Three entries mentioned periods in which mining activity had high transaction fees, a hash rate decline, or halving days. For example, one participant found high volatility in the transaction fee during the 3rd halving period in May 2020. The transaction fee was tripled in the halving month and then decreased in the next month before it turned to increase again. Surprisingly, he did not observe the same pattern during the 2nd halving period. Four other entries looked at the shift in pool characteristics (i.e., location and payout scheme) that dominated the mining ecosystem over time. One entry observed that BTC.com became the top pool in Sep. 2020. He then looked at the mining pool detail and found that BTC.com's pool fee is 4%, which is higher than other pools. This finding is counter-intuitive to the assumption that a lower pool fee should lead to a higher market share. Exceptionally, one entry explored the news related to "exchange" services and found interesting information on how bitcoins are used to buy other cryptocurrencies.

8.3.3 Value of the Visualization

We evaluated the value of the implemented visualizations with an adaptation of the ICE-T questionnaire [65]. We used a scale of five to score the heuristics (from 1–Strongly Disagree to 5–Strongly Agree) and calculated the average score for each value component.

The tool was rated high on its ability to provide *insights* (average: 4.30, sd: 0.36) on Bitcoin mining pools and *essence* (4.05, sd: 0.43) to see overall mining pool evolution. In the interview participants mentioned in particular the good analytic structure of the tool and the diversity of available Bitcoin mining indicators. Participants also gave a high score for *confidence* (3.73, sd: 0.54), except for "the visualization helps to understand data quality" (2.8, sd: 0.75). This is fair as we did not specifically reveal quality-related information in the tool. *Time*-related heuristics were scored the lowest (but still above average) (3.70, sd: 0.56). In the interview, participants commented on latency issues when the tool processed data of the entire Bitcoin history.

9 Discussion

Our study results highlight several aspects related to the value of our tool and our methodologies.

Value of the tool: Throughout the development of MiningVis, we were able to show the value of dedicated visual analytics approaches to understanding mining. In particular, based on the prototypes, our collaborator made scientific discoveries reported in two analysis articles [60, 62]. Still today, our economist collaborator actively uses the tool to develop a miner's decision model. From this past work, we found that mining activity constantly evolves and is volatile relative to multiple factors. Therefore, the tool will be helpful to monitor and analyze the behavior of mining pools in the long term.

The value of our tool for Bitcoin miners was a little less obvious from our user study. While the overall feedback on usability and visualization design was positive, and we could confirm our design choices, participants in particular wanted real-time data in order to make small-scale personal decisions. In the follow-up interview, four participants suggested that the tool has a potential benefit to researchers and mining companies to analyze historical trends of Bitcoin mining. We concluded that our current tool is most useful for modeling and long-term decision making from a historical perspective.

Integration with real-time data: Real-time updates of Bitcoin transaction data requires attributing each new block to a mining pool. The calculation of cross-pooling poses a bottleneck as we need to track transaction flows to identify individual miners. For future work, we plan to update the Bitcoin transaction data in real-time and integrate a script to crawl external data. Pool characteristics are collected and cleaned by hand but could be extracted automatically from the Bitcoin Wiki [6] and Bitcoin news APIs with natural language processing techniques.

Generalization to other cryptocurrencies: While our tool is focused on Bitcoin, other cryptocurrencies implement the same mining protocol. Those cryptocurrencies also require miners to solve a computationally-expensive puzzle and provide financial rewards as an incentive to mine. Mining pools also emerged in those cryptocurrencies. For example, miners can join the BTC.com mining pool to mine Ethereum, Bitcoin Cash, Litecoin, or Monero. Therefore, the tool can be easily adapted to analyze mining patterns of those cryptocurrencies.

Reflection on micro-entries. We were disappointed by the low response rate and quality of the micro entries we collected. We thought that the light-weight format to collect even simple "findings" we had formulated, would encourage participants to report more. Participants reported that they neither had the time nor specific questions they wanted to investigate in depth. We understood that they did not consider adding entries about data they were already familiar with through their daily activities with Bitcoin. In the first interview, we observed that participants were intimately familiar with the transaction fee spike we asked about in Task 4 and that they were not interested in investigating familiar phenomena such as this one in depth. It would have been good to give these domain experts areas of the data to make reports about, in case they were not inspired to find their own. For example, asking participants to make one report each about pool fees, pool hopping or rising or declining pools would have led to uptake—but given less useful information about their personal interests. Some participants used the form to provide feedback rather than report findings. Another mechanism should be found for participants to report feedback.

10 Conclusion

We present the MiningVis tool to analyze the Bitcoin mining pool emergence and evolution. Mining is a critical activity in cryptocurrencies that involves multiple longitudinal factors, both internal and external to the Bitcoin ecosystem. MiningVis allows analysts to relate multiple measures from complementary datasets in a multi-coordinated view. The tool helped our economist collaborator discover new findings and is currently used to develop an economic model. We conducted a user study with Bitcoin miners to study both the usability and insights obtained from the tool. By observing their usage patterns and through interviews we confirmed the design of our main views but also learned that pool hopping was less interesting to general miners than our collaborator. Participants rated the tool highly for usability and visualization value and we were able to make improvements based on participants' detailed feedback. For the first time, we used the micro-entries methodology and report ideas on how to improve its utility for studying a deployed tool. In the end, we discuss the potential of applying the tool to analyze mining patterns in other related cryptocurrencies.

Acknowledgments

The authors wish to thank all participants in the user study, Catherine Plaisant for valuable feedback to improve our work and Teppakorn Thanuthanad for technical advice on developing the tool. 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*.

REFERENCES

[1] A. M. Antonopoulos. *Mastering Bitcoin: Programming the Open Blockchain*. O'Reilly Media, Inc., 2nd ed., 2017.

[2] L. Badea and M. C. Mungiu-Pupăzan. The economic and environmental impact of bitcoin. *IEEE Access*, 9:48091–48104, 2021. doi: 10.1109/ACCESS.2021.3068636

[3] M. Belotti, S. Kirati, and S. Secci. Bitcoin pool-hopping detection. In *IEEE International Forum on Research and Technology for Society and Industry (RTSI)*, pp. 1–6. IEEE, 2018. doi: 10.1109/RTSI.2018.8548376

[4] M. Bevand. Cambridge bitcoin electricity consumption index, Jun 2021. `https://cbeci.org/`.

[5] Bitcoin Wiki. Comparison of mining pools. `https://en.bitcoin.it/wiki/Comparison_of_mining_pools`.

[6] Bitcoin Wiki contributors. Comparison of mining pools, 2020. Accessed on: Dec 12, 2020.

[7] Bitcointalk contributors. Bitcoin mining pools discussion forum. `https://bitcointalk.org/index.php?board=41.0`.

[8] Bitcointalk contributors. Bitcoin press discussion forum. `https://bitcointalk.org/index.php?board=77.0`.

[9] J. E. Block and E. D. Ragan. Micro-entries: Encouraging deeper evaluation of mental models over time for interactive data systems. In *IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pp. 38–47. IEEE, 2020. doi: 10.1109/BELIV51497.2020.00012

[10] Blockchain.info. Blockchain charts, 2020. `https://www.blockchain.com/charts`.

[11] R. Böhme, N. Christin, B. Edelman, and T. Moore. Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, 29(2):213–38, 2015. doi: 10.1257/jep.29.2.213

[12] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten. Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *IEEE Symposium on Security and Privacy*, pp. 104–121. IEEE, 2015. doi: 10.1109/SP.2015.14

[13] D. Bradbury. The problem with bitcoin. *Computer Fraud & Security*, 2013(11):5–8, 2013. doi: 10.1016/S1361-3723(13)70101-5

[14] W. C. Brinton. *Graphic methods for presenting facts*. Engineering magazine company, 1919.

[15] J. Brooke. Sus: a quick and dirty usability. *Usability evaluation in industry*, 189, 1996.

[16] BTC.com. Pool distribution, 2020. `https://btc.com/stats/pool`.

[17] S. Fairchild, T. Farrell, B. Gunton, A. Mackinnon, C. McNamara, and R. Trachtman. Design and data in balance: Using design-driven decision making to enable student success. *New Visions for Public Schools*, 2014.

[18] S. Few. Dual-scaled axes in graphs—are they ever the best solution? In *Visual Business Intelligence Newsletter*. Perceptual Edge, Mar. 2008.

[19] S. Foteinis. Bitcoin's alarming carbon footprint. *Nature*, 554(7690):169–170, 2018. doi: 10.1038/d41586-018-01625-x

[20] J. Fulda, M. Brehmer, and T. Munzner. Timelinecurator: Interactive authoring of visual timelines from unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):300–309, 2015. doi: 10.1109/TVCG.2015.2467531

[21] P. Giungato, R. Rana, A. Tarabella, and C. Tricase. Current trends in sustainability of bitcoins and related blockchain technology. *Sustainability*, 9(12):2214, 2017. doi: 10.3390/su9122214

[22] K. W. Haemer. Double scales are dangerous. *The American Statistician*, 2(3):24, June 1948.

[23] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002. doi: 10.1109/2945.981848

[24] F. Heimerl, Q. Han, S. Koch, and T. Ertl. Citerivers: Visual analytics of citation patterns. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):190–199, 2015. doi: 10.1109/TVCG.2015.2467621

[25] N. Houy. The economics of bitcoin transaction fees. *GATE WP*, 1407, 2014. doi: 10.2139/ssrn.2400519

[26] S. Huron, R. Vuillemot, and J.-D. Fekete. Visual sedimentation. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2446–2455, 2013. doi: 10.1109/TVCG.2013.227

[27] International Monetary Fund. Primary commodity price system: Fuel (energy) index, Jun 2021. `https://data.imf.org/?sk=471DDDF8-D8A7-499A-81BA-5B332C01F8B9&sId=1547558078595`.

[28] P. Isenberg, A. Bezerianos, P. Dragicevic, and J.-D. Fekete. A study on dual-scale data charts. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2469–2487, Dec. 2011. doi: 10.1109/TVCG.2011.160

[29] S. Jiang, Y. Li, Q. Lu, Y. Hong, D. Guan, Y. Xiong, and S. Wang. Policy assessments for the carbon emission flows and sustainability of bitcoin blockchain operation in china. *Nature communications*, 12(1):1–10, 2021. doi: 10.1038/s41467-021-22256-3

[30] C. Kinkeldey, J.-D. Fekete, T. Blascheck, and P. Isenberg. Bitconduite: Exploratory visual analysis of entity activity on the bitcoin network. *IEEE Computer Graphics and Applications*, 2021. doi: 10.1109/MCG.2021.3070303

[31] S. Köhler and M. Pizzol. Life cycle assessment of bitcoin mining. *Environmental science & technology*, 53(23):13598–13606, 2019. doi: 10.1021/acs.est.9b05687

[32] M. Krstajic, E. Bertini, and D. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2432–2439, 2011. doi: 10.1109/TVCG.2011.179

[33] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9):1639–1645, 2009. doi: 10.1101/gr.092759.109

[34] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *IEEE Pacific Visualization Symposium (PacificVis)*, pp. 117–121. IEEE, 2015. doi: 10.1109/PACIFICVIS.2015.7156366

[35] S. Küfeoğlu and M. Özkuran. Bitcoin mining: A global review of energy and power demand. *Energy Research & Social Science*, 58:101273, 2019. doi: 10.1016/j.erss.2019.101273

[36] X. Liu, W. Wang, D. Niyato, N. Zhao, and P. Wang. Evolutionary game for mining pool selection in blockchain networks. *IEEE Wireless Communications Letters*, 7(5):760–763, 2018. doi: 10.1109/LWC.2018.2820009.

[37] X. Liu, A. Xu, L. Gou, H. Liu, R. Akkiraju, and H.-W. Shen. Socialbrands: Visual analysis of public perceptions of brands on social media. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 71–80. IEEE, 2016. doi: 10.1109/VAST.2016.7883513

[38] 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.

[39] E. Masanet, A. Shehabi, N. Lei, H. Vranken, J. Koomey, and J. Malmodin. Implausible projections overestimate near-term bitcoin CO_2 emissions. *Nature Climate Change*, 9(9):653–654, 2019. doi: 10.1038/s41558-018-0321-8

[40] C. Mora, R. L. Rollins, K. Taladay, M. B. Kantar, M. K. Chock, M. Shimada, and E. C. Franklin. Bitcoin emissions alone could push global warming above 2°C. *Nature Climate Change*, 8(11):931–933, 2018. doi: 10.1038/s41558-018-0321-8

[41] M. Möser and R. Böhme. Trends, tips, tolls: A longitudinal study of bitcoin transaction fees. In *International Conference on Financial Cryptography and Data Security*, pp. 19–33. Springer, 2015. doi: 10.1007/978-3-662-48051-9_2

[42] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008. `http://bitcoin.org/bitcoin.pdf`.

[43] C. T. Nguyen, D. T. Hoang, D. N. Nguyen, D. Niyato, H. T. Nguyen, and E. Dutkiewicz. Proof-of-stake consensus mechanisms for future blockchain networks: fundamentals, applications and opportunities. *IEEE Access*, 7:85727–85745, 2019. doi: 10.1109/ACCESS.2019.2925010

[44] M. Oppermann and T. Munzner. Data-first visualization design studies. In *IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pp. 74–80, 2020. doi: 10.1109/BELIV51497.2020.00016

[45] F. Parino, M. G. Beiró, and L. Gauvin. Analysis of the bitcoin blockchain: socio-economic factors behind the adoption. *EPJ Data Science*, 7(1):38, 2018. doi: 10.1140/EPJDS/S13688-018-0170-8

[46] C. Perin, J. Boy, and F. Vernier. Using gap charts to visualize the temporal evolution of ranks and scores. *IEEE Computer Graphics and Applications*, 36(5):38–49, 2016. doi: 10.1109/MCG.2016.100

[47] S. Rendgen. *The Minard System: The Complete Statistical Graphics of Charles-Joseph Minard*. Chronicle Books, 2018.

[48] P. Riehmann, M. Hanfler, and B. Froehlich. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization*, pp. 233–240. IEEE, 2005. doi: 10.1109/INFVIS.2005.1532152

[49] 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.

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

[51] M. Rosvall and C. T. Bergstrom. Mapping change in large networks. *PloS one*, 5(1):e8694, 2010. doi: 10.1371/journal.pone.0008694

[52] M. Salimitari, M. Chatterjee, M. Yuksel, and E. Pasiliao. Profit maximization for bitcoin pool mining: A prospect theoretic approach. In *IEEE International Conference on Collaboration and Internet Computing*, pp. 267–274. IEEE, 2017. doi: 10.1109/CIC.2017.00043

[53] J. Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC, 2011.

[54] C. F. Schmid. *Statistical Graphics: Design Principles and Practices*. Krieger Pub Co, 1992.

[55] M. Schmidt. The sankey diagram in energy and material flow management: part ii: methodology and current applications. *Journal of industrial ecology*, 12(2):173–185, 2008. doi: 10.1111/j.1530-9290.2008.00015.x

[56] C. Shi, W. Cui, S. Liu, P. Xu, W. Chen, and H. Qu. Rankexplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2669–2678, 2012. doi: 10.1109/TVCG.2012.253

[57] C. Stoll, L. Klaaßen, and U. Gallersdörfer. The carbon footprint of bitcoin. *Joule*, 3(7):1647–1661, 2019. doi: 10.1016/j.joule.2019.05.012

[58] Y. Sun, H. Xiong, S. M. Yiu, and K. Y. Lam. Bitvis: An interactive visualization system for bitcoin accounts analysis. In *Crypto Valley Conference on Blockchain Technology*, pp. 21–25. IEEE, 2019. doi: 10.1109/CVCBT.2019.000-3

[59] N. Tovanich, N. Heulot, J.-D. Fekete, and P. Isenberg. Visualization of Blockchain Data: A Systematic Review. *IEEE Transactions on Visualization and Computer Graphics*, 2020. doi: 10.1109/TVCG.2019.2963018

[60] N. Tovanich, N. Soulié, and P. Isenberg. Visual Analytics of Bitcoin Mining Pool Evolution: On the Road Toward Stability? In *IFIP NTMS International Workshop on Blockchains and Smart Contracts (BSC)*. Paris, France, Apr. 2021.

[61] N. Tovanich, N. Soulié, N. Heulot, and P. Isenberg. Dataset: An empirical analysis of pool hopping behavior in the bitcoin blockchain, Dec 2020. doi: 10.5281/zenodo.4671055

[62] N. Tovanich, N. Soulié, N. Heulot, and P. Isenberg. An empirical analysis of pool hopping behavior in the bitcoin blockchain. In *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 2021.

[63] J. Truby. Decarbonizing bitcoin: Law and policy choices for reducing the energy consumption of blockchain technologies and digital currencies. *Energy research & social science*, 44:399–410, 2018.

[64] H. Vranken. Sustainability of bitcoin and blockchains. *Current opinion in environmental sustainability*, 28:1–9, 2017. doi: 10.1016/j.cosust.2017.04.011

[65] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2018. doi: 10.1109/TVCG.2018.2865146

[66] C. Wang, X. Chu, and Y. Qin. Measurement and analysis of the bitcoin networks: A view from mining pools. In *International Conference on Big Data Computing and Communications (BIGCOM)*, pp. 180–188. IEEE, 2020. doi: 10.1109/BigCom51056.2020.00032

[67] L. Wang and Y. Liu. Exploring miner evolution in bitcoin network. In *International Conference on Passive and Active Network Measurement*, pp. 290–302. Springer, 2015. doi: 10.1007/978-3-319-15509-8_22

[68] J. Xia, Y. Hou, Y. V. Chen, Z. C. Qian, D. S. Ebert, and W. Chen. Visualizing rank time series of wikipedia top-viewed pages. *IEEE Computer Graphics and Applications*, 37(2):42–53, 2017. doi: 10.1109/MCG.2017.21

[69] 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*, 21:507–523, 2020. doi: 10.1631/FITEE.1900532

[70] X. Yue, X. Shu, X. Zhu, X. Du, Z. Yu, D. Papadopoulos, and S. Liu. Bitextract: Interactive visualization for extracting bitcoin exchange intelligence. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):162–171, 2018. doi: 10.1109/TVCG.2018.2864814

[71] Y. Zuo, J. Zhao, and K. Xu. Word network topic model: a simple but general solution for short and imbalanced texts. *Knowledge and Information Systems*, 48(2):379–398, 2016. doi: 10.1007/s10115-015-0882-z