Visual Analytics for Monitoring and Exploration of Bitcoin Blockchain Data

Natkamon Tovanich

Visual Analytics for Monitoring and Exploration of Bitcoin Blockchain Data

Analyse visuelle pour la surveillance et l'exploration des données de la blockchain Bitcoin

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 580 : Sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat : Informatique
Graduate School : Informatique et Sciences du Numérique
Référent : Faculté des sciences d'Orsay

Thèse préparée au **Laboratoire interdisciplinaire des sciences du numérique** (Université Paris-Saclay, CNRS, Inria), et dans l'**IRT SystemX**, sous la direction de **Jean-Daniel FEKETE**, Directeur de recherche, sous le co-encadrement de **Petra ISENBERG**, Chargée de recherche, et le co-encadrement de **Nicolas HEULOT**, Ingénieur de recherche

Thèse soutenue à Paris-Saclay, le 8 février 2022, par

Natkamon TOVANICH

Composition du jury

Sylvain CONCHON Professeur, Univ. Paris-Saclay, CNRS, Inria, LISN	Président
Ross MACIEJEWSKI Professeur, Arizona State University	Rapporteur & Examinateur
Giuseppe DI BATTISTA Professeur, Università degli Studi Roma Tre	Rapporteur & Examinateur
William KNOTTENBELT Professeur, Imperial College London	Examinateur
Enrico BERTINI Professeur agrégé, NYU Tandon School of Engineering	Examinateur
Jean-Daniel FEKETE Directeur de recherche, Univ. Paris-Saclay, CNRS, Inria, LISN	Directeur de thèse

THESE DE DOCTORAT

NNT : 2022UPASG010

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Analyse visuelle pour la surveillance et l'exploration des données de la blockchain Bitcoin

Mots clés : Bitcoin, minage de Bitcoin, Pools de minage, Pool Hopping, Analyse Visuelle, Blockchain

Résumé : Bitcoin est une crypto-monnaie pionnière qui enregistre les transactions dans un registre public et distribué appelé blockchain. Il est utilisé comme support pour les paiements, les investissements et plus largement la gestion d'un portefeuille numérique qui n'est pas administré par un gouvernement ou une institution financière. Au cours de ces dix dernières années, l'activité transactionnelle de Bitcoin a rapidement et largement augmenté. La volumétrie ainsi que la nature évolutive de ces données posent des défis pour l'analyse et l'exploration des usages ainsi que des activités sur la blockchain. Le domaine de l'analyse visuelle travaille sur la conception de systèmes analytiques qui permettent aux humains d'interagir et d'obtenir des informations à partir de données complexes.

Dans cette thèse, j'apporte plusieurs contributions à l'analyse des activités de minage sur la blockchain Bitcoin. Tout d'abord, je propose une caractérisation des travaux passés et des défis de recherche liés à l'analyse visuelle pour les blockchains. À partir de cette étude, j'ai proposé un outil d'analyse visuelle pour comprendre les activités de minage qui sont essentielles pour maintenir l'intégrité et la sécurité des données sur la blockchain Bitcoin. Je propose une méthode pour extraire l'activité des mineurs à partir des données de transaction et tracer leur comportement de bascule d'un pool de minage à un autre. L'analyse empirique de ces données a notamment révélé que les nouveaux pools de minage offraient une meilleure incitation et attiraient davantage de mineurs. Cette analyse a également montré que les mineurs choisissaient stratégiquement leur pool de minage dans le but de maximiser leur profit. Pour explorer l'évolution et la dynamique de cette activité sur le long terme, j'ai développé un outil d'analyse visuelle, appelé MiningVis, qui intègre des données liées au comportement des mineurs avec des informations contextuelles issues des statistiques et de l'actualité de Bitcoin. L'étude avec des utilisateurs démontre que les participants au minage de Bitcoin cherchent à utiliser l'outil pour analyser l'activité globale plutôt que pour étudier les détails d'un pool de minage. Les commentaires des participants prouvent que l'outil les a aidés à mettre en relation plusieurs informations et à découvrir les tendances dans l'activité de minage de Bitcoin.

Title : Visual analytics for monitoring and exploration of Bitcoin blockchain data

Keywords : Bitcoin, Bitcoin Mining, Mining Pools, Pool Hopping, Visual Analytics, Blockchain

Abstract : Bitcoin is a pioneer cryptocurrency that records transactions in a public distributed ledger called the blockchain. It has been used as a medium for payments, investments, and digital wallets that are not controlled by any government or financial institution. Over the past ten years, transaction activities in Bitcoin have increased rapidly. The volume and evolving nature of its data pose analysis challenges to explore diverse groups of users and different activities on the network. The field of Visual Analytics (VA) has been working on the development of analytical systems that allow humans to interact and gain insights from complex data.

In this thesis, I make several contributions to the analysis of Bitcoin mining activity. First, I provide a characterization of the past work and research challenges related to VA for blockchains. From this assessment, I proposed a VA tool to understand mining activities that ensure data integrity and security on the Bitcoin blockchain. I propose a method to extract miners from the transaction data and trace pool hopping behavior. The empirical analysis of this data revealed that emerging mining pools provided a better incentive to attract miners. Simultaneously, miners strategically chose mining pools to maximize their profit. To explore the evolution and dynamics of this activity over the long term, I developed a VA tool called MiningVis that integrates mining behavior data with contextual information from Bitcoin statistics and news. The user study demonstrates that Bitcoin miner participants use the tool to analyze higher-level mining activity rather than mining pool details. The evaluation of the tool proves that it helped participants to relate multiple information and discover historical trends of Bitcoin mining.

SYNTHÈSE DE THÈSE

Bitcoin est une crypto-monnaie qui enregistre les transactions dans un registre public et distribué appelé blockchain. Elle est utilisée comme un support pour les paiements, les investissements et plus largement la gestion d'un portefeuille numérique qui n'est pas administré par un gouvernement ou une institution financière. Au cours de ces dix dernières années, l'activité transactionnelle de Bitcoin a rapidement et largement augmenté. La volumétrie ainsi que la nature évolutive de ces données posent des défis pour l'analyse et l'exploration des usages ainsi que des activités sur la blockchain. Le domaine de l'analyse visuelle travaille sur la conception de systèmes analytiques qui permettent aux humains d'interagir et d'obtenir des informations à partir de données complexes.

Dans cette thèse, j'apporte plusieurs contributions à l'analyse des activités de minage sur la blockchain Bitcoin. Les travaux de recherche se sont jusque-là concentrés sur des analyses théoriques de l'activité, des analyses des réseaux de transactions ainsi que des applications de l'apprentissage non supervisé pour classer différents types d'entités ou des transactions malveillantes. Pour comprendre l'état de l'art de la visualisation des données appliquée aux blockchains ainsi que la façon dont les visualisations sont utilisées pour analyser l'activité blockchain, j'ai réalisé *une revue systématique*. En particulier, j'ai proposé un schéma de classification permettant de caractériser les différents outils et visualisations existants. Ensuite, j'ai recensé les défis restant à lever ainsi que les opportunités de recherche associées. J'ai constaté que les outils de visualisation dédiés à des questions avancées et axées sur un domaine en particulier sont rares malgré qu'ils puissent être utiles à des experts et à des analystes afin d'étudier et comprendre en profondeur une activité spécifique de la blockchain Bitcoin telle que le minage.

Les principales parties de cette thèse se concentrent ainsi sur le minage de Bitcoin qui est un mécanisme critique pour la validation des transactions dans un réseau décentralisé. Il a également des implications sur le modèle économique de la blockchain car les mineurs sont sensibles aux incitations et ils participent principalement pour maximiser leur gain. Les pools miniers sont rapidement apparus comme des organisations permettant d'assurer un revenu régulier aux mineurs. En utilisant une approche d'analyse visuelle, j'ai mené *une étude de conception de deux ans avec mon collaborateur économiste* pour étudier différentes questions de recherche et développer des prototypes de visualisation afin de proposer des réponses concernant l'émergence et l'évolution des pools miniers Bitcoin.

Dans la première itération de conception, j'ai *extrait plusieurs sources de données liées au développement des parts de marché et aux caractéristiques des pools miniers* comme la localisation, le système de paiement et les frais de pool. *Avec mon collaborateur économiste, nous avons présenté une première analyse exploratoire de l'évolution des pools miniers couvrant toute l'histoire de Bitcoin*. Nous avons découvert que les premiers pools miniers proposaient différents systèmes de paiement très variés avant qu'ils ne convergent vers le paiement à la part (PPS) et le paiement intégral à la part (FPPS). L'étude de la localisation des pools miniers montre que les pools chinois sont devenus des pools internationaux vers septembre 2017, lorsque le gouvernement chinois a réglementé le commerce des crypto-monnaies.

Lors de la deuxième itération de conception, nous avons étudié plus avant la migration des mineurs entre les pools miniers. J'ai développé *une méthodologie pour extraire les flux de paiement des récompenses des transactions de Coinbase vers les mineurs individuels*. Ensuite, j'ai défini trois mesures principales de flux de paiement : les nouveaux et les abandons, les entrées et les sorties, et le cross-pooling pour comprendre les schémas de mobilité des mineurs. J'ai montré que la méthode pouvait extraire des pools miniers basés sur des modèles de flux de paiement différents. À l'aide de différents prototypes de visualisation, *nous avons pu étudier et proposer des explications sur la motivation des mineurs à rejoindre, quitter ou se déplacer entre les pools miniers.* Nous avons constaté par exemple que les mineurs se dirigent toujours vers le pool qui offre des frais de transaction moins élevés et un revenu plus régulier aux mineurs. Nous avons également constaté que les mineurs utilisent le cross-pooling pour diversifier leur risque entre PPS (sans risque) et PPLNS (risqués).

Nous avons découvert que la compétition entre les pools de minage était en constante évolution et qu'elle devait faire l'objet d'un suivi constant sur une longue période. De plus, nous avons montré que le prix du marché du bitcoin et les événements de division par deux de la rémunération (halving days) avaient un impact sur la décision des mineurs de continuer à miner, et donc sur la durabilité de cette activité. J'ai développé *un outil d'analyse visuelle appelé* MiningVis *qui permet à des analystes de mettre en relation plusieurs facteurs liés aux pools de minage et aux comportements des mineurs*. Mon collaborateur économiste a utilisé cet outil pour étudier et développer un modèle permettant d'expliquer l'économie du minage. Nous avons également réalisé *une étude en ligne auprès de huit mineurs de bitcoins* et nous avons constaté qu'ils étaient plus intéressés par les informations en temps réel que par les données historiques à long terme. Les résultats de notre étude avec les utilisateurs ont confirmé nos choix de conception pour l'outil. Il est à noter que celui-ci peut également être étendu à d'autres crypto-monnaies basées sur le même protocole de *preuve de travail*.

PUBLICATIONS

Journal articles

1. Natkamon Tovanich, Nicolas Heulot, Jean-Daniel Fekete, and Petra Isenberg. "Visualization of Blockchain Data: A Systematic Review." In: *IEEE Transactions on Visualization and Computer Graphics* 27.7 (2019), pp. 3135–3152.
   DOI: 10.1109/TVCG.2019.2963018. HAL: hal-02426339

2. Natkamon Tovanich, Nicolas Soulié, Jean-Daniel Fekete, and Petra Isenberg. "MiningVis: Visual Analytics of the Bitcoin Mining Economy." In: *IEEE Transactions on Visualization and Computer Graphics* (2021).
   DOI: 10.1109/TVCG.2021.3114821. HAL: hal-03348145

Conference proceedings

3. Natkamon Tovanich, Nicolas Soulié, and Petra Isenberg. "Visual Analytics of Bitcoin Mining Pool Evolution: On the Road Toward Stability?" In: *11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. 2021.
   DOI: 10.1109/NTMS49979.2021.9432675. HAL: hal-02902465v2

4. Natkamon Tovanich, Nicolas Soulié, Nicolas Heulot and Petra Isenberg. "An Empirical Analysis of Pool Hopping Behavior in the Bitcoin Blockchain." In: *IEEE International Conference on Blockchain and Cryptocurrency*. 2021.
   DOI: 10.1109/ICBC51069.2021.9461118. HAL: hal-03163006v3

Poster communications

5. Natkamon Tovanich, Nicolas Heulot, Jean-Daniel Fekete, and Petra Isenberg. "A Systematic Review of Online Bitcoin Visualizations." In: *Posters of the European Conference on Visualization (EuroVis)*. 2019.
   DOI: 10.2312/eurp.20191148. HAL: hal-02155171

6. Natkamon Tovanich, Nicolas Soulié, Nicolas Heulot, and Petra Isenberg. "Interactive Demo: Visualization for Bitcoin Mining Pools Analysis." In: *Demo of the IEEE International Conference on Blockchain and Cryptocurrency*. 2021.
   DOI: 10.1109/ICBC51069.2021.9461124. HAL: hal-03163005

Other publications outside the scope of this thesis

7. Natkamon Tovanich, Pierre Dragicevic, and Petra Isenberg. "Gender in 30 Years of IEEE Visualization." In: *IEEE Transactions on Visualization and Computer Graphics* (2021). DOI: 10.1109/TVCG.2021.3114787. HAL: hal-03348149

8. Natkamon Tovanich, Alexis Pister, Gaëlle Richer, Paola Valdivia, Christophe Prieur, Jean-Daniel Fekete and Petra Isenberg. "VAST 2020 Contest Challenge: GraphMatchMaker: Visual Analytics for Graph Comparison and Matching." In: *IEEE Computer Graphics and Applications* (2021). DOI: 10.1109/mcg.2021.3091955. HAL: hal-03283200

9. Natkamon Tovanich, Simone Centellegher, Nacéra Bennacer Seghouani, Joe Gladstone, Sandra Matz and Bruno Lepri. "Inferring psychological traits from spending categories and dynamic consumption patterns." In: *EPJ Data Science* 10:24 (2021). DOI: 10.1140/epjds/s13688-021-00281-y.

If you can't support me, or you can't endorse me, get out of my way,
You do your thing and I'll do mine.

— Shirley Chisholm (1924–2005)

ACKNOWLEDGMENTS

My supervisor, Petra Isenberg, is the most wonderful doctoral thesis advisor I could ever imagine. I have learned a lot from her on how to do good research and write good publications. I am truly grateful for her support, kindness, and empathy, both academically and personally, over three years of my doctoral thesis. I wish that all doctoral thesis advisors would dedicate more time to their students and support them to become good researchers for generations to come. Nicolas Heulot helped set up the BlockSci server and has always been patient with my illiteracy on the back-end side. I would have had a much worse time solving technical issues without his aids. Jean-Daniel has always been an inspiration for me in his wisdom, grey hair, and dedication to persistently researching while solving everyone's issues in his laboratory.

My doctoral thesis is centered on the Bitcoin mining pool economy. I could not get the work done properly without the help of my economist collaborator, Nicolas Soulié. I feel thankful that he spent a lot of time with me working on this project and provided me with the requirements, feedback, and interpretation I needed on my visualization prototypes. I think we did get a good job done together. I would like to also thank my participants in the user study for spending time with the MiningVis tool and sharing their experience on Bitcoin mining.

During my Ph.D. study, I also had an opportunity to work with many collaborators on the VAST Challenge 2020. I would like to praise the effort and endeavor my co-authors invested in our project: Alexis Pister, Gaëlle Richer, Paola Valdivia, and Christophe Prieur. I also appreciate their patience with my commanding personality. Pierre Dragicevic had never run out of insights and ideas to our discussions when we worked on analyzing gender data in the IEEE VIS publication project. I have learned a lot about communicating statistics and HCI research stuffs from him and wishes to spend more time with him. Most importantly, I will miss the time when I played with his and Yvonne's son, Léo, and watched him grow up over the years.

I cannot say thank you enough to Teppakorn Thanuthanad for his responses to any issues, small or large, that I came up with in my source code. He is truly my reliable mentor when it comes to web development. Catherine Plaisant was very kind and kept asking about the progress of my work. She also kindly read the first draft of my

MiningVis paper and gave me many insightful comments that helped me to improve my IEEE VIS article. Moreover, I would like to express my gratitude toward my friends who spent time reading the first drafts of my manuscripts and gave me valuable feedback: Lucas Benmouffok, Lorenzo Candeago, Antoine Durand, Kalpana Singh, Ahmed Souissi, and Khalifa Toumi.

My Thai friends in Paris have always been good supporters as well as bringing some spice into my lonely life in Paris. I enjoyed every moment when we talked while eating Thai food, Chinese hot pot, Korean BBQ, Japanese ramen, Taiwanese bubble tea, and so on together. It has been my great pleasure to be your friend and spend some time with you: Chollada Boriboonmangsa, Chontisa Charnrit, Sathapana Chawananon, Siriwan Chongketkon, Plaifa Intavong, Sittiwan Jaiwaree, Airada Juntrawadee, Suprawee Mekrungruangkul, Athikom Oubomkul, and Wenita Srinulemak. I always enjoy their company and never feel lonely when I am with you guys. Thank you so much for letting me be a small part of your life. Additionally, I appreciate Ayothaya restaurant's owner who drove from Paris to deliver me Thai food every month during the lockdown. It was simple happiness that motivated me to continue working (after eating it all, of course).

There are thousands of moments when I feel discouraged about my work and feel like no one cares. Fortunately, I received little help from my friends in Thailand, who replied to my messages and talked to me on the video call occasionally. I want to express my gratitude toward those people, including Watsawan Harnnarong, Suwat Lorjaichon, Chanaphan Prasomwong, Pimploy Puwavicheanchai, Akraradet Sinsamerusk, Natthasasi Suma, Pongsakorn Thatsanarapan, Thipwimon Theekaporn, Thanida Thongmak, Mattawan Thongngok, and Sumana Wongrattanakul. I am wholeheartedly grateful for your company, time, and kindness in listening to my struggles throughout the years. Special thanks to Narong Sereepookkana for his tarot card prophecy that guided me to choose this doctoral thesis. He is also my guidance when I feel lost and insecure.

I am so fortunate to have learned with two great teachers from Thailand. Dr. Kitsana Waiyamai is my first data mining professor from my undergraduate study. He regularly keeps in touch with me in order to check up on the progress of my thesis and to exchange current research trends. Additionally, I have had many phone calls with my high school teacher, Kru Patsaya Sarutisut. She always cheers me up every time I feel down. Both of them gave me the courage to carry on my work amid downtimes in my life.

It has been my pleasure to join the company of doctoral students from the DOCSX association. Those fantastic people include Pascal Un, Adrien Touboul, Victor Pellegrain, Kévin Pasini, Tjark Gall, and Maria Hanini, to name a few. We have spent some time socializing on coffee breaks, lunch, and after-work beer together. In particular, Julien

Vidalie is a kind of friend in need who listened to both good and bad stories on my side and encouraged me to carry on with my thesis. My fellow Ph.D. students at Aviz lab are full of talents, and I wish them success in their thesis and future career: Tingying He, Jiayi Hong, Alaul Islam, Alexis Pister, and Lijie Yao. I am appreciated their help in proofreading my thesis at the last minute. Last but not least, I feel gratitude to administrative assistants from LISN and Inria, especially Katia Evrat, Stéphanie Druetta, and Valérie Berthou, for helping me through the fussy French administrative procedures. My immigration life in France could be much worse without their support.

I also would like to express gratitude toward developers behind data visualization libraries that I heavily used in this thesis: Altair, Vega, and D3.js. I would have a more challenging time developing my tool without examples and solutions from bl.ocks.org, Observable and StackOverflow. My thesis relies upon the BlockSci library that I used to access Bitcoin transaction data and address clustering. I feel indebted to the developer team behind the library, even though it was complicated to install firsthand. Special thanks to Malte Möser, who maintains BlockSci, and Aleš Janda (WalletExplorer), Marc Jourdan, and Francesco Zola for sharing the known entity dataset.

I am so fortunate to have a family that always supports me to pursue whatever I want. I always appreciate my parents who allow me to have the freedom to think, be myself, and follow my own dreams. I wish they would be proud of what I have achieved so far and I promise that I will try to do better every day. Finally, I dedicate this thesis in memory of my grandmother who passed away while I was writing this thesis. I wish she rests in peace and sees that I finish my doctoral degree from heaven. I will be seeing you again.

En revanche, que des contrariétés et un chagrin sans espoir aient enlevé à un homme tout goût de vivre, si le malheureux, à l'âme forte, est plus indigné de son sort qu'il n'est découragé ou abattu, s'il désire la mort et cependant conserve la vie sans l'aimer, non par inclination ni par crainte, mais par devoir, alors sa maxime a une valeur morale.

— Emmanuel Kant, Fondements de la métaphysique des mœurs (1785)

REMERCIMENTS

Durant mon séjour de 5 ans en France, de nombreuses personnes de mon entourage et moi-même avons souffert de l'inefficacité du service d'immigration de la sous-préfecture et de l'OFII. Dans ces remerciements, je souhaiterais mentionner que leur inhumanité m'a rendu la vie difficile et misérable. En rétrospective, je déplore que ce service ne soit jamais amélioré depuis mon arrivée en 2017. Je chérissais l'idée que le tissu social nous liait avec une compassion mutuelle et de l'empathie ; je leur en suis reconnaissant de m'avoir prouvé le contraire. À la longue, j'ai perdu toute dignité et tout espoir de solidarité dans cette société. Comme les autres citoyens français, je travaille dur, je paye mes impôts sur le revenu et je m'investis dans la société : Pourquoi n'aurais-je pas le droit à la reconnaissance et au salaire que je mérite ? Mon expérience m'a fait réaliser que je ne devrais pas croire en la bonté humaine et que je ne devrais pas ressentir d'empathie envers ceux qui n'essaient même pas d'effectuer leur travail correctement.

Je crois à l'effet papillon que tout ce que nous faisons affecte les autres et même la société dans son ensemble. Je suis sûr que je ne suis pas le seul à avoir eu cette mauvaise expérience de cette procédure administrative. Je ne peux m'empêcher de penser à la façon dont cette expérience négative s'est accumulée pour produire une externalité négative envers la société égoïste et insouciante. Enfin et surtout, je suis honnêtement et profondément reconnaissant envers ceux qui font leur travail et rendent ma vie plus facile en France, de l'agent d'entretien, au conducteur de bus, en passant par le personnel du service administratif de l'Université. Je suis reconnaissant de rencontrer encore ces personnes tous les jours. Ils m'ont fait sentir que la vie vaut la peine d'être vécue et me rappellent qu'une si petite gentillesse peut faire une bonne journée. J'aimerais pouvoir encore voir du bien chez tout le monde, mais en fin de compte, je m'excuse de ne pas avoir pu éviter d'envisager des choses comme la stratégie du tit-for-tat. Donc je souhaite juste que je ne deviendrais pas ce genre de personne insouciante et méchante que j'ai rencontrée.

Merci pour la leçon mais… honte à vous !

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

VA	Visual Analytics
P2P	Peer-to-Peer
dApps	Decentralized Applications
EDA	Exploratory Data Analysis
GDP	Gross Domestic Product
HDI	Human Development Index
UMAP	Uniform Manifold Approximation and Projection
UTXOs	Unspent Transaction Outputs
DDoS	Distributed Denial-of-Service
Prop	Proportional
PPLNS	Pay Per Last N Shares
DGM	Double Geometric Method
PPS	Pay Per Share
FPPS	Full Pay Per Share
SUS	System Usability Scale
ICE-T	Value-driven Visualization Evaluation

Let me remind you of the particular characteristics of all of these behavior systems that I am trying to focus on. It is that people are impinging on other people and adapting to other people. What people do affects what other people do.

— Thomas C. Schelling, Micromotives and Macrobehavior (1978)

1 INTRODUCTION

In the past years, blockchain has gained attention as a new technology to revolutionize transactions and exchanges on the internet [150]. Blockchain is a distributed peer-to-peer network storing append-only transaction data. The advantage of this technology lies in the decentralized system governed by autonomous logic rather than being controlled by any government or financial institution. Nowadays, most blockchain applications are used for cryptocurrencies such as Bitcoin [109] and Ethereum [170]. However, its use in practice is still evolving and poorly understood. In order to adopt blockchain technology in a wider set of domains, we will need to explore and analyze transaction data to better understand emergent user behavior and mechanisms in blockchain systems. As such, Visual Analytics (VA) tools can support human analysts in deriving hypotheses and models of blockchain use.

Bitcoin is the first and so far the highest valued cryptocurrency blockchain. It was proposed by a pseudonym named Satoshi Nakamoto in a seminal article, "Bitcoin: A peer-to-peer electronic cash system" [109]. The article proposed a way to prevent double-spending of digital currency transactions without requiring a trusted third party. Since Bitcoin was initially released in 2009, the activities in its blockchain increased substantially. As of September 2021, Bitcoin users create 300k transactions to exchange a total of 100k Bitcoin currency every day. Bitcoin also attracts investors who speculate to make a profit from trading in the cryptocurrency exchange market with $170B market capitalization.

In this thesis, I focus on the internal activities of Bitcoin users in the transaction data. Internal activities are the fundamental factors to understand the adoption and practical usages of Bitcoin. Cryptocurrency trading is an external activity that gives Bitcoin valuation to fiat currencies (e.g., the U.S. Dollar, the Euro) and is outside the scope of this thesis.

1.1 THE BITCOIN BLOCKCHAIN

Bitcoin records transactions in the public distributed ledger (i.e., database) called the blockchain. A blockchain consists of a chain of blocks that records all previous transactions, hence the term blockchain. Transactions are registered, validated, maintained, and distributed across the entire network of users in the peer-to-peer network. Unlike traditional financial data, Bitcoin transactions are pub-

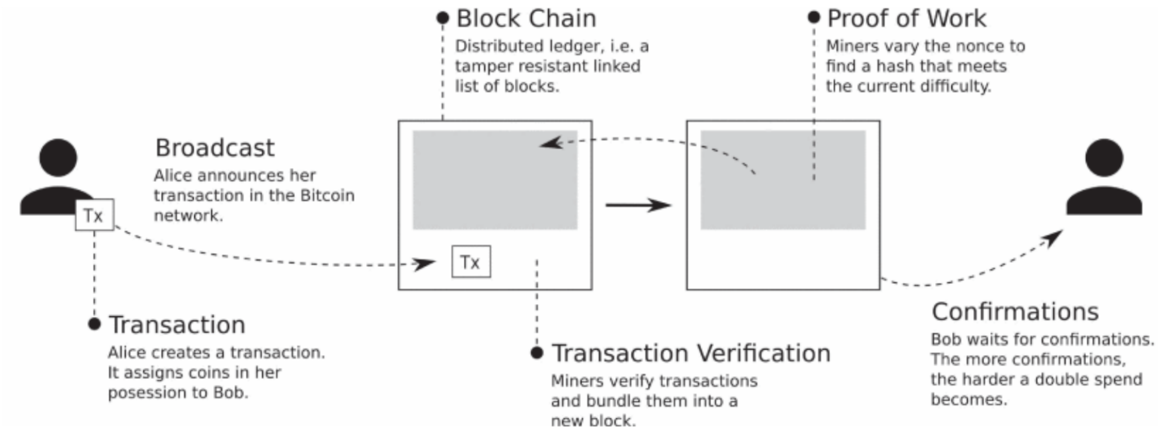

Figure 1: Building Blocks of the Bitcoin Blockchain (from Tschorsch et al. [158] © 2015 IEEE)

licly available on the blockchain but the owner of the address cannot be inferred directly from the address, hence they are referred to as pseudonymous. Chapter 2 provides an overview of Bitcoin, its components, and connections among them. I identify three groups of activities that happened on 1) financial transactions, 2) mining (or consensus protocol), and 3) peer-to-peer networks. Figure 1 summarizes the process to append transactions to the blockchain network.

Bitcoin has been used to transfer digital currency between users, payment to merchants, or as an investment. As the Bitcoin data is constantly growing (>350 GB of raw data, as of September 2021 [O10]), it offers a unique opportunity to study the evolution of the transaction data as well as the interactions of users in the network. The Bitcoin network involves diverse groups of users (e.g., individuals, enterprises, miners, and exchanges), and their activities are influenced by multiple factors from both internal (e.g., Bitcoin protocol, frauds, and cyber-attacks) and external (e.g., news and market price) historical events.

To ensure the security and sustainability of Bitcoin and blockchain technology, in general, we need to analyze transaction data to understand its practical uses and the economic incentives of different kinds of users and activities. As I review in Chapter 2, many existing works present empirical analyses of Bitcoin data. However, the results reported in those works have limitations because they 1) report on the aggregated information, 2) focus on a particular historical event or group of users, or 3) analyze the data within a limited time frame. Nonetheless, as Bitcoin activities keep evolving over time, we need methods for exploring diverse kinds of activities or monitoring the Bitcoin blockchain in the long term.

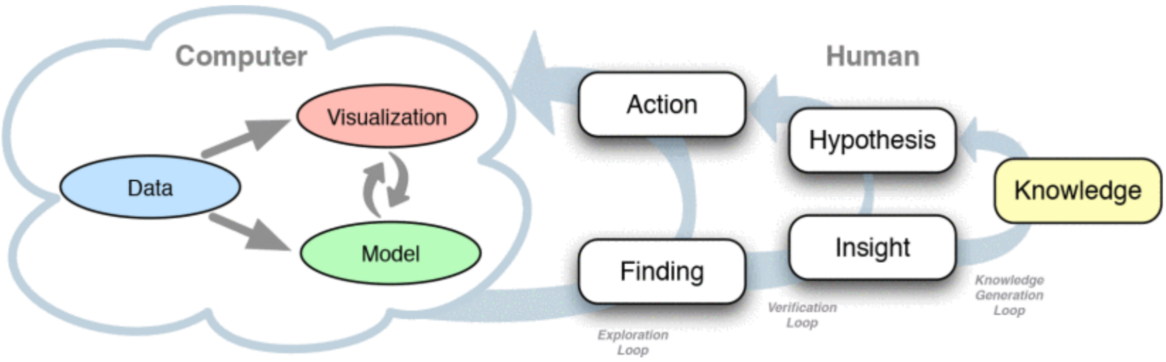

Figure 2: Knowledge generation model for visual analytics (from Sacha et al. [130] © 2014 IEEE)

1.2 VISUAL ANALYTICS

Visual analytics (referred to as "VA") is "the science of analytical reasoning supported by interactive visual interfaces" [31]. The field has worked on combining information visualization with automated data analysis methods and allows users to be involved in the analytical process to gain insights into large and complex datasets [31]. Figure 2 depicts the VA model that involves the users interacting with the visualization, and giving feedback to the data analysis model in order to generate new insights from the data [130]. The VA model also aims to study human factors by conducting design studies [139] to understand research questions, design the visualization tool, get the feedback, and evaluate the tool with real users [88].

VA is a suitable approach for analyzing the Bitcoin blockchain due to the large scale of transaction data and evolving activities of different groups of Bitcoin users. But in this case, the aggregated information cannot show different activities, user groups, or events in the blockchain data. Chapter 3 presents a comprehensive review of the past work and research challenges in blockchain data visualization. In contrast to previous works presenting Bitcoin data on the aggregated information, Bitcoin VA allows experts and researchers to 1) deeply analyze the data at different levels of aggregation and time scales, 2) detect patterns and anomalies on a particular activity or group of Bitcoin users, and 3) provide related information to describe the behavior or patterns of the activity.

1.3 THESIS STATEMENT

Blockchain is an emerging research field that involves researchers from various disciplines in computer science and economics. In computer science, the majority of the work, mainly from cryptography and computer networks, proposed new blockchain systems and protocols that give better performance and/or are more secure. Economic researchers mainly studied the incentives of users both on the consensus protocol and the market. The empirical analysis takes an important role in both fields to understand what happened in the blockchain and assess security issues to develop better blockchain solutions. VA is a promising approach to facilitate blockchain analytics in the long term but the advanced tools dedicated to Bitcoin and blockchain are still rarely found in the existing work (Chapter 3).

Bitcoin is currently the most widely-used and longest-running blockchain, and its mechanism is also applied to other cryptocurrency blockchains. For this reason, I chose this blockchain as the subject of study. In Bitcoin, there are many activities on the different layers of the blockchain (Chapter 2). In this thesis, I focused on the mining economy because it is the consensus mechanism that ensures the integrity and stability of the blockchain network. Bitcoin mining involves economic agents called miners who work on verification and appending transactions to the blockchain network. Miners expect to receive financial rewards as incentives to perform mining operations. My economist collaborator is particularly interested in this topic and would like to understand the evolution of mining activity and the decision model of miners. Working closely with the economist, I proposed a data analysis method to analyze this activity on the transaction level (Chapter 5) and developed a VA tool to support the economist's research questions (Chapter 6).

My thesis contributes to 1) a systematic review of past work and research challenges on Bitcoin and blockchain VA, 2) a data processing and algorithm to extract miners from transaction flow and detect miners' mobility, 3) an empirical analysis on pool hopping behavior, and 4) a new VA tool, called *MiningVis*, to analyze mining behavior from a long-term historical perspective.

1.4 THESIS OVERVIEW

The thesis is structured into seven chapters. The title and short descriptions of each chapter are listed as follows:

- **Chapter 2 Background on Bitcoin and Blockchain Technology** describes the background details of Bitcoin blockchain and reviews past work related to blockchain data analysis and visualization.

- **Chapter 3 Systematic Review on Blockchain Data Visualization** provides a comprehensive review of past work in blockchain data analysis and visualization. I proposed the classification scheme to group those works and listed research challenges in blockchain VA.

- **Chapter 4 The Emergence and Evolution of Bitcoin Mining Pools** presents data collection and first visualization prototypes to understand the competition of mining pools over a long history of Bitcoin mining. Based on these prototypes, my economist collaborator and I reported some findings on the factors that affect market share evolution in Bitcoin mining.

- **Chapter 5 Empirical Analysis of Bitcoin Mining Pools** presented a data analysis framework to attribute mining pools for each block, identify miners from the transaction graph, and detect miners' migration between pools. I justified that this method can extract miners from various mining pools and. I also conducted an empirical analysis to characterize the payout flow patterns and understand possible factors related to mining pool evolution and pool hopping behavior. Chapter 4 and Chapter 5 can be seen as the *Data* and *Model* part of the VA approach.

- **Chapter 6 MiningVis: Visual Analytics of the Bitcoin Mining Economy** presented a VA tool to analyze the Bitcoin mining pool. The tool is designed based on the research questions of the economist. I performed a user study to evaluate the tool with real Bitcoin miners. This chapter describes the details of the *Visualization* and *User* parts in the VA model.

- **Chapter 7 Conclusion** provides the summary of my work and outlook at the possible future work on Bitcoin and blockchain VA.

2 BACKGROUND ON BITCOIN AND BLOCKCHAIN TECHNOLOGY

Blockchain was introduced in the early 1990s as a theoretical system to store a time-stamped digital document that cannot be modified [10, 56]. The articles proposed data structures and algorithms to store non-modifiable data and maintain trust in decentralized systems without centralized control. Bitcoin was initially released in 2009 and considered the first open-source implementation of the blockchain concept for digital payment [109]. In 2015, Ethereum was released as a blockchain that implements smart contract functionality [170]. A smart contract is a piece of computer code guaranteed to run in the same way on all peers. It has been used to build Decentralized Applications (dApps) that run on the Peer-to-Peer (P2P) network.

In this chapter, I describe the background of blockchain technology and focus on Bitcoin blockchain mechanisms. I also summarize literature reviews to understand research challenges in blockchain and look at dedicated surveys on blockchain data analysis and visualization domains.

2.1 TYPES OF BLOCKCHAIN

Blockchains can be categorized into three types: public blockchains, consortium blockchains, and private blockchains [29, 177].

- **Public blockchains** are open blockchains in which any participant can read, write, and submit transactions to the ledger. Any participant can join the consensus process to determine whether to add blocks and transactions to the ledger. Public blockchains are suitable for applications that are open for everyone and need fully decentralized systems. Bitcoin and Ethereum are well-known examples of this type of blockchain.

- **Consortium blockchains** are semi-private blockchains that restrict the consensus process to the selected group of participants that are trusted by the system. This reduces the time to verify transactions and blocks but also makes the systems partially centralized to selected nodes. The overseeing group of organizations grants permission to operate a node on a consortium blockchain.

- **Private blockchains** are fully controlled by an organization that determines the consensus of the blockchain ledger. The private

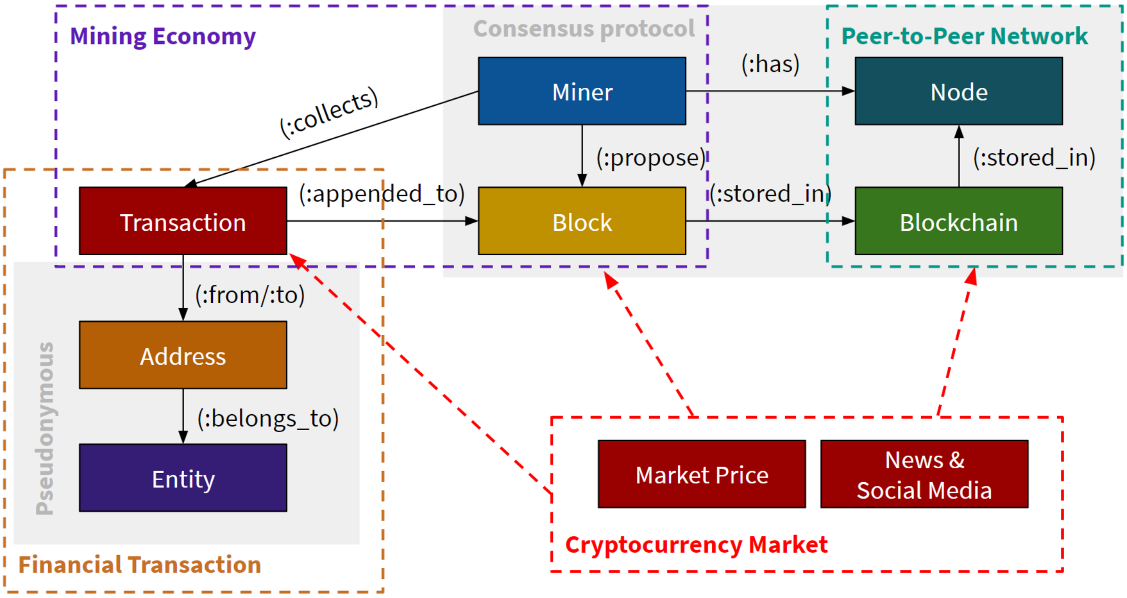

Figure 3: An overview of Bitcoin blockchain components

blockchain owner has the authority to allow or restrict the read permission to participants. Private blockchains are centralized systems, similar to database systems, and usually suitable for applications that require high trust and privacy.

In this thesis, I focus on a public blockchain, in particular Bitcoin, as the transaction data of this kind of blockchain is publicly available. Some of the mechanisms described in the next section can differ for consortium and private blockchains, where, for example, the consensus is determined by selected nodes that can be trusted. Therefore past records could theoretically be tampered with.

2.2 HOW THE BITCOIN BLOCKCHAIN WORKS

Since the Bitcoin blockchain is currently the most well-known and widely-used public blockchain in the public domain, I will explain the mechanism behind the Bitcoin system as many of its concepts also similarly apply to other blockchains. I refer readers who are interested in technical details to the original Bitcoin paper [109] and books [5]. Technical details of blockchain technology, in general, can be found in two articles [12, 158].

Bitcoin consists of multiple components, as depicted in Figure 3. The arrow indicates interaction between two components. The components can be grouped into three different layers: 1) financial transactions, 2) mining economy, and 3) Peer-to-Peer (P2P) network. The cryptocurrency market is an external activity of the Bitcoin network but indirectly impacts activities in the Bitcoin network.

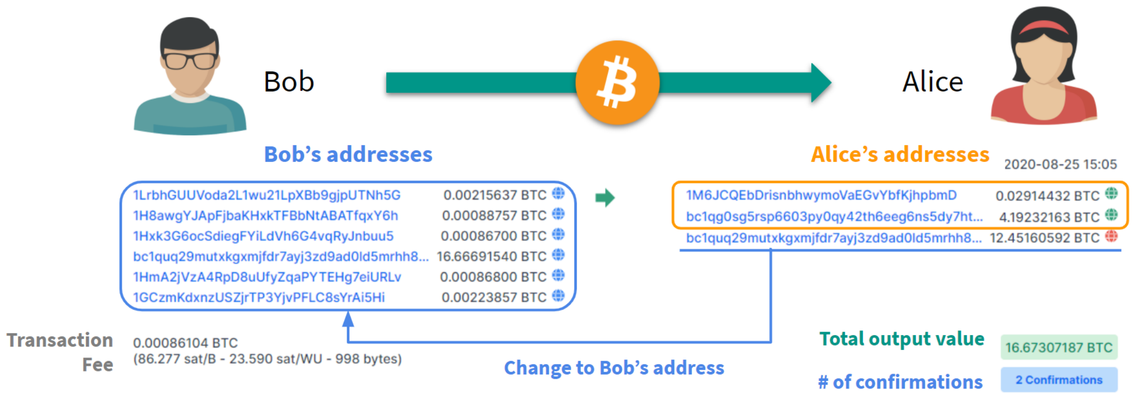

Figure 4: An example of a Bitcoin transaction. In this example, Bob spent his bitcoin stored in different addresses to Alice and kept the changes back to one of his addresses. Transaction fees were paid to the miner as the difference between the total input and the total output. (The transaction was retrieved from Blockchain.info. Icons were created by Muhammad Haq under CC BY 3.0 license.)

2.2.1 *Financial Transaction*

The *transaction* is the most granular data of the Bitcoin blockchain. Each transaction records the bitcoin value transfer from the input address(es) to the output address(es). Bitcoin uses cryptographic proof to verify the ownership of transactions—which explains why it is called a cryptocurrency. An *address* is represented as a long string with cryptographic properties that can be signed by its owner. Figure 4 shows an example of a Bitcoin transaction. All the value of input addresses is sent to the output addresses. The owner can send the change back to any of his or her addresses if he or she wants to transfer less than the total value of inputs. The owner also pays *transaction fees* as the difference between input and output values.

> Apart from financial transactions, blockchains can store any kind of transaction rather than financial transactions. For example, a smart contract is a kind of transaction with computer code that can be executed in a peer-to-peer network. It has been implemented in many public blockchains such as Ethereum, EOS.IO, and Tezos.

Transactions are publicly available on the blockchain but the owner of the address cannot be inferred directly from the address, hence they are referred to as *pseudonymous*. A common practice is that owners should regularly change addresses to hide their identities. Yet, we can still trace the activities of entities from address clustering heuristics. Multiple input address is a common heuristic that assumes all input addresses of the transaction belong to the same entity [123].

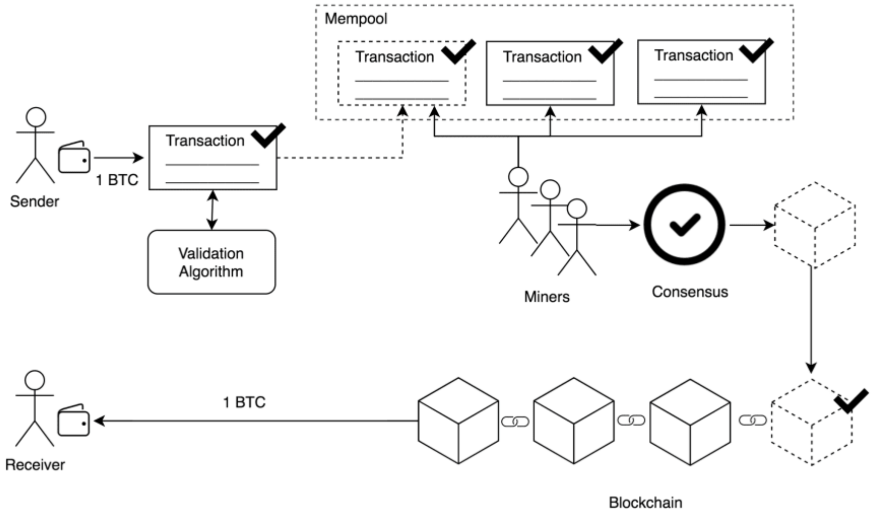

Figure 5: Bitcoin mining process (from Guilherme de Freitas Castro [45])

The surveys of clustering heuristics can be found in four articles [16, 81, 93, 167]. Many blockchain analytics companies maintain the list of addresses that belong to the known entity but keep the dataset for private use. WalletExplorer.com is the rare public source that provides this kind of dataset, but the list of entities is not updated from 2016 [71].

2.2.2 *Mining Economy*

As Bitcoin and many other cryptocurrencies operate on decentralized networks, the *double-spending problem* is a possible threat to the blockchains. Users can fool the network that a unit of digital money has never been spent before while sending other transactions which contain the exact same value to different addresses simultaneously. To solve this problem, Satoshi Nakamoto applied the *proof-of-work* protocol to validate Bitcoin transactions and append them to the blockchain [109]. This mechanism—commonly refers as *mining*—is critical to the integrity and trustworthiness of the blockchain data.

In Bitcoin mining, transactions are validated and appended into a block by a pool of people called *miners*. Figure 5 explains the mining process to add transactions to the blockchain. Miners participate in collecting pending transactions in the blockchain network, validating them, and collecting them into a block. At the same time, miners perform a *proof-of-work* consensus protocol that involves solving the computational-intensive puzzle to obtain the right to append the new block to the blockchain ledger. The more computational power a miner has, the more chance that they successfully mine a block. The block will be broadcast to the network and permanently appended to the blockchain.

The miner who successfully proposed a new block can then reclaim a coinbase transaction that includes newly generated bitcoin values (*block reward*) and *transaction fees* from every transaction in a block. The block reward is initially set to 50 bitcoin for each block mined and halved every 210,000 blocks (around 4 years) to control the supply of Bitcoin circulating in the system. The *difficulty* of the puzzle is decided by the total computation power in the blockchain network called *hash rate*. The difficulty is adjusted for every 2,016 blocks (around 2 weeks) to reach the desired rate of adding a new block every 10 minutes. The more computational power in the Bitcoin network, the more difficult it is to mine a new block.

Due to the rapid growth of the total hash rate in the network, nowadays, individual miners are hardly expected to receive a mining reward in the short term. They also need to bear the cost of purchasing specific hardware for mining and electricity costs. In practice, mines, thus, pool their computational resources to mining pools to receive a more stable and predictable income. The fierce competition of Bitcoin mining also raises the carbon footprint concern of this activity because tremendous amounts of electricity are consumed to run mining hardware [14, 87, 146].

> Other kinds of blockchain can implement other consensus protocols (i.e., *proof-of-x*) to decide who can obtain the right to propose a new block to the network. For example, proof-of-stake protocol decides who has the right to propose a new block based on the coins owned by miners. Other consensus protocols can be found in dedicated articles on this topic [12, 174].

2.2.3 *Peer-to-Peer Network*

Once the new block is created, it will be broadcast to all nodes in the network. Nodes in the Bitcoin blockchain are connected in the decentralized peer-to-peer network as depicted in Figure 6. Each node stores its own version of blockchain data and broadcasts their blockchain data to its closest neighbors. In this way, the blockchain data will propagate through the entire network via a gossip protocol.

When each node receives the new block, it will verify all transactions in the block to check that they have not been spent before in the previous blocks. In other words, they check the absence of double-spending. If the block is validated, the node will then append this block to the chain. Figure 7 illustrates the blockchain where each square represents a block. The longest chain from the *genesis block* (block 0, in green) is considered the main chain. If there are multiple

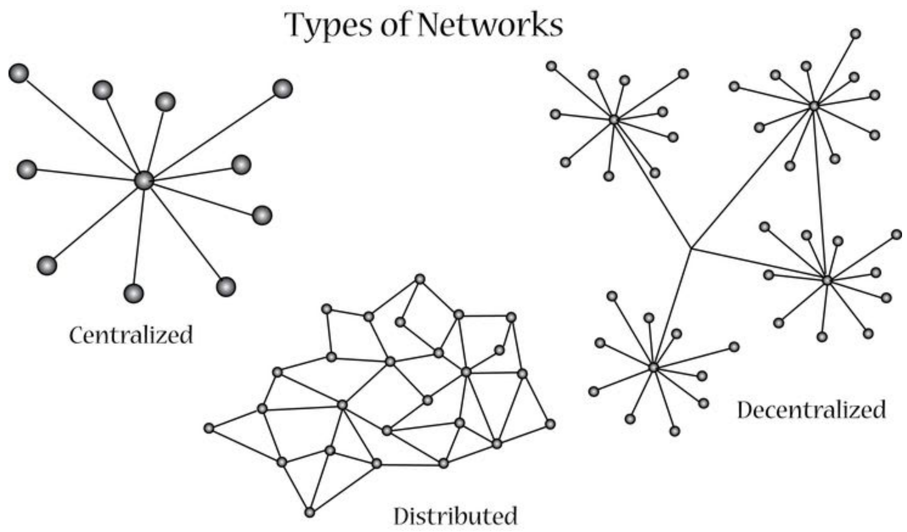

Figure 6: Comparison of different network types: centralized, distributed, and decentralized networks. Bitcoin nodes are connected in the decentralized network. (from Guilherme de Freitas Castro [45])

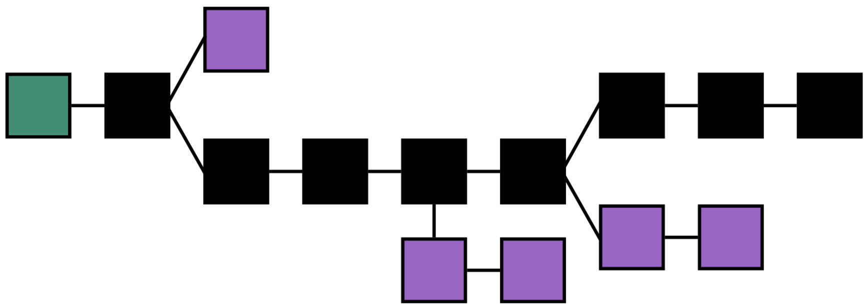

Figure 7: The illustration of blockchain. The main chain (black) consists of the longest series of blocks from the genesis block (green) to the latest block. Orphan blocks (purple) exist outside the main chain. (Theymos, CC BY 3.0)

alternative chains to extend the chain, nodes will choose a new block on the main chain to maximize the chance that their blockchain is synchronized with the peers. Other blocks outside the main chain are considered orphan blocks (in purple) which will be removed from the blockchain.

2.2.4 *Security Attacks in the Bitcoin Blockchain*

Because Bitcoin is a decentralized network, it is vulnerable to cyber attacks from criminals or dishonest users that exploit the consensus protocol to manipulate or disfunction the network. The attacks can happen in multiple layers in the Bitcoin network: transactions, mining, or the peer-to-peer network. Some malicious users may try to double-spend their bitcoin (e.g., Finney attack and Brute force attack).

Bitcoin money can be stolen from a security breach as it happened to the Mt.Gox exchange company. In the consensus layer, if a few miners control more than 50% of the total mining power, they can perform 51% majority attacks to alter the record of transactions [6, 134, 175]. The peer-to-peer network is also vulnerable to network attacks (e.g., Distributed Denial-of-Service (DDoS) and Sybil attacks). I refer readers to the dedicated survey on this issue for more detail [30].

2.2.5 *Cryptocurrency Market*

Apart from internal components in the Bitcoin blockchain, a cryptocurrency exchange market provides platforms to exchange Bitcoin value to fiat currencies (e.g., US Dollar and Euro) and therefore determine the market price of Bitcoin. The exponential growth of Bitcoin value in recent years is partially explained by the spread of narratives from news and social media to a wider public [143, Chapter 1]. This growth also raises the question of the true value of Bitcoin and the price inflation [149]. When analyzing activities in Bitcoin, it is important to consider these external elements as contextual information because they have some potential influences on internal activities in the Bitcoin blockchain.

2.3 OVERVIEW OF BITCOIN AND BLOCKCHAIN RESEARCH

Blockchain is a vast research field involving algorithms, cryptography, formal software verification, database systems, computer security, system architecture, data security, and economics. To understand the scope of this research domain, I describe previous surveys that related to the blockchain. I started with previous surveys on blockchain technology to identify possible research challenges. After that, I looked at blockchain analysis work to extract research questions and analysis methods that researchers have explored. As this thesis focuses on Visualization and Visual Analytics (VA), I identified the work that reviews visualizations and tools to explore blockchain data.

2.3.1 *Reviews on blockchain research challenges*

I found a total of 15 existing literature reviews on blockchain research. Five articles described the research challenges of blockchain in general [12, 48, 97, 103, 177]. Two articles focused on Bitcoin and cryptocurrency blockchains [23, 158]. Four articles provide bibliographic analysis [55, 100, 159, 176]. The other four surveys were dedicated to the security and privacy issues [30, 49, 81, 167]. I found four main research challenges repetitively identified in those surveys.

SYSTEM PERFORMANCE concerns the scalability and availability of the blockchain network when the size and the volume of transactions grow [48]. Scalability problems occur from the growing size of the blockchain, the block size limitation, and the time interval between blocks. Availability concerns the increased volume of transactions on throughput (i.e., how many transactions the blockchain can process per second) and latency (i.e., waiting time for adding a block to the blockchain).

SECURITY ISSUES are critical to blockchain systems as transactions need to be stored correctly in the peer-to-peer network. Security problems involve double-spending, mining-related attacks, client-side security threats, and network attacks [30]. Mining attacks were listed directly as a challenge in several surveys as mining is the consensus protocol of many blockchain networks. For example, 51% of computational power (or even less than that) can alter transaction history in the blockchain. Miners perform selfish mining by withholding validated blocks to gain more profits but waste the overall resources.

PRIVACY AND ANONYMITY issues are related to the fact that the identity of users should be anonymous, and their personal data should be protected. Users are pseudonymous in blockchain as they use public addresses to conceal their true identities. However, address clustering heuristics can group addresses that may belong to the same user [30]. Mixing services can tackle privacy and anonymity by hiding the trace of users' transactions [158].

LAWS AND REGULATIONS have been actively developed by many governments to regulate cryptocurrencies and blockchains in recent years. Cryptocurrencies are the main focus of the current rules to prevent money laundering and impose taxation on blockchains' financial activities. Some countries even made cryptocurrency tradings illegal as they threatened the authority of central banks [103]. As the new blockchain applications emerged, especially smart contracts, the regulation guidelines are not yet available, which may delay adopting the blockchain technology [12].

Three dedicated surveys highlighted the importance of blockchain analysis in practice. Bonneau et al. [23] argued that current work did not provide adequate tools to assess blockchain in practice. Tschorsch and Scheuermann [158] described alternative approaches to solve the technical challenges above, but the consequence in practice remains unclear. Merediz-Solà and Bariviera [100] suggested that interdisciplinary work on inefficiencies in the Bitcoin market and in-depth behavioral analysis were missing in the current work. This thesis proposes both the in-depth analysis result and the tool for studying mining behavior in the Bitcoin blockchain to address these concerns.

2.3.2 *Reviews on blockchain analysis*

Blockchain analysis work aims to analyze and detect interesting behaviors and possible vulnerabilities in the current blockchain solutions. The results from this kind of work suggest evidence and solutions to those problems and challenges described in the previous section. Researchers applied data analysis methods to investigate blockchain data. I summarized three main methods usually applied: game theory, graph analysis, and data mining.

GAME THEORY has been used widely to analyze the incentives of theoretical agents and assess mostly security issues in blockchain protocols. Azouvi and Hicks [7] surveyed game-theoretic models applied to cryptographic and distributed systems and described existing work that proposed models to cryptocurrency blockchains and their limitations. Liu et al. [96] reviewed existing works analyzing privacy, mining, and economic issues on blockchains. Most of the works they found heavily focused on the incentive of miners and mining pools as they were the leading players in the consensus process.

GRAPH ANALYSIS is a primary technique to analyze the interconnection of transactions, addresses, or entities as blockchain data are represented in the graph data structure. Ankora et al. [2] provided extensive technical backgrounds on graph analysis in multiple blockchains, including Bitcoin, Monero, Zcash Ethereum, Ripple, and Iota. Wu et al. [171] summarized the past work into three steps: network modeling, network profiling, and network-based detection. First, cryptocurrency transactions were represented as graph data *(network modeling)*. Then, researchers can extract features such as clustering coefficient, centrality, and network motif from the graph representation *(network profiling)*. Finally, researchers can use those features for many analysis tasks, including entity recognition, transaction pattern recognition, illicit activity detection, and transaction tracking *(network-based detection)*. Interestingly, the authors reported that current research did not explore much on dynamic networks, mixing services, and early warning of misbehavior.

DATA MINING TECHNIQUES allow researchers to discover patterns from large blockchain transaction data. Liu et al. [93] described data preparation techniques in cryptocurrency transaction data and summarized previous works into three research questions: 1) traceability and linkability issues, 2) collective user behavior, and 3) individual user behavior. In the survey, data mining techniques have been used to 1) predict entity types (e.g., exchange, wallet, mining), 2) predict fraudulent entities, 3) forecast cryptocurrency price, and 4) anomaly detection. Li et al. presented a dedicated survey to anomaly detection methods. They summarized four main techniques (i.e., statistical, machine learning, deep learning, and graph learning) and discussed the advantages and drawbacks of each technique.

As these techniques are commonly used to analyze the blockchain data, research articles usually reported the analysis results at the aggregated level or limited study time interval. Visual Analytics (VA) framework combines analysis methods and visualization that allow analysts to adjust model parameters and analyze the result in different levels of detail. However, blockchain VA tools were still rarely found and included in the current surveys.

2.3.3 *Reviews on blockchain visualization tools*

Blockchain tools facilitate the exploration, analysis, and monitoring of activities in the blockchain. In contrast to analysis literature, users can interact with tools to filter the data they want to investigate, drill down, roll up, or jump into the related information. I found three existing surveys dedicated to this topic.

- Balaskas and Franqueira [8] examined 13 tools for the Bitcoin blockchain available on the internet. They proposed a taxonomy based on analysis themes: analysis of entity relationships, metadata, money flows, user behavior, transaction fee, and market/wallets. The found tools are mainly able to track and monitor cryptocurrency values, and therefore help detect fraudulent transactions.

- Bartoletti et al. [9] surveyed 15 Bitcoin and cryptocurrency tools found in academic articles and websites. The tools in their survey were classified based on analysis goals: anonymity, market analytics, cybercrime, metadata, and transaction fees. For each analysis goal, the authors further specified the kind of blockchain-related data used in the tools, such as transaction graphs, address tags, IP addresses, mining pools, exchange rates, and lists of DDoS attacks.

- Sundara et al. [148] reviewed 8 Bitcoin tools available on the internet and briefly described visual representations and implementation techniques. Most tools in their survey performed real-time monitoring for Bitcoin transactions. Nonetheless, the authors neither performed an exhaustive search nor proposed a method to classify the tools they found.

The first two surveys proposed a classification of blockchain tools based on analysis tasks. The last article is the only previous work that surveyed visualization tools. All of the surveys include only Bitcoin blockchain tools which are available at their publication date. Many aspects of the surveys on this topic are still missing, including tools targeted at different types of blockchain, blockchain data, and visualization.

2.4 CONCLUSION

This chapter describes three main types of blockchain and provides a detailed explanation of the mechanism behind the Bitcoin blockchain. I divided the elements of Bitcoin into three layers: 1) financial transactions, 2) mining economy, and 3) the peer-to-peer network. To understand the scope of blockchain research, I summarized the research challenges in previous surveys on Bitcoin and blockchains. I then focused on data analysis and visualization, which is the topic of the thesis. Visual Analytics (VA) is a promising method to apply for Blockchain analysis, but the comprehensive survey was still missing. In the next chapter, I extend the survey to include more tools in other blockchains and propose a new classification scheme to consider data, analysis, visualization, and users.

3

SYSTEMATIC REVIEW OF BLOCKCHAIN DATA VISUALIZATION

To assess the current state-of-the-art in blockchain Visual Analytics (VA), I conducted a comprehensive review from visualization articles, Exploratory Data Analysis (EDA) articles, and online sources. I systematically assessed the motivations and characteristics of each source. Then, I defined a classification scheme to group visualizations based on five aspects: target blockchains, blockchain data, task domains, target users, and visualization types. In the end, I summarized the start-of-the-art and present research challenges in blockchain VA that would benefit from future research.

> This chapter is an updated version of my original article published at *IEEE Transactions on Visualizationand Computer Graphics (TVCG)* [153]. The article was led by myself and co-authored with my supervisors: Nicolas Heulot, Jean-Daniel Fekete, and Petra Isenberg.

3.1 DATA COLLECTION

I collected articles and online sources from the internet using specific keywords and manually filtered sources relevant to blockchain visualization. I worked on searching blockchain visualization sources in April 2019 and constantly updated the sources after publishing the systematic review article.

3.1.1 *Identifying search idioms*

I used a combination of blockchain-related terms with visualization-related terms to retrieve relevant sources for blockchain visualization. I also added analysis-related terms because these sources often used blockchain visualization to report empirical findings.

- **Blockchain-related terms:** I chose four keywords that commonly refer to blockchain technology: "blockchain," "bitcoin," "cryptocurrency," and "ethereum."
- **Visualization-related terms:** To narrow down the search result to tools related to visualization techniques for blockchains, I used the character sequence "visual" to cover keywords such as "visualization," "visual analytics," "visualizing," etc.

- **Analysis-related terms:** I used the character sequences "data analy" and "graph" to retrieve sources that did not specifically use any "visual"-related key terms.

The search idioms are relatively broad to optimize for recall rather than precision of the search results.

3.1.2 *Searching academic articles*

First, I searched relevant academic articles with the search idioms defined above in six scientific databases: 1) IEEE Xplore, 2) ACM Digital Library, 3) ScienceDirect, 4) DBLP, 5) Springer Link, and 6) Google Scholar. Then, I combined search results from these six databases and removed duplicate articles in multiple databases. Next, I screened the returned results by reading the title of returned articles one by one and selected articles that seemed to include blockchain visualizations beyond simple charts. If the title did not clearly describe an article's relevance, I also read the abstract before deciding on inclusion in our survey. Inclusion criteria were: a) the article is related to VA on any blockchain, and b) the article includes EDA on any blockchain technology and uses visualization to communicate results. I decided to include EDA articles to understand possible questions that researchers are interested in and common visualization types they used to convey their results.

3.1.3 *Searching online web-based visualization*

To collect blockchain visualization tools that are available on the internet, I search idioms from the combination of ("blockchain" OR "bitcoin" OR "cryptocurrency" OR "ethereum") AND ("analysis" OR "analytics" OR "visualization" OR "visual analytics" OR "graph" OR "chart") on Google Search and retrieved the first 100 results. I looked at each web page one by one and checked whether the web page contained blockchain visualizations. In the case of web pages that contained links to other visualization tools, I followed each link in the web page and added the link to our list. The web page had to contain interactive graphics showing raw or aggregated data stored on a blockchain to be selected. I excluded web pages that showed only market data on cryptocurrency exchanges (e.g., the current $ value of a Bitcoin).

At the end of the data collection phase, I collected a total of 110 blockchain visualization sources: 20 visualization articles (18%), 43 Exploratory Data Analysis (EDA) articles (39%), and 47 online web-based visualizations (43%). I include references to all visualization sources: visualization articles (annotated with [V#]), EDA articles ([A#]), and online sources ([O#]) in Table 2, Table 3, and Table 4, respectively.

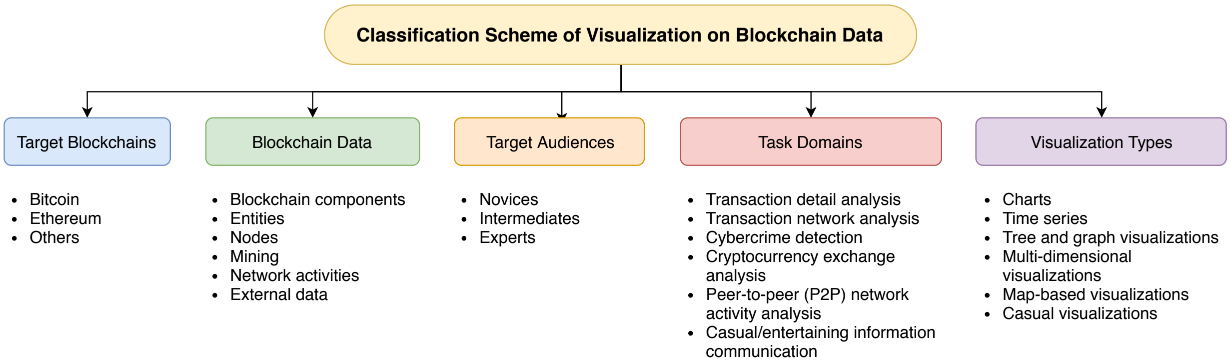

Figure 8: The classification scheme of blockchain data visualizations

3.2 CLASSIFICATION SCHEME AND METHODOLOGY

I considered many visualization-related categories such as data, tasks, types of visualizations, or end-users. After several rounds of open coding with an evolving code-set, I converged on five main aspects for delineating blockchain visualization sources: 1) target blockchains, 2) blockchain data, 3) target audiences, 4) task domains, and 5) visualization types. Figure 8 gives an overview of the classification scheme, and Table 1 summarizes the number of sources in each category by source type. Notice that the total counts and percentages do not necessarily correspond to 110 sources (100%) as sources may have included multiple types of visualizations in the classification scheme.

3.2.1 *Target Blockchains*

I extracted the blockchain that each visualization source targeted. The **Bitcoin** blockchain (69%) was the most common to be visually represented. This is not surprising because Bitcoin is the oldest running cryptocurrency blockchain and is still widely used nowadays. The **Ethereum** blockchain (28%) was the second-most common visually represented blockchain. I found that **other** blockchains (15%) were mostly cryptocurrency blockchains, such as *Namecoin*, *Litecoin*, *Dogecoin*, and *Dash*. Only two sources visualized the data on *Hyperledger*, an open-source consortium blockchain.

3.2.2 *Blockchain Data*

I categorized seven different types of blockchain data that appeared in visualization sources.

BLOCKCHAIN COMPONENTS (80%) are fundamental data types stored in public ledgers, including transactions, addresses, blocks, and smart contracts.

Table 1: The number of blockchain data visualization sources for each data source.

Source	Blockchain			Data						Audience			Task Domain						Visualization Type					
	Bitcoin	Ethereum	Others	Blockchain Components	Entities	Nodes	Mining	Network Activities	External Data Source	Novices	Intermediates	Experts	Transaction Detail	Transaction Network	Cybercrime Detection	Cryptocurrency Exchange	P2P Network Activity	Casual / Entertaining	Charts	Time Series	Tree & Graph Vis.	Multi-dimensional Vis.	Map-based Vis.	Casual Vis.
Visualization Articles	16	2	2	19	6	1	2	2	3	0	4	16	11	11	9	0	2	0	5	10	16	5	0	0
EDA Articles	26	12	6	35	12	2	7	6	15	0	0	43	11	28	11	2	11	0	38	36	21	1	2	0
Online Sources	34	17	9	34	3	10	15	24	14	16	30	1	19	5	1	9	23	12	19	22	9	2	8	12
Total	76	31	17	88	21	13	24	32	32	16	34	60	41	44	21	11	36	12	62	68	46	8	10	12
Percent	69	28	15	80	19	12	22	29	29	15	31	55	37	40	19	10	33	11	56	62	42	7	9	11

ENTITIES DATA (19%) contains the identity of users who own the addresses or accounts in the blockchain. In Bitcoin, it requires address clustering heuristics or external sources to identify entities from anonymous addresses.

NODES (12%) ensure consensus through mining to verify transactions and store them in the public ledger. The data from nodes includes the IP address of nodes and inferred locations.

MINING ACTIVITY STATISTICS (22%) can be directly calculated from the blockchain, such as, for Bitcoin, the average miner's speed to solve the proof-of-work problem (hash rate), mining difficulties over time, and the amount of reward to the successful miners.

NETWORK ACTIVITIES (29%) displayed aggregated statistics of the whole blockchain network. Network activity data usually included the number of unique addresses used, the total number of transactions recorded in a given time period, the number of transactions waiting to be confirmed (mempool), and the number of Unspent Transaction Outputs (UTXOs).

EXTERNAL DATA SOURCES (29%) convey meaningful contexts such as cryptocurrency exchange rates, online news, socio-economic data (e.g., percentage of internet users, Gross Domestic Product (GDP) per capita, or the Human Development Index (HDI)), social media information, or even Google Trends data.

3.2.3 *Task Domains*

I categorized blockchain visualization sources into six task domains to detect goals for developing, analyzing, and exposing existing tools.

Table 2: List of visualization article sources

	Blockchain			Data						Audience			Task Domain						Visualization Type					
Source	Bitcoin	Ethereum	Others	Blockchain Components	Entities	Nodes	Mining	Network Activities	External Data Source	Novices	Intermediates	Experts	Transaction Detail	Transaction Network	Cybercrime Detection	Cryptocurrency Exchange	P2P Network Activity	Casual / Entertaining	Charts	Time Series	Tree & Graph Vis.	Multi-dimensional Vis.	Map-based Vis.	Casual Vis.
[V1] Ahmed et al.	x			x								x		x	x						x			
[V2] BlockChainVis	x			x								x		x	x						x			
[V3] Bogner		x					x	x			x						x		x	x				
[V4] Chawathe	x			x								x	x									x		
[V5] Bitconeview	x			x								x	x		x						x			
[V6] Hao et al.			x	x								x	x	x					x		x			
[V7] GraphSense	x			x	x				x			x	x	x							x			
[V8] Isenberg et al.	x			x	x							x	x								x			
[V9] BitConduite	x			x	x							x	x						x	x		x		
[V10] Blockchain explorer	x			x	x							x	x		x					x	x			
[V11] McGinn et al. 2016	x			x								x		x	x					x	x			
[V12] Norvill et al.		x		x								x		x							x			
[V13] BiVA	x			x								x		x	x						x			
[V14] HyperSec			x	x		x		x			x		x		x		x			x	x			
[V15] Schretlen et al.	x			x							x		x							x				
[V16] BitVis	x			x	x							x		x	x					x	x			
[V17] Chronograph	x			x					x			x		x	x						x			
[V18] SuPoolVisor	x			x			x					x		x					x	x	x	x		
[V19] BitExTract	x			x	x				x			x	x	x						x	x	x		
[V20] SilkVisor	x			x							x		x						x	x	x	x		

Table 3: List of exploratory data analysis article sources

	Blockchain			**Data**						**Audience**			**Task Domain**						**Visualization Type**					
Source	Bitcoin	Ethereum	Others	Blockchain Components	Entities	Nodes	Mining	Network Activities	External Data Source	Novices	Intermediates	Experts	Transaction Detail	Transaction Network	Cybercrime Detection	Cryptocurrency Exchange	P2P Network Activity	Casual / Entertaining	Charts	Time Series	Tree & Graph Vis.	Multi-dimensional Vis.	Map-based Vis.	Casual Vis.
[A1] Akcora et al.	x			x								x		x	x				x	x	x			
[A2] Alqassem et al.	x			x	x							x		x					x	x				
[A3] Anoaica and Levard		x		x					x			x	x			x			x	x				
[A4] Aw et al.	x				x				x			x		x					x	x				
[A5] Badev and Chen	x							x	x			x				x	x		x	x				
[A6] De Balthasar et al.	x			x	x							x			x					x	x			
[A7] Bartoletti et al.		x		x					x			x			x				x	x				
[A8] Bartoletti and Pompianu	x							x				x					x		x	x				
[A9] Bistarelli et al. 2020		x		x								x	x						x	x				
[A10] Bistarelli et al. 2018	x			x								x		x							x			
[A11] Cao et al.	x			x			x		x			x					x		x					
[A12] Chang and Svetinovic			x	x								x		x					x	x				
[A13] Ting Chen et al.		x		x								x		x					x		x			
[A14] Weili Chen et al.		x		x								x	x	x					x	x	x			
[A15] Di Battista et al.	x			x								x		x					x	x				
[A16] Fujiwara and Islam	x			x								x		x					x	x		x		
[A17] Gao et al.		x		x					x			x	x	x	x				x	x	x			
[A18] Gebraselase et al.	x			x			x	x	x			x	x				x		x	x				
[A19] Guo et al.		x		x								x		x					x	x	x			
[A20] Huang et al.			x	x					x			x	x	x	x				x	x	x			
[A21] Jiang and Liu			x	x								x		x					x	x	x			
[A22] Kappos et al.			x	x	x							x	x		x				x	x	x			
[A23] Kondor et al.	x	x		x								x	x						x	x				
[A24] Lee et al.		x		x					x			x		x					x	x				
[A25] Li et al.	x			x			x		x			x			x		x		x	x				
[A26] Lischke and Fabian	x			x	x	x		x	x			x		x			x		x	x	x		x	
[A27] Maesa et al.	x				x							x		x					x	x				
[A28] McGinn et al. 2018	x			x								x		x	x				x		x			
[A29] Meiklejohn et al.	x			x	x							x		x			x			x	x			
[A30] Möser et al. 2013	x			x	x							x		x					x		x			
[A31] Möser et al. 2018			x	x								x		x					x	x				
[A32] Norbutas			x		x							x			x				x	x	x			
[A33] Parino et al.	x					x			x			x					x		x	x	x		x	
[A34] Phetsouvanh et al.	x			x								x	x	x	x				x		x			
[A35] Pinna et al.		x		x					x			x	x	x					x	x				
[A36] Reid and Harrigan	x			x	x							x		x	x					x	x			
[A37] Romiti et al.	x			x	x		x		x			x		x			x		x	x	x			
[A38] Ron and Shamir	x			x				x				x		x							x			
[A39] Vallarano et al.	x			x	x							x		x					x	x	x			
[A40] Wang et al.	x						x	x	x			x					x		x	x				
[A41] Wang and Liu	x						x					x					x		x	x				
[A42] Zhao et al.		x		x								x		x					x	x				
[A43] Zheng et al.		x		x			x					x	x	x					x	x				

Table 4: List of online web-based sources

Source	Blockchain			Data						Audience			Task Domain						Visualization Type					
	Bitcoin	Ethereum	Others	Blockchain Components	Entities	Nodes	Mining	Network Activities	External Data Source	Novices	Intermediates	Experts	Transaction Detail	Transaction Network	Cybercrime Detection	Cryptocurrency Exchange	P2P Network Activity	Casual / Entertaining	Charts	Time Series	Tree & Graph Vis.	Multi-dimensional Vis.	Map-based Vis.	Casual Vis.
[O1] EthStats.io		x				x	x	x			x						x		x				x	
[O2] Alethio		x		x							x		x						x		x			
[O3] BitBonkers	x			x						x								x						x
[O4] Bitcoin Globe	x					x				x							x						x	
[O5] BitcoinWisdom	x		x				x	x	x		x					x	x			x				
[O6] Etherchain		x		x				x	x		x		x			x	x		x	x				
[O7] ChainFlyer	x			x							x		x					x						x
[O8] BitForce5	x			x						x				x							x			
[O9] BitInfoCharts	x	x	x	x				x	x		x		x			x			x	x				
[O10] Blockchain.info	x	x		x	x		x	x	x		x		x			x	x		x	x	x			
[O11] Blockchair	x	x	x	x			x	x			x		x				x		x	x				
[O12] BlockSeer	x			x								x			x						x			
[O13] BTC.com	x	x	x	x	x		x	x	x		x		x				x		x	x				
[O14] Bitcoinity	x						x	x	x		x					x	x		x	x				
[O15] Coin Dance	x						x	x	x		x						x		x	x			x	
[O16] CoinDesk	x	x	x					x	x		x					x				x		x		
[O17] DailyBlockchain	x			x						x				x							x			
[O18] DashRadar			x	x		x	x	x			x		x	x			x			x	x			
[O19] Dune Analytics		x		x			x	x	x		x		x				x		x	x				
[O20] Bitcoin Big Bang	x				x						x			x							x			
[O21] Ethernodes.org		x				x		x	x		x		x				x		x				x	
[O22] Etherscan		x		x		x	x	x	x		x		x			x	x		x	x			x	
[O23] EtherView		x		x							x		x						x					
[O24] Ethviewer		x		x				x			x		x								x			
[O25] Ethplorer		x		x					x		x		x						x	x				
[O26] Plantoids	x			x						x								x						x
[O27] Gastracker.io		x		x		x		x			x		x				x		x	x				
[O28] Interaqt	x			x						x								x						x
[O29] Federal Bitcoin	x			x						x								x						x
[O30] Johoe's Mempool	x		x					x			x						x			x				
[O31] Symphony	x			x						x								x						x
[O32] OXT	x			x			x	x			x		x				x		x	x		x		
[O33] Bitcoin Visuals	x		x				x	x	x		x					x	x			x				
[O34] Hyperledger Explorer			x	x		x		x			x		x				x		x	x				
[O35] BitListen	x			x						x								x						x
[O36] BitcoinCity	x			x						x								x						x
[O37] EthStats.net		x				x	x	x			x						x		x				x	
[O38] Blockchain 3D Explorer	x			x							x			x							x			
[O39] Statoshi.info	x						x	x			x						x			x				
[O40] On Brink	x			x						x								x						x
[O41] TradeBlock	x	x		x			x	x	x		x		x			x	x		x	x				
[O42] TX Highway	x			x						x								x						x
[O43] Bitcoin Monitor	x			x						x			x							x				
[O44] Bitcoinrain	x			x				x		x								x						x
[O45] Bitcoin VR	x			x						x								x						x
[O46] Wizbit	x			x		x				x							x						x	
[O47] BitNodes	x					x					x						x			x			x	

These task domains are not mutually exclusive; in other words, a visualization source could target multiple task domains.

TRANSACTION DETAIL ANALYSIS (37%): Transaction detail analysis tools often expose basic statistics on the level of individual transactions, of blocks, and sometimes related to individual blockchain users (entities) such as individual people, exchange platforms, dark marketplaces, gambling services, or companies.

TRANSACTION NETWORK ANALYSIS (40%): A blockchain transaction network is a bipartite graph connecting addresses through transactions. Most of the sources targeting this task domain are EDA articles that analyzed transaction networks and described the structures and dynamics of blockchain transaction networks. Visualization articles and online sources allowed for interactive exploration of transaction networks based on specific events or groups of entities.

CYBERCRIME DETECTION (19%): Cybercrime is a serious threat to the use of blockchains. This task domain is particularly common for the cryptocurrency community because of the historical frequency of fraudulent activities (e.g., money laundering and illegal trading) as well as cyberattacks (e.g., denial-of-service and Sybil attacks) on most cryptocurrency blockchains.

CRYPTOCURRENCY EXCHANGES ANALYSIS (10%): Cryptocurrency exchanges are an important target domain, particularly cryptocurrency blockchains such as Bitcoin. Tools in this target domain present exchange market statistics and financial data related to different cryptocurrencies, such as the exchange rate between a cryptocurrency value and the US Dollar. In this review, I did not systematically collect all tools focused on market-related data without including some data stored on a blockchain. A comprehensive review of online cryptocurrency exchange sources can be found in my EuroVis poster [152].

P2P NETWORK ACTIVITY ANALYSIS (33%): This target domain concerns the presentation of aggregated statistics that gives an overview of activities in the P2P network, such as mining, transaction rates, transaction volume, mempool statistics, sometimes coupled with inferred geographic locations.

CASUAL/ENTERTAINING INFORMATION COMMUNICATION (11%): This kind of visualization was built on the web to attract the attention of novice audiences to blockchain technologies and engage them through casual information visualization.

3.2.4 *Visualization types*

I categorized blockchain visualization sources into six common visualization types.

CHARTS (56%) showed predominantly two or three (never four) data dimensions using basic representation types such as bar charts, pie charts, histograms, scatterplots, word clouds, and heatmaps.

TIME SERIES (62%) were the most common visualization type because timestamps are essential in blockchain data. Most commonly, time series showed the activity of a blockchain address or entity summarized across different time granularities. Time series were often presented as line plots and bar graphs with a temporal x-axis. Another time-oriented data visualization I found is tilemaps [101], a heat map with calendar divisions to encode activity statistics with one or two temporal dimensions.

TREE AND GRAPH VISUALIZATION (42%) was common to represent the blockchains' money flows and transaction networks. These representations typically showed the connection of transactions from input addresses to output addresses. Node-link diagrams were the most common technique to show the connectivity of blockchain components. A few sources used different graph visualization techniques, such as an adjacency matrix, a Circos diagram [85], or customized visualizations.

MULTI-DIMENSIONAL VISUALIZATIONS (7%) are designed for showing data of higher dimensions than basic charts, including multiple glyphs, self-organizing maps, classification trees, 3D scatterplots, spider charts, and parallel coordinates.

MAP-BASED VISUALIZATION (9%) was a common technique to display geographical information associated with the blockchain. I found point maps, density maps, choropleth maps, and 3D virtual globes among all map-based sources.

CASUAL VISUALIZATIONS (11%) were a set of non-standard, custom-made graphical representations of blockchain data as casual information visualizations [117]. These sources did not use common charts or plots as described above. Instead, they depicted basic blockchain components in unique ways to attract attention.

3.3 DETAIL ANALYSIS PER TASK DOMAIN

As task domains are an important distinguishing factor in the classification scheme, I describe patterns of sources for each task domain in greater detail and provide some representative examples.

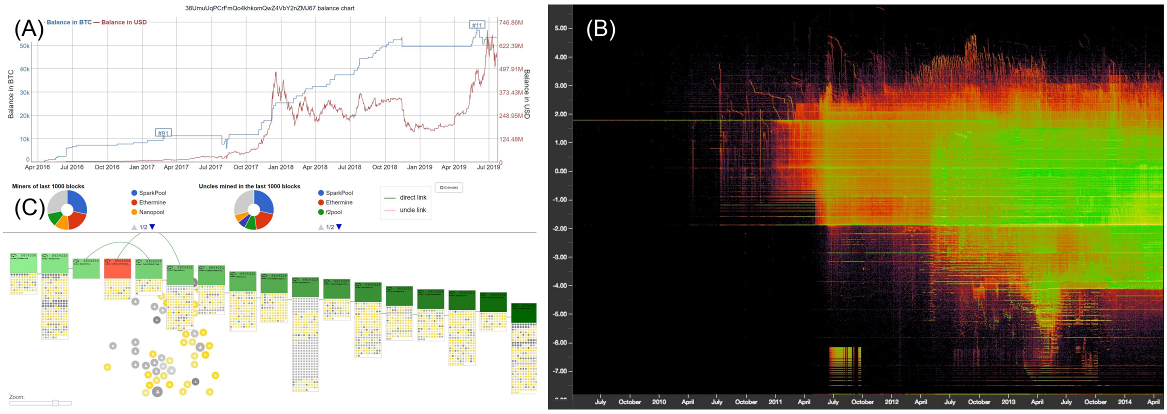

Figure 9: Representative examples of financial transactions and blocks visualizations: (A) BitInfoChart, (B) Schretlen et al. (Image from a public presentation, used with permission of Uncharted Software Inc.), and (C) Ethviewer.

3.3.1 *Transaction detail analysis*

Transaction detail analysis aims to analyze transaction patterns for individual blockchain components (i.e., transactions, addresses, and blocks) or derived entities in blockchain networks. Visualization sources in this task domain can be divided into three main groups based on the blockchain data visualized: 1) visualization of financial transactions, 2) visualization of blocks, and 3) visualization of multiple entities.

Visualization of financial transactions: Visualizations of this category allow intermediate users to search and explore the details of cryptocurrency value transactions, addresses, and blocks. Most tools in this task domain focus on representing financial transaction activities in a specific address or entity, such as the total received, sent, or the balance amount over time in the form of time series. BitInfoCharts [O9] uses line plot time series to show the balance amount of individual addresses for several cryptocurrencies, including a conversion rate to US Dollar (Figure 9 (A)). Other time series visualizations have also been used. For example, Blockchain Explorer [V10] visualizes weekly or monthly transaction volumes as a tilemap [101].

In contrast to most sources in this task domain that show aggregated statistics on transactions and addresses, Schretlen et al. [V15] proposed an interactive visualization for exploratory analysis of transaction data stored in the Bitcoin blockchain. It used a large-scale tilemap (Figure 9 (B)) to display the distribution of Bitcoin transaction values.

Visualization of blocks: Four sources displayed the content of blocks. Chawathe [V4] applied a self-organizing map to create a low-dimensional representation of transactions in a block. The self-organizing map is visualized as a hexagonal grid of wind rose plots to show the

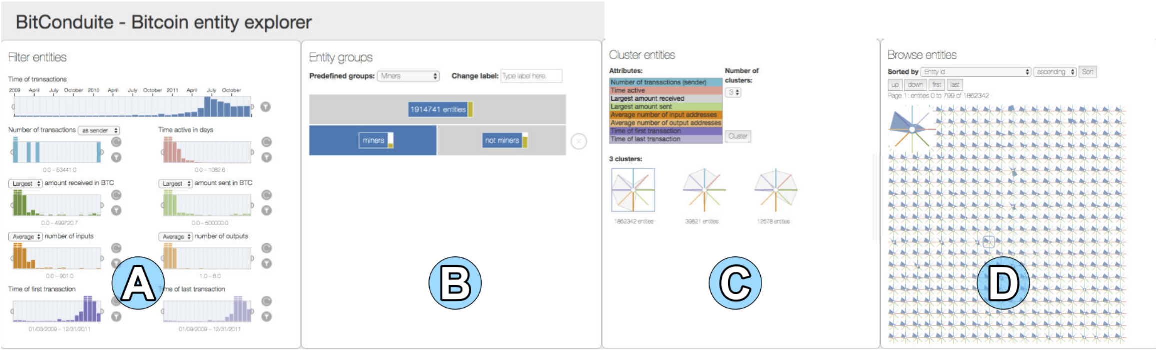

Figure 10: BitConduite is a VA tool to explore and filter entity groups in the Bitcoin blockchain. (© 2021 IEEE)

main characteristics of transaction groups in a block. Ethviewer [O24] shows the real-time transaction pool in Ethereum. The tool shows a chain of linked blocks as a node-link diagram (Figure 9 (C)). In SilkVisor [V20], blocks are illustrated as papers. The arrow connects papers to show the block order. OXT Landscapes [O32] is the only source that uses 3D scatterplots to represent attributes of blocks.

Visualization of multiple entities: Two sources presented financial information that allows experts to explore a single or a group of entities and drill down to see transaction behavior. Attributes that characterize entities were usually represented as multiattribute visualizations. BitConduite [V9] is a visual analytics tool for exploring entities in Bitcoin using multiple views (Figure 10). It allows analysts to filter groups of entities visually from value range selectors (A) and a classification tree (B). Then, the tool clusters entities with similar activity patterns. It encodes them as radar charts to represent quantitative attributes of entities, such as the number of transactions, time active, and the average number of input addresses per transaction (C–D). BitExTract [V19] is another visual analytics tool that also falls in entity visualizations, focusing on the analysis of activities among Bitcoin exchanges, including transactional volume, market share, and connectivity between exchanges.

3.3.2 *Transaction network analysis*

Transaction network analysis sources generally showed three kinds of information: 1) transaction networks, 2) the network of entities, and 3) value flow tracing the transfer of cryptocurrency values through transactions over time. These sources were always represented as tree and network visualizations. In particular, node-link diagrams were most often used to show the connectivity among blockchain components. A common technique to arrange nodes was the force-directed graph layout.

Figure 11: Representative examples of transaction network visualizations: (A) BlockchainVis, and (B) The Bitcoin Blockchain Entity Explorer.

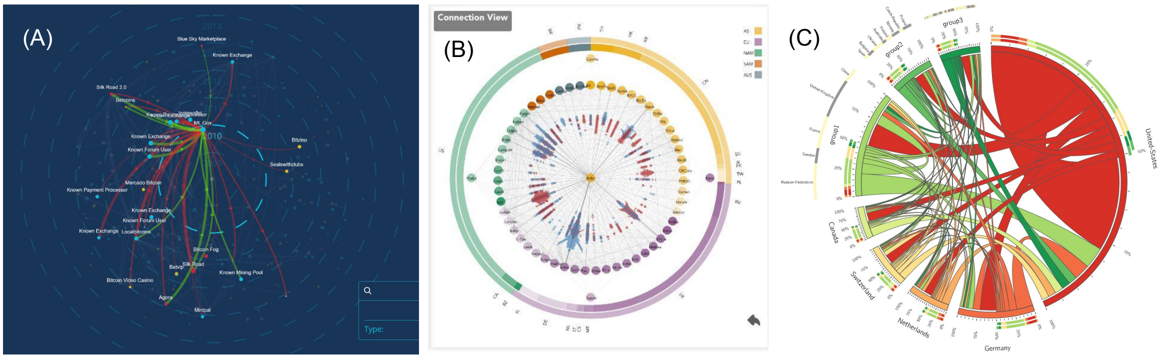

Figure 12: Representative examples of entity network visualizations: (A) The Bitcoin Big Bang, (B) The connection view in BitExTract, and (C) Circos diagram in Parino et al.

A transaction network is a directed bipartite graph connecting addresses via a transaction. There are two kinds of nodes: one type for addresses and one for transactions. Two kinds of directed edges exist in such a graph. Input edges connect input address(es) to a transaction, and output edges connect to output address(es).

Several visualization articles proposed tools to explore transaction networks based on specific events. The BlockchainVis [V2] tool displays a fully connected network of a transaction or an address entered by the user (Figure 11 (A)). McGinn et al. [V11] proposed a system to display a transaction network on a large screen on which users can pan, zoom, and hover over to get a better overview or more detail. Blockchain 3D Explorer [O38] is the only tool in this domain that visualizes a transaction network as a 3D graph. It also supports virtual reality systems for *Google Cardboard* to explore the blockchain network in an immersive way. Instead of showing a static transaction network as a node-link diagram, Bitcoin Entity Explorer [V8] is an exception in that it presents a transaction activity timeline of a chosen entity with a timeline-based squared graph layout connecting input and output addresses over time (Figure 11 (B)).

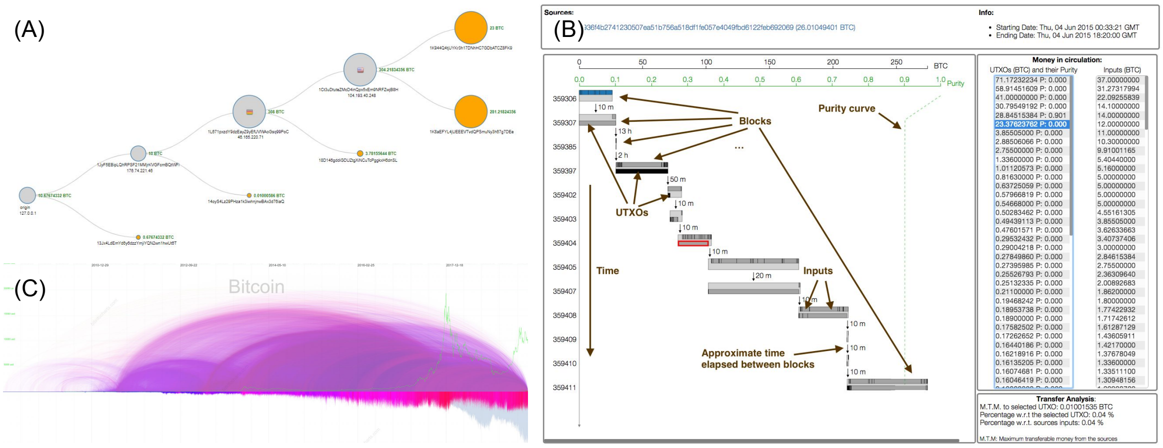

Figure 13: Representative examples of value flow visualizations: (A) Blockchain.info, (B) BitConeView (© 2015 IEEE), and (C) The Bitcoin visualization in BitInfoCharts.

A network of entities shows the connectivity between entities in the blockchain network. Nodes represent entities, and edges represent connectivity through transactions. For example, in Bitcoin, an edge represents the total amount of exchanged values between two entities and is absent if no value was exchanged.

The Bitcoin Big Bang [O20] is an online visualization presenting a network of entities as a node-link diagram connecting well-known wallets and highlighting the transaction volume between them. It adds a temporal dimension to the node-link diagram by arranging the node distance from the center based on their first appearance (Figure 12 (A)). BitExTract [V19] has a connection view that shows the relationship of exchange entities using a circular network layout to investigate the interaction of entities in the blockchain network (Figure 12 (B)). Parino et al. [A33] describe a flow network of Bitcoin transactions aggregated by country. The authors use a Circos diagram [85], also known as dependency wheel, to visualize major countries' total transactions (Figure 12 (C)).

A value flow presents traces of cryptocurrency value given a particular transaction or address of interest. Sources visualizing a value flow usually used as a tree diagram to connect the values flowing in chronological order. In this layout, a node represents a transaction or address, and an edge represents the amount of value exchanged. For example, Blockchain.info [O10] provides a tree diagram in which users can click through tree levels to follow value flow from connected input and output addresses (Figure 13 (A)). Instead of presenting the value flow as a tree structure, BitConeView [V5] provides a unique diagram showing the value flow of a seed transaction as it appears in blocks from top to bottom (Figure 13 (B)).

All of the examples above present static graphs that do not consider the timestamp of transactions. BitInfoCharts [O9] provides a unique

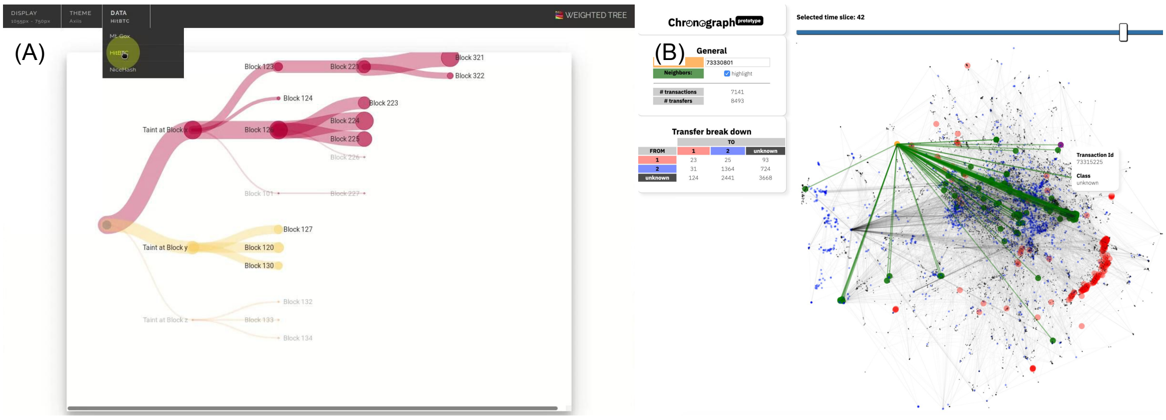

Figure 14: Representative examples of cybercrime detection visualizations: (A) Ahmed et al. and (B) Chronograph.

visualization that shows the flow of transactions over the entire history of a cryptocurrency blockchain as a kind of node-link diagram arranged using a linear layout (Figure 13 (C)). The same kind of visualization also appeared in McGinn et al. [A28] as an adjacency matrix representation.

3.3.3 *Cybercrime detection*

The cybercrime detection task domain includes tools that can detect suspicious transactions and entities or investigate cyber-attack events. Sources in this task domain also include shared characteristics with sources in the transaction detail analysis and transaction network analysis domains. Additionally, they have particular user tasks and subsequently focused features for cybercrime detection. Current blockchain visualization tools for cybercrime detection focus on two questions: 1) value flow analysis to see how cryptocurrency value is propagated and 2) transaction network analysis to see how the blockchain network reacted in light of cybercrime events.

One way to detect fraudulent financial activities in cryptocurrency blockchains is to **analyze the value flows of transactions**. Di Battista et al. [A15] and Ahmed et al. [V1] proposed transaction graph tools to analyze Bitcoin stolen money mix in the transaction flow (i.e., taint analysis). To analyze the degree of money mixing from the original transaction, Di Battista et al. [A15] introduced a purity measurement, the degree that a seed transaction is mixed with other transactions. Ahmed et al. [V1] developed an interactive visualization tool to display taint propagation as a node-link tree visualization (Figure 14 (A)). They use a First-In-First-Out (FIFO) algorithm to track the diffusion of tainted transactions in both forward (i.e., starting from a stolen coin to the following transactions) or backward (i.e., tracing

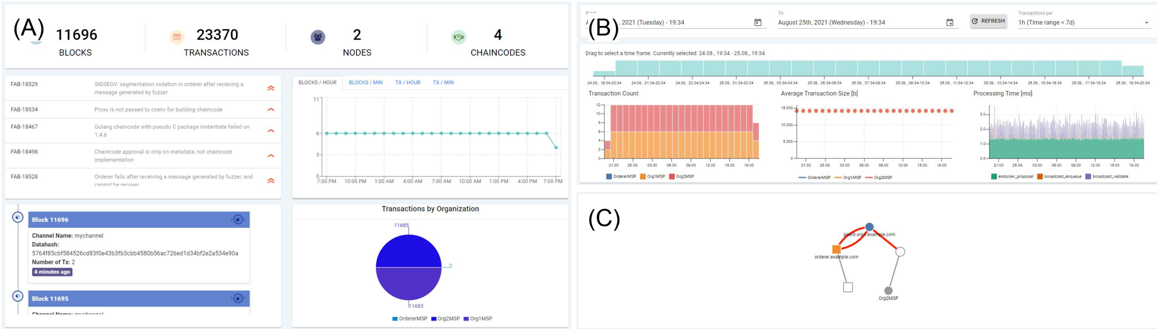

Figure 15: HyperSec is a tool to monitor security issues on the Hyperledger blockchain. (A) The dashboard shows network statistics, recent blocks, and open issues. (B) The transaction view displays transaction volume, size, and processing time. (C) The network view shows traffic between nodes.

the previous transactions until the origin of a tainted coin is found) directions.

Another kind of tool in this domain aims for **transaction network analysis on specific events or groups of entities**. For example, BitVis [V16] uses multiple graph visualizations with a filtering panel to display transaction networks for detecting abnormal and suspicious Bitcoin entities. Two articles from McGinn et al. [A28, V11] show how their tool can be used to visualize a transaction network during cybercrime attacks, including denial-of-service attacks where an attacker tries to fill up a block with spam transactions. Chronograph [V17] proposed a graph visualization tool to detect money laundering transactions from a Graph Convolutional Network (Figure 14 (B)). The positions of the nodes are set with a 2-dimensional projection using Uniform Manifold Approximation and Projection (UMAP) technique. The BlockChainVis tool [V2] is another transaction network tool that allows filtering specific parts of the transaction network during an event of interest. It has been used to analyze the WannaCry ransomware incident on May 12th, 2017 [A10]. Other EDA articles performed ad-hoc analyses of transaction networks during attacks on the Bitcoin network, including money laundry services [A6], online drug marketplaces [A32], and Bitcoin thefts [A36].

I found only one tool dedicated to cybercrime detection in the consortium blockchain. HyperSec [V14] extends the Hyperledger Explorer tool [O34] to monitor security issues on the Hyperledger network (Figure 15). The tool consists of multiple dashboards to identify blockchain components, smart contracts, log files, configuration changes, and network activities that could threaten the blockchain network. Users can detect anomalies in blockchain activities from the chart and further investigate individual transaction details.

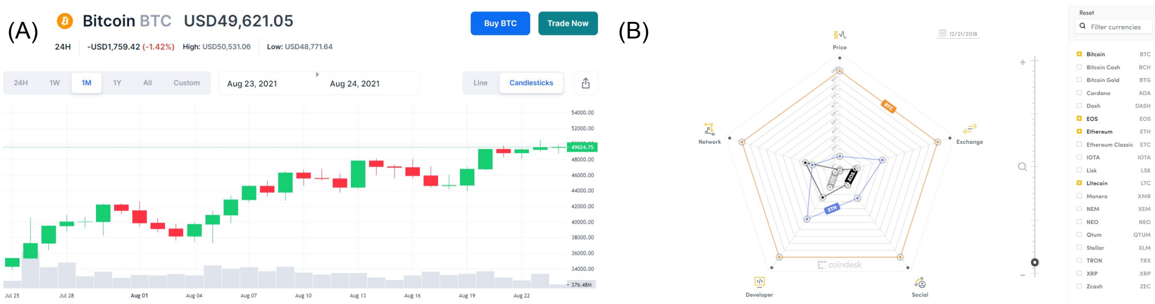

Figure 16: Representative examples of cryptocurrency exchange analysis visualizations: (A) Blockchain.info and (B) CoinDesk.

3.3.4 *Cryptocurrency exchange analysis*

Cryptocurrency exchanges convert cryptocurrency values into fiat currencies, such as US Dollars and Euros. This kind of data is not stored in the blockchain. Sources in this task domain either 1) cover the conversion of cryptocurrency value to US Dollar for blockchain components (Section 3.3.1), or 2) provide an additional view that relates information on blockchain components to market statistics, such as historical price, trading volume, and market capitalization.

The first type visualizes **conversion rates for cryptocurrency values on blockchain components**, such as individual addresses or blocks. These sources are already described in the "transaction detail analysis" task domain section (Section 3.3.1). The second type visualizes **cryptocurrency market statistics** in separate charts to provide contextual information. These sources mainly used time series to display the historical exchange rate and market volume for different time scales (i.e., hours, days, weeks, months) in addition to more detailed information on individual transactions, addresses, or blocks. For example, Blockchain.info [O10] provides a market view for various cryptocurrencies (Figure 16 (A)). CoinDesk [O16] is a unique online tool in this category that shows summarized measures of several cryptocurrencies' size and investment opportunities. It presents a spider chart to compare multiple measures related to price, exchanges, social media, developers, and the overall network size (Figure 16 (B)).

3.3.5 *P2P network activity analysis*

Blockchains are decentralized systems running with client nodes in a Peer-to-Peer (P2P) network architecture. Understanding the activities within the P2P network helps intermediate and expert users track the current status of a block due to overall activities among participants in the network. Sources in this task domain use two kinds of visualizations: 1) time series to show the aggregated statistics of the P2P

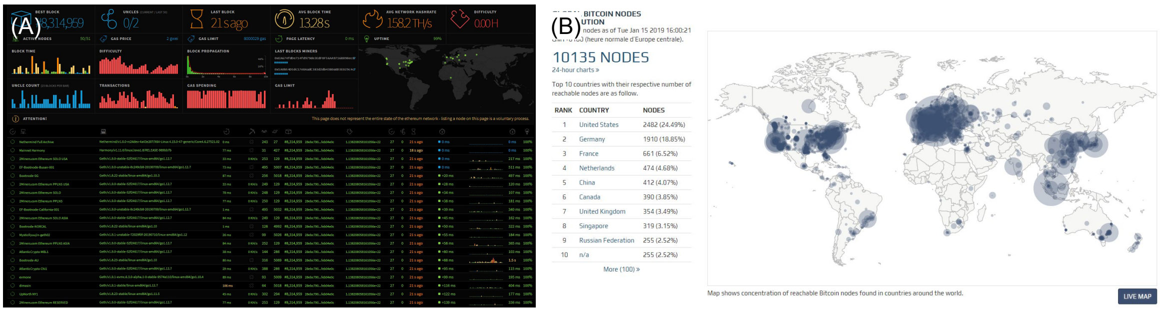

Figure 17: Representative examples P2P network activity visualizations: (A) EthStats.net and (B) BitNodes.

network, and 2) map-based visualizations to show the geographical distribution of blockchain usage around the world.

Most sources in this task domain present **P2P network statistics** calculated from aggregated node activities in the blockchain network over time. All sources use time series visualizations to represent changes in the blockchain network over time. Blockchain.info [O10] provides a long list of time series charts to display a wide range of Bitcoin network statistics, such as the total hash rate, average block size, total transaction fee, mining difficulty, etc. Bogner [O8] proposed a dashboard that presents time series and basic charts on Ethereum statistics and highlights outlier data using anomaly detection techniques. EthStats.net [O1] provides a real-time dashboard for monitoring network status and active nodes in the Ethereum blockchain (Figure 17 (A)).

Analyzing **the global distribution of a blockchain network** involves observing the geographical distribution of blockchain nodes and transactions. Public blockchain data does not inherently include geographic information about senders, receivers, or blockchain nodes. However, when nodes in the blockchain network have associated IP addresses, these can be used to infer the geographic location with a degree of uncertainty [15, 76]. The geographic origin of a transaction can then be inferred from the IP address of the first node that relayed it [79, 132]. They all display geography information in map-based visualizations—the only task domain that uses this kind of visualization type. I found 9 sources that display the number of nodes active in the blockchain P2P network ([O6, O9, O18, O21, A26, A33, O47]) and transaction origins ([O16, A38, O46]). For example, BitNodes [O47] implemented a node crawler to gather reachable node locations to estimate the global distribution of Bitcoin nodes (Figure 17 (B)).

3.3.6 *Casual/entertaining information communication*

Sources in this task domain generally provided original and experimental visualizations of blockchain components distinct from those used for the above task domains. These sources encoded, for example, attributes of transactions and blocks as custom objects—often in 3D—with animation and real-time updates. To show the wide variety of visual encodings in this category, I briefly discuss a few examples: BitBonkers [O3] shows live Bitcoin transactions as 3D balls falling on a plate each time a new transaction broadcasts to the network. BitcoinCity [O36] represents Bitcoin transactions as 3D toy models of buildings along the road moving as new transactions are created. BitListen [O35] presents transactions as animated bubbles floating on the screen, producing notes that combine into improvised music. Symphony of Blockchains [O31] includes a combination of interactive visual representations of Bitcoin data. It allows web visitors to browse blocks as a 3D visual representation and navigate through a flight-simulator mode, along with background audio representing the network hash rate and using a unique tone for each of the transactions in the block. Bitcoin VR [O45] is an open-source project that visualizes Bitcoin transactions as balloons flying over a 360-degree view. Plantoids [O26] and On Brink [O40] represent Bitcoin blockchain components as data physicalization [34].

3.4 DISCUSSION AND OPEN CHALLENGES

Even though the attention to the blockchain is increasing over the years, visualization and visual analytics tools on blockchain data are still rarely found in visualization research. There were only 20 visualization articles (18%) in our survey. Most of the works were online tools (43%) that display basic blockchain information or ad-hoc data analyses published in research articles (39%). Consequently, I saw opportunities for researchers to develop more advanced tools that support higher-level and more in-depth studies of blockchain data. I describe some open challenges in this emerging research field from three perspectives: data, users, and visualizations.

3.4.1 *Data*

Blockchain technology produces a large transaction dataset which provides opportunities to pseudonymize data in such granular details. However, the analysis and visualization of blockchain data are challenging because blockchain activities keep and involve various types of actors who use the blockchain in different ways. Blockchain networks are much more complex than most social networks due to their pseudonymous use of addresses and the heterogeneity of users' behaviors. Besides, blockchains are regulated by automated mechanisms, which are challenging to comprehend for non-technical audiences.

THE DOMINANCE OF CRYPTOCURRENCY BLOCKCHAINS: In this survey, most of the sources present cryptocurrency blockchain data, in particular, Bitcoin and alt-coins. Ethereum-related visualizations usually focus on the cryptocurrency aspect of value exchanges among entities, not the smart contract functionality that makes Ethereum different from Bitcoin. Some analysis articles present the analysis of smart contracts, but dedicated visualization tools are still largely missed. Consortium and private blockchains apply the same concept as public blockchains but have different mechanisms (e.g., the transaction data structure and mining protocol). So far, I found only Hyperfabric Explorer and HyperSec that visualize these kinds of blockchains. Therefore, a lot of exciting future work remains to be done in this domain.

MINING TOOLS IN SOME BLOCKCHAIN DATA TYPES: The majority of tools in the survey visualize blockchain components while we miss VA tools that help analyze nodes and mining data. In particular, mining is an critical activity that ensures the integrity and security of transaction data. This domain was rarely explored the past work. The existing tools on mining only show summary statistics on hash rate, mining difficulty, and market shares of mining pools. Nonetheless, the continuous growth of mining activites raise questions on the economic incentive as well as miners' actions taken to keep the blockchain network stable.

MISSING CONTEXT OF BLOCKCHAIN DATA: Most visualization sources presented details about blockchain components and overviews of network activities in the transaction detail analysis and P2P network analysis task domains. However, those tools did not provide contextual information for monitoring and analysis of activities in the blockchain, including the identification of entities, geographic information, social network activity, or historical events. For example, there is an opportunity for VA tools to help study the blockchain network in light of historical events, such as volatility of market prices, cyber-attacks (e.g., bitcoins stolen from Mt. Gox exchange, money laundering, and denial-of-service attacks), government regulation, or changes in mining rewards.

DEANONYMIZATION OF BLOCKCHAIN USERS: Blockchain users were originally meant to be anonymous through non-identifiable addresses to create and sign in transactions. However, the identity of users will be revealed if people or enterprises post their addresses openly and connect them to other pieces of identification, such as usernames in public forums, their websites, or data leaks from blockchain services. To understand how individual users use blockchains, information about which addresses belong to the same entities is required (e.g., individual users, businesses receiving Bitcoin for payment, or exchanges). Clustering heuristics are effective strategies to group addresses that may belong to the same entities [61, 107], while data mining models were proposed to classify different types of users [75, 179]. These methods can deanonymize users with a degree of uncertainty. However, entity-based visualizations are, nevertheless, rarely found in the current visualization work (e.g., [V9, V16, V19]).

BLOCKCHAIN DATA INFRASTRUCTURE: Accessing blockchain data requires installing a full blockchain node and some technical knowledge to parse the raw data. Data infrastructures should provide easier access, update, and preprocessing of blockchain data to facilitate blockchain visualization research. Some work in this direction has emerged, such as BlockSci [78] and GraphSense [V7], but it will need further development to become usable for visual analytics. Data infrastructures would allow researchers on Bitcoin analysis tools to focus on designing analysis tools rather than the data backend needed to extract blockchain data, compute usage metrics, and make them accessible for quick visual analysis.

3.4.2 *Users*

Blockchain enthusiasts and startups have developed online visualizations tools to understand what happens in and around blockchains. Most of the online sources aim to communicate transaction details and P2P network statistics for blockchain users. These sources only use simple charts and time series visualization to display data on the aggregated level. Nonetheless, many EDA articles in the survey indicate the demand from researchers and blockchain experts for visualization tools to analyze the blockchain data in deeper detail.

IN-DEPTH ANALYSIS OF BLOCKCHAIN DATA: Blockchain analysis articles often focused on a higher-level data analysis on the blockchain network (i.e., the global blockchain network or longitudinal study of P2P network analysis). The number of EDA articles indicates the demand for data analysis experts and decision-makers. To better understand blockchain activities, it would be an opportunity for the VA research to develop more advanced tools that support higher-level and more in-depth analyses of blockchain data. In particular, future VA tools should allow experts to take on specific viewpoints such as individual entities in the network (e.g., people, enterprises, miners), historical events (e.g., cyber-attacks), or network-related events (e.g., halving days or forks).

TRANSACTION NETWORK ANALYSIS TOOLS: Most analysis articles focused on transaction network analysis—indicating a current research focus. However, a few tools exist that experts could use to explore and monitor transaction networks in detail. Those analysis articles often did not conduct an in-depth analysis of the entire blockchain, probably because of the large data size and the lack of simple ways to explore and statistically analyze the data in its entirety. Therefore, new visualization tools would help analysts better understand different types of activities in the blockchain rather than global network statistics of the blockchain data.

TASK-FOCUSED TOOLS FOR BLOCKCHAIN EXPERTS: There is a lack of tools tailored to the specific needs of particular experts, including economists, regulators, and blockchain managers. Economists want to understand activities on blockchains and compare them with related economic activities in the real world. Regulators want to identify the real-world identity of users and detect possible money laundering transactions. Consortium blockchain managers need to monitor the network and compare its performance to other blockchains.

3.4.3 *Visualization*

Most visualization sources focused on using common chart types (i.e., time series and basic charts) with basic interaction techniques (i.e., querying and zooming) that give an overview of the blockchain network. Those charts and interactions are not sufficient for advanced analysis tasks, which require exploration on multiple levels of data aggregation and to relate different factors.

MULTIPLE-COORDINATED VIEWS VISUALIZATION: Sources from EDA articles and online sources usually provide many single view charts showing a particular blockchain measure over time. However, single disconnected views make it difficult to relate multiple blockchain characteristics to each other. As numerous factors and actors interconnect and evolve, multiple-coordinated visualizations would help analyze the cause and effect that drives changes in the blockchain network. These multiple-coordinated tools are mostly found in visualization articles. For example, BitExTract [V19] proposes a dashboard with multiple chart elements for analyzing transaction activities among Bitcoin exchange entities. Yet, additional sophisticated interaction techniques for visual comparison [51] would help to connect views and generate more comprehensive insights.

VISUAL REPRESENTATIONS FOR TRANSACTION NETWORKS: Existing network visualization sources present transaction networks and value flows as static graphs at specific points of interest (i.e., a time interval, a particular block, or a group of entities). In the survey, there are 16 network online tools and visualization articles. However, those tools, except [O9, O20], do not consider the temporal evolution of the network and, in other words, changes in blockchain connectivity over time. The advancement in dynamic network visualization [11] will add a time dimension into current network visualization sources and help experts explore the changes in blockchain activities.

Transaction networks can be very large and heterogeneous. Hence, they may not display precisely on the screen without filtering the network to a manageable size. Clustering heuristic can help simplify the raw graph into the entity network. To handle the large graph, blockchain networks could benefit from dedicated graph visualization layouts, such as Alvarez et al. [4], to interactively display and navigate from the overview of the network to focus on a particular subgraph of interests.

UNCERTAINTY VISUALIZATION: Much of the contextual information related to Bitcoin comes with a degree of uncertainty. For example, heuristics to cluster Bitcoin entities are not sure to capture Bitcoin entities with 100% accuracy, and IP addresses of nodes in the P2P network are not necessarily reliable indicators of the geographic location of a node. Moreover, analysis tools that may label specific transaction patterns as fraudulent or belonging to certain services (e.g., exchanges, mixing services, etc.) may induce false predictions. Any uncertainty in the data should be made evident in the visualization [72, 98, 113] and expose where viewers should be cautious about inferring insights and making decisions on the data.

PROGRESSIVE VISUAL ANALYTICS: Exploring Blockchain data involves navigating over large amounts of data for computing aggregated values on selections of the transactions or over time windows. These operations are usually simple to compute but take a long time. Current visualization sources in the survey mostly perform data computation offline. Doing the calculation offline means that the data exploration is limited to precomputed values. All the interactive sources were limited in that respect. Novel tools could rely on progressive data analysis and visualization to present the intermediate result while progressively compute the data [41, 147].

Techniques proposed by Boukhelifa et al. [24] could be applied to continuously compute derived data when the Blockchain evolves. Kinkeldey et al. report that BitConduite [V9] provides dynamic queries on time and attribute values to visualize aggregated information about Bitcoin transactions, but each filter operation takes a minute or so to complete depending on the amount of loaded data. Performing these operations iteratively using methods reported by Moritz et al. [105] would drastically reduce the interactive latency and significantly improve the efficiency of exploring Bitcoin data.

3.5 CONCLUSION

This chapter provides a systematic review of 110 blockchain data visualizations from both academic and online sources—20 visualization articles, 43 exploratory data analysis EDA articles, and 47 online web-based tools. I classified those sources based on blockchain data, task domains, and visualization types and described each task domain's different kinds of tools. Most of the online tools focus on P2P network analysis and transaction detail analysis task domains. EDA articles show the demands from data analysis experts for more advanced tools in transaction network analysis and cybercrime detection. Most of the sources I found reported aggregated information. VA tools have emerged to facilitate the exploration and monitoring of blockchain data from overviews to detailed investigations. As the need for more VA tools will grow, I outlined several opportunities and open challenges for future research in this domain.

In this thesis, I focus on mining which is an intriguing activity in the Bitcoin blockchain. However, as I showed, VA tools to deeply explore this data are missing. In the next chapters, I propose a multiple-coordinated visualization tool to relate various information sources in the context of Bitcoin mining. The following chapters describe a design study process with an economist expert, intermediate data analysis results, and a VA tool as the outcome from the design study.

4 THE EMERGENCE AND EVOLUTION OF BITCOIN MINING POOLS

In the previous chapter, I outlined research challenges in blockchain Visual Analytics (VA). One of the topics that are missed in current work is task-focused tools for blockchain experts. As a consequence, my work focuses on an in-depth analysis of an important activity in Bitcoin called mining. Mining is a backbone mechanism to ensure the integrity and security of transaction data in blockchains. Economists are interested in mining as the activity generates new bitcoins circulating in the network from block rewards. However, empirical research in Bitcoin mining needs more attention, and economists need a tool to help access the data. I present a data analysis methodology, results, and a VA tool in the following three chapters.

Bitcoin miners are assumed to be economic agents who invest in mining hardware and compete to profit from mining. As the competition among miners escalated, mining pools emerged to reduce miners' risk and share more regular rewards among pool participants. This chapter extracts both on-chain (i.e., transaction data and Bitcoin statistics) and off-chain data (i.e., market price and other external factors) to investigate Bitcoin mining as an economic ecosystem. The data and results help my economist collaborator empirically study the dynamics of mining pools and miners' migration among pools over the entire history of Bitcoin mining.

> This chapter is written mainly based on my original article published at the *Blockchains and Smart Contracts workshop (BSC 2020–2021)* in *IFIP International Conference on New Technologies, Mobility & Security* [156]. The work was led by myself in collaboration with an economist, Nicolas Soulié, and my supervisor, Petra Isenberg.

4.1 BITCOIN MINING AS AN ECONOMIC ECOSYSTEM

In Section 2.2.2, I described the Bitcoin mining mechanism as miners follow the proof-of-work protocol to obtain a mining reward and participate in mining pools to share more regular rewards among the mining pool participants. At the time of writing, known mining pools contribute about 99% of the total hash rate and have become the dominant actors of the mining activity. As miners and mining pools are driven by monetary incentives to perform mining operations, they consider multiple factors from the competition among them: the mar-

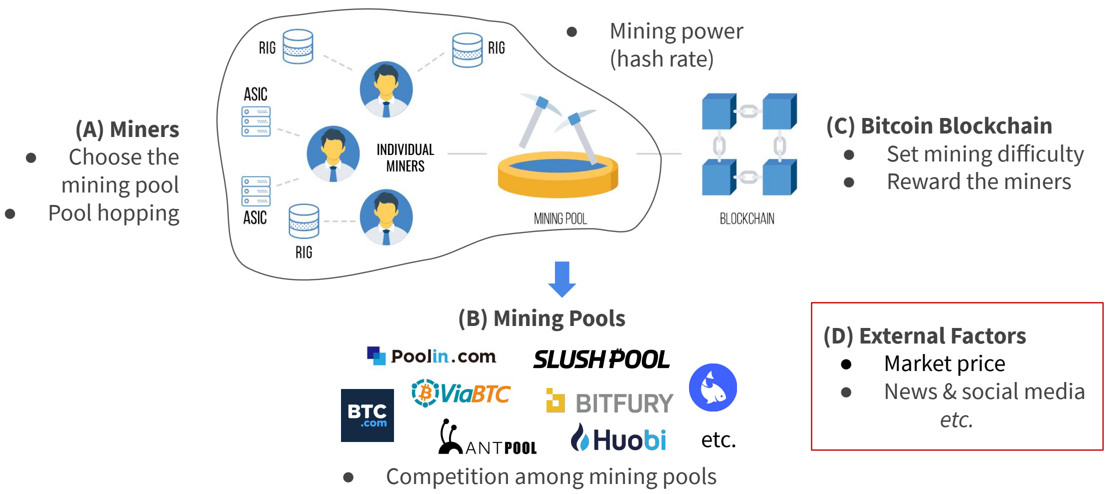

Figure 18: As an economic system, Bitcoin mining consists of three actors: (A) miners, (B) mining pools, and (C) the Bitcoin blockchain. (D) External factors, such as market price, news, and social media, can also affect miners' decisions to participate in this activity. (Image from https://miningpools.com/)

ket share of mining pools, the mining reward from the blockchain, the Bitcoin market price, and the cost of mining. Figure 18 models Bitcoin mining as an economic ecosystem considering interaction among three main actors and external factors.

- *Miners* participate in mining pools to get mining rewards. They invest in mining hardware and pay electricity costs in expectation to make a profit. Miners typically join a mining pool to increase their chance of a reward [94, 131]. Considering different mining pools, they decide to participate in mining pools or move to another pool—called *pool hopping*—to maximize their profit [13]. Besides, miners can also *cross pool* with more than one pool at the same time [127].

- *Mining pools* merge the computational resources from miners to maximize the chance to obtain mining rewards frequently and share them among miners. The market share indicates the relative size of the pool. Mining pools compete with other pools to attract more miners by providing better reward incentives. The reward incentives that I consider in this study include reward payout schemes [128] and imposed pool fees.

- *The Bitcoin blockchain* pays a mining reward when miners or mining pools successfully mine a new block and broadcast it to the blockchain network. The mining reward includes block reward, newly generated bitcoins from the network, and transaction fees from transactions included in the block. Its protocol regularly adjusts the mining difficulty to maintain the block discovery rate and half the block reward every 210,000 blocks (so-called *halving days*) to control the bitcoins supply in the system.

- Apart from the above actors of the Bitcoin mining activities, *external* factors also impact miners' decisions to mine Bitcoin. For example, market price determines the mining profit when exchanged to fiat currencies, such as US Dollars and Euros. News and social network sentiments can encourage or discourage potential miners from investing in this activity.

In this chapter, I explore the relationship between mining pools and the Bitcoin blockchain network. Chapter 5 will investigate the behavior of individual miners and their impact on mining pools. Finally, Chapter 6 combines data and analysis tasks from both preceding chapters and proposes a visual analytics tool to help economists and Bitcoin miners to understand the Bitcoin mining economy from these three main actors.

4.1.1 *The Dangers of Mining Pools*

In the Bitcoin mining ecosystem, three significant risks to its security and sustainability exist: 1) The risk that one pool becomes dominant, 2) The risk that mining stops carrying economic incentives and miners stop participating, and 3) The carbon footprint of Bitcoin mining.

As mining pools represent a large group of miners, they play a major role in ensuring miners' consent income and potentially dangerous position 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 to double spending their bitcoin values and prevent other miners from proposing new blocks [25]. As large and specific investments are needed to mine successfully, only a few large and persistent mining pools dominate the market share [127]. Hence, the danger of a 51% attack is real and needs to be monitored constantly.

Another possible attack is selfish mining or block withholding. The attack occurs when a mining pool or miner intentionally withholds a new block they found and continues mining from the withheld block to form a longer chain. Later, they broadcast their chain to the network. If their chain is the longest one, it will become the main chain, and the mining pool will obtain all rewards from all withholding blocks. In contrast, the block mined by an honest miner or pool becomes invalid in an orphan chain. A study shows that this attack is possible if a selfish mining pool has more than 33% of the market share [39]. This potential attack 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 incentives and mining rewards. On what is called a *"halving day"*, 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 [109]. A theoretical study previously cast doubt on the compensation of these freely fixed transaction fees [66]. In fact, early evidence showed that transaction fees are relatively deficient and seem to be driven by social norms rather than economic reason [106]. Analyzing miners' behavior around halving days and the evolution of transaction fees allows observing the consequence of mining rewards on the behavior of mining pools and individual miners.

The third issue concerns the carbon footprint of Bitcoin mining. Tremendous amounts of electricity are needed to run mining hardware [14, 87, 146, 161]. Bitcoin price increases have incentivized miners to compete in a prisoner's dilemma to upgrade their hardware continuously [163]. This violates the *one-CPU-one-vote* policy initially envisioned by Nakamoto [109] and causes a negative externality on climate change [44, 99, 104]. Alternative solutions, such as *proof-of-stake*, have been recently promoted to reduce the excessive energy consumption of cryptocurrencies [110]. Analyzing miners' revenue function and the environmental impact would help inform policymakers to develop sustainable mining regulation and policy [50, 73, 157, 161]

4.2 RELATED WORK

As mining is central to how Bitcoin and other cryptocurrencies work, researchers have started to analyze it more closely. Much of the research on Bitcoin mining and mining pools is theoretical in nature and focuses on analyzing reward rules [43, 128, 138], mining strategies [65, 90], and attacks [39, 74, 83, 89].

Yet, empirical analyses on the Bitcoin mining economy have recently appeared. On the development of Bitcoin mining activity, Prat and Walter [118] built a structural model to explain the rapid growth of the hash rate considering Bitcoin market price and mining hardware evolution. They showed that Bitcoin mining operates under competitive conditions. Song and Aste [144] estimated the mining cost from energy to run the hardware and found that the ratio between mining cost and transaction volume remained in the same range from 2010 to 2020. Möser and Böhme [106] concluded that transaction fees between 2011 and 2014 were not totally driven by bitcoin demands but followed the conventions of other bitcoin users.

The rapid growth of mining activity was associated with the rise of mining pools as intermediaries to collect mining rewards. Mining pools raised concerns on the sustainability and security of the Bitcoin blockchain as they posed a threat of the majority attack. Apart from theoretical studies on the mining pools, as comprehensively reviewed by Liu et al. [96], a few studies had analyzed the evolution of mining pools based on empirical data.

An early work from Wang and Liu [165] provided evidence that the top mining pools gained a larger market share while the hash rate grew exponentially between Mar. 2013 and Mar. 2014. The authors analyzed the mining profit regarding hardware cost and electricity price and concluded that the profit became negative when the hash rate increased faster than the Bitcoin price. Another work by Romiti et al. [127] analyzed the distribution of mining pools from Dec. 2013 to Dec. 2018 and found that 3–4 mining pools controlled >50% of the hash rate. The authors further analyzed the reward payout among the top-3 mining pools and found that a small number of members received a total of >50% of the reward from the pool. In addition, the authors detected cross-pool miners who received rewards from multiple mining pools and provided evidence that miners tend to transfer their rewards to exchange services and wallet providers. Wang et al. [164] analyzed the daily hash rate of the top mining pools from Feb. 2016 to Jan. 2019. They found that mining pools increased their hash rate exponentially to maintain their market share. Mining pools were caught in the Prisoner's Dilemma. They raise their hash rate to compete with other pools, but their mining profit diminishes. Finally, Wang et al.'s data confirmed that pools tended to collect transactions with higher transaction fees to maximize their profit.

My work studies mining pools more broadly using various parameters instead of focusing on just a few mining pools and a limited timeframe. As such, my work overlaps with previous results (e.g., Romiti et al.'s study of mining pool distribution), but extends the past work with custom visualizations. My work shows the distributions of mining pools' market shares and help detect changes in their ranks over time. I also provide a first exploratory analysis of the evolution of reward rules and locations, which has not been studied in previous empirical works. The findings provide new information toward more realistic theoretical models of the Bitcoin mining organization.

4.3 DATA PREPARATION

To analyze the evolution of Bitcoin mining over time, I retrieved coinbase transactions (the first transaction in each block) from the Bitcoin genesis block (the first-ever mined block) until the last block mined in Sep. 2021. From the coinbase transactions, I extracted several metrics as described next. Moreover, I collected monthly Bitcoin statistics from Blockchain.info and historical data about mining pools from the Bitcoin Wiki to give contextual information to the mining activity. The datasets about mining pool attribution and pool characteristics are available in a public repository: `https://zenodo.org/record/4342747`.

Mining pool attribution: First, I identified the mining pool that mined each block in the Bitcoin blockchain. When a mining pool mines a block, it receives a 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, I attributed the mining pool based on address matchings or coinbase string patterns. I initially used the dataset from Romiti et al. [127] that compiled known mining pool attribution until the block 556,400 (December 31, 2018). After this block, I continued their procedure and tagged pools until the block 700,714 (September 16, 2021) with the datasets from Blockchain.info [22] and BTC.com [28]. The blocks that did not match any known mining pool are labeled as "unknown."

Mining pool market shares: The market share of a mining pool indicates to which extent that mining pool dominates the Bitcoin mining activity. I calculated each pool's market share as the percentage of the blocks it mined compared to the total blocks mined in a month. To assess whether there are a few pools that dominated the market, I adopt the *Herfindahl index* (H) to measure the concentration of mining pools for each month.

Let n be the number of active mining pools and s_i indicates the market share of pool i as a percentage. The concentration index is defined as:

$$H = \sum_{i=1}^{n} s_i^2 \tag{1}$$

H is equal to 1 for a perfectly concentrated market and converges toward 0 for a very fragmented market.

Mining pool characteristics: I obtained information about pool characteristics, in particular payout schemes and pool fees, from the Bitcoin Wiki page [17] on the topic. I downloaded the page's edit history and manually cleaned the data for each month by comparing it with the information from the Bitcoin Forum [18]. As a result, I constructed panel data that includes all changes in pool characteristics over time, such as locations, reward payout schemes, and pool fees. My economist collaborator, Nicolas Soulié, manually verified pool characteristics data, particularly payout scheme and pool fees, from the Bitcointalk discussion forum [18] and mining pools' websites.

From the Bitcoin Wiki data, I grouped reward payout schemes into five main groups that are commonly found in mining pools: Proportional (Prop), Score, Pay Per Last N Shares (PPLNS), and Double Geometric Method (DGM), and Pay Per Share (PPS). For the full technical coverage and analysis of reward payout schemes, I refer readers to a detailed analysis article by Rosenfeld [128].

- *Proportional (Prop)* is the most straightforward payout policy in which mining pools distribute the rewards to their miners after receiving the block reward. Miners receive the reward shares in proportion to the computational power they contribute for that block reward (or round).

- *Score* distributes the block reward to miners based on the proportion of miners' scores in each round. Miners who contribute to the mining pool will receive a score calculated from the time elapsed from the previous round. The more time has passed without finding a new block, the higher score the miner receives.

- With *Pay Per Last N Shares (PPLNS)*, mining pools share mining rewards to only miners who submitted the last N shares in each round. This policy aims to prevent miners from profiting early in the round before moving to other pools (pool hopping).

- *Double Geometric Method (DGM)* distributes mining rewards to miners with the PPLNS payout scheme. However, mining pools keep varied fees depending on the time between blocks it successfully mined before paying the remaining amount to miners (geometric method). The longer the time gap between blocks, the fewer rewards it will keep. This method is a hopping-proof payout scheme.

- *Pay Per Share (PPS)* pools pay miners a certain amount for each share submitted to the pool, regardless of how many blocks the mining pool mined. Miners receive reward shares regularly (usually every day). With this method, mining pools need to have reserved funds to manage the risk of not obtaining enough rewards to pay miners.

I grouped the locations of mining pools on a continent-level: Asia (including Australia), Europe, America, and Africa. I assigned Australia to the Asian continent as only one pool (OzCoin) is listed with an insignificant influence on the overall market share. South Africa is the only country in the Africa continent for which I found a pool (ZA Bitcoin) in the Bitcoin Wiki dataset. I defined "global" for pools that operated in more than two continents while kept China separately as some large mining pools are located in the country.

Bitcoin network statistics: Apart from information about mining pools, I also obtained Bitcoin network statistics from Blockchain.info [O10]. I averaged those measures to a monthly time window. These measures are used to relate to the evolution of mining pools. Examples of network statistics for each month include the market price, total hash rate, total block reward, total transaction fees, mining difficulty, or the number of transactions. Other external data that I included in Bitcoin statistics are mining hardware evolution [82, 151], electricity consumption [14], and the global energy price index [68].

4.4 VISUALIZATION DESIGN

To understand what the economist wants to analyze d on Bitcoin mining pools, I collaborated with an economic lecturer, Nicolas Soulié, who is working on research in this domain. We identified data analysis goals, design requirements, and iteratively revised prototype visualization designs. The visualization design process loosely follows Tamara Munzner's four-level nested model [108]. We identified the first four analysis questions listed below.

- Q1 What is the evolution of mining pools over Bitcoin history?
- Q2 Are there any mining pools likely to dominate the market at a time?
- Q3 Are there any external factors from the Bitcoin network that affect the mining pool competition?
- Q4 What are the pool characteristics that dominate the mining market?

I developed visualization prototypes in Jupyter Notebook using Altair [160], an interactive visualization library in Python. Using Google Colaboratory allowed me to create rapid prototypes to share them with the economist, and improve the design based on his feedback. At the end of the design iterations, I developed three visualization views to help the economist explore the data and answer his analysis questions.

4.4.1 *Visualizing the emergence and evolution of mining pool market shares*

The first visualization allows economists to explore when mining pools started to gain market share and which pools dominated the market over time. To answer Q1 and Q2, Figure 19 (A) provides information on the historical evolution of the top 30 mining pools based on their market share for each month. The scatterplot-based chart gives an overview of the top mining pools and shows when each

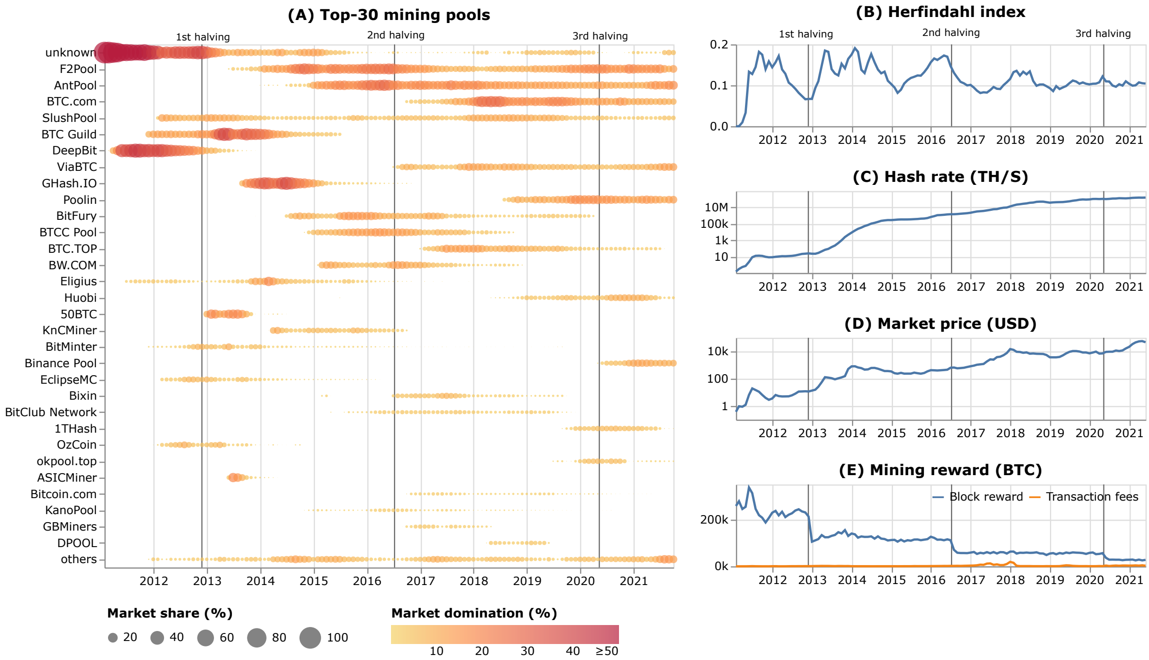

Figure 19: (A) The market share of the top 30 mining pools over time. (B) The concentration index of mining pools. (C) The total hash rate in tera hashes per second. (D) The average market price in the US Dollar. (E) The total mining reward from bitcoin is divided into block reward (blue line) and transaction fees (orange line).

mining pool started or stopped its operation and gained or lost mining power. The size of each circle indicates the market share for each month. I used a heated color scale to describe mining power domination (0%–>50%) and detect the pools that were likely to dominate the Bitcoin network at a time. In the case of 50% of the mining power (or even only >33% [39]), a mining pool can modify the transactions and perform majority attacks on the Bitcoin network. Three grey vertical lines in both charts indicate three halving days on Nov. 28, 2012, Jul. 9, 2016, and May 11, 2020. I chose this representation over a simple heat map alternative because it can represent two quantitative variables: market share and market domination.

To quantify the dominance of a few mining pools over time, Figure 19 (B) shows the Herfindahl index on the time series chart. The chart provides reference points to assess when there is a potential risk of a majority attack. Figure 19 (C–E) includes a list of chosen relevant Bitcoin statistics: hash rate, market price, and total mining rewards. These time series charts help answer Q3 by giving contextual information concerning the evolution of mining pools in (A). Hash rate is the total mining power in the Bitcoin network, indicating the competition in the activity. The valuation of mining income can be estimated from the mining reward and converted to the market price in the US Dollar.

4.4.2 *Visualizing mining pool characteristics influencing the evolution of market shares and their distribution*

Understanding when and why mining pool market shares change is critical to understanding the Bitcoin mining economy. The second visualization presents the evolution of mining pools' market shares in connection to the panel data we collected (e.g., reward payout schemes and location). Using the same visual representation from the previous section, Figure 20 presents the market share together with mining pool characteristics. For each mining pool, the first row shows the market share of the mining pools as the circle size. I changed the color of the circles to encode the location of the pool. As mining pools can implement different payout schemes and pool fees over time, I added the payout scheme in separate rows and encoded the pool fee as the color scale. I used two pastel colors in the background to distinguish whether the mining pool kept transaction fees or shared them with miners. However, this visual representation cannot fully help answer Q4 because it shows the details of individual mining pools rather than explores the big picture of which payout schemes or locations that dominate the mining market share over time.

In particular, the economist wanted to understand the overall evolution of pool characteristics and assess if mining pools converged to the reward payout schemes or locations that are the most attractive to miners to address Q4. Therefore, we wanted to analyze 1) which pool characteristics are commonly found in all mining pools and 2) how much the total market share for active pools is in each characteristic over time. I proposed a third visualization to explore the pool characteristics that dominate Bitcoin mining over time. I used a ribbon chart design—a stacked bar chart showing quantitative measurement over time with ribbons connecting the same data category. Figure 21 and Figure 22 show the temporal evolution of reward payout schemes and pool locations sorted by pools' market shares. Each mining pool is represented as a bar whose height is relative to the pool's market share in a month. The color indicates categorical data (i.e., mining pool, reward rule, or location). For each month, the bars are sorted by the total market share of the category from the highest to the lowest value. Within each category, the bars are sorted by their market shares. This sorting helps to identify large mining pools or categories that dominate mining power. Ribbons track a mining pool's ranking connection between months. This visualization shows overall distributions and detects the top mining pools, while the first visualization is better for observing the development of individual pools.

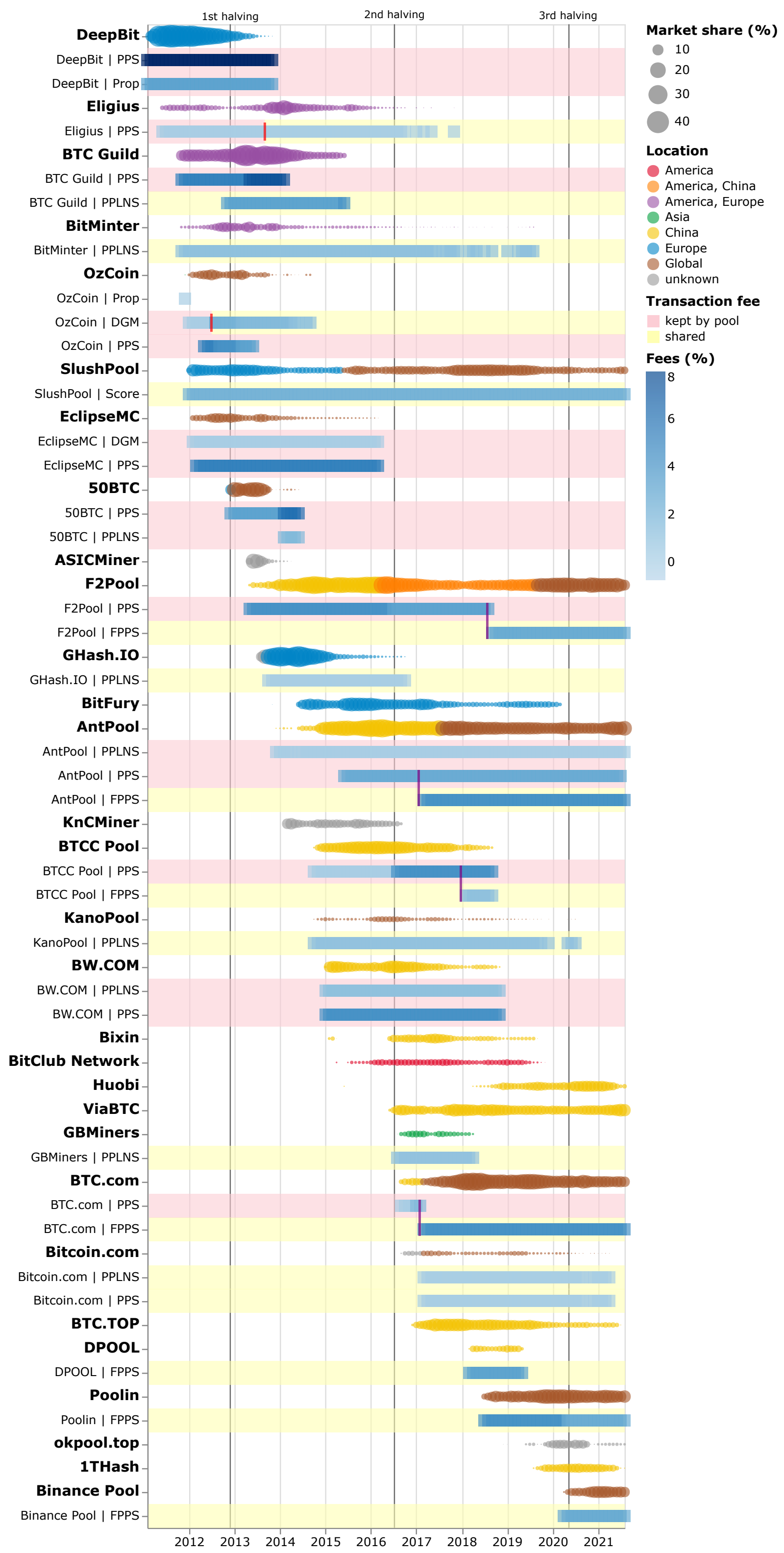

Figure 20: The market share of the top-30 mining pools with location (color on the market share rows), payout scheme (rows below each pool's market share), and pool fee (color scales in the payout scheme lines). The background color in pool fee rows indicate if the mining pool kept transaction fees to itself or shared them to miners.

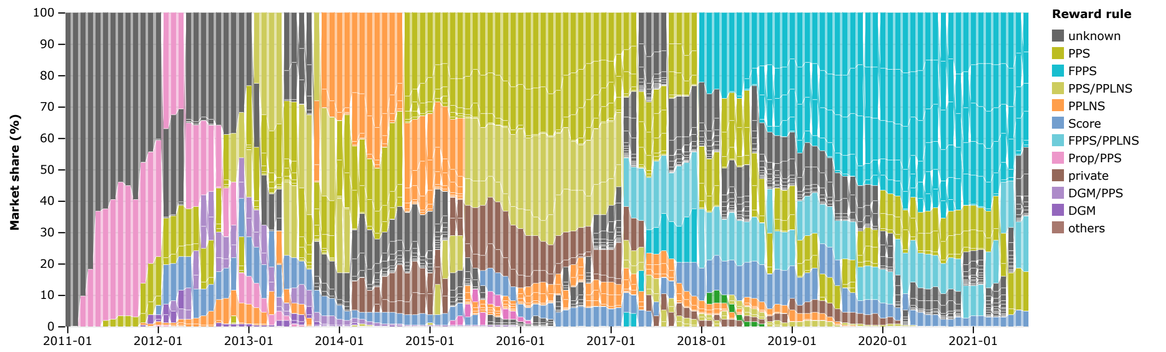

Figure 21: The market share distribution according to payout schemes.

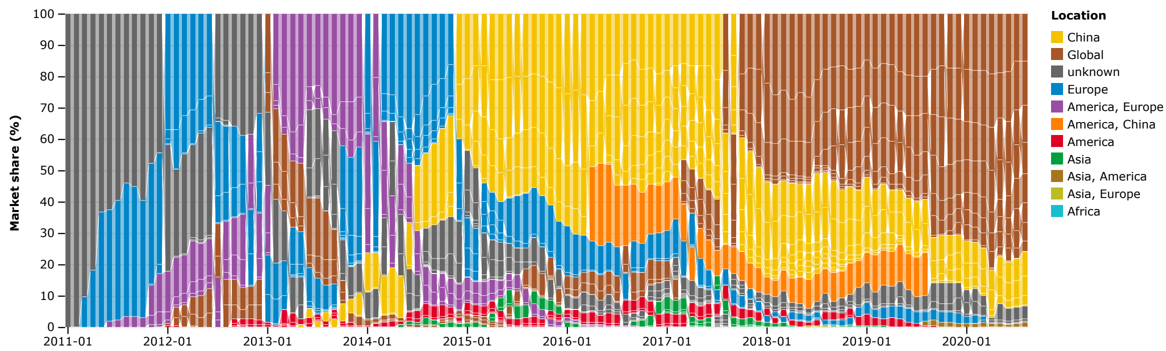

Figure 22: The market share distribution according to mining pool locations.

4.5 RESULT AND ANALYSIS

In this section, I present results and analysis of mining pools evolution over the Bitcoin history based on the four analysis questions I identified with the economist collaborator.

4.5.1 *Q1: What is the evolution of mining pools over Bitcoin history?*

Before 2011, most mining activities were performed by small independent miners following Bitcoin's original one-CPU-one-vote policy. Some pools might have existed between 2009 and 2011. However, they are not identifiable in our data, and their market shares were very small. DeepBit is the first mining pool detected in our dataset that operated from 2011 to 2014. It almost dominated 50% of the market share from mid-2011 to early 2012. After that, the market share continuously decreased while new pools named BTC Guild and SlushPool emerged and gained a larger market share. In late 2013, GHash.IO started its mining operation and replaced BTC Guild to be the top pool. The market share of GHash.IO has been growing and was likely to cross the 50% threshold before it suddenly dropped in 2014. Top mining pools in this period tended to be short-lived but solely dominated almost 50% of the market share. From this period, only SlushPool still operates and obtained constant rewards until now.

After 2015, many top mining pools emerged, including F2Pool, AntPool, BitFury, BTCC Pool, and BW.com. F2Pool and AntPool are two top mining pools that have remained the top mining pools until now. These mining pools tended to have a moderate market share. However, only 3–4 top mining pools have a total of >50% market share, which poses a risk to the majority attack. After the second halving day, we found three new mining pools, BTC.com, ViaBTC, and BTC.TOP that still operates and maintains the same market share over time. Poolin and Huobi emerged on the market and became the top pool after 2019. We observe that mining pools become less dominant in market shares as more mining pools enter the competition. Since 2020, the concentration of mining activities has been relatively stable, with five main pools accounting for around 10% of market share in a total of >50% market share: AntPool, BTC.com, F2Pool, Poolin, and ViaBTC.

4.5.2 *Q2: Are there any mining pools likely to dominate the market at a time?*

To detect the possibility that mining pools dominated the market, we observed five cycles of concentration index in Figure 19 (B). The peaks of the later cycles decreased, indicating that the market shares became less concentrated than before. All these cycles are associated with the increment of bitcoin's value, as investments in mining hardware are indeed strongly correlated with bitcoin's value [119]. By increasing the expected revenue of mining, a rise in bitcoin value gives an incentive to miners to enhance their investments and might also attract new miners into pools [164].

The first cycle of mining pool concentration started in early 2011 with the emergence of the first large and identifiable pools, namely DeepBit and Eligius. DeepBit's hash power grew during the first peak of bitcoin's value in Jul. 2011. At this moment, bitcoin's value reached more than $30, compared to roughly $1 a few months earlier. The expected growth in revenue gained from mining probably lead miners to create new large pools such as BTCGuild and SlushPool. These two pools contributed to the increase in the mining concentration index until the beginning of 2012.

The second cycle began in 2013. Two peaks occurred in Apr. and Nov., during which bitcoin's value respectively reached more than $260 and $1,200. During this period, existing pools increased their hash power (e.g., BTCGuild, SlushPool, or 50BTC), and new collections experienced significant growth (F2Pool and GHash.IO), especially in 2014. After the decrease of bitcoin's value in 2015, the market price raised again and reached peaks at roughly $450 in Dec. 2015 and even $750 in Jun. 2016. During this period, F2Pool and AntPool vastly increased their hash power. Newly created pools (Bitfury, BTCC Pool, and BW.COM) also enhanced their hash rate.

The fourth cycle was driven by the sharp rise of the Bitcoin market price at the end of 2017. Most existing large pools, and in particular, BTC.com and AntPool, increased their hash rates. Both mining pools are owned by BitMain, the Chinese Bitcoin mining hardware manufacturer. In the most recent cycle, the concentration index remains stable at around 0.1 for the entire period. We found that seven mining pools share around 10-20% market share: F2Pool, AntPool, BTC.com, ViaBTC, Poolin, Huobi, Binance. The evidence shows that Bitcoin mining has become less dominant when a higher number of mining pools gain a competitive market share and become top pools in the market.

It is noticeable that the five cycles have a decreasing amplitude over time, indicating that the market might progressively reach an equilibrium situation. This is due to the exponential increase of the hash power which involves a huge investment for an incumbent or a new pool to get a significant market share.

4.5.3 *Q3: Are there any external factors from the Bitcoin network that affect the mining pool competition?*

Bitcoin mining activity had increased exponentially over time, according to the total hash rate in the network (Figure 19 (C)). The market price growth was highly correlated with the hash rate, indicating that miners were motivated to participate in this activity in expectation of monetary profit (D). Over the years, the market price was increased sharply and partially compensated for block reward halving days (E). The total transaction fees did not increase as expected to substitute for diminished block rewards, except for an outlier period between 2017 and 2018.

We calculated the mining revenue per hash rate to explore the relationship between the hash rate, market price, and mining rewards. Figure 23 (A) shows that miner revenues per computational power unit had decreased drastically over time. It seems that the competition among miners drove the expected revenue to fall sharply. However, as the market prices have been volatile in recent years, we question the sustainability of Bitcoin mining rewards after the next halving day. To measure the efficiency of Bitcoin mining, we compared the hash rate growth with the energy consumed for mining operations. Figure 23 (B) shows the energy consumption per hash rate compared to the mining hardware evolution. The chart displayed the exponential decreased trend of the energy consumption per hash rate. The energy efficiency improved abruptly when the hardware upgraded from FPGA to ASIC and faster ASIC devices.

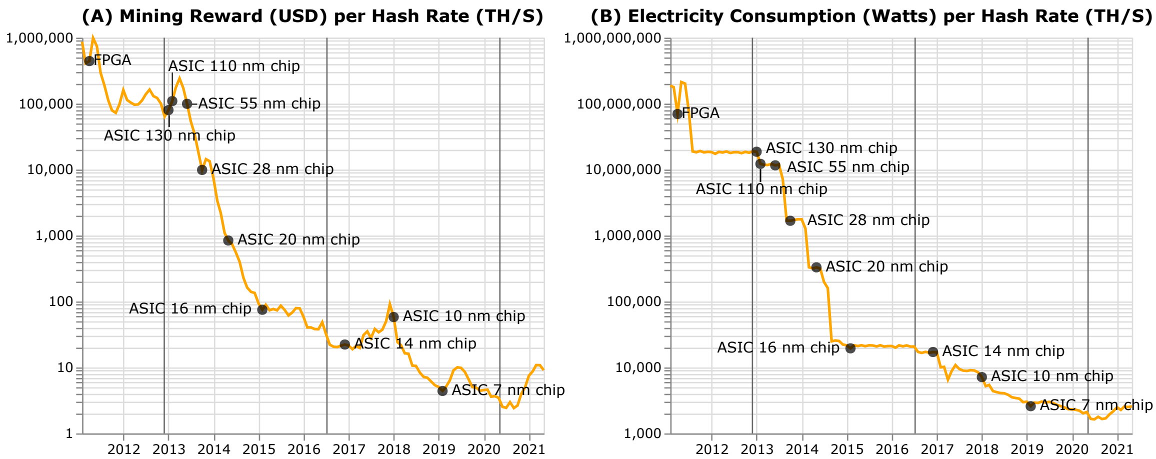

Figure 23: (A) Mining revenue in US Dollars per hash rate (TH/s). (B) Electricity consumption in Watts per hash rate (TH/s). I annotated new mining technology evolution as a dot on the charts.

Both charts in Figure 23 show that while mining hardware is more efficient over time, the profit per hash rate decreased as more miners upgraded their hardware to stay competitive in the market. Besides, supporting Bitcoin statistics from Figure 19 provide evidence that the increased market price cannot wholly compensate for the block reward dropped after halving days. As the market prices have been volatile and miner revenue has been relatively low in recent years, we question the sustainability of Bitcoin mining after the next halving day.

4.5.4 *Q4: What are the pool characteristics that make the mining pool dominate the market?*

We analyzed two main mining pool characteristics: reward payout schemes and pool locations. Reward payout schemes determine how the mining pool distributes the reward to miners and the risk of reward variance handled by the pool. To sustain the market share, pool managers adopt a payout scheme that attracts miners and prevents them from performing block withholding and pool hopping. In Figure 21, we observed heterogeneity of applied reward rules until 2015. DeepBit applied the proportional payout scheme to pay its miners. After the DeepBit market share declined, many mining pools entered the market with a variety of payout schemes. Noticeably, PPS/PPLNS and PPLNS have become common payout methods in many mining pools. In this period, we also observed that PPS pools emerged and gained a higher market share. After 2015, two payout schemes dominated the market: PPS and PPS/PPLNS. PPS/PPLNS offered two reward schemes that miners could choose. Miners obtain regular income with the PPS scheme. The PPLNS scheme is riskier because only miners in the last N shares are eligible to receive reward shares.

After 2017, we observed a decline in the market share of PPS/PPLNS, and PPS became the standard payout scheme for mining pools. The total market share of mining pools that applied the PPS payout scheme increased from around 20% in 2013 to at the peak of 40% in 2017. After 2017, PPS mining pools increasingly switched to Full Pay Per Share (FPPS) payout scheme, notably BTC.com and F2Pool in 2017 and 2018, respectively. FPPS have become the dominant payout scheme since 2019 with more than 50% of the total market share. It implements the same PPS protocol to share a constant income to miners while also sharing transaction fees with miners. AntPool is the only top pool that still offers the PPLNS payout scheme in addition to FPPS. This result implies that miners prefer to receive a regular payment, and mining pools have become the industry since they need to absorb mining variance and pay miners regularly.

From the economic perspective, both PPS (also applied for FPPS) and PPLNS rely on their different but complementary risk/return ratios. These payout schemes are more attractive for pool managers because they are more robust to pool hopping than the proportional reward [128]. While PPS provides risk-free but lower income, PPLNS generates higher revenue but more randomly. PPS pools pay miners in proportion to their contribution to the pool, thus providing risk-free, low income. The pools hold all the risks by creating a reserve of money 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 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 [128] 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* [140]. In financial markets, risky investments must have a higher expected return to be attractive. In the case of Bitcoin mining, Figure 20 shows that the more risky asset (PPLNS) is likely to have a lower fee (≈0%) compared to the risk-free one (PPS, ≈2-3%).

Next, we were interested in the continents on which mining pools operate. The mining pools' locations may collectively pose a risk to mining if a country controls the majority of mining power. Figure 22 shows the large evolution in mining pool locations. Until 2015, most pools were located in Europe and the US. Then, the market share of European pools rapidly decreased while the hash rate grew exponentially together with the rise of Chinese pools. The rapid growth of two important Chinese pools (F2Pool and AntPool) profoundly modified this landscape and made China the largest pool hosting nation from 2015 to 2018. According to Hileman and Rauchs [64], cheap electricity and land costs in remote Chinese areas (e.g., Sichuan and Xinjiang) are significant drivers of this location pattern. Between Mar. 2015 and Feb. 2017, the combination of Chinese pools exceeded the 51% mining power threshold and therefore posed the risk of a majority attack. This situation persisted until 2019 when global pools became dominant. This growth resulted from the increase of originally global pools (BTC.com and Poolin) and Chinese or European pools that turned into global ones (F2Pool and SlushPool). Currently, global pools are common to manage the risk in the case that some big Bitcoin countries, like China, suddenly made Bitcoin mining illegal or shut down mining farms.

4.6 CONCLUSION

Mining pools are the heart of the Bitcoin ecosystem's security and growth. Their evolution toward rational and stable organizations is critical for Bitcoin's future. I collected and joined multiple datasets about mining pools and developed two custom visualizations to investigate their evolution according to several variables. In particular, I combined aggregated measures of market share with pool-level information across time. The visualizations allowed me and my economist collaborator to document the rise of mining activity concentration and give rise to insights and further hypotheses about the cause of this evolution. Important events regarding mining activity profitability (e.g., halving days and bitcoin value) seem to be major drivers of changes to the Bitcoin ecosystem.

The visualizations based on ribbon charts show two other important trends regarding pool organization. First, we can observe the rise of PPS and PPS/PPLNS as the standard reward rules used by pools. These rules seem to provide attractive and stable incentives for miners. Secondly, ribbon charts highlight the emergence of China as the leading hosting country for mining pools and show that global pools have become dominant in recent times. These findings raise questions worth further examination, particularly the possible factors that stimulate the stability of Bitcoin mining. The visualizations are relatively simple but already provide the first practical tools for people who want to analyze the evolution of mining pools by highlighting relevant elements that affect these organizations.

In the next chapter, I investigate miners' behaviors to choose mining pools and how their migration from one pool to another impacts mining pool competitiveness.

5 EMPIRICAL ANALYSIS OF POOL HOPPING BEHAVIOR IN THE BITCOIN BLOCKCHAIN

As miners join mining pools to obtain reward shares, their behaviors directly affect the rise and fall of mining pools and the entire Bitcoin mining economy. Miners can be studied as economic agents who invest in mining hardware and join a mining pool to ensure a stable income. Joining or leaving a pool is a decision miners make to maximize their profit and counteract the possible domination of pools in the network. In return, miners' behaviors also affect how mining pools set their policies (e.g., payout schemes and pool fees) to compete in the market. Nonetheless, the internal dynamics of miners and the impact on the mining economy are not yet well understood. Few methods exist that allow studying a larger number of mining pools over the entire Bitcoin history.

In this chapter, I propose a data analysis method to detect miners and analyze miners that migrated among mining pools. First, I derived a heuristics algorithm identifying individual miners from the mining pools' reward payout flows. Next, I derived quantitative measurements to estimate miners who move to another pool (called *pool hopping*) or received reward shares from more than one pool (called *cross pooling*). Finally, I produced visualizations to explore patterns of miners' migrations compared to different payout schemes and pool fees. Working with my economist collaborator, we studied the possible factors that impact miners' decisions to join or leave mining pools over Bitcoin history. We found evidence that miners make economic decisions to select a pool and that mining pools compete to offer better reward incentives to attract miners.

> This chapter is an updated and extended version of my original article published at *IEEE International Conference on Blockchain and Cryptocurrency (ICBC 2021)* [154]. The work was led by myself in collaboration with Nicolas Soulié, Nicolas Heulot, and Petra Isenberg.

5.1 RELATED WORK

Bitcoin mining has become an industry where miners gather into pools to maximize their investments in mining devices [119]. Choosing a pool becomes a strategic economic decision for miners as a pool's characteristics (e.g., payout schemes and pool fees) greatly affect a miner's income. Some previous theoretical work related to pool hopping analysis has applied game theory to explain the motivation behind pool selection and miners' migration [96].

Lewenberg et al. [90] 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. [138] 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 where all miners received the best income. Liu et al. [94] considered the hash rate and the block propagation delay as metrics in their analysis and numerical simulation. They found that miners' strategies will converge at the market equilibrium when there is a dominant strategy and no miner can switch pools without undermining some other miner's payoff. Altman et al. [3] 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 entirely rational.

Even though there is a lot of work on the analyses of pool hopping behavior in theoretical studies, there are few works that provide an empirical analysis on this topic. Belotti et al. [13] investigated pool hopping between KanoPool and SlushPool from 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 higher profit. However, the result is varied among pool-hopping miners, and their profit gain is negligible. Romoti et al. [127] presented reward payout flow patterns of three pools: BTC.com, AntPool, and ViaBTC, between block 510,000 and 514,032 (1 month) and detected overlapped miners in those pairs. They noticed high cross-pooling between BTC.com and AntPool because the same company, Bitmain, owns both pools. Xia et al. [V18] developed a visualization tool showing the internal address networks of mining pools and the estimated number of pool hoppers for each pool. Xia et al. is closely related to my study but deviates in several areas. I propose a miners' migration flow model and measurements to detect different types of pool hoppers over long time intervals. The miners extraction method is also less computationally expensive, although it shares a similar underlying concept to detect miners.

This chapter contributes to an empirical analysis attempting to detect miners in mining pools and analyze miners' migration patterns among pools in the Bitcoin network. Compared to most past work, my proposed method can be used to analyze more mining pools across Bitcoin's mining history. Moreover, I compare the results with off-chain information (e.g., market shares, payout schemes, and pool fees) to help explain mining pools' evolution and pool hopping behavior.

5.2 DATA PREPARATION

I trace the reward payout from coinbase transaction to miners. As Bitcoin transactions do not include any identity of miners and mining pools, I extract off-chain information from external sources to identify mining pools that receive mining rewards and trace pools' reward sharing to detect miners who participated in each pool. I provide the data about mining pool attribution and miner's migration in the public repository https://zenodo.org/record/4342747.

5.2.1 *Mining pool payout flows*

Mining pools collect mining rewards from the Bitcoin network before distributing them to pool members. I track reward payouts from coinbase transactions and follow the money until it reaches the miners' addresses. I adopt a transaction flow graph and transaction purity notions to detect miners in the pool automatically.

A *transaction flow* is a directed graph of Bitcoin transactions from a seeding transaction. Each node represents a transaction $\mathtt{tx}$ in the transaction flow. *Transaction purity* measure determines 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., [1, 32]).

Let $\mathtt{tx.in}$ and $\mathtt{tx.out}$ be sets of receiving (inputs) and spending (outputs) edges of a transaction $\mathtt{tx}$ 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 $\mathtt{tx}$ can be expressed as follows:

$$\mathtt{purity}(\mathtt{tx}) = \frac{\sum_{e \in \mathtt{tx.in}} \mathtt{purity}(e.\mathtt{receive}) \cdot e.\mathtt{value}}{\sum_{e \in \mathtt{tx.in}} e.\mathtt{value}} \tag{2}$$

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

Reward payout flow model: Past work reported that mining pools distribute rewards to individual miners in different patterns [92, 127]. I propose that those different payout patterns can be generalized by annotating four transaction types in the payout flow: coinbase (● $\mathtt{tx}_{\mathtt{coinbase}}$), payout (● $\mathtt{tx}_{\mathtt{payout}}$), intermediate (● $\mathtt{tx}_{\mathtt{inter}}$), and miner (● $\mathtt{tx}_{\mathtt{miner}}$).

1. A mining pool receives mining rewards from coinbase transactions ● $\mathtt{tx}_{\mathtt{coinbase}}$ and collects them in a payout transaction ● $\mathtt{tx}_{\mathtt{payout}}$ before distributing them 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 a reward from tx_{inter} and spend it in a transaction called miner transaction tx_{miner}. I assume 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.

In short, the *reward payout flow* is the Bitcoin transaction flow from a payout transaction tx_{payout} to pool members tx_{miner}. I considered tx_{payout} as the seeding transaction because it collects every mining reward and distributes it to pool members. Figure 24 shows reward payout flow patterns extracted from the top 10 mining pools in my study. A payout flow was sampled for each pool in the month when it 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 values are the top and the right labels for each node, respectively. I omitted labels for combined values below 1 BTC.

Reward payout flows extraction: Based on the reward payout flow model, I present Algorithm 1 to automatically extract payout flows from the coinbase transaction and follow the transaction network until it finds miners. I used the BlockSci API [78] to access the transaction data. First, the list of tx_{payout} has been initiated from all outputs of $tx_{coinbase}$ as inputs of the algorithm. For each tx_{payout}, the algorithm traversed the transaction graph from tx_{payout} which has purity $= 1$ until the transaction has $purity_{tx} < 1$ (i. e. tx_{miner}). Finally, the algorithm returns a directed edge list that represents the payout flow.

I added two additional termination criteria $valid(tx)$ that stop following the current transaction tx flow: 1) when the 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 [13, V18].

Identifying individual miners: For each edge list obtained from Algorithm 1, I constructed a payout flow graph using the NetworkX library [58]. Next, I extracted tx_{miner} and derived the list of miners from each payout flow graph.

Miner transaction (tx_{miner}) is a transaction in the payout flow graph that does not have any output in the payout flow graph $|tx_{miner}.out| = 0$. I tagged all input edge(s) of tx_{miner} as owner

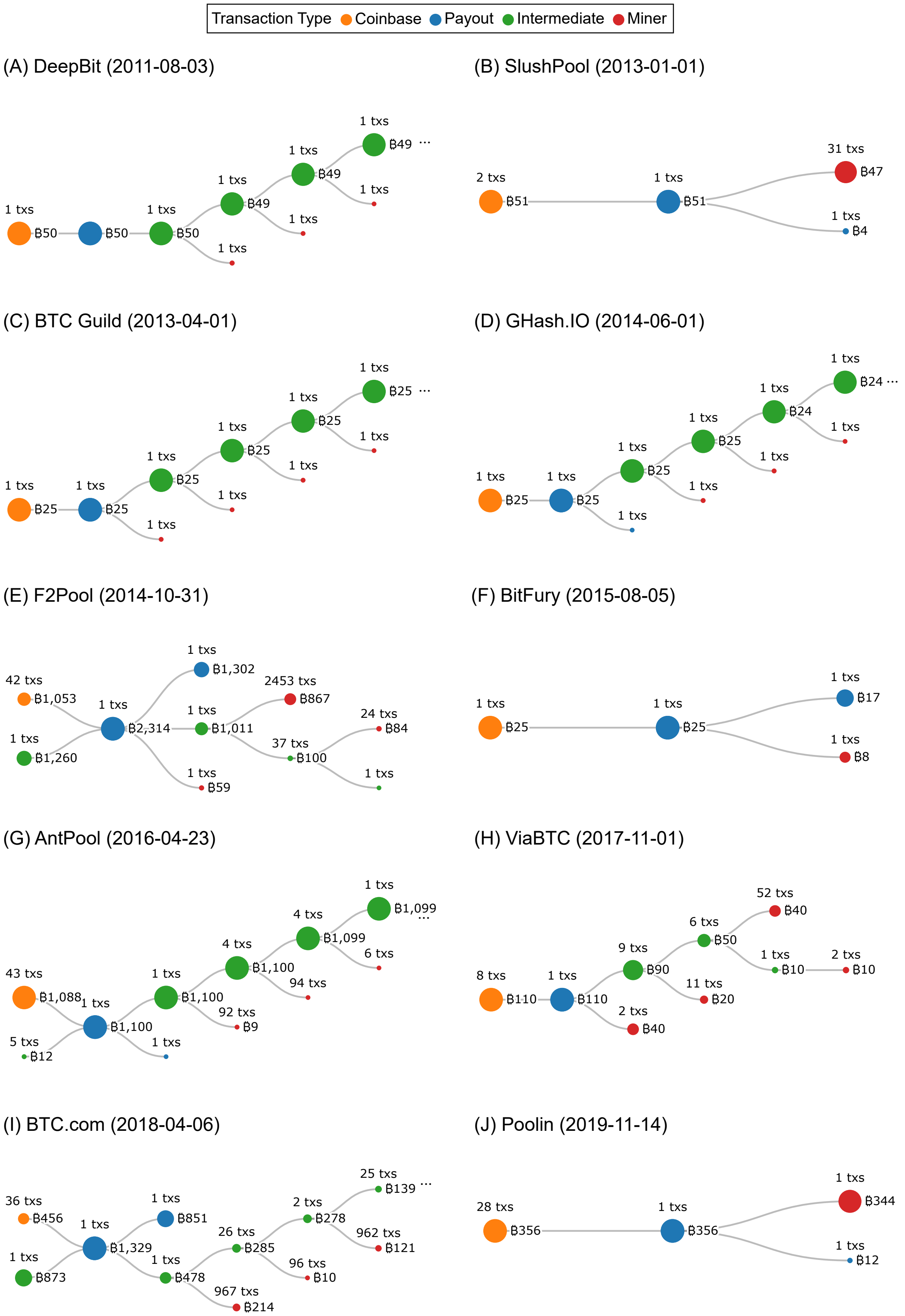

Figure 24: Reward payout flow patterns extracted from the top 10 mining pools in this study. Each node represents a transaction type with branches of similar patterns grouped together. The color of the node indicates the transaction type. The total value of transactions in each node is encoded by circle size in proportion to the tx_{payout} value.

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 for
    end if
end while
```

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 deposits it as reserve the reward to pay for the next payout, as illustrated in Figure 24 (E), (F), and (I). I detected ● tx_{miner} input edges with the same $owner$ addresses as the mining pool and assigned them as ● tx_{payout} to extract further reward payout flows.

5.2.2 *Miners' migration between mining pools*

To analyze miner migration between pools, I 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. I set the time interval to months to be able to analyze detailed patterns for the entire mining pool history.

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.

The *miner's migration flow* is modelled as a diagram in Figure 25. For each time interval t, the list of miners that migrate from/to a mining pool $pool$, annotated as M^t_{pool}, 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^t_{pool} but are also in M^{t-1}_{pool} (M^{t+1}_{pool}).

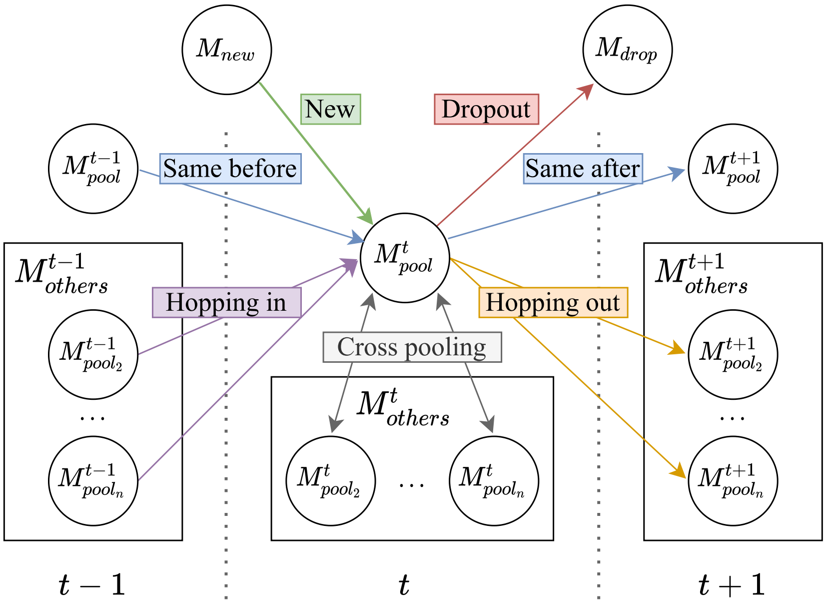

Figure 25: The miners' migration flow model of $\mathtt{pool}$ at time interval t. $t-1$ $(t+1)$ is the time interval before (resp. after) t. M_{new} (M_{drop}) is the list of miners not in any pool at $t-1$ (resp. is not found in any mining pool at $t+1$). The union of the list of miners from other pools is M_{others}.

- *Hopping in (Hopping out)* miners are in M^t_{pool} but move from (to) other pools $M^{t-1}_{others|pool}$ ($M^{t+1}_{others|pool}$).
- *Cross-pooling* miners are in M^t_{pool} but also found in other pools at the same t ($M^t_{others|pool}$).

The quantity of miners' migration is calculated from the percentage of the total reward of miners for each miner group. I report the percentage of value rather than the number of addresses because it gives more weight to miners with a high contribution to the pool. Therefore, the measure is more robust regarding small or occasional miners.

The *percentage of the total value of miners* (X) is the total value of M^t_{pool} associated with M_x, where x is a set of miners from *miner groups*, defined as:

$$X(M^t_{pool}, M_x) = \frac{\sum_{m \in M^t_{pool} \cap M_x} m.value}{\sum_{m \in M^t_{pool}} m.value} \quad (3)$$

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

For each mining pool, I obtained the monthly percentage of miners' migration for each miner group. As there are many variables to consider in the miners' migration flow, I summarized *miners' migration flows* into three flow types. Each type of the flow indicates a net gain or loss percentage of the miners in the mining pool from different flow types.

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.
3. *Cross-pooling:* the percentage of cross-pooling miners. This is the only flow that compared miners within the same time period.

Additionally, *the percentage of cross miners' rewards from the pool* is 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.

5.2.3 *Evaluation and discussion on the data preparation method*

This section discusses the validity and quality of the data preparation method and compares it to related work. Table 5 shows the median and median absolute deviation (MAD) on the number of miners, path length from payout transaction to miners, and the number of coinbase transactions per payout, for each mining pool in this study. I detected that these measures are non-normal distributions in many mining pools. Thus, I report robust statistics measures that provide a more reliable point estimation and variance of the distributions.

5.2.3.1 *Assumption on the payout flow model*

The primary assumption of our approach is that mining pools spend only the mining reward from coinbase transactions to distribute among their miners. Therefore, the purity of payout and intermediate transactions is set to 1. However, this assumption did not apply to 6 mining pools in our study: ASICMiner, Binance Pool, BitFury, DPOOL, KnCMiner, OzCoin, and Poolin. I observed that these pools sent all rewards to one or two miner addresses which are likely to be miners' addresses, for instance, Poolin in Figure 24 (J). This can be detected by a very low median number of miners in Table 5. Most of them were small pools and tended to operate in a short period (Figure 20). BitFury is likely to be a private pool as they keep the reward in their wallets and do not show an obvious payout pattern. Binance provides wallet and exchange services to miners in addition to mining pools.

Table 5: List of the top-30 mining pools with the number of blocks until August 31, 2021. I reported median and median absolute deviation (MAD) of the number of miners, miners' path lengths, and the number of $tx_{coinbase}$ per tx_{payout}. Mining pools that I cannot extract the payout pattern from the algorithm (resp. has no pool characteristics information) are highlighted in red (grey).

Mining Pool	# Blocks	# Miners		Path Length		# $tx_{coinbase}$	
		Median	MAD	Median	MAD	Median	MAD
F2Pool	65,772	3,351	1,469	3	0	5	4
AntPool	57,271	187	186	13	8	8	7
BTC.com	36,811	1,519	1,513	3	0	16	10
SlushPool	35,240	162	104	2	0	1	0
BTC Guild	32,936	38	29	4	2	1	0
DeepBit	31,107	30	12	23	13	1	0
ViaBTC	23,157	35	28	3	0	5	3
GHash.IO	23,083	6	4	2	0	1	0
Poolin	21,097	2	0	4	1	1	0
BitFury	20,901	1	0	3	0	1	0
BTCC Pool	18,036	1	0	4	2	1	0
BTC.TOP	16,748	184	183	2	0	7	5
BW.COM	12,733	1	0	16	12	4	2
Eligius	11,430	62	39	1	0	1	0
Huobi	9,155	16	7	4	1	1	0
50BTC	7,859	25	19	42	31	1	0
KnCMiner	7,477	2	1	4	2	1	0
BitMinter	6,464	87	46	4	2	1	0
Binance Pool	6,046	2	1	3	0	10	5
EclipseMC	6,024	43	23	6	3	1	0
Bixin	5,852	2	0	3	1	1	0
BitClub Network	5,672	2	1	7	3	29	27
1THash	4,967	177	175	2	0	5	3
OzCoin	4,845	1	0	10	7	1	0
okpool.top	3,153	9	4	4	1	1	0
ASICMiner	3,146	1	0	10	5	30	26
Bitcoin.com	2,475	11	6	3	1	1	0
KanoPool	2,432	795	253	2	0	1	0
GBMiners	2,093	2	1	8	3	4	2
DPOOL	1,918	1	0	3	0	6	2

The pool sent all the reward to their wallet adress and could draw bitcoin value from other services to pay miners. In these cases, I cannot track miners from payout flow because the pool has its own mechanism or use money outside the flow to pay miners.

The payout flow extraction algorithm stops crawling the payout flow because the pools may add bitcoin values outside the payout flow (e.g., their saving account), and therefore the purity of the transaction is < 1. To fix this problem, the purity threshold could be adjusted to less than 1 in case that mining pools combine bitcoin values from non-mining related sources. I tried a lower purity threshold and found that it is computationally much slower. Besides, it tends to classify miners as intermediate transactions and continues following the flow because miners can also combine the reward with other sources of bitcoin values to spend in Bitcoin marketplaces and services.

5.2.3.2 *Justification on using addresses to identify miners*

Another important assumption is that individual miners who receive a reward share will spend it in a transaction that includes input transactions 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 $tx_{inter}.out$.

Although the basic address clustering method [123] is an effective method to group the addresses that are likely to belong to the same entity [61], I 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. I expect that miners would participate in 1–2 pools at a time. I report the average number of mining pools that miners participated as the average weighted by their total reward with the 95% bootstrap confidence intervals. During the first halving (second halving) period, miners received rewards from 3.92 [3.12, 4.74] (3.06 [2.47, 3.68]) different pools per cluster compared to 1.46 [1.39, 1.56] (1.30 [1.24, 1.37]) pools per address. I also found that the percentage of cross-pooling per month using address clustering is higher than using solely miner addresses on the average of 25.9% [24.2%, 27.6%] (31.6% [29.7%, 33.7%]), using pairwise comparison for the same pool and month. According to this result, I decided to use miner addresses to avoid adding errors from the address clustering to the results.

5.2.3.3 *Miners addresses association with known entities*

Since there is no ground truth to evaluate the identity of individual miners, I indirectly validated whether our approach can identify individual miners correctly. I assumed that miners should receive a

Table 6: The percentage of addresses and total Bitcoin values associated with known entities from 2013 to 2016

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

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).

I used a known entity dataset from WalletExplorer.com with entity type classification from Zola et al. [178]. I studied the payout flows that were spent between 2013 and 2016 because the website stopped updating more known entities from 2016 [71]. The percentage of addresses and Bitcoin values for each entity type is reported in Table 6. The main findings from the result are as follows:

1. Miners detected from our algorithm mostly cannot be associated with any known address ("unknown" type in Table 6) (84.8%) as well as input and output address (96.1% and 91.5% resp.). However, when I measured the total value for each entity type, I found that 84.8% of miner rewards are from unknown addresses, compared to 44.0% for input and 68.6% for output addresses. The result shows 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 the assumption that miners should receive money from ● $tx_{\mathtt{inter}}$ 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 reward using unknown output addresses. I also detected that some miners spent their rewards on exchanges (18% of the total value) and wallet services (12.6%). This result provides evidence that regular miners con-

vert mining rewards to fiat currencies or deposit them to their Bitcoin wallets.

The pool hopping detection approach I developed differs from Xia et al.'s work [V18] as I did not filter out known entities after I extracted the payout flow based on three reasons: 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 did not update new entity labels after 2016 [71]. Hence, it cannot be applied to recent reward payout flows; and 3) The percentages of reward values are tolerant to possible misclassification of miners.

5.3 CHARACTERIZATION OF REWARD PAYOUT FLOW PATTERNS

Using the payout flows I extracted from the previous section, I analyzed the patterns of payout flows and reward payout regularity. Previous work has reported the payout patterns for some mining pools in a limited time interval. Romiti et al. explore the payout flow pattern of three miners: BTC.com, AntPool, and ViaBTC [127]. They reported that BTC.com and AntPool used an address to distribute rewards in multiple steps, while ViaBTC divided the reward to random addresses 10 BTC each before payout to miners. This work extracted the payout flow of a more significant number of mining pools over the entire Bitcoin history.

Liu et al. reviewed the payout patterns from previous studies and classified them into *direct* and *indirect* distributions [93]. The payout structure of indirect distributions can be divided into *tree-like* and *chain-like* structures. Yet, they did not consider the regularity that mining pools distribute the reward. In this work, I proposed two quantitative measures to characterize the reward payout flows based on reward payout regularity and payout flow structures.

1. **Payout regularity** can be measured from the correlation between blocks mined and the number of payouts per month. Mining pools that *pay per block*(s) (or per round) to miners should have a high positive correlation since they distributed rewards to miners when they obtained the reward from coinbase transaction. In contrast, mining pools that *pay miners regularly* (e.g., daily) will have a fixed number of payout transactions regardless of the block mined. Therefore, there should be no correlation among those regular payout pools.

2. **Payout flow structures** determine the complexity of reward payout flows. Mining pools can distribute rewards to miners directly from the payout transaction or forward them to intermediate addresses before paying to miners in multiple steps. I calculated the path length for each miner to see how many steps it

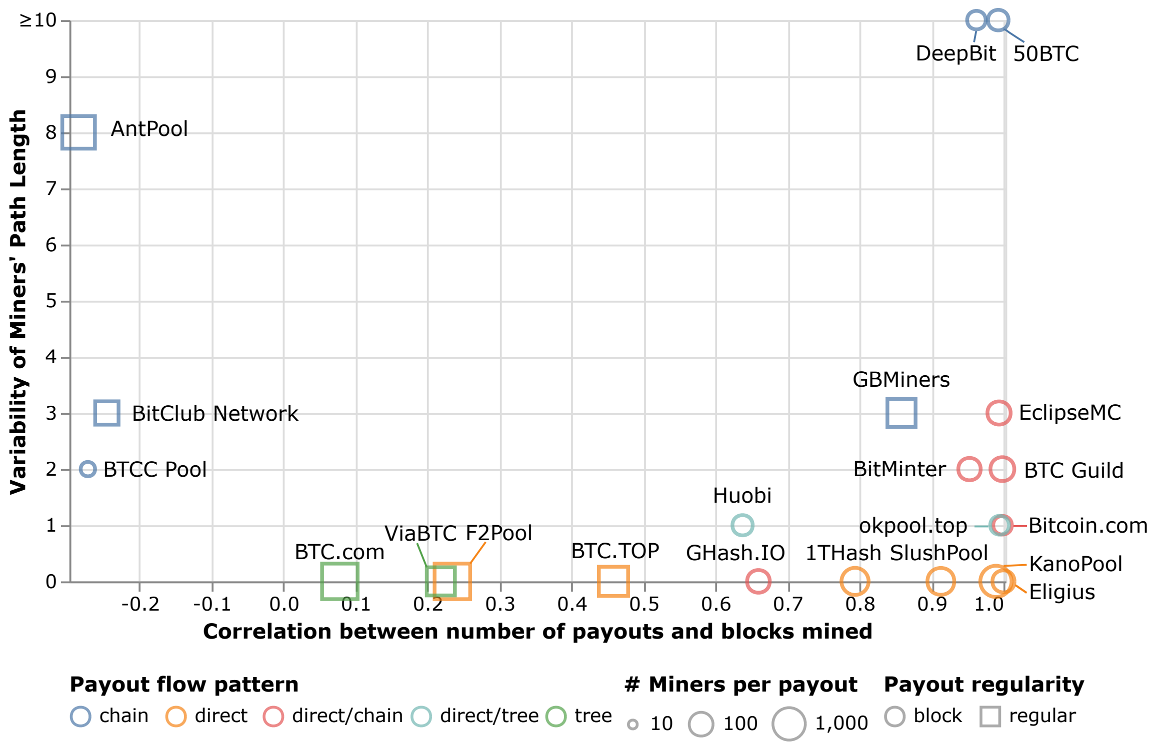

Figure 26: The scatter plot shows the payout flow pattern of mining pools regarding the payout regularity and payout flow structure. The color indicates the payout flow structure for each mining pool. The shape suggests whether a mining pool pay per block (round) or regularly (e.g., daily). The size of each point is Proportional to the median number of miners per payout on the logarithmic scale.

took for miners to receive the reward from the payout transaction. I measured the median absolute deviation (MAD) of path lengths to describe the payout pattern for each mining pool. The large MAD indicates a high variability of path length. It implies that miners received rewards in different steps, and therefore the payout flow looked like a *chain-like* or a *tree-like* structure. The low MAD means that miners were likely to receive the reward in the same step. Therefore, the mining pool tends to have a *direct* or *fixed-length* payout flow pattern.

Figure 26 reports payout regularity, and miners' path length variability for each mining pool in a scatter plot. I looked into samples of payout flows for each mining pool and annotated what I found as the shape (regularity) and color (payout flow structure). The payout patterns I observed align into four clusters based on two axes: 1) whether mining pools *pay per block* (per round) or *pay regularly* (e.g., per day) and 2) the payout of structure tends to be the *fixed-length* or *chain-like* structure.

5.3.1 *Payout regularity*

Mining pools in the early years between 2011 and 2013 commonly distributed reward shares to miners directly after they successfully mined a block. DeepBit is the first mining pool that dominated the market and spread the reward to miners for each block mined. After the inception of DeepBit, others mining pools also adopted the same policy to pay miners per block, such as Eligius, BTC Guild, BitMinter, and SlushPool. The correlation between payouts and blocks mined is close to 1, implying that mining pools pay miners as much as they receive from coinbase transactions.

The *pay-per-block* policy poses no risk for mining pools because they do not need to hold any funds to pay miners [128]. Instead, miners accepted uncertainty to keep a constant income as the expected reward time is in proportion to the market share of the pool. This policy induces some miners to mobilize between pools (cross-pool) to maximize their reward. After the first halving day, *pay-per-block* pools tended to be less successful in market share and eventually disappeared from the mining competition. Even though new mining pools emerged, they tended to be short-lived (e.g., 50BTC) due to the competition from mining pools which provided a more stable income to miners (e.g., F2Pool and AntPool). SlushPool is an exceptional pool that adopted this policy and is still active and has constantly mined new blocks from 2012 until now.

As the Bitcoin mining activity has grown exponentially since 2013, new mining pools offer miners a more *regular income* to attract miners. These mining pools collect mining rewards to their address before regularly distributing them to miners (i.e., daily). Therefore, the number of payout transactions remains constant over time and has a low correlation with blocks mined in the range of -0.3 and 0.6. F2Pool and AntPool are early pools that adopted this policy from 2013–2014. BTC.com and ViaBTC emerged around the second halving day in mid-2018 and adopted the same *regular payout* policy. Nowadays, these four mining pools remain the top pools in the mining market.

With the *regular payout* policy, miners are guaranteed to receive a predictable income from the pool. Mining pools need to cope with the risk from the uncertainty of Bitcoin mining and have some funds to pay miners when they do not manage to obtain enough rewards on bad luck days. This can be seen from the payout flow where mining pools send rewards to their addresses and spend them in subsequent payout transactions. Hence, mining pools with this risk tend to pose more fees to miners. The domination of regular payout pools shows that miners are willing to join mining pools that provide a steady income. Nowadays, the top mining pools distribute rewards to miners regularly, indicating that mining has become an industry since mining pools reserved some funds to pay miners and manage the risk from mining new blocks.

I spotted two mining pools whose correlations do not match our expectations in the study: BTCC Pool and GBMiners. BTCC Pool paid rewards to miners per block. However, the pool split coinbase transactions into numerous payout transactions with a tiny amount of bitcoins (e.g., `https://bit.ly/3FCcgPo`). Therefore, the number of payout transactions outnumber the block mined. GBMiners distribute, in multiple payout transactions (median = 2, MAD = 1), the reward to miners at the same hour every day. However, each payout transaction usually contains the median of 4 coinbase transactions (MAD = 2). This explains why the correlation of GBMiners is high even though it is a regular payout mining pool.

5.3.2 *Payout flow structure*

The variability of path length to miners (y-axis on Figure 26) can imply the complexity of payout flow structure ranging from the most straightforward pattern in which a mining pool distributes the reward to miners in a fixed path length to a more complex pattern in which it pays miners in multiple steps. The median path length variability of 0 implies that mining pools were likely to distribute rewards to miners at the same path length. I called this *"fixed-length"* payout structure. This pattern has been used among the pools that started operating between 2011 and 2014: SlushPool, GHash.IO, and KanoPool. During this time, I observed a constant increase in the transaction fees. Eligius is an exceptional pool in this study in which it paid miners *directly* from the output of coinbase transactions (median path length = 0).

Mining pools with this pattern send all of the mining rewards from coinbase transactions to their address and distribute them to all miners in the same payout transaction. Due to the simplicity of the payout flow, it minimizes the number of transactions for each payout. However, miners' addresses can be detected easily from the flow. Despite this concern, it has still been used among the recent small-size pools, such as KanoPool and 1THash. F2Pool and Binance Pool are an exception in that they offer a regular income to miners but still use this pattern to pay miners (path length = 3).

As the number of miners became higher (1,000–10,000 miners per payout), BTC.com and ViaBTC used a *tree-like* structure to distribute rewards to more miners in a few path lengths (path length = 3). They are currently the top mining pool with 12% and 9% of market shares, respectively (as of Aug. 2021). For each payout flow, the mining pools split rewards into multiple intermediate transactions before sending them to miners' addresses. These pools have a low miners' path length variability because miners received reward shares from the different intermediate transactions with the same path length.

In comparison, mining pools that distribute rewards to miners in a *"chain-like"* structure have a path length variability > 1. DeepBit distributed rewards to miners per block in a *long-chain* structure. A high path length variability characterizes this pattern. Three other pools, EclipseMC and 50BTC that came later, also used the same payout pattern. Mining pools sent the reward to a single miner in each step and kept the change to pay other miners in the subsequent transactions. Mining pools need to create as many transactions as many miners they need to pay. Nonetheless, it is feasible at this period due to the small number of miners and negligible transaction fees.

Active mining pools from 2012 to 2013 also adopt *chain-like* payout pattern to distribute rewards to multiple miners, including ElicpseMC, BTC Guild, BitMinter, and Bitcoin.com. At each step, these mining pools spread rewards to some miners due to the limitation of transaction size. Then, they forwarded the remaining reward in the following steps to pay to the remaining miners. Compared to the *long-chain*, this pattern helps reduce the number of intermediate transactions, and therefore, transaction fees to distribute rewards. Table 5 shows that the median path length of these pools is around 3–5 compared to 23 for DeepBit and 42 for 50BTC. Besides, I observed that these pools also sometimes distribute rewards to miners directly. As a result, the variability of their payout flow dropped to around 1–3 path lengths.

AntPool is the only regular payout mining pool that adopted the *chain-like* payout structure (Figure 27 (A)). The pool distributed the reward to 50–100 miners at each step. According to the icicle plot on the right, there is one ● tx_{inter} that holds a large reward. Even though I found more than one ● tx_{inter} at each step is shown in the node-link diagram, they are likely to be ● tx_{miner} that the algorithm misclassified as ● tx_{inter}. The median path length is highly correlated with the number of miners (path length = 13). Nonetheless, I detected that AntPool changed the pattern to *tree-like* payout flow from November 2019 (Figure 27 (B)). The icicle plot shows that AntPool distributes the reward into five equal ● tx_{payout}. After that, each ● tx_{payout} distributes the reward to miners in a *chain-like* structure, as the ● tx_{payout} in the following steps have a slightly less amount of Bitcoin than the previous one.

From this analysis, I witnessed the development of payout flow structure from a *chain-like* in the first mining pools to *fixed-length* patterns. Mining pools pay miners from the reward obtained from coinbase transactions and adopt a payout structure that minimizes transaction fee cost at a time. Therefore, most mining pools use simple payout flow patterns as indicated by the path length variability. Active large mining pools nowadays use *tree-like* distributions to pay a

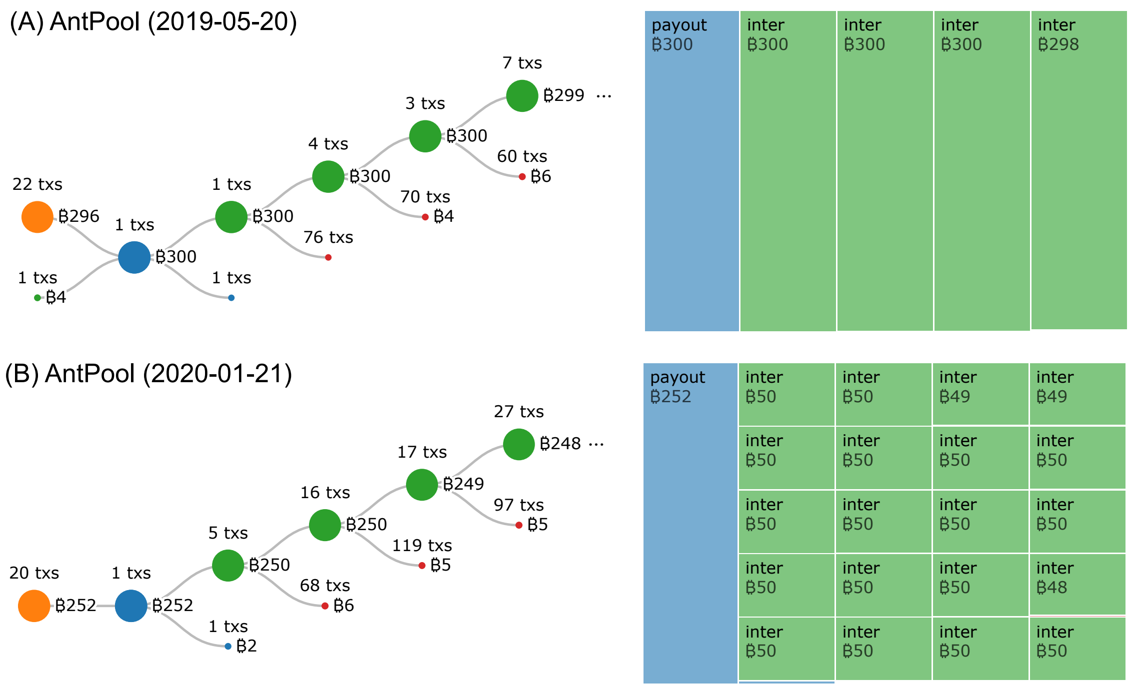

Figure 27: Reward payout flow patterns of AntPool. (A) AntPool paid miners in a chain-like payout structure. (B) After November 2019, AntPool adopt a tree-like payout structure to distribute the reward. The icicle plots on the right display the distribution of tx_{payout} and tx_{inter} for the first five path lengths in the payout flow.

large number of miners, such as and ViaBTC and BTC.com (Figure 26 H–I).

5.4 VISUALIZATION DESIGN

The miners' migration flow provides in-detail information on miners' mobility incoming from and outgoing to each mining pool and hopping with other pools. The economist wanted to assess whether the mobility of miners affects market shares and pool characteristics or vice versa. I provided two interactive visualizations based on a Jupyter Notebook that offered him an overview of three main miner flows, zoom and filter on a specific time period, and detail of the percentage of mobility between pools.

5.4.1 *Visualizing the overview of the miners' migration flow*

The first set of visualizations helps analysts understand the net flow of miners for each pool over time. As there are 28 mining pools to observe at a glance, I used the same bubble chart visualization design as in Figure 20. Each row represents the development of miners' flow for a particular mining pool. Mining pools are sorted by the time sequence when they mined the first block. The new and dropout miners flow statistics (resp. the hopping in and out flow) are displayed in Figure 28 (Figure 29). The absolute percentage difference is encoded as the circle size. The color indicates the direction of the difference: green for a positive and red for a negative number. For the cross-pooling flow, the circle size is proportional to the percentage of cross-pooling miners. I used a sequential single-hue color scheme to represent the percentage of reward that cross-pool miners received from the pool. The result from the cross-pooling flow is shown in Figure 31. These three figures are concatenated horizontally in the notebook.

5.4.2 *Visualizing the detail of pool hopping and cross-pooling miners*

After the economist figured out the time period of interest, his research questions was 1) how many miners migrate from one pool to another or cross cross with other pools and 2) can we detect significant miners' flow between pools? The overview visualization could not display detailed information on the relationship among pools. Hence, I developed a compound chart to allow the economist to look in detail at the pool hopping and cross pooling over a time period of interest. The compound chart consists of three elements: 1) a stacked bar showing the total payout amount to miners and cross-pooling miners; 2) a heatmap showing the percentage of cross-pooling between mining pools; and 3) context information about reward payout schemes and transaction fees for each mining pool. I showed four insights that this visualization produced in Section 5.5.4.

5.5 RESULT AND ANALYSIS

Based on miners' migration flow visualizations, we explored whether the collective behavior of miners affects the evolution of market share and pool characteristics (i.e., payout schemes and pool fees). First, we focused on the competition between pools based on payout schemes and transaction fees. Then, we investigated market entry and the expected revenue of new miners. Finally, we analyzed miners' cross-pooling behavior that helps to diversify income and risks.

5.5.1 *Miners' migration flow and Bitcoin market price*

Overall, the new and dropout miners' flow (Figure 28) shows that more new miners were joining the top mining pools over time. The evidence aligns with the total hash rate that has been growing very fast over time (Figure 19). Compared to the Bitcoin market price, we observed that many new miners entering pools and pool creations happened on the local peaks in the market price: June 2011 ($19), April 2013 ($130), February 2013–March 2014 ($800), June 2014 ($600), December 2015 ($420), June 2016 ($630), May–December 2017 ($15,000) and June–December 2019 ($10,500). We also observed a significant miner dropout from mining pools close to halving days, which correspond to periods of the sharp decrease in mining revenue.

5.5.2 *Mining 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. Two prime payout schemes (PPS and PPLNS) provide different reward incentives, either risk-averse or risk-seeking strategies. Pool fees are used as a competitive advantage for mining pools. The lower the pool fee, the more attractive it should be to miners. We compared the reward payout scheme (Figure 20) of mining pools with the migration flows to examine whether miners were affected by the mining pools' competition.

Within each payout scheme type, Figure 20 illustrates that 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 (Figure 28) and hopping-in miners (Figure 29 (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

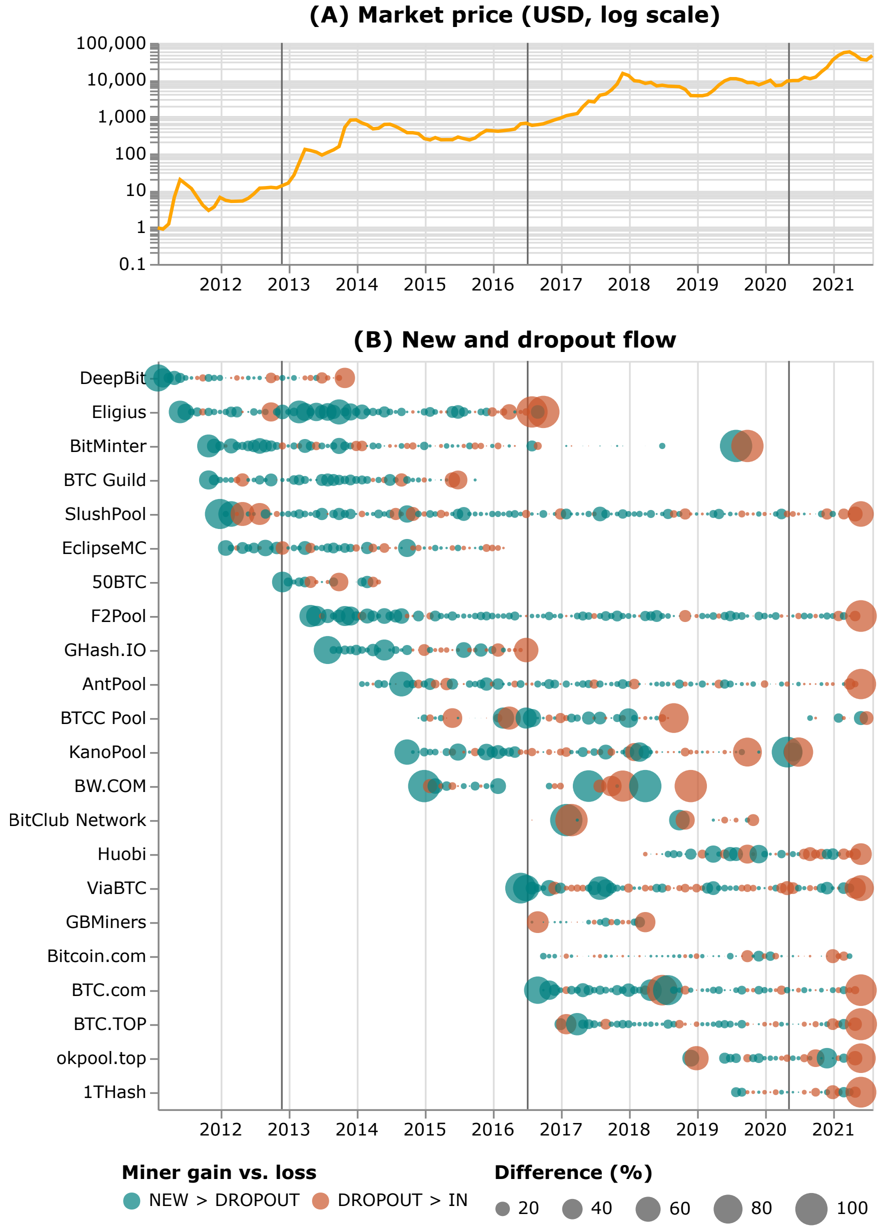

Figure 28: New and dropout flow. (A) The line chart displays the market price of Bitcoin in US Dollar on the log scale. (B)The size of the circle represents the absolute difference between new and dropout miners. New miners' positive (or negative) flow is encoded with the green (or red) color.

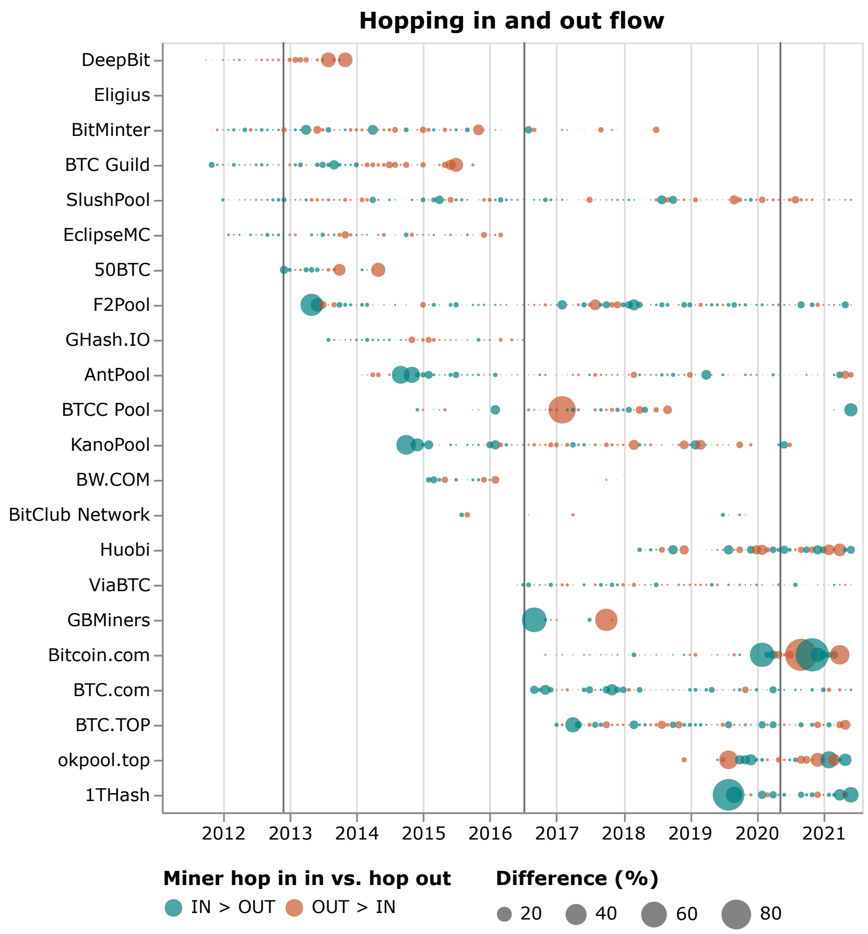

Figure 29: Hopping in and out flow. The size of the circle represents the absolute difference between hopping in and hopping out miners. The color indicates more hop-in miners than hop-out miners (green) or vice versa (red).

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 pools' average PPS fees, 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 (Figure 28) and hopped out from older pools (Figure 29).

In 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 for pool-hoping we observed in our study is the gap between pool fees 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-hopping flow decreased.

5.5.3 *The impact of Chinese government policy on Bitcoin mining pools*

Chinese mining pools (e.g., AntPool and F2Pool) emerged in the market around 2013 and increasingly became the dominant source of Bitcoin mining power. As shown in Figure 22, the total market share of Chinese mining pools exceeded the 51% majority attack threshold from March 2015 and continued until early-2017. At this point, the mining community raised concerns that Chinese pools had relatively too much computational power [114, 122] and could collectively pose a threat to the Bitcoin network [77, 134].

The Chinese authorities monitored activities in Bitcoin closely [70]. The People's Bank of China banned cryptocurrency trading and peer-to-peer lending in September 2017 [**Investopedia_2019-06-25**]. Compared to our pool characteristics information in Figure 20, many Chinese pools in this period transformed themselves into global pools, including three top mining pools: BTC.COM (March 2017), AntPool (August 2017), and F2Pool (September 2019). We also saw two Chinese pools, BTCC Pool and BW.COM, that lost their market share and stopped their operation during this time (Figure 20).

In May 2021, Chinese authorities shut down Bitcoin trading and mining by cutting electricity supply to mining farms [141]. The total hash rate dropped suddenly in this month (Figure 30 (A)). From the miners' migration flow in Figure 28, we found a large number of dropout miners for most of the active pools, such as BTC.COM, BTC.TOP and okpool.top. Interestingly, the market shares of Chinese

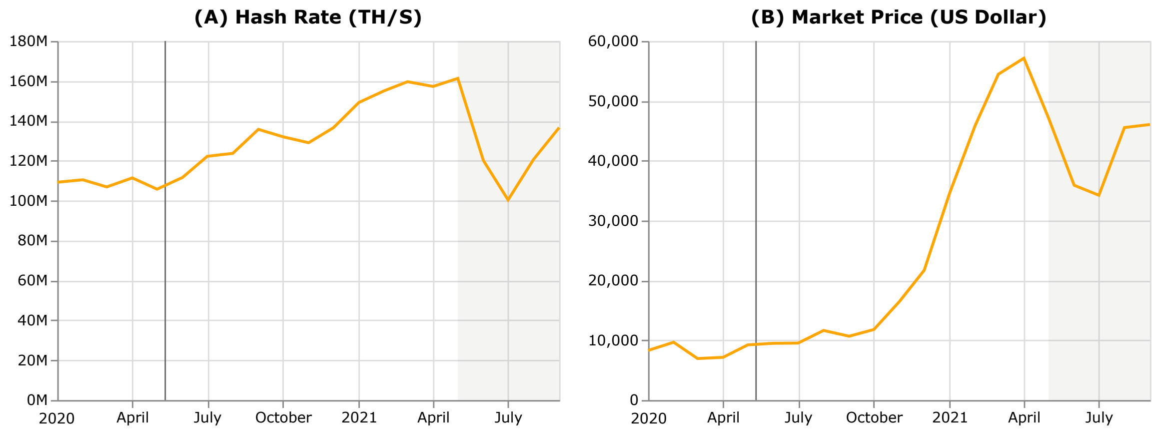

Figure 30: (A) The total hash rate (in TH/S) and (B) the market price (in US Dollar) of Bitcoin from January 2020 to September 2021. The time period that the Chinese government shut down mining farms is highlighted in grey.

pools (e.g., Huobi, 1THash, and BTC.TOP) suddenly dropped after this event (Figure 30 (B)). Nowadays, miners moved their facilities to America, and Central Asia [125], while the total hash rate gradually recovered (Figure 30 (A)).

5.5.4 *Payout scheme, income optimization, and cross-pooling*

Cross-pooling should imply that miners try to diversify risks and optimize income from mining pools. Figure 31 shows that cross-pooling tends to be very important at the beginning of many new pools. This cross-pooling occurs from incumbent pools toward the new pools, especially if the latter offers similar or better expected income. We identified four periods of intense cross-pooling shown in Figure 31 and further explored cross-pooling and pool hopping behaviors between pools. We chose four time periods where we detected a high percentage of cross-pooing to investigate what happened in further detail.

April 2013 – November 2013: This period corresponded to the switch from PPS toward the PPLNS scheme for BTC Guild as we discovered from Figure 32. BTC Guild applied PPS until February 2013, then proposed PPS and PPLNS until March 2014. After that, it offered PPLNS uniquely. Before that time, cross-pooling was very low. F2Pool proposed remunerating miners using a PPS reward scheme with a 4% pool's fee when entering the mining market in 2013. At that time (May 2013), the three biggest pools were BTC Guild (PPS, 5% fee), 50 BTC (PPS, 3%), and SlushPool (Score, 2%). The entry of F2Pool is associated with an important and cross-pooling between those pools, especially with BTC Guild, which provides a higher transaction fee.

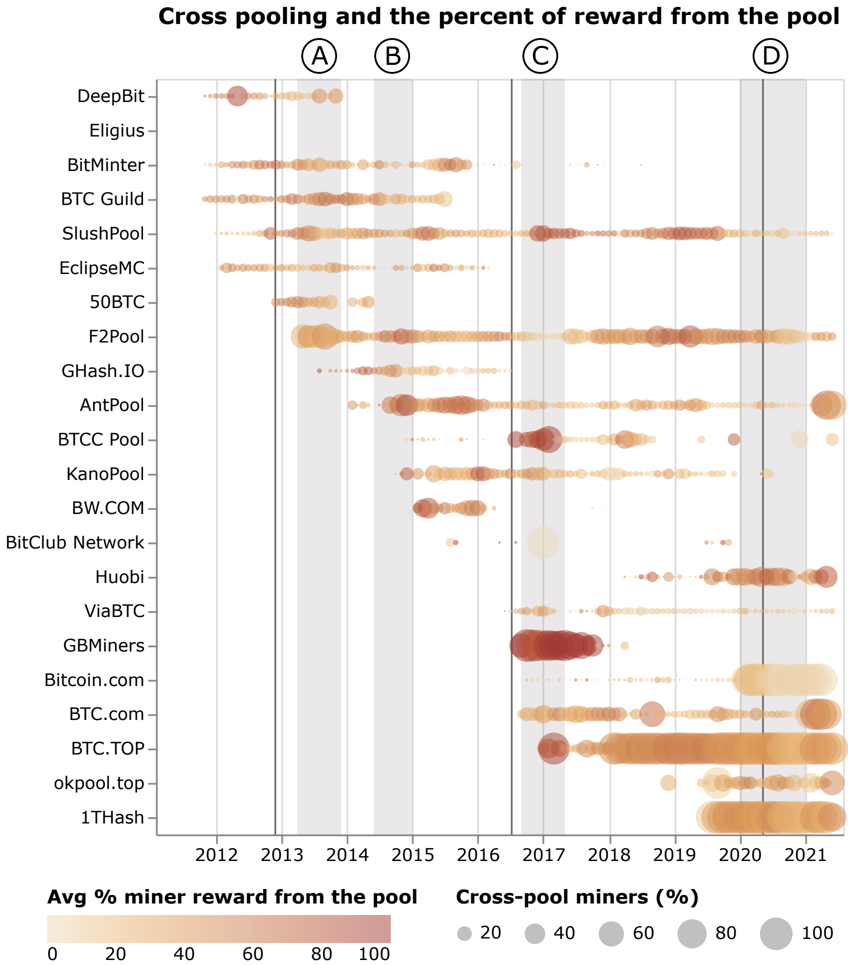

Figure 31: Cross pooling flow. The size of the circle corresponds to the percentage of cross-pooling of all miners in a given month. The percentage of reward that cross-pooling miners obtained from the pool is encoded using a brown color scale. Four periods with a high cross-pooling are highlighted in the grey background: (A) April 2013 – November 2013, (B) June 2014 – December 2014, (C) September 2016 – April 2017, and (D) January 2020 – December 2020.

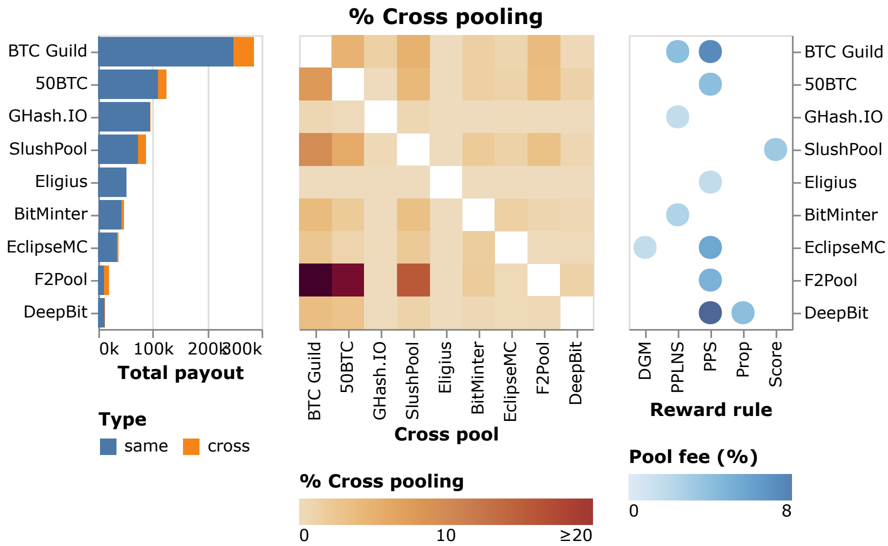

Figure 32: The compound chart displays the total payout (left), reward payout scheme (right), and cross-pooling between mining pools (middle) between April 2013 and November 2013.

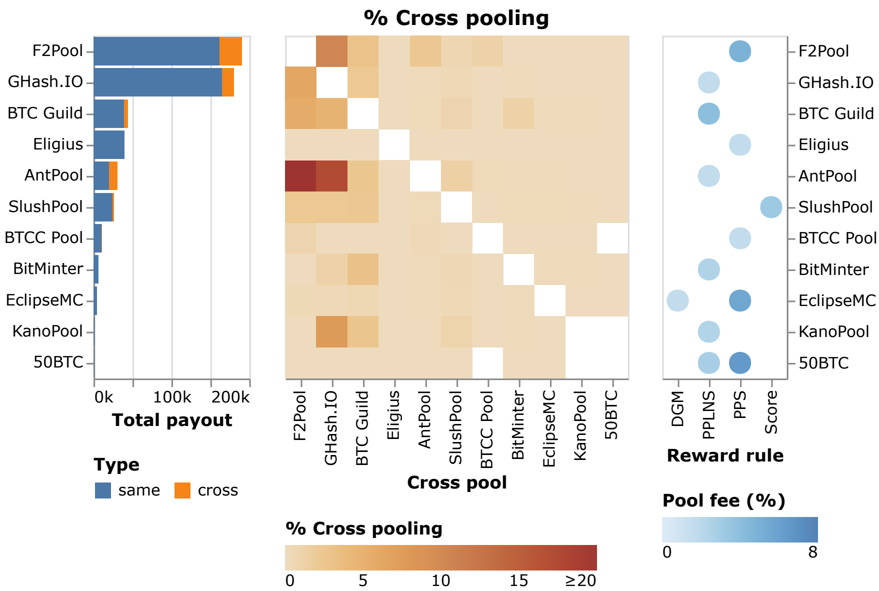

Figure 33: The compound chart displays the total payout (left), reward payout scheme (right), and cross-pooling between mining pools (middle) between June 2014 and December 2014.

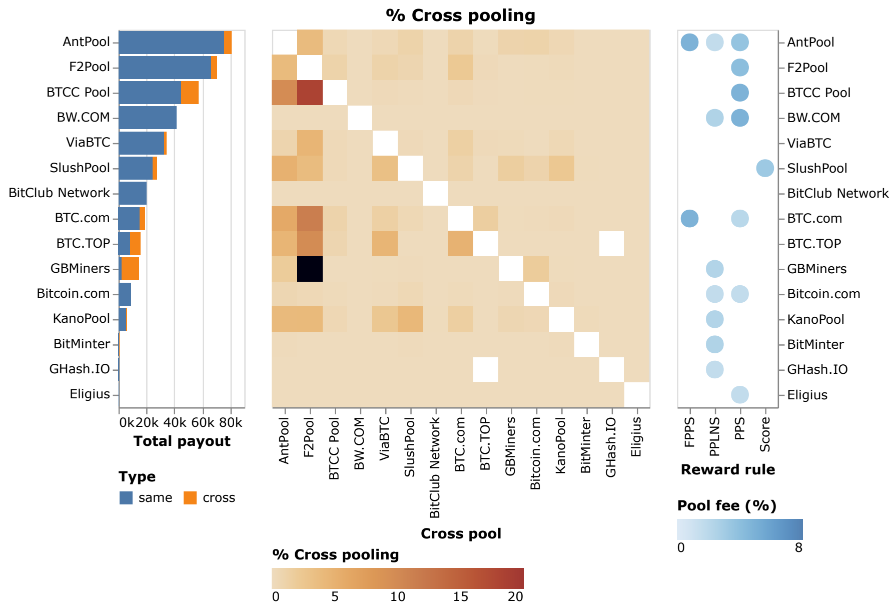

Figure 34: The compound chart displays the total payout (left), reward payout scheme (right), and cross-pooling between mining pools (middle) between September 2016 and April 2017.

June 2014 – December 2014: The period followed a similar pattern as the previous one, except for AntPool, a newly created pool in this period. Before the creation of AntPool, cross-pooling was limited. Once AntPool launched the PPLNS payout scheme (in addition to PPS), cross-pooling rose considerably. Figure 33 shows that AntPool (PPLNS, 0%) generated large cross-pooling with the main existing pools: GHash.IO (PPLNS, 0%), F2Pool (PPS, 4%), and Eligius (PPS, ~0%).

September 2016 – April 2017: 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). Figure 34 shows that a large cross-pooling flow existed between AntPool and BTCC Pool, and also other PPS pools (BTC.com or F2Pool). BW.com which entered the market in 2015 (PPS, 4%; PPLNS, 1%) is also associated with cross-pooling with the currently largest pools—F2Pool (PPS, 4%) and AntPool (PPLNS, 0%).

Surprisingly, we also found that GBMiners has a very high cross-pooling with F2Pool. We checked the payout flow of GBMiners during this period and found that it has two different payout flows, as shown in Figure 35 (A–B). On one hand, the pool distributed rewards in a *chain-like* payout flow pattern (A). On the other hand, GBMiners sent all payout rewards to a single miner. We suspect that a miner address is the external address of the mining pool (B). We detected that those miners happen to be an address of the F2Pool miners. This

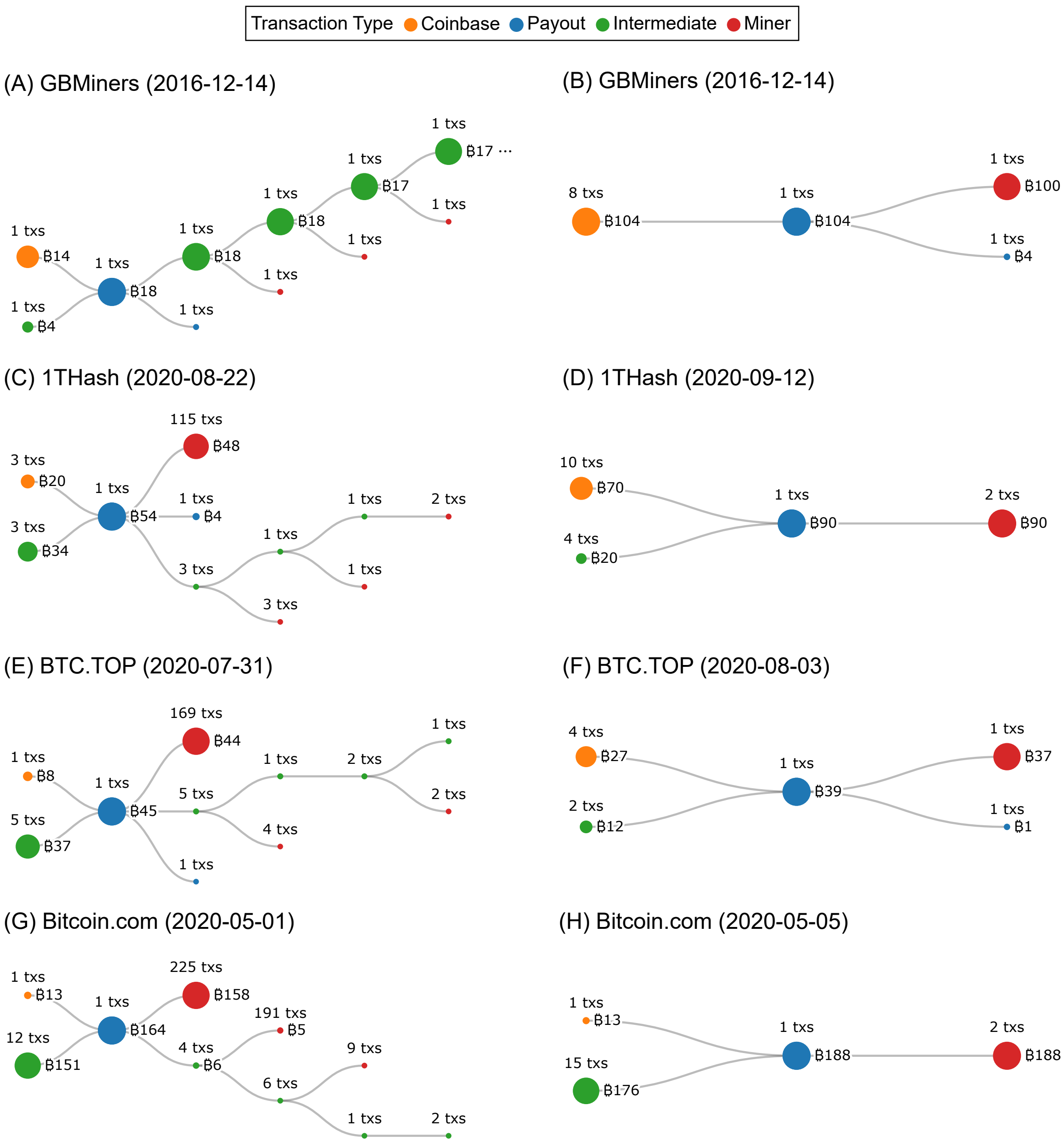

Figure 35: Reward payout flow patterns of mining pools that has high cross-pooling miners: (A–B) GBMiners, (C–D) 1THash, (E–F) BTC.TOP, and (G–H) Bitcoin.com. I detected that these mining pools has two payout flow patterns: (Left) mining pools pay reward directly to miners; (Right) mining pools forward the reward to their own addresses.

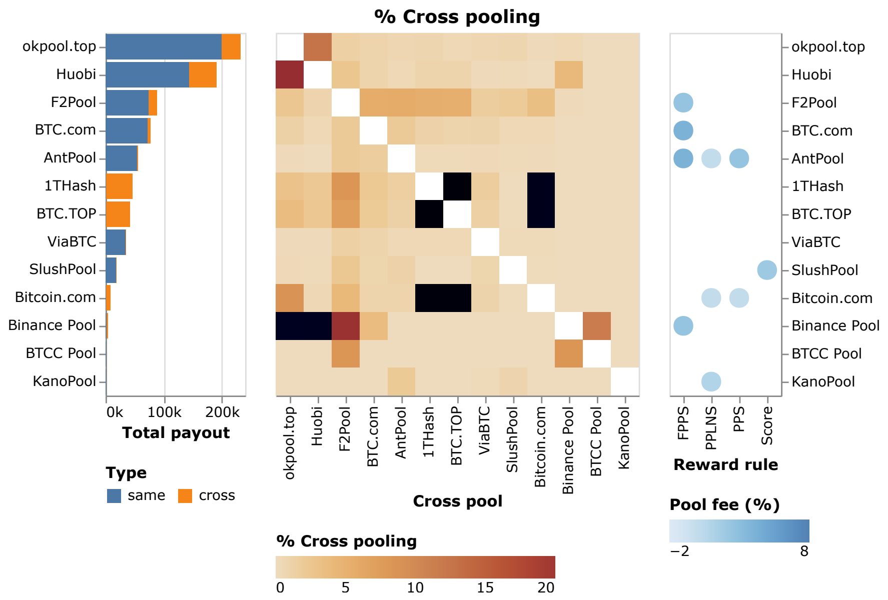

Figure 36: The compound chart displays the total payout (left), reward payout scheme (right), and cross-pooling between mining pools (middle) between January 2020 and December 2020.

result offers evidence that there is a connection between GBMiners and F2Pool.

All three cases reviewed above 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 portfolio manager who optimizes their income concerning the risk associated with each type of asset.

January 2020 – December 2020: Though we observed a large percentage of cross pooing among new mining pools operating around the 3rd halving period. To study this, we plotted a chart to see the detailed cross-pooling in 2020 (Figure 36). 1THash, BTC.TOP, and Bitcoin.com miners are largely cross-pooling. This result generates a first hypothesis that these mining pools are probably associated as they totally have the same miners. Figure 35 (C), (E), and (G) show that payout flow patterns of these pools look very similar. Besides, we found that, like GBMiners, these pools have many payouts flows that pay a considerable amount to a few miners, probably to their own wallet addresses as displayed in Figure 35 (D), (F), and (H). This pattern violates one assumption of our payout flow model. We assume that the practice of mining pools nowadays tends to operate together with the wallet and exchange services, in which they can draw other sources of money to pay miners. As a consequence, we have a new challenge to automatically track the miners directly from the transaction flow.

5.6 CONCLUSION

This chapter presented a new approach to assess the flow of incoming, outgoing, pool-hopping, or cross-pooling miners among mining pools. The approach consists of extracting reward payout flows of mining pools and detecting individual miners from each mining pool. I provide rationale and evidence that it can be used to extract payout flows with different payout patterns. Based on the top-30 mining pools, the payout flows from the algorithm is consistent with previous findings. They can be characterized into four clusters based on the payout frequency and miners' path length variability.

The miners' migration flow measures the net amount of mining rewards given to 1) new and dropout miners, 2) pool-hopping miners, and 3) cross-pooling miners. I developed visualizations that allow me and my economist collaborator to study the impact of payout schemes, pool fees, and Bitcoin values on miners' decisions. The main findings are:

1. The competition of mining pools leads to a lower fee which converged toward an equilibrium after 2016. Payout schemes can be seen as a risk index for mining activities. PPLNS pools are considered more risky assets, so they compensate with lower fees than PPS for attracting miners.

2. Miners perform cross-pooling between PPS and PPLNS pools to diversify their risk to maximize their income and act as financial asset portfolio managers.

3. Bitcoin value significantly affects the expected income and then entry or exit decisions. New miners enter the market when the market price is high, while sharp decreases of the reward, such as halving days, lead many miners to exit this activity.

The empirical analysis provides an insight into the development of mining pools which is crucial for improving regulations and policies in cryptocurrency. As the mining pool competition is constantly changing, I spotted an opportunity to develop a visual analytics tool to explore and continuously monitor this activity in the long term.

6 MININGVIS: VISUAL ANALYTICS OF THE BITCOIN MINING ECONOMY

Based on analysis questions that we came up with, we wanted to integrate our visualization prototypes into a unified view to help explore the interconnection of multiple factors in Bitcoin mining activity. The results and analysis discussed in the two previous papers built the foundation for our Visual Analytics (VA) tool but only focused on specific narrow elements of Bitcoin mining. Some questions could not be solved with simple charts and single datasets. For example, the economist's questions regarding pool hopping required the analysis of various internal and external factors that affect the evolution and dynamics of the activity. Therefore, we found a need to integrate those visualization prototypes into a single muti-coordinated view to help explore the interconnection between three main actors in mining activity.

Over two years, I iteratively developed a visual analytics tool called *MiningVis* for the analysis of long-term historical trends of mining pools for economic analysis. The tool allows analysts to detect interesting time periods and engage in a multi-variate exploratory analysis of pool characteristics. In this chapter, I describe analysis factors and motivations from the economist's collaboration that drove the development of our Visual Analytics (VA) tool. Additionally, I conducted a two-week online user study with eight Bitcoin miners aiming to understand research questions of interest to them and insights they were able to make with the tool. I reported the qualitative and quantitative results from the user study and compare differences between the types of questions relevant to miners and economic researchers.

This chapter is written mainly based on my original article presented at *IEEE Visualization & Visual Analytics conference (VIS 2021)* [155]. The work was led by myself in collaboration with Nicolas Soulié, Nicolas Heulot, and Petra Isenberg.

6.1 DESIGN STUDY

The data-first design study [112] is a research approach that modified the design study methodology [139] to put data acquisition in the early process before eliciting tasks from stakeholders. Motivated by the impact and necessity for understanding Bitcoin mining and mining pools, I loosely follow this approach by initially studying the domain and iteratively collecting mining-related data from various sources. I collaborated with an economist working on the Bitcoin mining economy, who was considered the primary stakeholder. To study this kind of activity, the economist needed to analyze multiple internal and external factors that affect the evolution and dynamics of Bitcoin mining. I organized regular meetings, mostly virtual, with my economist collaborator to discuss my visualization prototypes and discover new findings from the prototype. I talked with the economist to brainstorm what we wanted to analyze based on our data. After that, I developed a visualization prototype in a *Jupyter Notebook* and demonstrated it to the economist. The economist used the prototype to analyze the data, annotate interesting observations or patterns, and give feedback about the design. I improved the prototype from his feedback and observed how he used the tool. We also discovered many new findings and developed further research questions that started subsequent design iterations. At the end of each iteration, we identified additional data we should collect to help explain our findings in the next iteration.

After the first year of collaboration, we made progress in collecting data, creating visualization prototypes, and producing exploratory data analysis results. I describe the scientific discoveries we made during the design study in the two previous chapters—showing the potential success of close collaborations for VA researchers. As Bitcoin market price is highly volatile and the block reward will be halved around every four years, little is still known about miners' behaviors and whether this activity remained and will remain stable in the future. Therefore, we identified the need to develop a VA tool to help researchers and miners explore and monitor this activity's development over the long term. 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. I developed the tool with constant feedback from the economist.

At the end of the design study, we performed a user study to validate the tool prototype with eight Bitcoin miners recruited from the internet. The user study consists of three phases: 1) the tool demo and the first interview, 2) free exploration of the tool and 3) final interview and survey. Users in the first interview gave me some feedback to improve the functionalities of the tool. In the free exploration phase, we adopt the micro-entries [21] method to capture users' insights and analyze their usage patterns. Finally, we conducted a final interview with users to ask about the impression and practical usages of the tool. We reported the quantitative results from System Usability Scale (SUS) [27] and the slightly adapted version of the Value-driven Visualization Evaluation (ICE-T) [162] questionnaires.

6.2 DATASETS

My economist collaborator and I progressively collected the data over the design iterations. I described datasets D1–D4 in the two previous chapters. For the MiningVis tool, I additionally crawled news related to Bitcoin (D5) to help users describe mining pools' evolution from contextual information not captured in Bitcoin statistics (e.g., the legality of Bitcoin, ransomware, and social sentiments). The datasets are updated until September 2021 and publicly available at `https://zenodo.org/record/4342747`. However, I did not include D3 and D5 datasets that can be retrieved from the original sources.

D1: The mining pool distribution dataset gives information about the evolution of mining pools over time. The dataset consists of computed quantitative measures related to mining power: hash rate, market share (normalized hash rate), and total reward received from mining (in BTC and USD). I also estimated the electricity consumption of each pool from their hash rate. (described in Section 4.3)

D2: The mining pool characteristics dataset consists of external data about mining pools from public sources [17, 19] and manually cleaned by the economist. 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 with miners (binary). (described in Section 4.3)

D3: The Bitcoin network statistics dataset contains multiple quantitative measurements of the Bitcoin network for each month. Examples of network statistics for each month include the market price, total hash rate, total block rewards, total transaction fees, mining difficulty, the number of transactions, electricity consumption [14], and the global energy price index [68]. (described in Section 4.3)

D4: The miners' migration dataset describes miners' migration between mining pools over months. The dataset includes the total rewards paid to miners who participate in more than one pool (*cross-pooling*). Besides, it also contains the total reward of miners who join a mining pool for the first time (*enter*), move from a pool to another (*pool hopping*), and leave the mining pool (*exit*) between months. (described in Section 5.2)

D5: The Bitcoin news dataset lists headlines from the Press forum in Bitcointalk.org [20], where users posted links to news articles related to Bitcoin. Each news contained information about the published date from the source, the news headline, and the number of replies and views. There are 33,325 news items in the dataset, as I crawled the last news at the end of September. Assuming that important news should have many viewers and encourage discussions in the Bitcoin community, I defined an importance score of each news item as $views \times (replies + 1)$. This factor is used as a criterion to select the news to display in the tool.

The news headlines covered various Bitcoin-related topics ranging from articles on the market price, regulation, or specifically about mining pools. However, the dataset does not group similar types of news into the same category. Various topic modeling techniques are proposed to infer the latent topics on short texts [121]. I used spaCy [38] library to process each headline into bag-of-words, filter stop words, and get the lemma for each word. Then, I examined many methods implemented in the STTM library [120] to group news headlines into different topics.

I manually checked the top-10 topic keywords for each model to find the model which can easily distinguish topics from the keywords list. In the end, I decided to use the Word Network Topic Model (WNTM) model for this dataset. The WNTM method is a simple and effective approach that learns the latent features of the topics from the word co-occurrence network, rather than the topics of documents [180]. After training the WNTM model with different configurations, I divided the headlines into 15 topics and manually labeled the topic name (e.g., illegal activities, market price, and mining) based on the top-10 keywords and a list of news headlines for each topic.

6.3 ANALYSIS FACTORS AND MOTIVATION

Here, I characterize the three main analysis factors that drove the development of MiningVis—mainly based on the previous research questions derived during the design study.

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 better insights into the impact of regulatory changes in some countries.

AF2: Pool managers' behaviors and the competition to attract miners. Mining pools share mining rewards to their miners with different payout schemes [128] that define when and how much miners are paid. They also kept some amount for operating costs by setting *pool fees* [17]. These mining pools compete to gain higher market share by offering better reward incentives to attract miners. Pool competition might be visible by convergence toward a limited number of (best) payout schemes and reduced pool fees. Such information is crucial for economic debates around pool viability as these fees correspond to a significant part of their income. Analyzing these pool characteristics information might provide additional evidence of the importance of payout schemes and pool fees in attracting miners and the evolution of market shares.

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 and might impact the behavior of miners. To understand the effect of market share, pool characteristics, and external events on miners, miners' migration data (i.e., *pool hopping* and *cross-pooling*) helps to test assumptions and explore the drivers behind collective miners' pool choice. Such analyses are critical to analyzing a mining 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 security and sustainability of Bitcoin mining in the long term. This analysis factor can also benefit Bitcoin miners and users who want to explore mining pool evolution from a long-term historical perspective.

6.4 TASK ANALYSIS

I derived visualization analysis tasks from my observations on actions the economist repeatedly performed on visualization prototypes. The visualization tasks on the tool 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

Table 7: Mapping design levels of MiningVis tool from analysis direction, data, analysis tasks, to visualization (view).

Analysis Factor	Data	Analysis Task	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
	D5	T1.5: Browse Bitcoin news that explain the mining activity evolution in a chosen period.	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 dominate the market share over time	V2, V3, V4
	D3	T2.4: Compare mining power and pool characteristics across multiple pools	V3
AF3	D4	T3.1: Identify a pair of mining pools that have a significant miners' migration flow.	V6
	D4, D2	T3.2: Lookup the mining pool characteristics of the pools that have a high miner's migrations.	V6, V3

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) 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 economist 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.

In Table 7, I summarize the relationship of analysis factors (AF1–AF3), the datasets (D1–D5), and the visualization tasks (T1.1–T3.2). The last column lists the views in the MiningVis tool using specified data and tasks.

6.5 RELATED WORK

In Chapter 3, I reported that most blockchain visualizations came from online sources and used basic representations of blockchain network statistics and individual transaction details. However, dedicated visualization tools for in-depth analyses and task-focused for expert users are still rare and should be challenges for future research. In this work, I focused on mining activities in Bitcoin rather than proposing a general-purpose blockchain exploration tool. Khairuddin et al. [80] documented the practice of Bitcoin mining from interviews with 20 miners to understand their motivation and characterize different types of miners. Instead, our work focuses on the economic analysis of Bitcoin mining over the long term and applies our tool with the real Bitcoin miners. Besides, they also raised trust and centralization issues in the Bitcoin mining activity, which our tool facilitates the monitoring of these issues.

Even though mining is an important activity in the Bitcoin blockchain, most existing visualizations simply show mining pool market shares as pie and area charts (e.g., [O10, O13]). SuPoolVisor [V18] is the only prior work in the visualization literature on Bitcoin mining. It presents a visual analytics system that tracks mining pools' daily computational power and their 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, I conducted a user study to understand the usability and value of our tool for Bitcoin miners and economic researchers.

Related to the analysis tasks in Section 6.4, I also reviewed some related visualization work on similar tasks, including rank comparison, temporal text visualizations, and abstract flow visualizations.

6.5.1 *Rank visualization*

Tasks T1.1 and T1.2 require the visualization of mining pool rankings over time. I 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 [26, 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 [53]. 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) [115]. 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 [173].

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 analysis tasks is RankExplorer, which applies a stream chart variant in which quantities are stacked per ranked item [142]. I decided to use a ribbon chart to encode mining pools with stacked bars for each time point and ribbons connecting the same items. This choice allowed me to show different quantitative measures as bars and their changes in rank and quantity by ribbons.

6.5.2 *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 [126, 137] display stock and flow data as nodes and links. The stocks are represented as nodes while the flows from one stock to another are shown as links. A node is usually encoded as the rectangular bar where the height indicates the node value. A link connects between nodes where the width of links 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 [124].

It encodes the flow of men in both space and time. Telly's Stock and Flow Visualization [40] divides students into groups and tracks their high school advancement. Rosvall and Bergstrom [129] 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 indicates the node's value, while the ribbon's size represents the flow's value between two stocks. Circos [85] 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, such as bar charts, time series, or customized visualizations. Among Bitcoin empirical research, Parino et al. [A33] use the Circos diagram to show the total amount of Bitcoin values (flow) transfer between countries (stock). In our analysis tool, I represent the flow of miners between pools (cross-pooling). The stock equivalent for us is the size of the pool. I chose the chord dependency diagram because it effectively shows the crossing amount among many pools (as well as self-loops) with a high data-to-ink ratio.

6.5.3 *Text visualization*

MiningVis includes a view to visualize news headlines over time in relation to mining pool rankings. The Text Visualization Browser [86] 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 [62], is one method to encode evolving topics in large-scale documents. CiteRivers [63] 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 [84] 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, I focused on visualizing individual news items rather than topics or themes. As such, my work is more closely related to past work on exploring documents over time.

TimeLineCurator [47] 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 documents. BrandSediment [95] adopted a visual sedimentation [67] metaphor to show brand perception (document) distribution in each trait (topic) as a facet of beeswarm charts.

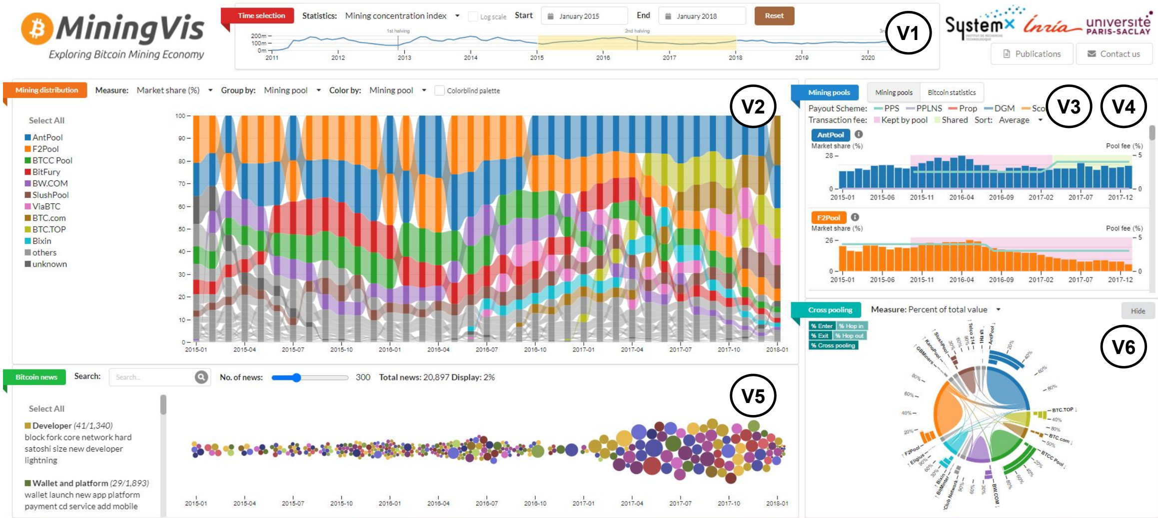

Figure 37: The MiningVis tool consists of six coordinated views: (V1) the time filter view, (V2) the mining distribution view, (V3) the mining pool details view, (V4) the Bitcoin statistics view (toggle with V3, see Figure 41), (V5) the Bitcoin news view, and (V6) the cross pooling view.

I adopted the same concept in our Bitcoin news view because it allows to encode data with two additional visual channels (hue, size) per news item.

The Bitcoin news dataset contains a list of news headlines that also have a time attribute. To display text documents over time, Thread River [46] displays sequences of events in an individual thread of a discussion forum on a compact timeline. The timeline consists of posts as nodes and arcs showing a reply-to relationship. E-Comp [166] includes a temporal view of restaurant reviews (document) as a stack of papers (y-axis) over months (x-axis). For each stack, visual elements (i.e., shape, color, and size) encode different aspects of the review. The time layouts of those two previous works are on the x-axis, which is similar to the MiningVis tool, except that I use a beeswarm plot to increase the visibility of smaller items.

6.6 VISUALIZATION DESIGN

Given the variety of information we wanted to make available for exploratory analysis, I designed MiningVis to include six coordinated views; each showing multiple different metrics. Figure 37 presents a screenshot of the tool. 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. The tool is publicly available at `http://miningvis.fr/`.

6.6.1 *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.

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. 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 critical events related to mining rewards. The halving day lines reference critical days in Bitcoin when the mining reward is reduced by half, directly impacting miners' reward function.

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.

Use: My economist collaborator used this view to see historic Bitcoin statistics and look for time periods based on a specific event, outlier, or trend in the selected Bitcoin mining or network statistics.

6.6.2 *V2: Mining distribution view*

The mining distribution view allows analysts to detect the dominating mining pool, the rise, and decline of mining pools, and to find characteristics that possibly lead to changes in ranking. The view allows analysts to detect the evolution of mining pools, especially the dominated mining pool at the time, and the concentration of mining power. These factors are critical and need to be monitored regularly to ensure the integrity and security of the Bitcoin network.

Visual Encoding: I used the same temporal ribbon chart design from Section 4.4.2. The ribbon chart allows analysts to see both the changing rank and a quantitative metric related to each mining pool. The time axis is filtered to the time period selected in V1. Analysts can choose among multiple statistical measures and how to group and color the pools on the input panel. For each stacked bar, 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. The top-10 mining pools for the selected time and measure are encoded in distinct colors, while the remaining pools are colored in grey.

Figure 38 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 choose mining pool characteristics to display in different color hue

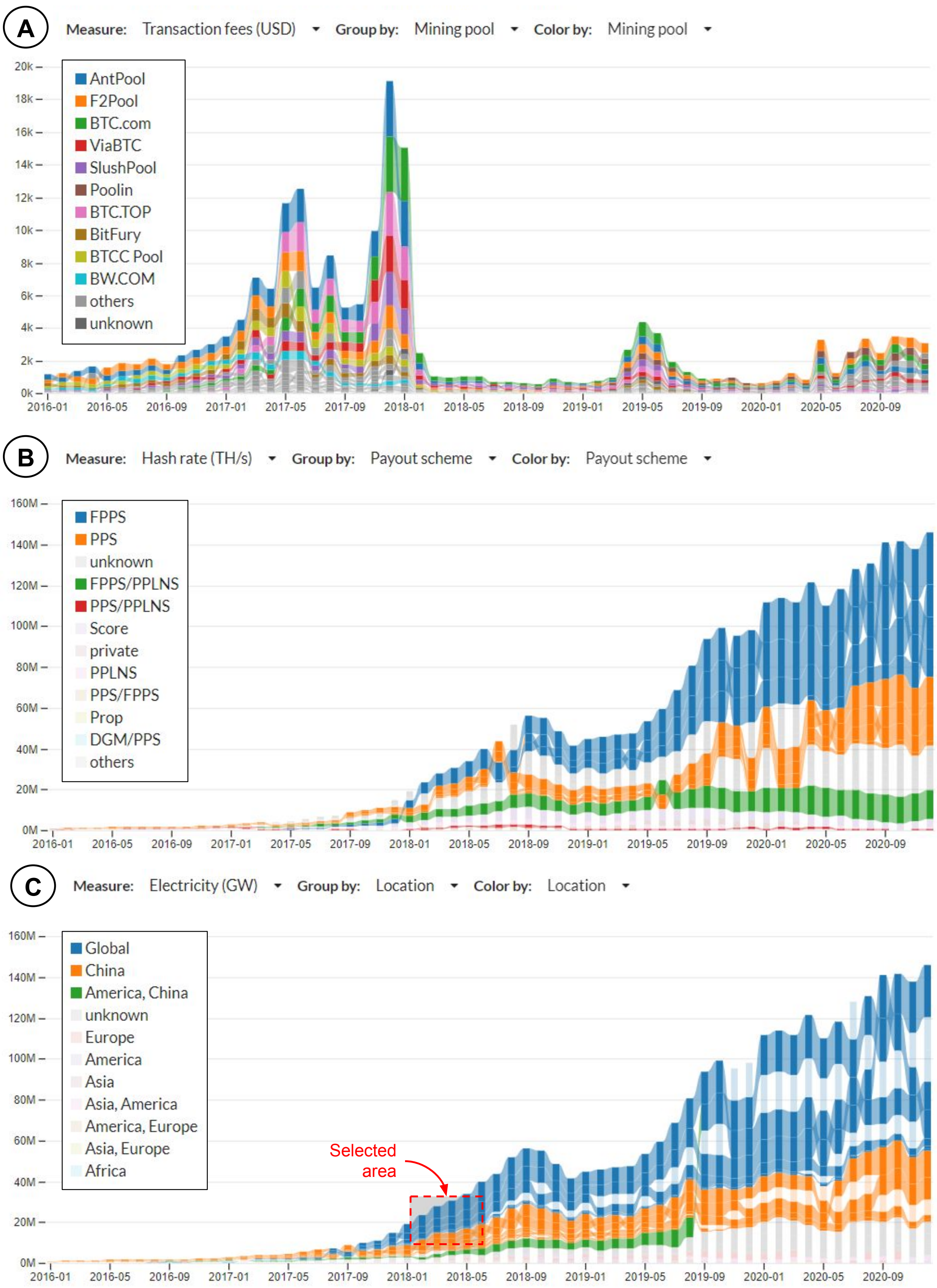

Figure 38: The mining pool distribution view (V2) with different configurations. (A) Measure: Transaction fees (USD), Group by Mining pool, and Color by: Mining pool. (B) Measure: Hash rate (TH/s), Group by: Payout scheme, and Color by: Payout scheme with label selection. (C) Measure: Electricity (GW), Group by: Location, Color by: Location with brush selection.

scales, i.e., the mining pool's name, 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 factors 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, I provide three ways to highlight mining pools or their characteristics:

1. Analysts can click on the left side labels, increasing the transparency of the unselected pools and highlighting those that fall into the selection (e.g., Figure 38 (B)). The selection also affects and filters the mining pools displayed on the mining pool details view (V3).

2. Analysts can draw a brush on the ribbon chart to filter both mining pools and highlight a specific time range (e.g., Figure 38 (C)). The highlighted timeframe is also represented in views V3–V6.

3. Like in all other views, detail-on-demand is available on hover via tooltips that show the mining pool name and the exact value of the selected measure for this pool.

Use: My economist collaborator used this view primarily to study the emergence and evolution of mining pools (T1.1) as intended. Primarily, he aimed to detect which mining pool dominates the market and how much mining power a pool has compared to other pools over time (T1.3). After that, he investigated which characteristics of mining pools might have led to mining power domination in the selected time interval (T2.3).

Design Alternatives: In the early prototypes, I used a bubble chart to show the mining power over time (Figure 19). 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. I 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, I opted for the ribbon chart design because it allows analysts to examine the total mining power while also showing the mining power and rank changing for each pool. The bar length is a more accurate channel to estimate the value with bar height than circle area [145]. Besides, the ribbon chart displays total mining power in the same view, which requires an additional chart in the bubble chart.

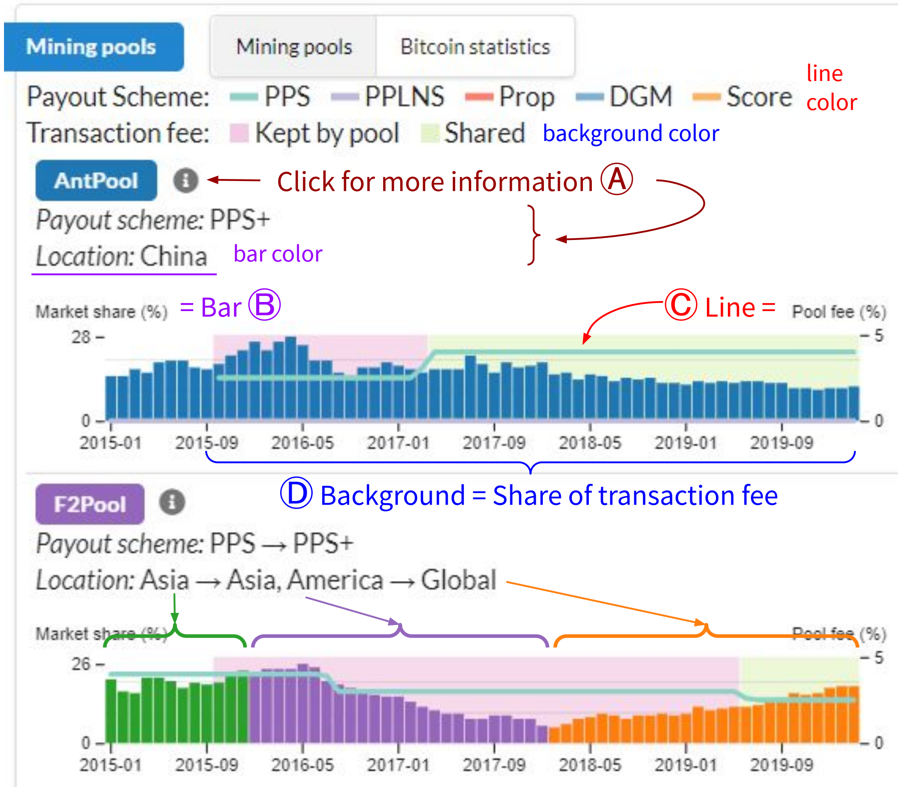

Figure 39: The mining pool details view (V3) shows the details of each mining pool as a dual-axis chart. The measure on the left y-axis and the bar color are selected from V2: *Measure: Market share (%)* and *Color by: Location*, respectively.

6.6.3 *V3: Mining pool details view*

The mining distribution view (V2) focused on the overall evolution of mining pool measures and detecting the dominant mining pools. However, the view makes it difficult to see and compare the development of an individual pool. The mining pool details (V3) view displays the details for each mining pool in multiple small charts to address this challenge. 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.

Visual Encoding: I used a temporal bar chart to encode aggregated (per month) mining pool measures (Figure 39 (B)). Each bar chart is normalized to the maximum measure of the individual pool to help see the mining power of small pools more clearly. The color of each bar corresponds to the one in V2 to help cross-comparison between these two views and detect changes in a single pool's characteristics.

To address task T2.3, I 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 the limited screen space, I chose a dual-axis encoding instead

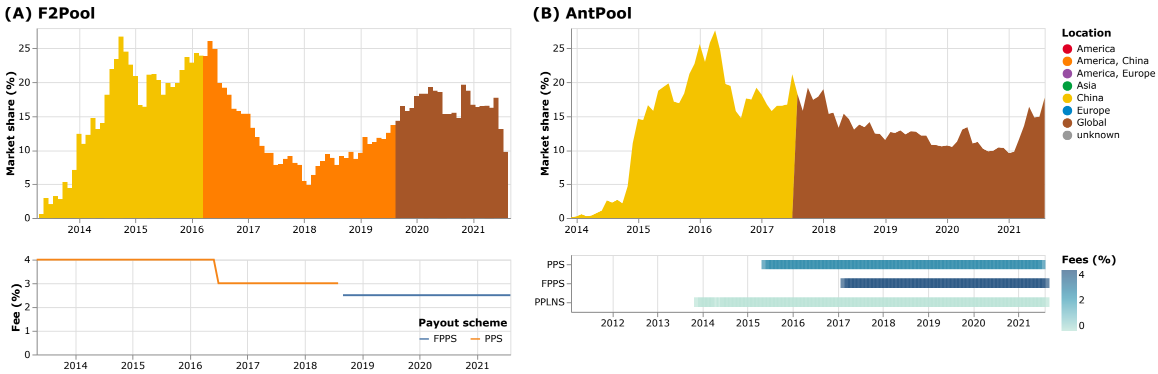

Figure 40: Design alternatives for the mining pool details view: (A) F2Pool—The bar chart displays market share, and the line chart below shows pool fees; (B) AntPool—The area chart displays market share, and the heat map below encodes pool fees for each payout scheme.

of an additional chart and showed the pool fee as a line chart (Figure 39 (C)). I used different line colors for reward schemes because a mining pool can have more than one payout scheme at a time. 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 (Figure 39 (D)) to detect 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 (Figure 39 (A)). Additional interactions are tooltips for detail-on-demand.

Use: The economist looked at the individual mining pool and assessed the correlation between mining power, payout fee, and share of the transaction fee (T2.2). Then, he compared across pools in the list to validate if a similar pattern holds for other pools (T2.4). He also used this view to locate if pools with the lower fee attract more miners and validate whether they obtained a higher market share.

Design Alternatives: I selected bar charts as the visual encoding in V3 as it is closely related to V2 and added lines in the second axis to encode pool fees. The dual-axis chart provides three pieces of data in a single chart, which may be overwhelming at first glance. Dual-axis encodings have been extensively critiqued and careful design has been advised [42, 57, 69, 136]. The alternative design is an additional line chart or heat map below each pool's bar chart to encode pool fees (). After some tests, I decided against it due to screen-space issues. These additional charts took away to focus on the display of the selected measure. In addition, my collaborator preferred the dual-axis chart as he was familiar with this chart type from the economic literature. I opted for overlaying a line chart on the bar chart to emphasize

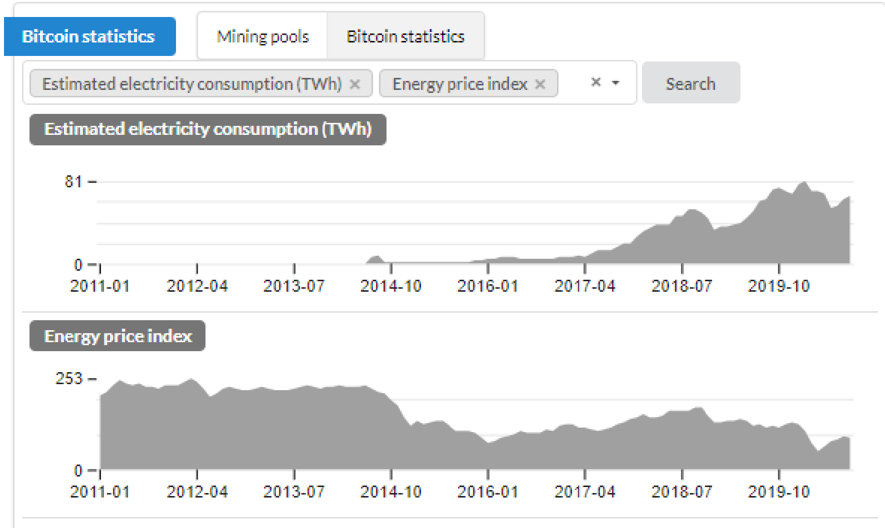

Figure 41: The Bitcoin statistics view (V4) shows each Bitcoin statistics measure as the area chart.

two different data types and provide summary data in the text form. Nevertheless, I acknowledge that this chart is relatively complex, but analysts can directly correlate mining power and pool fee.

6.6.4 *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. The economist tried to discover factors that could explain the phenomena they observed in the Bitcoin mining and then generalize those observations into the model. I 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 (Figure 41) is located in the same position as V3, reachable via a toggle bar. I encoded each of the 20 Bitcoin statistics measures as a gray temporal area chart, including Bitcoin-internal statistics. These include the total number of blocks mined, the total hash rate, the median confirmation time for a block, total rewards paid out, or the total number of transaction fees. In addition, I 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: Analysts can search and select statistics measures with the dropdown selection.

Use: The economist uses this view to search and browse statistics measures and compare with the mining distribution view (V2) to find the correlation between the development of mining activity and other factors, such as block size, market price, energy consumption, etc.

Design Alternatives: I considered coloring the area chart instead of using gray but opted against it to avoid confusing the view with V2 and V3 that include an explicit color encoding. Representing the data as line charts instead of area charts or bar charts would also be possible without compromising the readability of the data. Again, I opted for a slightly different chart type to make it visually obvious which view was currently shown.

6.6.5 *V5: Bitcoin news view*

News and social media are good sources for additional information to give context about historic mining behavior and its relation with other factors, such as competing blockchains, new mining hardware, or a surprise ban of Bitcoin from some governments. The Bitcoin news view allows analysts to both get overviews of the news over time and further browse news headlines that are potentially relevant to Bitcoin mining activities. Individuals and groups of news headlines might explain trends or patterns in the Bitcoin measures directly or might themselves have influenced miners' and mining pools' behavior. I used short-text topic modeling to structure and group the relevant news in our large corpus of documents (D4) and designed a beeswarm plot with search functionality. The dataset I collected included >30,000 postings that cannot be displayed in one chart. Therefore, the news is selected from its importance score and balanced by the news frequency in each month.

Visual Encoding: I used a swarm plot to display each news item in a compact fashion across a timeline. It trades off an 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. Each news has a different color to indicate topic membership.

The list of all news topics and their top-10 keywords is displayed on the left panel. The numbers behind the topic label indicate the number of news shown in the swam plot versus the total news on that topic. The topics are sorted by the number of news displayed in the swarm plot.

Interaction: An example use case for the news view is demonstrated in Figure 42. Analysts can browse the news by hovering circles to see tooltips with news headlines on the top-left of the chart. When

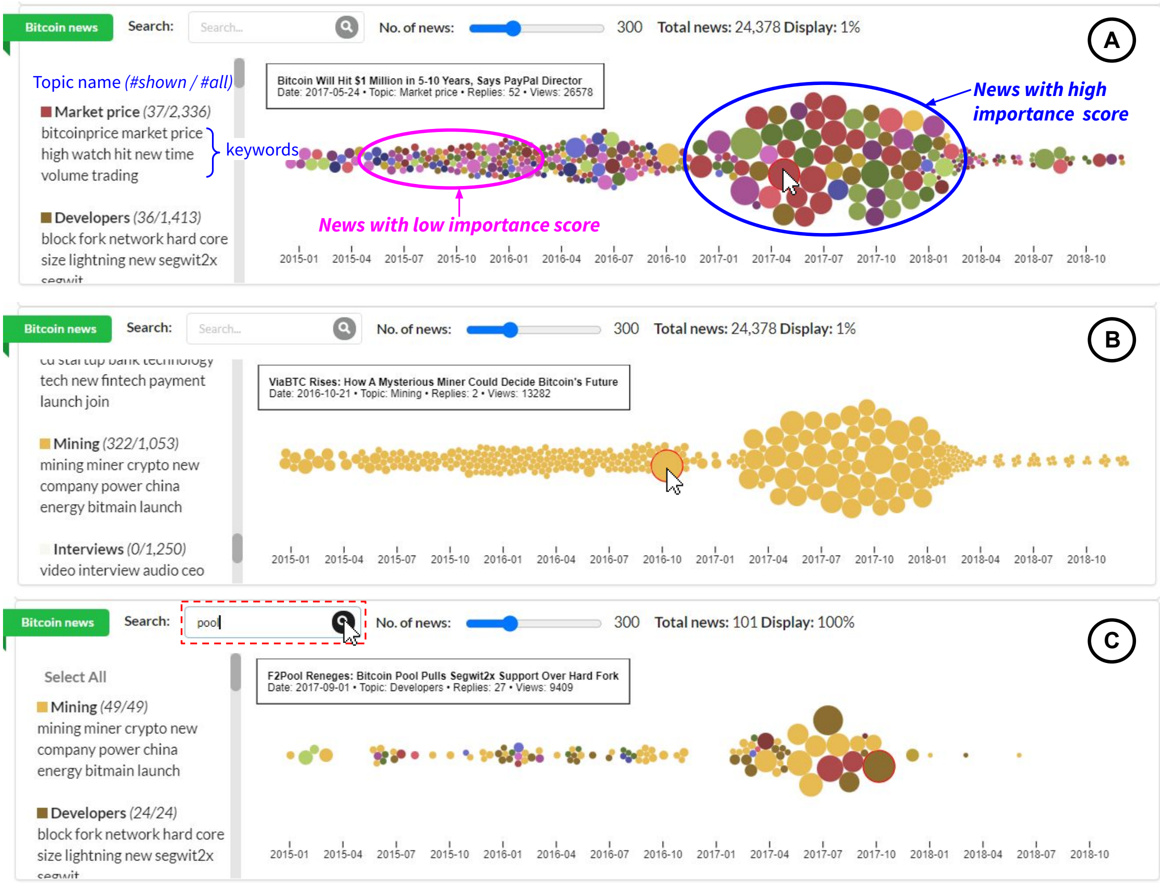

Figure 42: Examples of the Bitcoin news view configurations. (A) Top 300 news over the time period. (B) Selected news from the "mining" topic. (C) Search for news that contain the term "pool" in their headline.

the analyst clicks on the circle, it will open a new tab to the news source. Analysts can use the left panel to select news on the topic of their interests.

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 the news displaying in the chart. The panel also shows the number of total news and the percentage of news that the chart currently displays.

Use: The economist usually used the news view to search for news headlines about specific mining pools (e.g., AntPool and F2Pool) and specific terms indicating the external event (e.g., law, regulation, Bitcoin ban). Then, he focused on the time of changes in market share and looked for the plausible news that explains why the pools' market share was rising or declining.

Design Alternatives: The first design used a stream chart and heat map to focus on news topic distribution over time (Figure 43). However, the economist wanted to browse individual news items to find

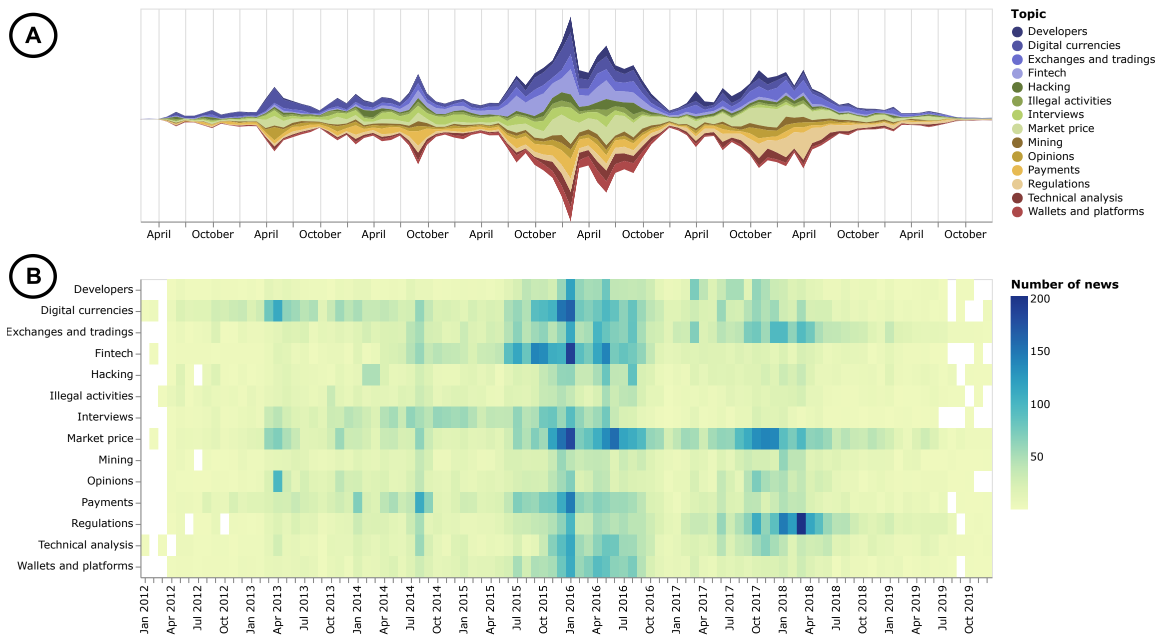

Figure 43: Design alternatives for Bitcoin news view: (A) stream chart and (B) heat map.

detailed information that would explain the mining distributions he observed. I 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. In the end, I decided to use a beeswarm design because it uses a compact layout to show the distribution of news while still encoding topic membership and importance with the color and size of the circle.

6.6.6 *V6: Cross pooling view*

Exploring miners' migration between pools is crucial to understanding why they gain or lose their market share over time. At a selected time period, the economist wanted to investigate whether there was any significant miners' migration among pools (T3.1), and if so, what factors might influence miners' decision to switch (T3.2). 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. I designed the cross pooling view to represent and summarize migration patterns. Compared to the past work [172], I proposed cross-pooling as an additional measure to track miners who received the mining reward from multiple mining pools. This measure allows the economist to understand miners' strategies to maximize their profit at the focused time interval.

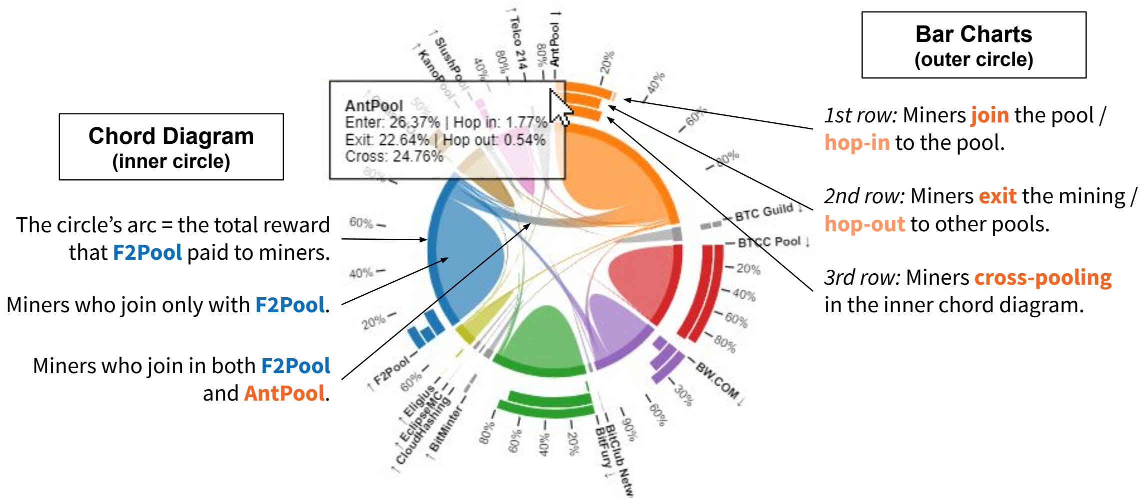

Figure 44: The miners' migration flows are represented as a chord diagram to visualize the amount of cross-pool miners between pools. For each pool, the bar chart displays the percentage of new, pool hopping, cross pooling, and dropout miners.

Visual Encoding: The cross-pooling miners during a selected time interval can be considered as a flow of miners between mining pools. Figure 44 shows the visual encoding of the view, which consists of two parts: chord diagram and bar chart. I 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 outer arcs' length. The flow between mining pools represents the total amount of the metric for those miners who cross-pooled.

Around the outer ring of the chord diagram, the average percentages of the miner's migration statistics per month are represented as stacked bar charts (i.e., new, exit, hopping in, hopping out, cross pooling). For each pool, three rows of the bar chart represent the percentages of miners incoming (new and hopping in), outgoing (dropout and hopping out), and cross-pooling with the mining pool.

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.

Use: The view allows the economist to look up cross-pooling flows directly in the chord diagram and overall miners' flow in the bar chart. He used the view to detect the pairs of mining pools with high cross-pooling miners and then focused on those pools in other views (T3.1). He also selected the color-by dropdown menu in V2 to directly check if any pool characteristics affect miners' flow (T3.2).

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 [172] to show numbers over time but does not highlight the association between pools. Due to the fluctuations of miner counts in the data [154], I chose to show aggregated data to reduce outliers. A heatmap matrix is an alternative to display the percentage of cross-pooling miners like in Figure 32. Still, the color encoding turned out to be problematic due to the different largely varying scales 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.

6.7 user study

MiningVis was designed with an economist and focused on analysis questions about the Bitcoin mining economy. To evaluate whether the tool would also apply to other user groups, we identified Bitcoin miners as other potential users with economic interests in this activity. Therefore, we conducted a multi-week online user study with eight Bitcoin miners to evaluate the usability and relevance of the tool. The research ethics board of the Université Paris-Saclay approved this study, *CER-Paris-Saclay-2020-062*.

The study consisted of four phases. First, we recruited Bitcoin miners from an online questionnaire posting on Bitcoin forums. Then, we conducted a 1-hour introductory session to introduce the tool and the first interview to get usability feedback. After that, participants were asked to use the tool for around two weeks and report findings on the micro-entries form. Finally, we asked them to fill the questionnaires to evaluate the tool and conducted a follow-up interview on the practical usages of the tool. The background and survey responses of all participants are published in the OSF repository https://osf.io/ud2c9/.

6.7.1 *Participant recruitment*

We asked potential participants to sign up for our user study in the popular Bitcointalk.org[1] forum as well as Reddit channels dedicated to Bitcoin (*r/Bitcoin* and *r/BitcoinMining*)[2]. 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

1 https://bitcointalk.org/index.php?topic=5311277.0
2 https://www.reddit.com/r/BitcoinMining/comments/l1zmgn/call_for_participation_in_a_user_study_of_bitcoin/

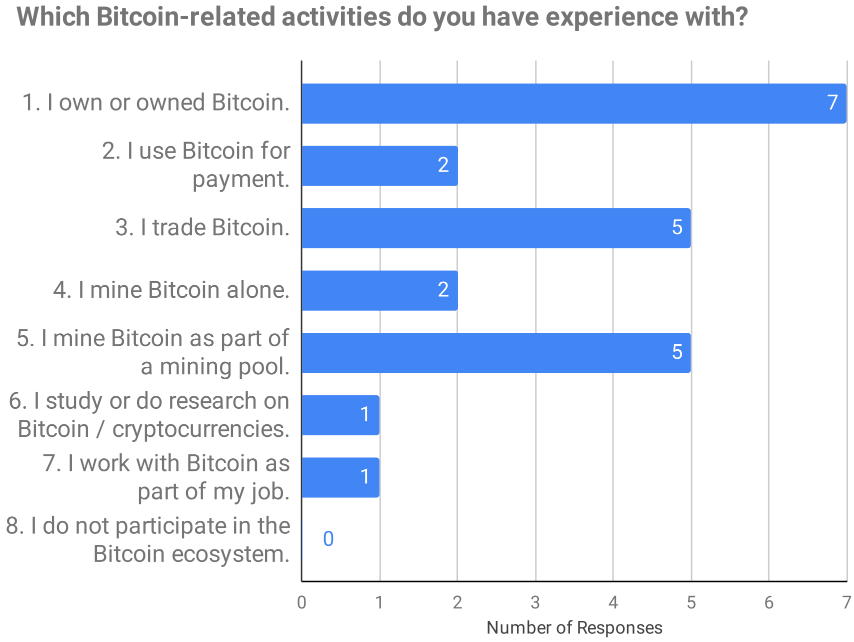

Figure 45: Participant background about Bitcoin and mining activity.

in their country. We received 11 complete responses and were able to recruit eight participants for the first interview.

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. Their age ranges are 25–34 years old (3 participants), 35–44 (3), 18–24 (1), and 45–54 (1). Their experiences with Bitcoin-related activities is reported in Figure 45. Participants had 4.38 years of experience on average in Bitcoin mining before the user study (min: 1, max: 8, median: 4.5, sd: 2). Three participants read Bitcoin news several times a week, followed by every day (2), several times a month (2), and once a week (1). I will refer to each participant with a unique ID (i.e., P1–P8).

During the interview, I also asked participants to freely express their Bitcoin mining experience and motivation to use our tool. P4, P6, and P7 owned mining hardware and mined Bitcoin as their hobby. P1 had experience working with a Chinese mining company. P5 was developing a small mining farm and looked for the optimal mining pool to join. Participants' motivation to participate in our user study was curiosity about our tool. P1, P3, and P7 wrote that they were interested in exploring the empirical data on Bitcoin mining pools. P2 expressed an interest in understanding the shifting of computational power and miners participating in mining pools. P8 worked on research related to blockchains.

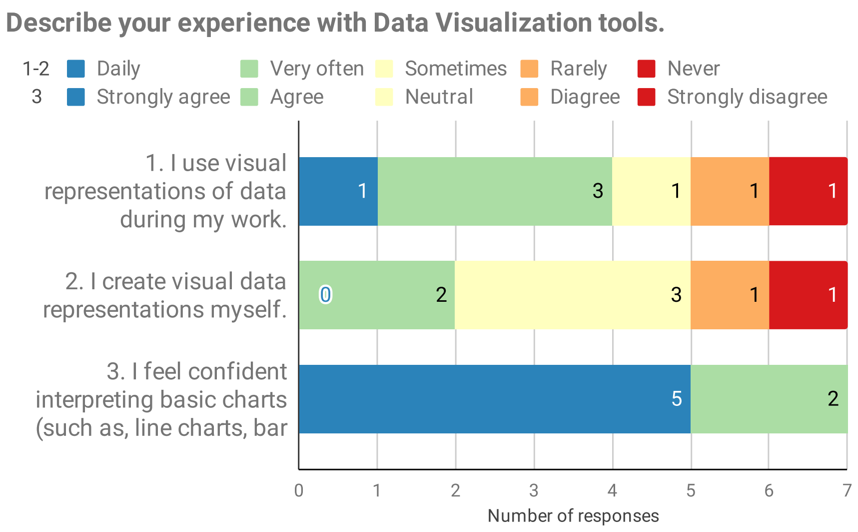

Figure 46: Participant experience on data visualization tools.

6.7.1.1 *Introductory Session*

The study began with a 1-hour 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 that they were familiar and comfortable with reading data visualizations, as exhibited in Figure 46.

In the first 15 minutes, we asked participants about their interests in Bitcoin mining and 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 the MiningVis tool for around 30 minutes with a set of short exploratory tasks we had prepared. The tasks aimed to get participants familiar with various aspects of the tool and to see how participants would intuitively go about answering them with MiningVis.

TASK 1 Which mining pool received the most BTC in August 2019?

TASK 2 Can you explain why GHash.IO dominated the Bitcoin mining activity and then stopped its operation?

TASK 3 Can you identify when top mining pools nearly dominated the market and posed a risk to the 51% attack?

TASK 4 Can you find a historical anomaly in the transaction fees? If yes, can you explain what else happened in the same period?

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.

6.7.1.2 *Usage patterns from the warm-up tasks*

During the think-aloud session in the first interview, we observed participants performing the four warm-up tasks. When participants got stuck, we explained views they had not considered yet to increase familiarity with the tool. I tracked the sequence of views participants interacted with for each task, summarized the patterns to see which ones they considered important, and verified our visualization design choices. Table 8 reports these exploration patterns for every participant.

TASK 1 All 8 participants quickly used V1 to zoom close to the month and then select the "Total Reward (BTC)" measure in V2.

TASK 2 7 participants detected the GHash.IO pool from V2 quickly, then selected the pool in the legend, and looked at the market share domination and decline. After that, 4 of them filtered the time to see patterns more clearly in V1. Finally, 7 participants searched for the news with "GHash.IO" that could explain the decline.

TASK 3 7 participants searched for pools with outstanding market shares in V2 and looked for detail in V3 later and 1 participant looked at V3 directly. 6 of them were satisfied with naming the pools and stopped looking for mining pool details. Exceptionally, 1 participant searched for the news about "51% attack."

TASK 4 All 8 participants changed the measure to "Transaction Fee" and spotted the anomaly in V2. Then, 4 of them proceeded to use V1 to zoom-in on the time with extreme transaction fees. Another 4 participants browsed the news with the "Transaction Fee" keyword before finding the Bitcoin statistics in V4 that correlated with this event.

Participants reported that they 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 expectations. V5 was also used frequently to look for explanations in the news. Participants often overlooked the detail provided by V3 and V4 and I had to remind them about those views and visual encoding details occasionally. Participants never used V6 to investigate miners' mobility, even though it might have provided information for Task 2. Pool hopping was not a familiar metric related to participants' work or Bitcoin usage, and participants reported that the metric is non-intuitive to interpret.

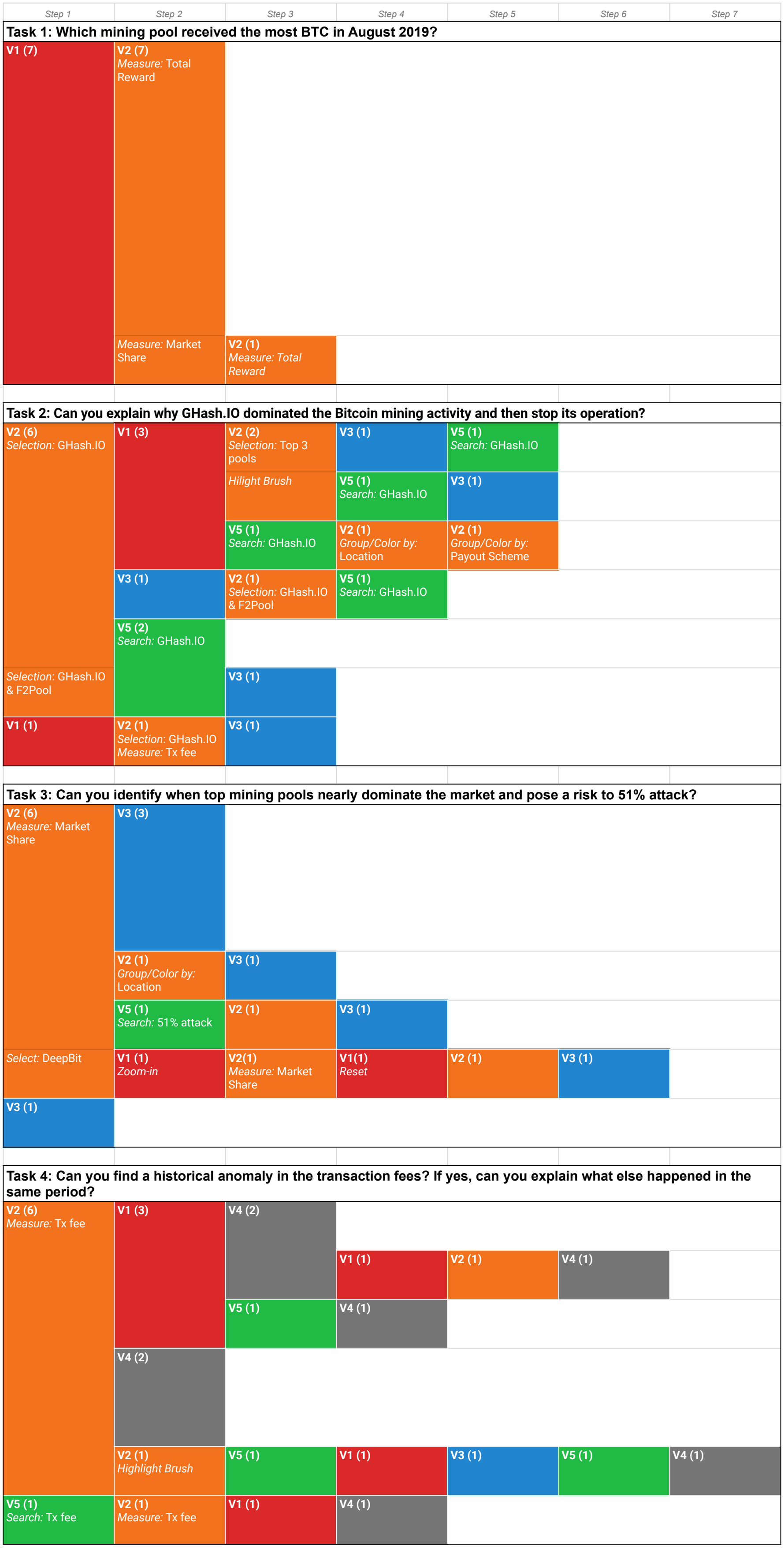

Table 8: The sequence of views that participants used in the MiningVis tool to explore the four warm-up tasks.

6.7.2 *Usability Improvements*

After the warm-up task, we interviewed participants to get their first impression and feedback to improve the usability of the tool. Overall, they reported being slightly overwhelmed by the amount of information in the tool. After the first two tasks, they became more familiar with the tool and adopted their strategy to solve the following two tasks. Nonetheless, participants reported some issues that helped me improve the tool's usability.

- Several participants expected the highlight brush in V2 to zoom in on a specific time interval instead of only highlighting mining pools and time intervals. I improved the tool by adding a highlight brush on V1 and a "Focus" button to allow users to focus on a highlighted time interval.
- The mining pool color legend changed every time participants changed the time interval because it assigned colors to the top-10 pools of the selected time interval. A future improvement would be adding a feature to allow users to assign a persistent color to specific mining pools they want to focus on. However, this feature was not my priority because only 4–5 top mining pools dominate the market and impact the mining economy.
- Participants wanted to specify additional ways to sort the mining pools in V3. I added a sorting dropdown that allows the user to sort mining pools according to the total, average, minimum, or maximum of the selected pool measurement (e.g., market share, hash rate, and total rewards).
- Participants also wanted to look at their own mining pool's performance, even though a small pool never significantly impacted the market. Here, the search function in the web browser was sufficient to navigate to a specified mining pool in V3.
- There are a lot of Bitcoin statistics measures to compare with the evolution of mining pools. I added a search panel in V4 to allow users to filter specific Bitcoin measures.

6.7.3 *Free-Exploration Phase*

At the end of the first interview session, we gave participants free access (login and password) to the tool and asked them to use it at their own pace for two weeks. To learn which questions our participants were trying to answer, we encouraged them to report any finding or insight from the tool in the short form inspired by the micro-entries [21] methodology (Figure 47). The Micro-entries method intends to capture insights from users while they are using the tool. In this study,

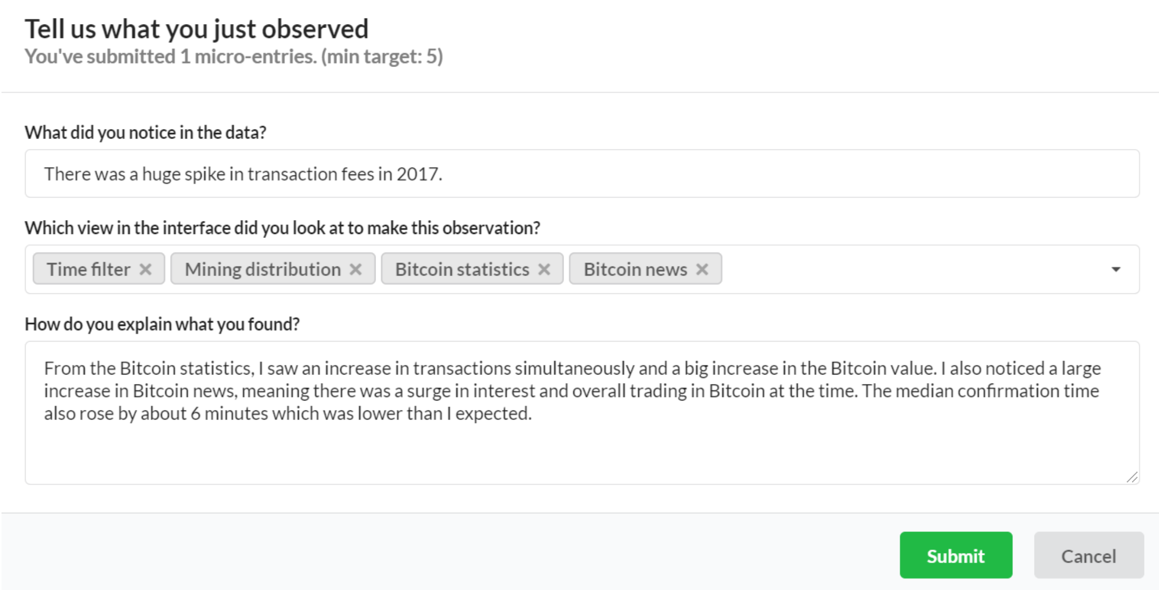

Figure 47: The micro-entries form to report insight-based findings in the free-exploration study.

we tried to minimize the time of participants to report each finding. The form includes three short questions about a) any interesting pattern they found using the tools, b) the views in the tool they used, and c) the short explanation about what they observed. we sent them an e-mail notification to remind them to submit at least five entries in the middle of this phase.

6.7.3.1 *Insights and findings from micro-entries*

We received 12 micro-entries from four participants. We coded each entry according to analysis factors from Section 6.3. 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). Among 12 micro-entires, we found that some entries are not recording insights as expected: three were feature suggestions, and an exceptional one was a general tool review.

Overall, the 8 insight-based micro-entries from three participants focused on higher-level mining activity evolution (i.e., mining pool domination and Bitcoin statistics) rather than on a specific mining pool (i.e., via payout schemes, pool fees, and cross pooling). In comparison, our economist delved deeper into the data and used the tool to detect pools that dominated the market, then looked at the pool characteristics and noticed a high cross pooling amount between the pools. 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 details and discovered 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.

6.7.4 *Final Interview and Survey*

After two weeks, we contacted all eight participants to schedule a follow-up interview and quantitative questionnaires. Five participants responding to two questionnaires: System Usability Scale (SUS) [27], and Value-driven Visualization Evaluation (ICE-T) [162]. One of them was not available for the interview but sent his responses. The other three participants did not respond to my e-mails at all after the first interview. The SUS questionnaire was intended to evaluate the general usability of the tool, including effectiveness, efficiency, and evaluation. We used a slightly modified version of the ICE-T questionnaire to assess the visualization tool on the four perspectives: time, insight, essence, and confidence.

Four participants agreed to be interviewed. The session took around one hour for each participant. We asked participants open-ended questions about their experience using the tool, what they found the most considerable value, and what variables were missing.

6.7.4.1 *Results*

Participants reported that they logged in 5.2 times on average (sd: 4.38, median: 3) within the two weeks. All but one participant mentioned spending only a limited amount of time with the tool during the free-exploration phase and therefore did not make deep insights. Next, I reported the quantitative results from SUS and ICE-T questionnaires and related them to follow-up responses from the interviewees.

System Usability Scale: After the two-week free exploration, we asked participants to fill the SUS survey. The responses showed an average score of 73.5/100 (sd: 13.3, median: 72.5), which is an above-average (68) usability score [133]. However, the score still shows room for improvement. The average score for each question is reported in Figure 48. No individual question scores were below average (on a nega-

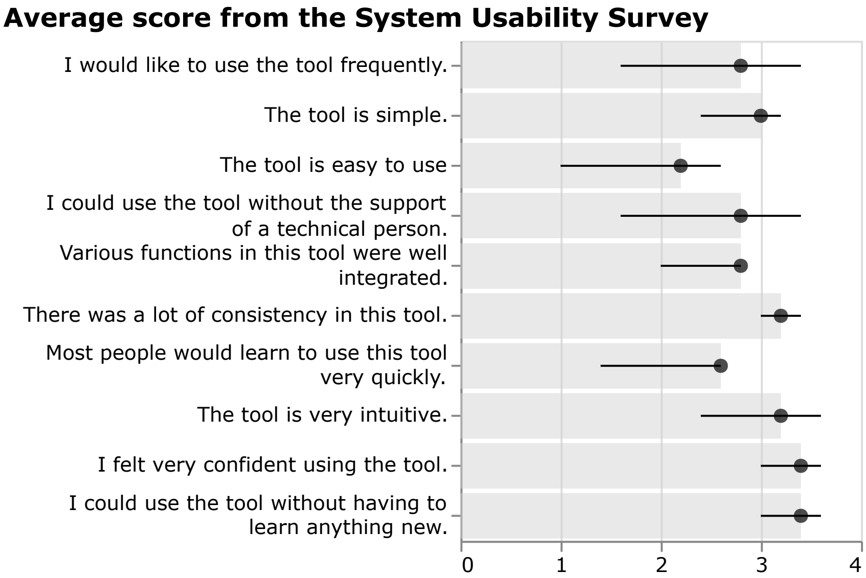

Figure 48: The average score from the System Usability Survey with the 95% bootstrap confidence interval.

tive to positive response scale). The highest scores were given for confidence of use and learnability. The lowest (but still positive) scores were given for ease of use. From feedback on the final interview, the unavailability of real-time mining data kept users 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.

Value of the Visualization: We evaluated the value of the implemented visualizations with an adaptation of the ICE-T questionnaire [162]. Our modified questionnaire is shown in Table 9 compared to the original version. We changed some wordings to be more specific on the mining domain and removed three heuristics that are vague to measure. 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 through the hierarchy from heuristics to guidelines and then from guidelines to components. Table 9 reports the average score and standard deviation for each heuristics while Figure 49 displays the average score with 95% confidence interval for each visualization value guideline.

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 particularly mentioned the good analytic structure of the tool and the diversity of available Bitcoin mining indicators. One participant said that the tool "*helps understand well the unfolding of the competition between mining pools and can go from a broad overview of the historical un-*

Origial Heuristics	Our Adapted Version	Average	SD
Insight		4.30	0.36
The visualization facilitates answering questions about the data.		3.90	0.66
The visualization exposes individual data cases and their attributes.	*Not included*		
The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables.	The visualization facilitates perceiving relationships like patterns & distributions of the variables related to mining pools.	4.40	0.49
The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases.	The visualization helpes to compare individual mining pools as well as different groups of pools	3.40	1.02
The visualization provides a new or better understanding of the data.		4.40	0.20
The visualization helps generate data-driven questions.	The visualization helps generate data-driven questions.	4.20	0.40
The visualization helps identify unusual or unexpected, yet valid, data characteristics or values.	The visualization helps identify unusual or unexpected, yet valid, data characteristics or values.	4.60	0.49
The visualization provides opportunities for serendipitous discoveries.		4.60	0.37
The visualization provides useful interactive capabilities to help investigate the data in multiple ways.	The visualization provides useful interactive capabilities to help investigate the data in multiple ways.	4.60	0.49
The visualization shows multiple perspectives about the data.	The visualization shows multiple perspectives about the data.	4.60	0.49
The visualization uses an effective representation of the data that shows related and partially related data cases.	*Not included*		
Time		3.70	0.56
The visualization affords rapid parallel comprehension for efficient browsing.		3.70	0.51
The visualization provides a meaningful spatial organization of the data.	The visualization uses a meaningful layout of the data on the screen.	3.60	0.49
The visualization shows key characteristics of the data at a glance.	The visualization shows key characteristics of the data at a glance.	3.80	0.75
The visualization provides mechanisms for quickly seeking specific information.		3.70	0.87
The interface supports using different attributes of the data to reorganize the visualization's appearance.	*Not included*		
The visualization supports smooth transitions between different levels of detail in viewing the data.	The visualization supports smooth transitions between different levels of detail in viewing the data.	3.40	1.02
The visualization avoids complex commands and textual queries by providing direct interaction with the data representation.	The visualization avoids complex commands and textual queries by providing direct interaction with the data representation.	4.00	1.10
Essence		4.05	0.43
The visualization provides a big picture perspective of the data.		4.20	0.40
The visualization provides a comprehensive and accessible overview of the data.	The visualization provides a comprehensive and accessible overview of the data.	4.20	0.40
The visualization presents the data by providing a meaningful visual schema.	The visualization presents the data by providing a meaningful visual schema.	4.20	0.40
The visualization provides an understanding of the data beyond individual data cases.		3.90	0.58
The visualization facilitates generalizations and extrapolations of patterns and conclusions.	The visualization facilitates generalizations and extrapolations of patterns and conclusions.	4.00	0.63
The visualization helps understand how variables relate in order to accomplish different analytic tasks.	The visualization helps understand how different mining pool metrics relate in order to accomplish different analytic tasks.	3.80	0.75
Confidence		3.73	0.54
The visualization helps avoid making incorrect inferences.		4.20	0.51
The visualization uses meaningful ~~and accurate~~ visual encodings to represent the data.	The visualization uses meaningful visual encodings to represent the data.	4.00	0.63
The visualization avoids using misleading representations.	The visualization avoids using misleading representations.	4.40	0.49
The visualization facilitates learning more broadly about the domain of the data.		4.20	0.75
The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes.	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes.	4.20	0.75
The visualization helps understand data quality.		2.80	0.75
If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues.	If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight those issues.	2.80	0.75

Table 9: Our adapted version of ICE-T questionnaire compared to the original version. We reported the average and standard deviation for each heuristic.

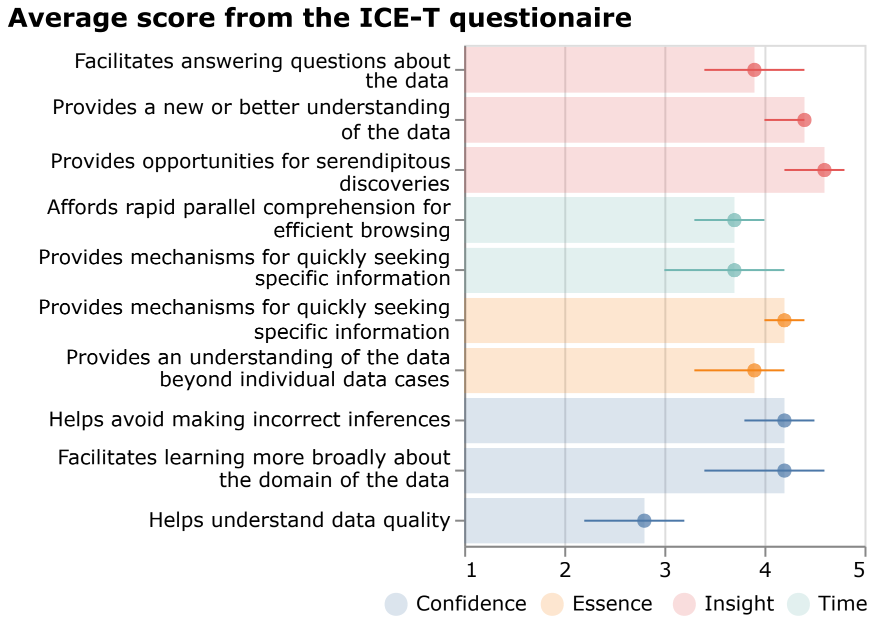

Figure 49: The average score from the ICE-T questionnaire with the 95% bootstrap confidence interval.

folding to a deep zoom into a very specific data point". One participant further said that he did not find any other tools that give complete data about Bitcoin mining.

Participants also gave a high score for *confidence* (3.73, sd: 0.54), except for "the visualization helps to understand data quality" (2.80, sd: 0.75). The result reflects that the tool does not specifically reveal the quality of the datasets, which is correct because our tool focuses on data exploration tasks rather than assessing data quality. However, our economist collaborator used the tool to cross-check if the pool characteristics information aligned with the Bitcoin mining pool statistics. *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 the entire Bitcoin history data. This probably explained why they gave a relatively low score on "the visualization supports smooth transitions between different levels of detail in viewing the data" (3.40, sd: 1.02). Hence, the time-related heuristics can be improved by refactoring the source code on both the front-end and back-end to optimize the data retrieval performance.

6.8 DISCUSSION

The outcomes from the design iterations with the economist and the user study with Bitcoin miners highlight the tool's usefulness to analyze the Bitcoin mining economy while also pointing out future improvements from the practitioners' perspectives. The result from the user study also raised some discussion on the user study methodology we applied in this work.

6.8.1 *Value of the visualization tool*

Throughout the design iterations of the MiningVis tool, we used visual analytics approaches to help my economist collaborator understand the evolution of mining pools and the Bitcoin mining economy as a whole. My economist collaborator and I were able to make discoveries reported in two analysis articles [153, 154] based on the visualization prototypes. The economists used the tool to identify time periods to investigate pool hopping behavior, compared the characteristics between pools, and discovered contextual information to explain the behavior of actors in the activity, including individual miners, mining pools, and the Bitcoin blockchain. 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 still benefit both analysts and practitioners to monitor and analyze the behavior of mining pools in the long term.

In the user study, participants complimented that the tool provided complete information on Bitcoin mining. Still, we found that the value of our tool was a little less obvious for them. Participants, who were all Bitcoin miners, can detect interesting time periods and explore the changes in mining pool domination. However, they did not have specific research questions like the economist and did not use all the features presented in the tool. In particular, participants rarely used the cross-pooling view (V2) because it is unfamiliar to them or unrelated to their work. They wanted to decide on a pool to join. Therefore, they demanded real-time data focusing on recent information rather than the long-term overview of the mining activity we provided. In the follow-up interviews, four participants suggested that the tool potentially benefits researchers and mining companies to analyze historical trends of Bitcoin mining. According to the user study result, We concluded that the current version of the MiningVis tool is most useful for modeling and long-term decision-making from a historical perspective.

6.8.2 *Future improvements*

The overall feedback on usability and visualization design was positive and confirmed our visualization design choices. However, we identified some possible improvements of the tool from the feedback of participants in the user study.

Integration with real-time data: Many participants demanded real-time data to monitor the market share of the mining pool and decide on which mining pool to choose or move to. To do so, real-time updates of Bitcoin transaction data require attributing a mining pool to each new block. 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 [17] and Bitcoin news APIs with natural language processing techniques. Moreover, the calculation of cross-pooling poses a bottleneck as the tool needs to track transaction flows to identify individual miners. Because participants did not focus on the cross-pooling data, a simplified version of the tool without the cross pooling view could be more easily deployed to present real-time data.

Generalization to other cryptocurrencies: While the MiningVis tool is focused on Bitcoin, other cryptocurrencies implement the same proof-of-work (PoW) mining protocol, such as Bitcoin, Ethereum, Monero, and Dash. Mining pools also emerged in those cryptocurrencies. For example, miners joining BTC.com mining pool can also mine Ethereum, Bitcoin Cash, Litecoin, or Monero. Those cryptocurrencies also require miners to solve a computationally-expensive puzzle and provide financial rewards as an incentive to mine. Therefore, the tool can be easily adapted to analyze the mining patterns of those cryptocurrencies. The future tool would be interesting to study miners' behaviors and assess the risk of mining pools across cryptocurrencies.

6.8.3 *Reflections on the micro-entries*

In the user study, we applied the micro-entry methodology to record the insight when participants discovered any findings from the tool. We assumed that the lightweight format to collect even simple "findings" using this methodology would encourage participants to report more. However, we were disappointed by the low response rate and quality of the micro-entries from participants.

In the follow-up interview, we specifically asked participants about the micro-entries method and what led them to submit only a few or no findings. Participants reported that they neither had the time nor specific questions they wanted to investigate in depth. They did not add entries even about data they were already familiar with through their daily activities with Bitcoin. This evidence shows challenges to conducting a micro-entries study in a free exploration phase and convincing Bitcoin miners to spend more time exploring the tool and coming up with more in-depth findings. Besides, some participants used the form to provide usability feedback on the tool rather than report findings. An additional feedback form should solve this problem to allow participants to report any bugs or feedback on the user interface.

We found that the interest of miners was in monitoring the recent activities in mining but not an in-depth analysis of mining pools on the historical data as my economist did. In the first interview, participants were intimately familiar with the transaction fee spike we asked about in Task 4. However, they were not spending more time looking for relevant information nor investigating familiar phenomena such as this one in-depth. If participants did not have their own questions, it would have been good to give specific analysis tasks and ask them to report answers to these tasks as in micro-entries. For example, asking participants to make one report each about pool fees, pool hopping, or rising and declining pools would have likely helped increase the number of responses—but given less useful information about their personal interests.

6.9 CONCLUSION

In this chapter, I proposed a visual analytics tool called MiningVis to understand the Bitcoin mining economy. Mining is a critical activity in cryptocurrencies that involves multiple longitudinal factors, both internal and external to the Bitcoin ecosystem. The tool allows analysts to relate multiple measures from complementary datasets in a multi-coordinated view. During the design iterations, the tool helped my economist collaborator discover new findings and he currently uses it to develop an economic model. We conducted a user study with Bitcoin miners to study both the usability and insights obtained from the tool. The user study confirmed the design of our main views but we also learned that pool hopping was less attractive 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 reported 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.

7 CONCLUSION

The Bitcoin blockchain provides a unique big dataset for the study of anonymous financial transactions in a decentralized network. Bitcoin consists of multiple components that can be studied in various aspects as categorized in the task domains of our systematic review in Chapter 3. The main challenge is that the activities related to Bitcoin, or blockchain in general, are diverse and organized not by centralized control but from the participants in the network (called *self-organization*). Bitcoin can be considered as a complex system because there are heterogeneous and autonomous entities in complex interaction [116]. On the contrary, Dos Santos argued that it should be viewed as an algorithmically complicated system unlikely to enter a chaotic regime [33]. Nonetheless, the Bitcoin economy is an exciting research field for the study of interaction between economic agents in a decentralized system governed by autonomous mechanisms.

Due to the complexity of the data, most of the existing works provide analyses on aggregated information of blockchain data. Most current tools show basic information visualizations that allows people to see aggregated information or granular transaction details. In this thesis, I applied a Visual Analytics (VA) approaches to to explore and monitor diverse activities in blockchain data with human-in-the-loop data analytics. The overarching goal of my thesis is to explore collective phenomena of the Bitcoin blockchain from collective *macrobehavior* analyses and allow analysts to drill down to *micromotives* behaviors of a diverse group of actors in the blockchain network [135].

This chapter summarizes the work of my thesis and my contributions to both blockchain and visual analytics research. I also propose some perspective toward future work in data-driven blockchain analysis and visualization.

7.1 SUMMARY AND CONTRIBUTIONS OF THE THESIS

In this section, I provide a short summary and describe the main contributions of my thesis. The contributions of my work are highlighted in *italics*.

The Bitcoin blockchain has a complicated mechanism to validate and store diverse transaction activities. Past research focused on theoretical analysis, analyzing transaction graphs, and applying machine learning to classify entity types or malicious transactions Chapter 2. To assess the current state of blockchain visualization and how they are used to understand activities in the blockchain, I conducted *a sys-*

tematic review to understand the state-of-the-art in blockchain data visualization (Chapter 3). I proposed a classification scheme to characterize tools I found. In addition, I assessed open challenges and research opportunities. I found that visualization tools dedicated to advanced and domain-focused questions are rare but will be helpful for experts and analysts to understand a particular activity in the Bitcoin blockchain in depth.

In the remainder of the thesis, I focused on Bitcoin mining which is a critical mechanism for the validation of transaction in the decentralized network. It also has implications on economic theory as miners react to reward incentives to continue perform their tasks. Mining pools emerged as organizations to ensure regular income among miners. Using a VA approach, I conducted *a two-year design study with my economist collaborator* to discover research questions and develop visualization prototypes to address questions regarding the emergence and evolution of Bitcoin mining pools.

In the first design iteration, I *extracted multiple data sources related to the development of market shares and characteristics of mining pools* (i.e., location, payout scheme, and pool fees) as explained in Chapter 4. Then, I created *interactive visualization prototypes in Jupyter Notebooks* that allowed my economist collaborator to analyze the evolution of mining pools regarding external factors from pool characteristics and Bitcoin network statistics. Based on these visualization prototypes, *we reported a first exploratory analysis on the evolution of mining pools over the entire Bitcoin history*. We discovered that early mining pools proposed various payout schemes before converging to PPS and FPPS because they pay regular income to miners. The primary location of mining pools shows that Chinese pools became global pools around September 2017 as the Chinese government regulated cryptocurrency trading.

After that, my economist collaborator and I further investigated the migration of miners among mining pools in the second design iteration described in Chapter 5. I developed *a methodology to extract reward payout flows from coinbase transactions to individual miners*. Then, I defined three main payout flow measures: new and dropout, hopping in and out, and cross-pooling to understand miners' mobility patterns. I showed that the method could extract various mining pools with different payout flow patterns. Using visualization prototypes, *we were able to report evidence about the motivation of miners to join, leave, or move in mining pools.* We found that miners move to the pool that provides lower transaction fees and more regular income to miners. We also detected that miners use cross-pooling to diversify their risk between PPS (risk-free) and PPLNS (risky) reward schemes.

During two design iterations, we discovered that mining pools competitions were evolving and needed to be constantly monitored over a long period of time. Moreover, Bitcoin market price and halving days in the future will impact the decision of miners to continue mining, and therefore the sustainability of this activity. Due to the demand to monitor this activity constantly, I developed *a visual analytics tool called* MiningVis *that allows analysts to relate multiple factors related to mining pools and miners' behaviors*. The VA approach visualizes detailed information and allows analysts to detect trends, outliers, or data heterogeneity from visualization views that cannot be detected directly from summary statistics. The details and visualization designs of the tool are described in Chapter 6. My economist collaborator uses the tool to derive interesting findings and develop a model to explain the mining economy. We also conducted *an online user study with eight Bitcoin miners* and found that they were more interested in real time information than long-term historical data. Nonetheless, the result from our user study supports our design choices for the tool. The tool can also be extended to other cryptocurrencies with the same *proof-of-work* protocol.

7.2 REFLECTIONS, LESSONS LEARNED, AND LIMITATIONS

In this thesis, I closely collaborated with an economist to discover research questions and developed a visualization tool to analyze the mining pool ecosystem. We started with rough research questions and continuously collected and analyzed data for the past two years while refining our questions further. Besides, I proposed a heuristic algorithm to extract miners from the payout flow to answer questions on miners' mobility. There are many interesting findings and practices from my thesis that I think it is worth discussing here as well as the limitations of my work.

A needle in the haystack. Bitcoin transaction data is vast and keeps growing. It contained >350 GB of raw data at the time of writing. Anyone can create a transaction for any purpose they want. This blockchain data pose analysis challenges because it has heterogeneous and pseudonymous properties. Without knowing the real-world information of users, it is difficult to track who exchanges with whom and extract meaningful patterns from such massive data. Data mining techniques are effective in finding frequent patterns from historical data. However, the lack of labeled datasets causes the limitation to training accurate supervised learning models. Weak signals on the sparse data like the Bitcoin transaction network also pose a difficulty in implementing unsupervised pattern mining. Moreover, the behaviors of Bitcoin users are diverse and changed over time, as we found some evidence in our Bitcoin mining activity studies. Therefore, data mining models should consider heterogeneous and paradigm shifts in blockchain data. In this case, VA should be an additional feature that helps researchers to detect the changes in the blockchain data and give feedback to tuning the model.

Other activities in the Bitcoin blockchain. Based on the demand of my economist collaborator, I focused on Bitcoin mining in this thesis. Bitcoin mining activity can be tracked from the origin of Bitcoin values from coinbase transactions. Besides, the topic has economic implications for the Bitcoin and blockchain ecosystems. However, other activities in Bitcoin are also worth investigating (e.g., money laundering, dark marketplaces, and the use of Bitcoin in some countries with high inflation rates). This could be a future challenge for visual analytics research to develop tools to investigate these activities.

Collaboration with the economist. During the design study, I had worked with one economist, which limits my ability to generalize observations from my work to the economists' community. Nonetheless, the outcome of my thesis is another piece of evidence that design study methodology can help not only to develop a VA tool but also to allow us to discover exciting results from exploratory data analysis. We were able to publish new findings in publications in our data-first visualization design study. With this respect, the MiningVis tool proved to be useful to our economist to investigate a lot of information on Bitcoin mining in multiple coordinated views and facilitated him to observe the data from both the overview and to drill down to the granular detail.

Quantitative evaluation of reward payout flow. In Section 5.2, I derived a heuristic algorithm to identify miners from the reward payout flow. However, there is no ground truth dataset that I can use to evaluate the method I proposed. Instead, I justified my approach indirectly by manually exploring the algorithm's payout flow result and calculating summary statistics to remove mining pools that do not satisfy the assumptions. Mining pool companies should have this information but could not share their customer data due to privacy concerns. Without this ground truth data, it will not be possible to directly evaluate the miners' detection model in the real world. Blockchain intelligence companies have been collecting the known entity dataset associated with addresses which are also helpful to assess the model I proposed. Still, they used that dataset as their competitive advantage and did not share the data with the public. Due to these limitations, I think the research community needs to develop a benchmark or hypothetic dataset to facilitate evaluation and comparison among different models.

The devil is in the details. My economist collaborator and I use many visualization prototypes to explore the data in the most granular detail possible. We found that the further we examined the data, the more we saw the heterogeneity and outliers in the details. This can possibly happen due to various reasons, including the noise in the data, errors from our miners' extraction approach, and some special events that drive miners to behave differently. Using the VA approach, we were able to investigate the data from aggregated-levels to the detailed information and vice versa to find explanations and contextual information as well as to detect interesting time periods that behave differently from what we expected. Nonetheless, we found it difficult to generalize the patterns we found. However, data visualizations keep us aware that granular detail data can be different from aggregated information.

Using micro-entries with Bitcoin miners. In our MiningVis user study, we were disappointed that Bitcoin miners did not actively report their findings as micro-entries, even though we tried to make the form lightweight. After the first interview, we lost contact with three participants without providing the reasons. As they were Bitcoin miners who wanted to take advantage of the competition, it is possible that they opted out of our study if they did or did not find something useful for their work. Five remaining participants who remained in our study reported that they did not have any specific questions to investigate and dedicated time to use the tool. Even though micro-entries are a helpful methodology for capturing user insights, we need to figure out how to motivate our participants in the online study to report findings to the remote study actively.

Instead of writing about their new observations, many participants reported bugs or suggestions about new features in micro-entries. In the future user study, I will be more careful to clarify what "finding" we expected and provide another feedback form for participants to report any bug or feature recommendations.

7.3 FUTURE WORK

In this section, I propose some possible future directions for blockchain visual analytics research.

7.3.1 *Characterization of entities in the Bitcoin network*

In contrast to traditional financial transactions, the identities of Bitcoin users (entities) are normally pseudonymous, in other words, cannot be derived directly from transaction data. The challenge is to characterize different types of users in the Bitcoin network as well as the pattern of their usages over time. Previous visualization works proposed tools to explore activities of a single entity [V8] or groups of entities [V9] based on clustering heuristics. Recent works about Bitcoin entities classification applied machine learning techniques to classify different types of entities (e.g., exchanges, miners, gambling, and marketplace) [60, 75, 91, 178, 179]. Those black-box classification models gave high accuracy but are difficult to interpret. Interpretable machine learning [54, 59, 102] and dimensionality reduction techniques [36, 37, 111] should help VA researchers develop tools to visually explore feature attributes (i.e., summary statistics, temporal pattern, and graph attributes) that distinguish different entity types while also displaying the degree of uncertainty in classification models.

7.3.2 *Illegal activity tracking in the Bitcoin blockchain*

Bitcoin and other cryptocurrencies have been used to transfer money for illegal activities, such as dark markets, gambling, mixing services, terrorist organizations, scams, ransomware, and Ponzi schemes, since they are not regulated by the government. Financial regulators would be interested to analyze transaction networks and detect suspicious illegal transactions. BitConeView [V5] and Tendril of Crime [V1] are the existing tools that trace suspicious transactions. Elliptic, a cryptocurrency intelligence company, published a public data set of the Bitcoin transaction network [35]. Each transaction is annotated as either containing licit or illicit addresses. The dataset has been used for graph machine learning classification tasks [168]. The result shows that the classification model did not perform well when activities in Bitcoin are abruptly changed (e.g., dark market shutdown and government regulation). To help money laundering investigators improve the automatic classification, it would be interesting to develop a new VA tool to explore illicit transactions in the large network as well as capture the abrupt changes of transaction patterns at different time intervals that could not be captured in the machine learning model.

7.3.3 *Smart contracts and decentralized finance*

Ethereum is one of the widely used blockchains that implements smart contract functionality [170]. A smart contract is a piece of computer code guaranteed to run in the same way on all peers. Developers can develop smart contracts to build decentralized applications that run on the Ethereum network. Nowadays, smart contracts have been used extensively to create *Decentralized Finance* (DeFi) that provide financial services (e.g., lending, borrowing, exchanges, derivatives) on the decentralized network [169]. Dune Analytics provides an online platform that allows users to create a dashboard to monitor aggregated statistics of measures over time [O19]. For more advanced analytics, The Graph provides a decentralized API to access transactions on DeFi protocols [52]. These open financial datasets offer researchers opportunities to analyze traders' behaviors, dynamics of exchange activities, and systematic risks of the protocols in granular detail. However, the transaction data is challenging to comprehend and model due to the complicated mechanisms that govern the protocol. Therefore, VA would be a useful approach to help visually present the transaction data that is both understandable to the general audience and facilitate advanced analytics to researchers.

7.4 CONCLUSION

This thesis studies the Bitcoin mining economy from comprehensive data sources and uses a visual analytics approach to facilitate data analysis tasks. I conducted a systematic review of visualizations of blockchain data and proposed a classification scheme to understand what is missing in the current work. For two years, I conducted a design study with an economist expert to acquire research questions regarding Bitcoin mining pools. During this process, we obtained some new results and reported them in two articles [154, 156]. The empirical analysis provides evidence that Bitcoin mining became an industry that mining pools compete to offer a regular reward while miners behave like rational agents to maximize their income. We also found that the top mining pools and their pool characteristics keep changing over the years. At the end of the design study, I proposed a visual analytics tool called MiningVis to monitor the Bitcoin mining economy in the long term. I conducted an online user study to study the usability of the tool. MiningVis tool helps the economist and Bitcoin miners in our user study to relate multiple factors in Bitcoin mining and look into a specific time frame of interests. The outcomes of my thesis prove that the visual analytics approach is promising to explore complex and heterogeneous data for a sophisticated mechanism like the Bitcoin blockchain. The result from the user study confirms that the value of the MiningVis tool lies in helping analysts identify interesting findings and relate multiple economic factors from numerous sources of data.

REFERENCES

[1] Mansoor Ahmed, Ilia Shumailov, and Ross Anderson. "Tendrils of Crime: Visualizing the Diffusion of Stolen Bitcoins." In: *Graphical Models for Security*. Springer, 2019, pp. 1–12. DOI: 10.1007/978-3-030-15465-3_1.

[2] Cuneyt Gurcan Akcora, Murat Kantarcioglu, and Yulia R. Gel. *Blockchain Networks: Data Structures of Bitcoin, Monero, Zcash, Ethereum, Ripple and Iota*. 2021. arXiv: 2103.08712 [cs.CR].

[3] Eitan Altman, Daniel Sadoc Menasche, Alexandre Reiffers, Mandar Datar, Swapnil Dhamal, Corinne Touati, and Rachid El-Azouzi. "Blockchain competition between miners: a game theoretic perspective." In: *Frontiers in Blockchain* 2 (2019), p. 26. DOI: 10.3389/fbloc.2019.00026.

[4] J Ignacio Alvarez-Hamelin, Luca Dall'Asta, Alain Barrat, and Alessandro Vespignani. "Large scale networks fingerprinting and visualization using the k-core decomposition." In: *Proceedings of the 18th International Conference on Neural Information Processing Systems*. 2005, pp. 41–50. URL: https://proceedings.neurips.cc/paper/2005/file/b19aa25ff58940d974234b48391b9549-Paper.pdf.

[5] Andreas M. Antonopoulos. *Mastering Bitcoin: Programming the Open Blockchain*. 2nd. O'Reilly Media, Inc., 2017. ISBN: 978-1491954386.

[6] Fredy Andres Aponte-Novoa, Ana Lucila Sandoval Orozco, Ricardo Villanueva-Polanco, and Pedro Wightman. "The 51% Attack on Blockchains: A Mining Behavior Study." In: *IEEE Access* 9 (2021), pp. 140549–140564. DOI: 10.1109/ACCESS.2021.3119291.

[7] Sarah Azouvi and Alexander Hicks. "SoK: Tools for Game Theoretic Models of Security for Cryptocurrencies." In: *Cryptoeconomic Systems* 0.1 (Apr. 5, 2021). DOI: 10.21428/58320208.8e7f4fab.

[8] Anastasios Balaskas and Virginia NL Franqueira. "Analytical tools for blockchain: Review, taxonomy and open challenges." In: *International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. IEEE. 2018, pp. 1–8. DOI: 10.1109/CyberSecPODS.2018.8560672.

[9] Massimo Bartoletti, Stefano Lande, Livio Pompianu, and Andrea Bracciali. "A general framework for blockchain analytics." In: *Proceedings of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers*. 2017, pp. 1–6. DOI: 10.1145/3152824.3152831.

[10] Dave Bayer, Stuart Haber, and W Scott Stornetta. "Improving the efficiency and reliability of digital time-stamping." In: *Sequences II: Methods in Communication, Security, and Computer Science*. Springer, 1993, pp. 329–334. DOI: 10.1007/978-1-4613-9323-8_24.

[11] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. "A taxonomy and survey of dynamic graph visualization." In: *Computer Graphics Forum*. Vol. 36. 1. Wiley. 2017, pp. 133–159. DOI: 10.1111/cgf.12791.

[12] Marianna Belotti, Nikola Božić, Guy Pujolle, and Stefano Secci. "A vademecum on blockchain technologies: When, which, and how." In: *IEEE Communications Surveys & Tutorials* 21.4 (2019), pp. 3796–3838. DOI: 10.1109/COMST.2019.2928178.

[13] Marianna Belotti, Sofiane Kirati, and Stefano Secci. "Bitcoin pool-hopping detection." In: *IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI)*. IEEE, 2018, pp. 1–6. DOI: 10.1109/RTSI.2018.8548376.

[14] Marc Bevand. *Cambridge Bitcoin Electricity Consumption Index*. Accessed on: Aug 26, 2021. 2021. URL: https://cbeci.org/.

[15] Alex Biryukov, Dmitry Khovratovich, and Ivan Pustogarov. "Deanonymisation of clients in Bitcoin P2P network." In: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*. 2014, pp. 15–29. DOI: 10.1145/2660267.2660379.

[16] Stefano Bistarelli, Ivan Mercanti, and Francesco Santini. "A suite of tools for the forensic analysis of bitcoin transactions: Preliminary report." In: *European Conference on Parallel Processing*. Springer. 2018, pp. 329–341. DOI: 10.1109/COMST.2018.2818623.

[17] Bitcoin Wiki contributors. *Comparison of mining pools*. Accessed on: Dec 12, 2020. URL: https://en.bitcoin.it/wiki/Comparison_of_mining_pools.

[18] Bitcointalk contributors. *Bitcoin Mining Pools Discussion Forum*. Accessed on: Dec 12, 2020. URL: https://bitcointalk.org/index.php?board=41.0.

[19] Bitcointalk contributors. *Bitcoin Mining Pools Discussion Forum*. https://bitcointalk.org/index.php?board=41.0. URL: https://bitcointalk.org/index.php?board=41.0.

[20] Bitcointalk contributors. *Bitcoin Press Discussion Forum*. URL: https://bitcointalk.org/index.php?board=77.0.

[21] Jeremy E Block and Eric 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)*. IEEE. 2020, pp. 38–47. DOI: 10.1109/BELIV51497.2020.00012.

[22] Blockchain.info. *Blockchain Known Pools*. Accessed on: Dec 12, 2020. URL: https://raw.githubusercontent.com/blockchain/Blockchain-Known-Pools/master/pools.json.

[23] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A. Kroll, and Edward W. Felten. "SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies." In: *IEEE Symposium on Security and Privacy*. 2015, pp. 104–121. DOI: 10.1109/SP.2015.14.

[24] Nadia Boukhelifa, Fanny Chevalier, and Jean-Daniel Fekete. "Real-time aggregation of wikipedia data for visual analytics." In: *IEEE Symposium on Visual Analytics Science and Technology (VAST)*. IEEE. 2010, pp. 147–154. DOI: 10.1109/VAST.2010.5652896.

[25] Danny Bradbury. "The problem with Bitcoin." In: *Computer Fraud & Security* 2013.11 (2013), pp. 5–8. DOI: 10.1016/S1361-3723(13)70101-5.

[26] Willard Cope Brinton. *Graphic methods for presenting facts*. Engineering magazine company, 1919.

[27] John Brooke. "SUS: a quick and dirty usability." In: *Usability evaluation in industry* 189 (1996).

[28] BTC.com. *Blockchain Known Pools*. Accessed on: Dec 12, 2020. URL: https://raw.githubusercontent.com/btccom/Blockchain-Known-Pools-BCH/master/pools.json.

[29] Vitalik Buterin. *On Public and Private Blockchains*. 2015. URL: https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/.

[30] Mauro Conti, E Sandeep Kumar, Chhagan Lal, and Sushmita Ruj. "A survey on security and privacy issues of bitcoin." In: *IEEE Communications Surveys & Tutorials* 20.4 (2018), pp. 3416–3452. DOI: 10.1109/COMST.2018.2842460.

[31] Kristin A Cook and James J Thomas. *Illuminating the path: The research and development agenda for visual analytics*. Tech. rep. Pacific Northwest National Lab. (PNNL), Richland, WA (United States), 2005.

[32] Giuseppe Di Battista, Valentino Di Donato, Maurizio Patrignani, Maurizio Pizzonia, Vincenzo Roselli, and Roberto Tamassia. "Bitconeview: visualization of flows in the bitcoin transaction graph." In: *Symposium on Visualization for Cyber Security (VizSec)*. IEEE, 2015, pp. 1–8. DOI: 10.1109/VIZSEC.2015.7312773. URL: https://doi.org/10.1109/VIZSEC.2015.7312773.

[33] Renato P Dos Santos. "On the philosophy of Bitcoin/Blockchain technology: is it a chaotic, complex system?" In: *Metaphilosophy* 48.5 (2017), pp. 620–633. DOI: 10.1111/meta.12266.

[34] Pierre Dragicevic, Yvonne Jansen, and Andrew Vande Moere. "Data Physicalization." In: *Springer Handbook of Human Computer Interaction*. Springer, 2021. URL: https://hal.inria.fr/hal-02113248.

[35] Elliptic. *Elliptic Data Set: Bitcoin Transaction Graph*. 2019. URL: https://www.kaggle.com/ellipticco/elliptic-data-set.

[36] Daniel Engel, Lars Hüttenberger, and Bernd Hamann. "A survey of dimension reduction methods for high-dimensional data analysis and visualization." In: *Visualization of Large and Unstructured Data Sets-Proceedings of IRTG 1131 Workshop 2011*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2012. DOI: 10.4230/OASIcs.VLUDS.2011.135.

[37] Mateus Espadoto, Rafael M Martins, Andreas Kerren, Nina ST Hirata, and Alexandru Cristian Telea. "Towards a quantitative survey of dimension reduction techniques." In: *IEEE Transactions on Visualization and Computer Graphics* (2019). DOI: 10.1109/TVCG.2019.2944182.

[38] ExplosionAI GmbH. *spaCy: Industrial-Strength Natural Language Processing*. Accessed on: Oct 24, 2021. URL: https://spacy.io/.

[39] Ittay Eyal and Emin Gün Sirer. "Majority is Not Enough: Bitcoin Mining is Vulnerable." In: *Commun. ACM* 61.7 (June 2018), pp. 95–102. ISSN: 0001-0782. DOI: 10.1145/3212998. URL: https://doi.org/10.1145/3212998.

[40] Susan Fairchild, Timothy Farrell, Brad Gunton, Anne Mackinnon, Christina McNamara, and Roberta Trachtman. "Design and Data in Balance: Using Design-Driven Decision Making to Enable Student Success." In: *New Visions for Public Schools* (2014). URL: https://files.eric.ed.gov/fulltext/ED547420.pdf.

[41] Jean-Daniel Fekete, Danyel Fisher, Arnab Nandi, and Michael Sedlmair. "Progressive Data Analysis and Visualization (Dagstuhl Seminar 18411)." In: *Dagstuhl Reports* 8.10 (2019), pp. 1–40. DOI: 10.4230/DagRep.8.10.1.

[42] Stephen Few. "Dual-Scaled Axes in Graphs—Are They Ever the Best Solution?" In: *Visual Business Intelligence Newsletter*. Perceptual Edge, Mar. 2008. URL: http://www.perceptualedge.com/articles/visual_business_intelligence/dual-scaled_axes.pdf.

[43] Ben Fisch, Rafael Pass, and Abhi Shelat. "Socially Optimal Mining Pools." In: *Web and Internet Economics*. Springer, 2017, pp. 205–218. DOI: 10.1007/978-3-319-71924-5_15.

[44] Spyros Foteinis. "Bitcoin's alarming carbon footprint." In: *Nature* 554.7690 (2018), pp. 169–170. DOI: 10.1038/d41586-018-01625-x.

[45] Guilherme de Freitas Castro. *The Bitcoin's lifecycle overview*. 2018. URL: https://dev.to/gmfcastro/the-bitcoins-lifecycle-overview-1fld.

[46] Siwei Fu, Jian Zhao, Weiwei Cui, and Huamin Qu. "Visual analysis of MOOC forums with iForum." In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2016), pp. 201–210. DOI: 10.1109/TVCG.2016.2598444.

[47] Johanna Fulda, Matthew Brehmer, and Tamara Munzner. "TimeLineCurator: Interactive authoring of visual timelines from unstructured text." In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (2015), pp. 300–309. DOI: 10.1109/TVCG.2015.2467531.

[48] Weichao Gao, William G Hatcher, and Wei Yu. "A survey of blockchain: Techniques, applications, and challenges." In: *27th International Conference on Computer Communication and Networks (ICCCN)*. IEEE. 2018, pp. 1–11. DOI: 10.1109/ICCCN.2018.8487348.

[49] Simin Ghesmati, Walid Fdhila, and Edgar Weippl. "Studying Bitcoin Privacy Attacks and Their Impact on Bitcoin-Based Identity Methods." In: *International Conference on Business Process Management*. Springer. 2021, pp. 85–101. DOI: 10.1007/978-3-030-85867-4_7.

[50] Pasquale Giungato, Roberto Rana, Angela Tarabella, and Caterina Tricase. "Current trends in sustainability of bitcoins and related blockchain technology." In: *Sustainability* 9.12 (2017), p. 2214. DOI: 10.3390/su9122214.

[51] Michael Gleicher. "Considerations for visualizing comparison." In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2017), pp. 413–423. DOI: 10.1109/TVCG.2017.2744199.

[52] The Graph. *The Graph: APIs for a vibrant decentralized future*. 2021. URL: https://thegraph.com/en/.

[53] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. "LineUp: Visual analysis of multi-attribute rankings." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2277–2286. DOI: 10.1109/TVCG.2013.173.

[54] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. "A survey of methods for explaining black box models." In: *ACM computing surveys (CSUR)* 51.5 (2018), pp. 1–42. DOI: 10.1145/3236009.

[55] Yi-Ming Guo, Zhen-Ling Huang, Ji Guo, Xing-Rong Guo, Hua Li, Meng-Yu Liu, Safa Ezzeddine, and Mpeoane Judith Nkeli. "A bibliometric analysis and visualization of blockchain." In: *Future Generation Computer Systems* 116 (2021), pp. 316–332. DOI: 10.1016/j.future.2020.10.023.

[56] Stuart Haber and W Scott Stornetta. "How to time-stamp a digital document." In: *Proc. Conference on the Theory and Application of Cryptography*. Springer. 1990, pp. 437–455. DOI: 10.1007/3-540-38424-3_32.

[57] K. W. Haemer. "Double Scales are Dangerous." In: *The American Statistician* 2.3 (June 1948), p. 24.

[58] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. "Exploring Network Structure, Dynamics, and Function using NetworkX." In: *Proceedings of the Python in Science Conference*. 2008, pp. 11–15.

[59] Patrick Hall. *On the Art and Science of Machine Learning Explanations*. 2020. arXiv: 1810.02909 [stat.ML].

[60] Mikkel Alexander Harlev, Haohua Sun Yin, Klaus Christian Langenheldt, Raghava Mukkamala, and Ravi Vatrapu. "Breaking bad: De-anonymising entity types on the bitcoin blockchain using supervised machine learning." In: *Proceedings of the 51st Hawaii International Conference on System Sciences*. 2018. DOI: 10.24251/HICSS.2018.443.

[61] Martin Harrigan and Christoph Fretter. "The unreasonable effectiveness of address clustering." In: *IEEE Conferences 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. DOI: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0071.

[62] Susan Havre, Elizabeth Hetzler, Paul Whitney, and Lucy Nowell. "Themeriver: Visualizing thematic changes in large document collections." In: *IEEE Transactions on Visualization and*

Computer Graphics 8.1 (2002), pp. 9–20. DOI: 10.1109/2945.981848.

[63] Florian Heimerl, Qi Han, Steffen Koch, and Thomas Ertl. "CiteRivers: Visual analytics of citation patterns." In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (2015), pp. 190–199. DOI: 10.1109/TVCG.2015.2467621.

[64] Garrick Hileman and Michel Rauchs. "Global cryptocurrency benchmarking study." In: *Cambridge Centre for Alternative Finance* (2017).

[65] Nicolas Houy. "The Bitcoin mining game." In: *Ledger* 1 (2016), pp. 53–68. DOI: 10.5195/ledger.2016.13.

[66] Nicolas Houy. "The economics of Bitcoin transaction fees." In: *GATE WP* 1407 (2014). DOI: 10.2139/ssrn.2400519.

[67] Samuel Huron, Romain Vuillemot, and Jean-Daniel Fekete. "Visual sedimentation." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2446–2455. DOI: 10.1109/TVCG.2013.227.

[68] International Monetary Fund. *Primary Commodity Price System: Fuel (Energy) Index*. Accessed on: Aug 26, 2021. 2021. URL: https://data.imf.org/?sk=471DDDF8-D8A7-499A-81BA-5B332C01F8B9&sId=1547558078595.

[69] Petra Isenberg, Anastasia Bezerianos, Pierre Dragicevic, and Jean-Daniel Fekete. "A Study on Dual-Scale Data Charts." In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2469–2487. DOI: 10.1109/TVCG.2011.160.

[70] Jacob J. *China Tightens Online Finance Regulations*. 2016. URL: https://cointelegraph.com/news/china-tightens-online-finance-regulations?ref=5823.

[71] Aleš Janda. *Information about WalletExplorer.com*. Accessed on: Dec 12, 2020. URL: https://www.walletexplorer.com/info.

[72] Amit Jena, Ulrich Engelke, Tim Dwyer, Venkatesh Raiamanickam, and Cecile Paris. "Uncertainty visualisation: An interactive visual survey." In: *IEEE Pacific Visualization Symposium (PacificVis)*. IEEE. 2020, pp. 201–205. DOI: 10.1109/PacificVis48177.2020.1014.

[73] Shangrong Jiang, Yuze Li, Quanying Lu, Yongmiao Hong, Dabo Guan, Yu Xiong, and Shouyang Wang. "Policy assessments for the carbon emission flows and sustainability of Bitcoin blockchain operation in China." In: *Nature communications* 12.1 (2021), pp. 1–10. DOI: 10.1038/s41467-021-22256-3.

[74] Benjamin Johnson, Aron Laszka, Jens Grossklags, Marie Vasek, and Tyler Moore. "Game-Theoretic Analysis of DDoS Attacks Against Bitcoin Mining Pools." In: *Financial Cryptography and Data Security*. Springer, 2014, pp. 72–86. DOI: 10.1007/978-3-662-44774-1_6.

[75] Marc Jourdan, Sebastien Blandin, Laura Wynter, and Pralhad Deshpande. "Characterizing entities in the bitcoin blockchain." In: *IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018, pp. 55–62. DOI: 10.1109/ICDMW.2018.00016.

[76] Péter L Juhász, József Stéger, Dániel Kondor, and Gábor Vattay. "A bayesian approach to identify bitcoin users." In: *PloS one* 13.12 (2018), e0207000. DOI: 10.1371/journal.pone.0207000.

[77] Ben Kaiser, Mireya Jurado, and Alex Ledger. *The Looming Threat of China: An Analysis of Chinese Influence on Bitcoin*. 2018. arXiv: 1810.02466 [cs.CR].

[78] Harry Kalodner, Malte Möser, Kevin Lee, Steven Goldfeder, Martin Plattner, Alishah Chator, and Arvind Narayanan. "BlockSci: Design and applications of a blockchain analysis platform." In: *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 2020, pp. 2721–2738. ISBN: 978-1-939133-17-5. URL: https://www.usenix.org/conference/usenixsecurity20/presentation/kalodner.

[79] Dan Kaminsky. *Black Ops of TCP/IP 2011 (Black Hat USA 2011)*. 2011. URL: https://www.slideshare.net/dakami/black-ops-of-tcpip-2011-black-hat-usa-2011.

[80] Irni Eliana Khairuddin and Corina Sas. "An Exploration of Bitcoin mining practices: Miners' trust challenges and motivations." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–13. DOI: https://doi.org/10.1145/3290605.3300859.

[81] Merve Can Kus Khalilov and Albert Levi. "A survey on anonymity and privacy in bitcoin-like digital cash systems." In: *IEEE Communications Surveys & Tutorials* 20.3 (2018), pp. 2543–2585. DOI: 10.1109/COMST.2018.2818623.

[82] Christine Kim. *The Rise of ASICs: A Step-by-Step History of Bitcoin Mining*. 2020. URL: https://www.coindesk.com/tech/2020/04/26/the-rise-of-asics-a-step-by-step-history-of-bitcoin-mining/.

[83] Joshua A Kroll, Ian C Davey, and Edward W Felten. "The economics of Bitcoin mining, or Bitcoin in the presence of adversaries." In: *Proceedings of Workshop on the Economics of Information Security (WEIS)*. Vol. 2013. 2013, p. 11.

[84] Milos Krstajic, Enrico Bertini, and Daniel Keim. "Cloudlines: Compact display of event episodes in multiple time-series." In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2432–2439. DOI: 10.1109/TVCG.2011.179.

[85] Martin Krzywinski, Jacqueline Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J Jones, and Marco A Marra. "Circos: an information aesthetic for comparative genomics." In: *Genome Research* 19.9 (2009), pp. 1639–1645. DOI: 10.1101/gr.092759.109.

[86] Kostiantyn Kucher and Andreas Kerren. "Text visualization techniques: Taxonomy, visual survey, and community insights." In: *IEEE Pacific Visualization Symposium (PacificVis)*. IEEE. 2015, pp. 117–121. DOI: 10.1109/PACIFICVIS.2015.7156366.

[87] Sinan Küfeoğlu and Mahmut Özkuran. "Bitcoin mining: A global review of energy and power demand." In: *Energy Research & Social Science* 58 (2019), p. 101273. DOI: 10.1016/j.erss.2019.101273.

[88] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. "Empirical studies in information visualization: Seven scenarios." In: *IEEE transactions on visualization and computer graphics* 18.9 (2011), pp. 1520–1536. DOI: 10.1109/TVCG.2011.279.

[89] Aron Laszka, Benjamin Johnson, and Jens Grossklags. "When Bitcoin Mining Pools Run Dry." In: *Financial Cryptography and Data Security*. Springer, 2015, pp. 63–77. DOI: 10.1007/978-3-662-48051-9_5.

[90] Yoad Lewenberg, Yoram Bachrach, Yonatan Sompolinsky, Aviv Zohar, and Jeffrey S Rosenschein. "Bitcoin Mining Pools: A Cooperative Game Theoretic Analysis." In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. AAMAS '15. Istanbul, Turkey: International Foundation for Autonomous Agents and Multiagent Systems, 2015, pp. 919–927. DOI: 10.5555/2772879.2773270.

[91] Yu-Jing Lin, Po-Wei Wu, Cheng-Han Hsu, I-Ping Tu, and Shih-wei Liao. "An evaluation of bitcoin address classification based on transaction history summarization." In: *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE. 2019, pp. 302–310. DOI: 10.1109/BLOC.2019.8751410.

[92] Xiao Fan Liu, Xin-Jian Jiang, Si-Hao Liu, and Chi Kong Tse. "Knowledge Discovery in Cryptocurrency Transactions: A Survey." In: *IEEE Access* (2021). DOI: https://doi.org/10.1109/ACCESS.2021.3062652.

[93] Xiao Fan Liu, Xin-Jian Jiang, Si-Hao Liu, and Chi Kong Tse. "Knowledge discovery in cryptocurrency transactions: a survey." In: *IEEE Access* 9 (2021), pp. 37229–37254. DOI: 10.1109/ACCESS.2021.3062652.

[94] Xiaojun Liu, Wenbo Wang, Dusit Niyato, Narisa Zhao, and Ping Wang. "Evolutionary Game for Mining Pool Selection in Blockchain Networks." In: *IEEE Wireless Communications Letters* 7.5 (2018), pp. 760–763. DOI: 10.1109/LWC.2018.2820009.

[95] Xiaotong Liu, Anbang Xu, Liang Gou, Haibin Liu, Rama Akkiraju, and Han-Wei Shen. "SocialBrands: Visual analysis of public perceptions of brands on social media." In: *IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE. 2016, pp. 71–80. DOI: 10.1109/VAST.2016.7883513.

[96] Ziyao Liu, Nguyen Cong Luong, Wenbo Wang, Dusit Niyato, Ping Wang, Ying-Chang Liang, and Dong In Kim. "A survey on blockchain: A game theoretical perspective." In: *IEEE Access* 7 (2019), pp. 47615–47643. DOI: 10.1109/ACCESS.2019.2909924.

[97] Yang Lu. "The blockchain: State-of-the-art and research challenges." In: *Journal of Industrial Information Integration* 15 (2019), pp. 80–90. DOI: 10.1016/j.jii.2019.04.002.

[98] Alan M MacEachren, Robert E Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan. "Visual semiotics & uncertainty visualization: An empirical study." In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2496–2505. DOI: 10.1109/TVCG.2012.279.

[99] Eric Masanet, Arman Shehabi, Nuoa Lei, Harald Vranken, Jonathan Koomey, and Jens Malmodin. "Implausible projections overestimate near-term Bitcoin CO_2 emissions." In: *Nature Climate Change* 9.9 (2019), pp. 653–654. DOI: 10.1038/s41558-018-0321-8.

[100] Ignasi Merediz-Solà and Aurelio F Bariviera. "A bibliometric analysis of bitcoin scientific production." In: *Research in International Business and Finance* 50 (2019), pp. 294–305. DOI: 10.1016/j.ribaf.2019.06.008.

[101] D Mintz, T Fitz-Simons, and M Wayland. "Tracking air quality trends with SAS/GRAPH." In: *Proceedings of the 22nd Annual SAS User Group International Conference (SUGI)*. 1997, pp. 807–812. URL: https://support.sas.com/resources/papers/proceedings/proceedings/sugi22/INFOVIS/PAPER173.PDF.

[102] Christoph Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/. 2019.

[103] Ahmed Afif Monrat, Olov Schelén, and Karl Andersson. "A survey of blockchain from the perspectives of applications, challenges, and opportunities." In: *IEEE Access* 7 (2019), pp. 117134–117151. DOI: 10.1109/ACCESS.2019.2936094.

[104] Camilo Mora, Randi L Rollins, Katie Taladay, Michael B Kantar, Mason K Chock, Mio Shimada, and Erik C Franklin. "Bitcoin emissions alone could push global warming above 2°C." In: *Nature Climate Change* 8.11 (2018), pp. 931–933. DOI: 10.1038/s41558-018-0321-8.

[105] Dominik Moritz, Bill Howe, and Jeffrey Heer. "Falcon: Balancing interactive latency and resolution sensitivity for scalable linked visualizations." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–11. DOI: 10.1145/3290605.3300924.

[106] Malte Möser and Rainer Böhme. "Trends, tips, tolls: A longitudinal study of Bitcoin transaction fees." In: *International Conference on Financial Cryptography and Data Security*. Springer. 2015, pp. 19–33. DOI: 10.1007/978-3-662-48051-9_2.

[107] Malte Möser and Arvind Narayanan. *Resurrecting Address Clustering in Bitcoin*. 2021. arXiv: 2107.05749 [cs.CR].

[108] Tamara Munzner. "A nested model for visualization design and validation." In: *IEEE transactions on visualization and computer graphics* 15.6 (2009), pp. 921–928. DOI: 10.1109/TVCG.2009.111.

[109] Satoshi Nakamoto. *Bitcoin: A peer-to-peer electronic cash system*. Tech. rep. 2008. URL: http://bitcoin.org/bitcoin.pdf.

[110] Cong T Nguyen, Dinh Thai Hoang, Diep N Nguyen, Dusit Niyato, Huynh Tuong Nguyen, and Eryk Dutkiewicz. "Proof-of-stake consensus mechanisms for future blockchain networks: fundamentals, applications and opportunities." In: *IEEE Access* 7 (2019), pp. 85727–85745. DOI: 10.1109/ACCESS.2019.2925010.

[111] Luis Gustavo Nonato and Michael Aupetit. "Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment." In: *IEEE Transactions on Visualization and Computer Graphics* 25.8 (2018), pp. 2650–2673. DOI: 10.1109/TVCG.2018.2846735.

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

[113] Alex T Pang, Craig M Wittenbrink, and Suresh K Lodha. "Approaches to uncertainty visualization." In: *The Visual Computer* 13.8 (1997), pp. 370–390. DOI: 10.1007/s003710050111.

[114] Morgen E. Peck. *Why The Biggest Bitcoin Miners are in China*. 2017. URL: https://spectrum.ieee.org/bitcoin-mining.

[115] Charles Perin, Jeremy Boy, and Frédéric Vernier. "Using gap charts to visualize the temporal evolution of ranks and scores." In: *IEEE Computer Graphics and Applications* 36.5 (2016), pp. 38–49. DOI: 10.1109/MCG.2016.100.

[116] Marc Pilkington. "Bitcoin through the lenses of complexity theory." In: *Handbook on the Geographies of Money and Finance*. Edward Elgar Publishing, 2017. DOI: 10.4337/9781784719005.00035.

[117] Zachary Pousman, John Stasko, and Michael Mateas. "Casual information visualization: Depictions of data in everyday life." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1145–1152. DOI: 10.1109/TVCG.2007.70541.

[118] Julien Prat and Benjamin Walter. "An equilibrium model of the market for bitcoin mining." In: *Journal of Political Economy* 129.8 (2021), pp. 2415–2452. DOI: 10.1086/714445.

[119] Julien Prat and Benjamin Walter. "An equilibrium model of the market for bitcoin mining." In: *Journal of Political Economy* 129.8 (2021), pp. 2415–2452. DOI: 10.1086/714445.

[120] Jipeng Qiang, Yun Li, Yunhao Yuan, Wei Liu, and Xindong Wu. *STTM: A Tool for Short Text Topic Modeling*. 2018. arXiv: 1808.02215 [cs.IR].

[121] Jipeng Qiang, Zhenyu Qian, Yun Li, Yunhao Yuan, and Xindong Wu. "Short text topic modeling techniques, applications, and performance: a survey." In: *IEEE Transactions on Knowledge and Data Engineering* (2020). DOI: 10.1109/TKDE.2020.2992485.

[122] Frederick Reese. *As Bitcoin Halving Approaches, 51% Attack Question Resurfaces*. 2016. URL: https://www.coindesk.com/tech/2020/04/26/the-rise-of-asics-a-step-by-step-history-of-bitcoin-mining/.

[123] Fergal Reid and Martin Harrigan. "An Analysis of Anonymity in the Bitcoin System." In: *Security and Privacy in Social Networks*. Springer, 2013, pp. 197–223. DOI: 10.1007/978-1-4614-4139-7_10.

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

[125] Reuters. *U.S. becomes largest bitcoin mining centre after China crackdown*. 2021. URL: https://www.reuters.com/technology/us-becomes-largest-bitcoin-mining-centre-following-china-ban-2021-10-13/.

[126] Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. "Interactive sankey diagrams." In: *IEEE Symposium on Information Visualization*. IEEE. 2005, pp. 233–240. DOI: 10.1109/INFVIS.2005.1532152.

[127] Matteo Romiti, Aljosha Judmayer, Alexei Zamyatin, and Bernhard Haslhofer. *A Deep Dive into Bitcoin Mining Pools: An Empirical Analysis of Mining Shares*. 2019. arXiv: 1905.05999 [cs.CR].

[128] Meni Rosenfeld. *Analysis of Bitcoin Pooled Mining Reward Systems*. 2011. arXiv: 1112.4980 [cs.DC].

[129] Martin Rosvall and Carl T Bergstrom. "Mapping change in large networks." In: *PloS one* 5.1 (2010), e8694. DOI: 10.1371/journal.pone.0008694.

[130] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey Ellis, and Daniel A Keim. "Knowledge generation model for visual analytics." In: *IEEE transactions on visualization and computer graphics* 20.12 (2014), pp. 1604–1613. DOI: 10.1109/TVCG.2014.2346481.

[131] Mehrdad Salimitari, Mainak Chatterjee, Murat Yuksel, and Eduardo Pasiliao. "Profit maximization for bitcoin pool mining: A prospect theoretic approach." In: *IEEE International Conference on Collaboration and Internet Computing*. IEEE. 2017, pp. 267–274. DOI: 10.1109/CIC.2017.00043.

[132] Marc Santamaria Ortega. "The Bitcoin transaction graph anonymity." In: (2013). URL: http://openaccess.uoc.edu/webapps/o2/bitstream/10609/23562/9/msantamariaoTFM0613memoria.pdf.

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

[134] Sarwar Sayeed and Hector Marco-Gisbert. "Assessing blockchain consensus and security mechanisms against the 51% attack." In: *Applied Sciences* 9.9 (2019), p. 1788.

[135] Thomas C Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[136] Calvin Fisher Schmid. *Statistical Graphics: Design Principles and Practices*. Krieger Pub Co, 1992.

[137] Mario Schmidt. "The Sankey diagram in energy and material flow management: part II: methodology and current applications." In: *Journal of industrial ecology* 12.2 (2008), pp. 173–185. DOI: 10.1111/j.1530-9290.2008.00015.x.

[138] Okke Schrijvers, Joseph Bonneau, Dan Boneh, and Tim Roughgarden. "Incentive Compatibility of Bitcoin Mining Pool Reward Functions." In: *Financial Cryptography and Data Security*. Springer, 2017, pp. 477–498. DOI: 10.1007/978-3-662-54970-4_28.

[139] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. "Design study methodology: Reflections from the trenches and the stacks." In: *IEEE transactions on visualization and computer graphics* 18.12 (2012), pp. 2431–2440. DOI: 10.1109/TVCG.2012.213.

[140] William Forsyth Sharpe. "The Sharpe Ratio." In: *The Journal of Portfolio Management* 21.1 (1994), pp. 49–58.

[141] Samuel Shen and Andrew Galbraith. *China's ban forces some bitcoin miners to flee overseas, others sell out*. 2021. URL: https://www.reuters.com/technology/chinas-ban-forces-some-bitcoin-miners-flee-overseas-others-sell-out-2021-06-25/.

[142] Conglei Shi, Weiwei Cui, Shixia Liu, Panpan Xu, Wei Chen, and Huamin Qu. "Rankexplorer: Visualization of ranking changes in large time series data." In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2669–2678. DOI: 10.1109/TVCG.2012.253.

[143] Robert J Shiller. *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press, 2019. ISBN: 978-0691210261.

[144] Yo-Der Song and Tomaso Aste. "The cost of Bitcoin mining has never really increased." In: *Frontiers in Blockchain* 3 (2020), p. 44. DOI: 10.3389/fbloc.2020.565497.

[145] Stanley Smith Stevens. *Psychophysics: Introduction to its perceptual, neural, and social prospects*. John Wiley & Sons, 1975.

[146] Christian Stoll, Lena Klaaßen, and Ulrich Gallersdörfer. "The carbon footprint of bitcoin." In: *Joule* 3.7 (2019), pp. 1647–1661. DOI: 10.1016/j.joule.2019.05.012.

[147] Charles D Stolper, Adam Perer, and David Gotz. "Progressive visual analytics: User-driven visual exploration of in-progress analytics." In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1653–1662. DOI: 10.1109/TVCG.2014.2346574.

[148] Tri Sundara. "Study on Blockchain Visualization." In: *International Journal on Informatics Visualization (JOIV)* 1.3 (2017), pp. 76–82. DOI: 10.30630/joiv.1.3.23.

[149] Nassim Nicholas Taleb. "Bitcoin, currencies, and fragility." In: *Quantitative Finance* 21.8 (2021), pp. 1249–1255. DOI: 10.1080/14697688.2021.1952702.

[150] Don Tapscott and Alex Tapscott. *Blockchain revolution: how the technology behind bitcoin and other cryptocurrencies is changing the world*. Portfolio, 2018. ISBN: 978-1101980149.

[151] Michael Bedford Taylor. "The Evolution of Bitcoin Hardware." In: *Computer* 50.9 (Sept. 2017), pp. 58–66. DOI: 10.1109/MC.2017.3571056.

[152] Natkamon Tovanich, Nicolas Heulot, Jean-Daniel Fekete, and Petra Isenberg. "A systematic review of online bitcoin visualizations." In: *Posters of the European Conference on Visualization (EuroVis)*. 2019. DOI: 10.2312/eurp.20191148.

[153] Natkamon Tovanich, Nicolas Heulot, Jean-Daniel Fekete, and Petra Isenberg. "Visualization of Blockchain Data: A Systematic Review." In: *IEEE Transactions on Visualization and Computer Graphics* 27.7 (2019), pp. 3135–3152. DOI: 10.1109/TVCG.2019.2963018.

[154] Natkamon Tovanich, Nicolas Soulié, Nicolas Heulot, and Petra Isenberg. "An empirical analysis of pool hopping behavior in the Bitcoin blockchain." In: *IEEE International Conference on Blockchain and Cryptocurrency*. 2021. DOI: 10.1109/ICBC51069.2021.9461118.

[155] Natkamon Tovanich, Nicolas Soulié, Nicolas Heulot, and Petra Isenberg. "MiningVis: visual analytics of the bitcoin mining economy." In: *IEEE Transactions on Visualization and Computer Graphics* (2021). DOI: 10.1109/TVCG.2021.3114821.

[156] Natkamon Tovanich, Nicolas Soulié, and Petra Isenberg. "Visual analytics of bitcoin mining pool evolution: on the road toward stability?" In: *11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE. 2021, pp. 1–5. DOI: 10.1109/NTMS49979.2021.9432675.

[157] Jon Truby. "Decarbonizing Bitcoin: Law and policy choices for reducing the energy consumption of Blockchain technologies and digital currencies." In: *Energy research & social science* 44 (2018), pp. 399–410. DOI: 10.1016/j.erss.2018.06.009.

[158] Florian Tschorsch and Björn Scheuermann. "Bitcoin and beyond: A technical survey on decentralized digital currencies." In: *IEEE Communications Surveys & Tutorials* 18.3 (2016), pp. 2084–2123. DOI: 10.1109/COMST.2016.2535718.

[159] Nitin Upadhyay. "Demystifying blockchain: A critical analysis of challenges, applications and opportunities." In: *International Journal of Information Management* 54 (2020), p. 102120. DOI: 10.1016/j.ijinfomgt.2020.102120.

[160] Jacob VanderPlas, Brian E Granger, Jeffrey Heer, Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. "Altair: interactive statistical visualizations for Python." In: *Journal of open source software* 3.32 (2018), p. 1057. DOI: 10.21105/joss.01057.

[161] Harald Vranken. "Sustainability of bitcoin and blockchains." In: *Current opinion in environmental sustainability* 28 (2017), pp. 1–9. DOI: 10.1016/j.cosust.2017.04.011.

[162] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert, and John Stasko. "A heuristic approach to value-driven evaluation of visualizations." In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2018), pp. 491–500. DOI: 10.1109/TVCG.2018.2865146.

[163] Canhui Wang, Xiaowen Chu, and Yang Qin. "Measurement and analysis of the bitcoin networks: A view from mining pools." In: *International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE. 2020, pp. 180–188. DOI: 10.1109/BigCom51056.2020.00032.

[164] Canhui Wang, Xiaowen Chu, and Qin Yang. "Measurement and Analysis of the Bitcoin Networks: A View from Mining Pools." In: (2020), pp. 180–188. DOI: 10.1109/BigCom51056.2020.00032.

[165] Luqin Wang and Yong Liu. "Exploring Miner Evolution in Bitcoin Network." In: *Passive and Active Measurement*. Springer, 2015, pp. 290–302. DOI: 10.1007/978-3-319-15509-8_22.

[166] Yong Wang, Hammad Haleem, Conglei Shi, Yanhong Wu, Xun Zhao, Siwei Fu, and Huamin Qu. "Towards easy comparison of local businesses using online reviews." In: *Computer Graphics Forum*. Vol. 37. 3. Wiley Online Library. 2018, pp. 63–74. DOI: 10.1111/cgf.13401.

[167] Yu Wang, Gaopeng Gou, Chang Liu, Mingxin Cui, Zhen Li, and Gang Xiong. "Survey of security supervision on blockchain from the perspective of technology." In: *Journal of Information Security and Applications* 60 (2021), p. 102859. DOI: 10.1016/j.jisa.2021.102859.

[168] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I. Weidele, Claudio Bellei, Tom Robinson, and Charles E. Leiserson. *Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics*. 2019. arXiv: 1908.02591 [cs.SI].

[169] Sam M. Werner, Daniel Perez, Lewis Gudgeon, Ariah Klages-Mundt, Dominik Harz, and William J. Knottenbelt. *SoK: Decentralized Finance (DeFi)*. 2021. arXiv: 2101.08778 [cs.CR].

[170] Gavin Wood. "Ethereum: A secure decentralised generalised transaction ledger." In: *Ethereum project yellow paper* 151 (2014), pp. 1–32. URL: https://gavwood.com/paper.pdf.

[171] Jiajing Wu, Jieli Liu, Yijing Zhao, and Zibin Zheng. "Analysis of cryptocurrency transactions from a network perspective: An overview." In: *Journal of Network and Computer Applications* (2021), p. 103139. DOI: 10.1016/j.jnca.2021.103139.

[172] Jia-zhi Xia, Yu-hong Zhang, Hui Ye, Ying Wang, Guang Jiang, Ying Zhao, Cong Xie, Xiao-yan Kui, Sheng-hui Liao, and Wei-ping Wang. "SuPoolVisor: a visual analytics system for mining pool surveillance." In: *Frontiers of Information Technology & Electronic Engineering* 21.4 (2020), pp. 507–523. DOI: 10.1631/FITEE.1900532. URL: https://doi.org/10.1631/FITEE.1900532.

[173] Jing Xia, Yumeng Hou, Yingjie Victor Chen, Zhenyu Cheryl Qian, David S Ebert, and Wei Chen. "Visualizing rank time series of wikipedia top-viewed pages." In: *IEEE Computer Graphics and Applications* 37.2 (2017), pp. 42–53. DOI: 10.1109/MCG.2017.21.

[174] Yang Xiao, Ning Zhang, Wenjing Lou, and Y Thomas Hou. "A survey of distributed consensus protocols for blockchain networks." In: *IEEE Communications Surveys & Tutorials* 22.2 (2020), pp. 1432–1465. DOI: 10.1109/COMST.2020.2969706.

[175] Congcong Ye, Guoqiang Li, Hongming Cai, Yonggen Gu, and Akira Fukuda. "Analysis of security in blockchain: Case study in 51%-attack detecting." In: *2018 5th International Conference on Dependable Systems and Their Applications (DSA)*. IEEE. 2018, pp. 15–24. URL: 10.1109/DSA.2018.00015.

[176] Jesse Yli-Huumo, Deokyoon Ko, Sujin Choi, Sooyong Park, and Kari Smolander. "Where is current research on blockchain technology?—a systematic review." In: *PloS one* 11.10 (2016), e0163477. DOI: 10.1371/journal.pone.0163477.

[177] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Xiangping Chen, and Huaimin Wang. "Blockchain challenges and opportunities: A survey." In: *International Journal of Web and Grid Services* 14.4 (2018), pp. 352–375. DOI: 10.1504/IJWGS.2018.095647.

[178] Francesco Zola, Jan Lukas Bruse, Maria Eguimendia, Mikel Galar, and Raul Orduna Urrutia. "Bitcoin and Cybersecurity: Temporal Dissection of Blockchain Data to Unveil Changes in Entity Behavioral Patterns." In: *Applied Sciences* 9.23 (Nov. 2019), p. 5003. DOI: 10.3390/app9235003.

[179] Francesco Zola, Maria Eguimendia, Jan Lukas Bruse, and Raul Orduna Urrutia. "Cascading machine learning to attack bitcoin anonymity." In: *IEEE International Conference on Blockchain (Blockchain)*. IEEE. 2019, pp. 10–17. DOI: 10.1109/Blockchain.2019.00011.

[180] Yuan Zuo, Jichang Zhao, and Ke Xu. "Word network topic model: a simple but general solution for short and imbalanced texts." In: *Knowledge and Information Systems* 48.2 (2016), pp. 379–398. DOI: 10.1007/s10115-015-0882-z.

[V1] Mansoor Ahmed, Ilia Shumailov, and Ross Anderson. "Tendrils of crime: Visualizing the diffusion of stolen bitcoins." In: *International Workshop on Graphical Models for Security*. Springer. 2018, pp. 1–12. DOI: 10.1007/978-3-030-15465-3_1.

[V2] Stefano Bistarelli and Francesco Santini. "Go with the-bitcoin-flow, with visual analytics." In: *Proceedings of the 12th International Conference on Availability, Reliability and Security*. 2017, pp. 1–6. DOI: 10.1145/3098954.3098972.

[V3] Andreas Bogner. "Seeing is understanding: anomaly detection in blockchains with visualized features." In: *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing and ACM International Symposium on Wearable Computers*. 2017, pp. 5–8. DOI: 10.1145/3123024.3123157.

[V4] Sudarshan Chawathe. "Monitoring blockchains with self-organizing maps." In: *17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE. 2018, pp. 1870–1875. DOI: 10.1109/TrustCom/BigDataSE.2018.00283.

[V5] Giuseppe Di Battista, Valentino Di Donato, Maurizio Patrignani, Maurizio Pizzonia, Vincenzo Roselli, and Roberto Tamassia. "Bitconeview: visualization of flows in the bitcoin transaction graph." In: *IEEE Symposium on Visualization for Cyber Security (VizSec)*. IEEE. 2015, pp. 1–8. DOI: 10.1109/VIZSEC.2015.7312773.

[V6] Zhihao Hao, Dianhui Mao, Bob Zhang, Min Zuo, and Zhihua Zhao. "A novel visual analysis method of food safety risk traceability based on blockchain." In: *International journal of environmental research and public health* 17.7 (2020), p. 2300. DOI: 10.3390/ijerph17072300.

[V7] Bernhard Haslhofer, Rainer Stütz, Matteo Romiti, and Ross King. *GraphSense: A General-Purpose Cryptoasset Analytics Platform*. 2021. arXiv: 2102.13613 [cs.CR].

[V8] Petra Isenberg, Christoph Kinkeldey, and Jean-Daniel Fekete. "Exploring entity behavior on the bitcoin blockchain." In: *Poster Proceedings of IEEE Information Visualization Conference (InfoVis)*. 2017. URL: https://hal.inria.fr/hal-01658500.

[V9] Christoph Kinkeldey, Jean-Daniel Fekete, Tanja Blascheck, and Petra Isenberg. "BitConduite: Exploratory Visual Analysis of Entity Activity on the Bitcoin Network." In: 2021, pp. 1–1. DOI: 10.1109/MCG.2021.3070303.

[V10] Hiroki Kuzuno and Christian Karam. "Blockchain explorer: An analytical process and investigation environment for bitcoin." In: *2017 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE. 2017, pp. 9–16. DOI: 10.1109/ECRIME.2017.7945049.

[V11] Dan McGinn, David Birch, David Akroyd, Miguel Molina-Solana, Yike Guo, and William J Knottenbelt. "Visualizing dynamic bitcoin transaction patterns." In: *Big Data* 4.2 (2016), pp. 109–119. DOI: 10.1089/big.2015.0056.

[V12] Robert Norvill, Beltran Borja Fiz Pontiveros, Radu State, and Andrea Cullen. "Visual emulation for Ethereum's virtual machine." In: *IEEE/IFIP Network Operations and Management Symposium (NOMS)*. IEEE. 2018, pp. 1–4. DOI: 10.1109/NOMS.2018.8406332.

[V13] Frédérique Oggier, Silivanxay Phetsouvanh, and Anwitaman Datta. "BiVA: Bitcoin network visualization & analysis." In: *IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE. 2018, pp. 1469–1474. DOI: 10.1109/ICDMW.2018.00210.

[V14] Benedikt Putz, Fabian Böhm, and Günther Pernul. "HyperSec: Visual Analytics for blockchain security monitoring." In: *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer. 2021, pp. 165–180. DOI: 10.1007/978-3-030-78120-0_11.

[V15] Peter Schretlen, Nathan Kronenfeld, Derek Gray, Jesse McGeachie, Eric Hall, Daniel Cheng, Nicole Covello, and William Wright. "Interactive Data Exploration with "Big Data Tukey Plots"." In: *Poster Proceedings of IEEE Visualization Conference (VIS)*. 2013. URL: https://uncharted.software/assets/interactive-data-exploration.pdf.

[V16] Yujing Sun, Hao Xiong, Siu Ming Yiu, and Kwok Yan Lam. "BitVis: An interactive visualization system for bitcoin accounts analysis." In: *Crypto Valley Conference on Blockchain Technology (CVCBT)*. IEEE. 2019, pp. 21–25. DOI: 10.1109/CVCBT.2019.000-3.

[V17] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. "Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics." In: *KDD Workshop on Anomaly Detection in Finance*. ACM. 2019, pp. 1–1. URL: https://arxiv.org/pdf/1908.02591.pdf.

[V18] Jia-zhi Xia, Yu-hong Zhang, Hui Ye, Ying Wang, Guang Jiang, Ying Zhao, Cong Xie, Xiao-yan Kui, Sheng-hui Liao, and Wei-ping Wang. "SuPoolVisor: a visual analytics system for mining pool surveillance." In: *Frontiers of Information Technology & Electronic Engineering* 21.4 (2020), pp. 507–523. DOI: 10.1631/FITEE.1900532.

[V19] Xuanwu Yue, Xinhuan Shu, Xinyu Zhu, Xinnan Du, Zheqing Yu, Dimitrios Papadopoulos, and Siyuan Liu. "BitExTract: Interactive Visualization for Extracting Bitcoin Exchange Intelligence." In: *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 25.1 (2018), pp. 162–171. DOI: 10.1109/TVCG.2018.2864814.

[V20] Zengsheng Zhong, Shuirun Wei, Yeting Xu, Ying Zhao, Fangfang Zhou, Feng Luo, and Ronghua Shi. "SilkViser: A Visual Explorer of Blockchain-based Cryptocurrency Transaction Data." In: *IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE. 2020, pp. 95–106. DOI: 10.1109/VAST50239.2020.00014.

DATA ANALYSIS ARTICLES

[A1] Cuneyt Gurcan Akcora, Yitao Li, Yulia R. Gel, and Murat Kantarcioglu. *BitcoinHeist: Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain*. 2019. arXiv: 1906.07852 [cs.CR].

[A2] Israa Alqassem, Iyad Rahwan, and Davor Svetinovic. "The anti-social system properties: Bitcoin network data analysis." In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50.1 (2018), pp. 21–31. DOI: 10.1109/TSMC.2018.2883678.

[A3] Andra Anoaica and Hugo Levard. "Quantitative Description of Internal Activity on the Ethereum Public Blockchain." In: *9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. 2018, pp. 1–5. DOI: 10.1109/NTMS.2018.8328741.

[A4] Ee Hong Aw, Ralucca Gera, Kenneth Hicks, Nicholas Koeppen, and Christopher Teska. "Analyzing preferential attachment in peer-to-peer bitcoin networks." In: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE. 2018, pp. 1242–1249. DOI: 10.1109/ASONAM.2018.8508273.

[A5] Anton I Badev and Matthew Chen. *Bitcoin: Technical background and data analysis*. FEDS Working Paper. 2014. DOI: 10.2139/ssrn.2544331.

[A6] Thibault de Balthasar and Julio Hernandez-Castro. "An analysis of bitcoin laundry services." In: *Nordic Conference on Secure IT Systems*. Springer. 2017, pp. 297–312. DOI: 10.1007/978-3-319-70290-2.

[A7] Massimo Bartoletti, Salvatore Carta, Tiziana Cimoli, and Roberto Saia. "Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact." In: *Future Generation Computer Systems* 102 (2020), pp. 259–277. DOI: 10.1016/j.future.2019.08.014.

[A8] Massimo Bartoletti and Livio Pompianu. "An analysis of Bitcoin OP_RETURN metadata." In: *International Conference on Financial Cryptography and Data Security*. Springer. 2017, pp. 218–230. DOI: 10.1007/978-3-319-70278-0_14.

[A9] Stefano Bistarelli, Gianmarco Mazzante, Matteo Micheletti, Leonardo Mostarda, Davide Sestili, and Francesco Tiezzi. "Ethereum smart contracts: Analysis and statistics of their source code and opcodes." In: *Internet of Things* 11 (2020), p. 100198. DOI: 10.1016/j.iot.2020.100198.

[A10] Stefano Bistarelli, Matteo Parroccini, and Francesco Santini. "Visualizing Bitcoin Flows of Ransomware: WannaCry One Week Later." In: *Italian Conference on Cybersecurity (ITASEC)*. 2018. URL: http://ceur-ws.org/Vol-2058/paper-13.pdf.

[A11] Tong Cao, Jérémie Decouchant, Jiangshan Yu, and Paulo Esteves-Verissimo. "Characterizing the Impact of Network Delay on Bitcoin Mining." In: *International Symposium on Reliable Distributed Systems*. 2021. URL: https://orbilu.uni.lu/bitstream/10993/47736/1/2021135560.pdf.

[A12] Tao-Hung Chang and Davor Svetinovic. "Data analysis of digital currency networks: Namecoin case study." In: *21st International Conference on Engineering of Complex Computer Systems (ICECCS)*. IEEE. 2016, pp. 122–125. DOI: 10.1109/ICECCS.2016.023.

[A13] Ting Chen, Zihao Li, Yuxiao Zhu, Jiachi Chen, Xiapu Luo, John Chi-Shing Lui, Xiaodong Lin, and Xiaosong Zhang. "Understanding ethereum via graph analysis." In: *ACM Transactions on Internet Technology (TOIT)* 20.2 (2020), pp. 1–32. DOI: 10.1145/3381036.

[A14] Weili Chen, Tuo Zhang, Zhiguang Chen, Zibin Zheng, and Yutong Lu. "Traveling the token world: A graph analysis of ethereum erc20 token ecosystem." In: *Proceedings of The Web Conference 2020*. 2020, pp. 1411–1421. DOI: 10.1145/3366423.3380215.

[A15] Giuseppe Di Battista, Valentino Di Donato, and Maurizio Pizzonia. "Long transaction chains and the bitcoin heartbeat." In: *European Conference on Parallel Processing*. Springer. 2017, pp. 507–516. DOI: 10.1007/978-3-319-75178-8_41.

[A16] Yoshi Fujiwara and Rubaiyat Islam. *Bitcoin's Crypto Flow Network*. 2021. arXiv: 2106.11446 [q-fin.GN].

[A17] Bingyu Gao, Haoyu Wang, Pengcheng Xia, Siwei Wu, Yajin Zhou, Xiapu Luo, and Gareth Tyson. "Tracking Counterfeit Cryptocurrency End-to-end." In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4.3 (2020), pp. 1–28. DOI: 10.1145/3428335.

[A18] Befekadu G. Gebraselase, Bjarne E. Helvik, and Yuming Jiang. *An Analysis of Transaction Handling in Bitcoin*. 2021. arXiv: 2106.10083 [cs.CR].

[A19] Dongchao Guo, Jiaqing Dong, and Kai Wang. "Graph structure and statistical properties of ethereum transaction relationships." In: *Information Sciences* 492 (2019), pp. 58–71. DOI: 10.1016/j.ins.2019.04.013.

[A20] Yuheng Huang, Haoyu Wang, Lei Wu, Gareth Tyson, Xiapu Luo, Run Zhang, Xuanzhe Liu, Gang Huang, and Xuxian Jiang. "Understanding (mis) behavior on the eosio blockchain." In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4.2 (2020), pp. 1–28. DOI: 10.1145/3392155.

[A21] Xin-Jian Jiang and Xiao Fan Liu. "CryptoKitties transaction network analysis: The rise and fall of the first blockchain game mania." In: *Frontiers in Physics* 9 (2021), p. 57. DOI: 10.3389/fphy.2021.631665.

[A22] George Kappos, Haaroon Yousaf, Mary Maller, and Sarah Meiklejohn. "An empirical analysis of anonymity in zcash." In: *27th {USENIX} Security Symposium ({USENIX} Security)*. USENIX Association, 2018, pp. 463–477. ISBN: 978-1-939133-04-5. URL: https://www.usenix.org/conference/usenixsecurity18/presentation/kappos.

[A23] Dániel Kondor, Nikola Bulatovic, József Stéger, István Csabai, and Gábor Vattay. *The rich still get richer: Empirical comparison of preferential attachment via linking statistics in Bitcoin and Ethereum*. 2021. arXiv: 2102.12064 [physics.soc-ph].

[A24] Xi Tong Lee, Arijit Khan, Sourav Sen Gupta, Yu Hann Ong, and Xuan Liu. "Measurements, analyses, and insights on the entire ethereum blockchain network." In: *Proceedings of The Web Conference 2020*. 2020, pp. 155–166. DOI: 10.1145/3366423.3380103.

[A25] Sheng-Nan Li, Zhao Yang, and Claudio J Tessone. "Mining blocks in a row: A statistical study of fairness in Bitcoin mining." In: *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE. 2020, pp. 1–4. DOI: 10.1109/ICBC48266.2020.9169436.

[A26] Matthias Lischke and Benjamin Fabian. "Analyzing the bitcoin network: The first four years." In: *Future Internet* 8.1 (2016), p. 7. DOI: 10.3390/FI8010007.

[A27] Damiano Di Francesco Maesa, Andrea Marino, and Laura Ricci. "Uncovering the bitcoin blockchain: an analysis of the full users graph." In: *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. 2016, pp. 537–546. DOI: 10.1109/DSAA.2016.52.

[A28] D McGinn, D McIlwraith, and Y Guo. "Towards open data blockchain analytics: a Bitcoin perspective." In: *Royal Society Open Science* 5.8 (2018), p. 180298. DOI: 10.1098/rsos.180298.

[A29] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. "A fistful of bitcoins: characterizing payments among men with no names." In: *Proceedings of the Conference on Internet Measurement Conference*. 2013, pp. 127–140. DOI: 10.1145/2504730.2504747.

[A30] Malte Möser, Rainer Böhme, and Dominic Breuker. "An inquiry into money laundering tools in the Bitcoin ecosystem." In: *APWG eCrime researchers summit*. IEEE. 2013, pp. 1–14. DOI: 10.1109/eCRS.2013.6805780.

[A31] Malte Möser et al. "An Empirical Analysis of Traceability in the Monero Blockchain." In: *Proceedings on Privacy Enhancing Technologies* 2018.3 (2018), pp. 143–163. DOI: 10.1515/popets-2018-0025.

[A32] Lukas Norbutas. "Offline constraints in online drug marketplaces: An exploratory analysis of a cryptomarket trade network." In: *International Journal of Drug Policy* 56 (2018), pp. 92–100. DOI: 10.1016/j.drugpo.2018.03.016.

[A33] Francesco Parino, Mariano G Beiró, and Laetitia Gauvin. "Analysis of the Bitcoin blockchain: socio-economic factors behind the adoption." In: *EPJ Data Science* 7.1 (2018), p. 38. DOI: 10.1140/EPJDS/S13688-018-0170-8.

[A34] Silivanxay Phetsouvanh, Frédérique Oggier, and Anwitaman Datta. "Egret: Extortion graph exploration techniques in the bitcoin network." In: *IEEE International conference on data mining workshops (ICDMW)*. IEEE. 2018, pp. 244–251. DOI: 10.1109/ICDMW.2018.00043.

[A35] Andrea Pinna, Simona Ibba, Gavina Baralla, Roberto Tonelli, and Michele Marchesi. "A massive analysis of ethereum smart contracts empirical study and code metrics." In: *IEEE Access* 7 (2019), pp. 78194–78213. DOI: 10.1109/ACCESS.2019.2921936.

[A36] Fergal Reid and Martin Harrigan. "An analysis of anonymity in the bitcoin system." In: *Security and Privacy in Social Networks*. Springer, 2013, pp. 197–223. DOI: 10.1109/PASSAT/SocialCom.2011.79.

[A37] Matteo Romiti, Aljosha Judmayer, Alexei Zamyatin, and Bernhard Haslhofer. *A Deep Dive into Bitcoin Mining Pools: An Empirical Analysis of Mining Shares*. 2019. arXiv: 1905.05999 [cs.CR].

[A38] Dorit Ron and Adi Shamir. "Quantitative analysis of the full bitcoin transaction graph." In: *International Conference on Financial Cryptography and Data Security*. Springer. 2013, pp. 6–24. DOI: 10.1007/978-3-642-39884-1_2.

[A39] Nicoló Vallarano, Claudio J Tessone, and Tiziano Squartini. "Bitcoin Transaction Networks: an overview of recent results." In: *Frontiers in Physics* 8 (2020), p. 286. DOI: 10.3389/fphy.2020.00286.

[A40] Canhui Wang, Xiaowen Chu, and Yang Qin. "Measurement and analysis of the bitcoin networks: A view from mining pools." In: *6th International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE. 2020, pp. 180–188. DOI: 10.1109/BigCom51056.2020.00032.

[A41] Luqin Wang and Yong Liu. "Exploring miner evolution in bitcoin network." In: *International Conference on Passive and Active Network Measurement*. Springer. 2015, pp. 290–302. DOI: 10.1007/978-3-319-15509-8_22.

[A42] Lin Zhao, Sourav Sen Gupta, Arijit Khan, and Robby Luo. "Temporal Analysis of the Entire Ethereum Blockchain Network." In: *Proceedings of the Web Conference*. 2021, pp. 2258–2269. DOI: 10.1145/3442381.3449916.

[A43] Peilin Zheng, Zibin Zheng, Jiajing Wu, and Hong-Ning Dai. "Xblock-ETH: Extracting and exploring blockchain data from Ethereum." In: *IEEE Open Journal of the Computer Society* 1 (2020), pp. 95–106. DOI: 10.1109/OJCS.2020.2990458.

ONLINE SOURCES

[O1] Aleth.io. *EthStats.io*. Accessed on: August, 2019. URL: https://ethstats.io/.

[O2] Alethio. *Alethio*. Accessed on: August, 2019. URL: https://aleth.io/.

[O3] BitBonkers. *BitBonkers - A Bitcoin Blockchain Transaction Visualisation*. Accessed on: August, 2019. URL: https://bitbonkers.com/.

[O4] Bitcoin Globe. *Bitcoin Globe*. Accessed on: August, 2019. URL: http://bitcoinglobe.com/.

[O5] BitcoinWisdom.com. *BitcoinWisdom*. Accessed on: August, 2019. URL: https://bitcoinwisdom.com/.

[O6] Bitfly GmbH. *Etherchain*. Accessed on: August, 2019. URL: https://www.etherchain.org/.

[O7] bitFlyer. *chainFlyer*. Accessed on: August, 2019. URL: https://chainflyer.bitflyer.jp/.

[O8] BitForce5. *BitForce5*. Accessed on: August, 2019. URL: http://bitforce5.com/.

[O9] BitInfoCharts. *BitInfoCharts*. Accessed on: August, 2019. URL: https://bitinfocharts.com/.

[O10] Blockchain Luxembourg S.A. *Blockchain Explorer*. Accessed on: August, 2019. URL: https://www.blockchain.com/explorer.

[O11] Blockchair. *Blockchair*. Accessed on: August, 2019. URL: https://blockchair.com/.

[O12] BlockSeer. *BlockSeer*. Accessed on: August, 2019. URL: https://www.blockseer.com/.

[O13] BTC.com. *BTC.com*. Accessed on: August, 2019. URL: https://btc.com/.

[O14] Kacper Cieśla. *Bitcoinity*. Accessed on: August, 2019. URL: https://data.bitcoinity.org/.

[O15] Coin Dance. *Coin Dance*. Accessed on: August, 2019. URL: https://coin.dance/.

[O16] CoinDesk, Inc. *CoinDesk*. Accessed on: August, 2019. URL: https://www.coindesk.com/data.

[O17] DailyBlockchain. *DailyBlockchain*. Accessed on: August, 2019. URL: https://dailyblockchain.github.io/.

[O18] DashRadar. *DashRadar*. Accessed on: August, 2019. URL: https://dashradar.com/.

[O19] Dune Analytics. *Dune Analytics*. Accessed on: August, 2021. URL: https://dune.xyz/.

[O20] Elliptic Enterprises Limited. *The Bitcoin Big Bang*. Accessed on: August, 2019. URL: https://info.elliptic.co/hubfs/big-bang/bigbang-v1.html.

[O21] ethernodes.org. *The Ethereum Nodes Explorer*. Accessed on: August, 2019. URL: https://www.ethernodes.org/.

[O22] Etherscan. *Etherscan*. Accessed on: August, 2019. URL: https://etherscan.io/.

[O23] EtherView. *EtherView*. Accessed on: August, 2019. URL: https://etherview.now.sh/.

[O24] Ethviewer. *Ethviewer*. Accessed on: August, 2019. URL: http://www.ethviewer.live/.

[O25] Everex. *Ethplorer*. Accessed on: August, 2019. URL: https://ethplorer.io/.

[O26] Primavera De Filippi. *Plantoids: Blockchain-based life forms*. Accessed on: August, 2019. URL: http://plantoid.org/.

[O27] GasTracker.io. *Gastracker.io*. Accessed on: August, 2019. URL: https://gastracker.io/.

[O28] Luc Hendriks. *Bitcoin Transaction Visualization*. Accessed on: August, 2019. URL: http://bitcoin.interaqt.nl/.

[O29] Chance Hines and Gregory Akerman. *Federal Bitcoin*. Accessed on: August, 2019. URL: https://federalbitcoin.herokuapp.com/.

[O30] Jochen Hoenicke. *Johoe's Bitcoin Mempool Statistics*. Accessed on: August, 2019. URL: https://jochen-hoenicke.de/queue/.

[O31] Input Output HK Limited. *Symphony of Blockchains: 3D Blockchain Explorer*. Accessed on: August, 2019. URL: https://symphony.iohk.io/.

[O32] Katana Cryptographic Ltd. *OXT*. Accessed on: August, 2019. URL: https://oxt.me/.

[O33] Edward Kerstein. *Bitcoin Visuals*. Accessed on: August, 2019. URL: https://bitcoinvisuals.com/.

[O34] Tracy Kuhrt. *Hyperledger Explorer*. Accessed on: August, 2021. URL: https://wiki.hyperledger.org/display/explorer/Hyperledger+Explorer.

[O35] Maximillian Laumeister. *BitListen*. Accessed on: August, 2019. URL: https://www.bitlisten.com/.

[O36] Pedro Marrucho. *BitcoinCity*. Accessed on: August, 2019. URL: http://bitcoincity.info/.

[O37] Marian Oancea. *Ethereum Network Status*. Accessed on: August, 2019. URL: https://ethstats.net/.

[O38] Kevin Small. *Blockchain 3D Explorer*. Accessed on: August, 2019. URL: https://blockchain3d.info/.

[O39] Statoshi.info. *Statoshi.info*. Accessed on: August, 2019. URL: https://statoshi.info/.

[O40] Dustin Stupp. *ON BRINK: Live Physicalization of the Bitcoin Blockchain*. Accessed on: August, 2019. URL: http://dataphys.org/list/on-brink-live-physicalization-of-the-bitcoin-blockchain/.

[O41] TradeBlock. *TradeBlock*. Accessed on: August, 2019. URL: https://tradeblock.com/bitcoin/.

[O42] TX Highway. *TX Highway*. Accessed on: August, 2019. URL: https://txhighway.com/.

[O43] Jan Vornberger. *Bitcoin Monitor*. Accessed on: August, 2019. URL: https://www.bitcoinmonitor.com/.

[O44] Greg Walker. *Bitcoinrain*. Accessed on: August, 2019. URL: http://bitcoinrain.io/.

[O45] Robin K. Wilson. *Bitcoin-VR*. Accessed on: August, 2019. URL: https://bitcoin-vr.github.io/.

[O46] Wizbit. *Wizbit Realtime Bitcoin Globe*. Accessed on: August, 2019. URL: https://blocks.wizb.it/.

[O47] Addy Yeow. *Bitnodes*. Accessed on: August, 2019. URL: https://bitnodes.earn.com/.