Decentralization Analysis on Cardano by Parsing Full-nodes

Shuren Miao

Master of Science
Cyber Security, Privacy and Trust
School of Informatics
University of Edinburgh
2023

Abstract

Blockchain technology has revolutionized the way we think about the Internet, with decentralization being a key feature of Web 3.0. Among the pioneers leading this change is Cardano, a third-generation proof-of-stake blockchain platform. However, like other networks, Cardano's level of decentralization has been subject to scrutiny. This dissertation focuses on examining Cardano's decentralization status within its consensus and tokenomics layers. To achieve this, I developed a parser that extracts on-chain data from full nodes, with data cleansing, transformation, mapping, and computation of decentralization indices being key features. Trend analyses of these indices provide insights into their changing dynamics, culminating in a comprehensive visual analysis report.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Shuren Miao*)

Acknowledgements

I would like to express my deepest gratitude to my parents for their unwavering support, encouragement, and belief in me throughout my academic journey. Your constant encouragement and sacrifices have been the driving force behind my achievements, and I am truly grateful for your love and guidance.

I am also indebted to my girlfriend for her patience, understanding, and encouragement during the ups and downs of this research journey. Your unwavering support and positivity have been a source of inspiration, and I am lucky to have you by my side.

I extend my heartfelt appreciation to my supervisors Daniel Woods and Christina Ovezik for their invaluable guidance, mentorship, and expertise. Your insightful feedback, constructive criticism, and dedication to my academic growth have been instrumental in shaping this study. Your commitment to excellence and your belief in my potential have motivated me to strive for the highest standards.

To all those who have played a role in my academic and personal journey, I am truly thankful for your presence in my life. Your support has been invaluable, and I look forward to continuing to learn, grow, and contribute with your guidance and encouragement.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

Blockchain technology has ushered in a paradigm shift in the realm of digital transactions and information exchange. Its core principle of decentralization has promised to reshape existing paradigms, enabling a trustless environment and redefining traditional centralized systems. One of the leading blockchain platforms in this arena is Cardano, a third-generation proof-of-stake platform that has garnered substantial attention due to its innovative research-driven approach. While Cardano boasts a robust theoretical foundation and advanced system architecture [3], the practical degree of decentralization within its network remains a crucial aspect requiring comprehensive analysis.

As the blockchain landscape evolves, questions about the effectiveness and true decentralization of platforms like Cardano have surfaced. Concerns arise regarding the distribution of resources, power, and control among network participants. The need to measure and quantify decentralization is paramount for validating claims, improving network design, and enhancing governance mechanisms. Further discussion on the significance of decentralization will be in section 2.1. The significance of this study lies in addressing these concerns by providing a thorough analysis of Cardano's decentralization, thus contributing to the broader discourse on blockchain governance and technology.

This research aims to shed light on the practical implications of its design choices by dissecting Cardano's decentralization across different layers and employing sophisticated metrics. Moreover, the findings of this study could potentially guide the development of strategies to enhance network resilience, user participation, and overall system integrity. By investigating Cardano's decentralization, this study endeavors to provide actionable insights for blockchain designers, developers, and enthusiasts alike.

In summary, this research is motivated by the critical need to bridge the gap between theoretical constructs and real-world implications of blockchain decentralization, particularly in the context of Cardano. Understanding the intricacies of decentralization in a pioneering platform like Cardano has far-reaching implications for the broader blockchain ecosystem and holds promise for shaping the future of digital trust and collaboration.

1.2 Problem Statement

The blockchain landscape is characterized by diverse designs and implementations, leading to challenges in comparing and quantifying critical properties such as security, stability, and privacy [23]. The linchpin to ensuring these essential attributes lies in the concept of decentralization. Within this context, Cardano, currently positioned as the 7th largest cryptocurrency by market capitalization [28], emerges as a third-generation blockchain platform celebrated for its robust theoretical underpinnings and meticulous system construction.

The landscape of blockchain platforms is marked by significant differences in decentralization due to variations in underlying technologies, rendering direct comparisons and inferences arduous. However, the majority of research efforts have predominantly centered around top platforms such as Bitcoin [27, 16, 25, 34, 26, 7, 15, 1, 12, 14, 2, 19, 4, 6, 31, 5] and Ethereum [27, 16, 25, 34, 14, 2, 8, 10, 4, 31, 5], leaving Cardano relatively unexplored [22, 30, 29]. Notably, existing analyses of Cardano's decentralization have been confined to discrete analysis systems, lacking a unified and methodical approach for comprehensive cross-blockchain evaluation.

Moreover, the limited research related to Cardano's decentralization analysis has largely overlooked the establishment of a systematic layered framework. The selection of entity-resource pairs has varied extensively, hindering effective comparative analysis across diverse projects. Furthermore, the process of selecting decentralization indices warrants meticulous empirical validation to discern the significance of computed metrics.

Furthermore, the continuous technological advancements and version iterations within Cardano, akin to other blockchain networks, raise concerns about the relevance and applicability of past data analyses. The evolution of the Cardano ecosystem, particularly with the emergence of its Shelley phase, introducing refined system functionalities and a dynamic ecosystem, underscores the urgency for contemporary research. Given that earlier analyses may have become obsolete over time and may no longer capture the true essence of the present ecosystem, there is a pressing need for real-time and up-to-date research to ensure the accuracy of research findings.

Finally, online resources like cryptocurrency exchanges, blockchain explorers, and developer forums that provide data on decentralization exhibit opacity, precluding the validation of authenticity and precision. Moreover, analytical tools such as BlockSci present inherent limitations in providing comprehensive insights into the decentralized landscape.

Hence, the primary objective of this study is to bridge this research gap by employing a systematic layering approach to scrutinize resource and entity distribution within each stratum of the Cardano blockchain. By quantifying the degree of decentralization, this study contributes to the broader Edinburgh Distribution Index initiative as a pivotal sub-project. Through this research endeavor, we aspire to enhance the comprehension of Cardano's decentralization, facilitate informed decision-making, and provide a comprehensive framework applicable to the evolving blockchain landscape.

1.3 Research Hypothesis

The research hypothesis of this study asserts that the precise measurement and evaluation of Cardano's degree of decentralization across specific layers can be effectively accomplished through the application of robust data analysis techniques. This approach relies on the utilization of on-chain data derived from full nodes, a resource accessible to any user with appropriately configured hardware and standard internet connectivity. This availability ensures that any user, equipped with a suitable device, can seamlessly tap into the Cardano blockchain network and retrieve unadulterated on-chain block data.

This hypothesis also hinges upon the dependability of the software stack underpinning the parser's functionality. Comprising elements such as cardano-node, cardano-db-sync, PostgreSQL, Pandas, and Matplotlib, this software stack is trusted to execute the desired tasks accurately, ensuring that the obtained results align with the intended outcomes of the analysis.

Moreover, the research hypothesis assumes the proven efficacy and significance of the applied decentralization metrics. These metrics are designed not only to capture essential features and properties but also to reveal the intricate layers of decentralization within Cardano's ecosystem. By applying established and effective decentralization metrics, this study aims to uncover insights into the distribution of resources, decision-making power, and overall network behavior.

1.4 Research Objectives

Nestled within the intricate architecture of Cardano's consensus and tokenomics layers, a rich tapestry of entity-resource pairs awaits exploration. These dynamic co-relationships encapsulate the block production, stake scale, and rewards of stake pools, as well as the genesis and ownership of tokens. The overarching objective is to untangle the complex interplay between Cardano's decentralization framework and the governing architecture, thereby shedding light on how the consensus and tokenomics layers harmonize to shape a sustainable allocation of decision-making influence and cryptocurrency holdings. This research initiative is driven by the following primary objectives:

- Establishing a Computation Platform: Create a robust computation platform endowed with ample hardware resources and a meticulously configured environment, outfitted with the requisite modules to facilitate seamless analyses.
- Parser Development and Implementation: Conceive and materialize a sophisticated parser, engineered to traverse the labyrinthine network of Cardano's full nodes. This parser will be designed to unearth nuanced block-level data, capturing the essence of the blockchain's structural intricacies.
- Parser Validation and Efficacy: Rigorously validate the parser's prowess through an arsenal of experiments. By subjecting the parser to comparative analyses against a diverse spectrum of methodologies and tools, the goal is to ascertain its efficacy, solidifying its role as a reliable conduit for data extraction.
- Application of Advanced Metrics: Harness the potential of state-of-the-art metrics to dissect the distribution of resources within the distinct strata of Cardano. This exploration will be laser-focused on the consensus and tokenomics layers. The ensuing discourse will unravel the insights gleaned from these metrics, casting a revealing spotlight on the dynamic interplay of structural facets and the trajectory of decentralization trends.

In synthesis, these objectives coalesce to forge a meticulous roadmap for unravelling the intricate tapestry of Cardano's decentralization paradigm. The culmination of these objectives will yield a comprehensive understanding of the network's resource allocation, governance mechanisms, and their reverberations within the broader blockchain domain.

1.5 Dissertation Structure

This dissertation provides a comprehensive analysis of Cardano's decentralization landscape. Chapter 2 Background explains the fundamentals of decentralization and Cardano in detail while contextualizing the research within existing works. Subsequently, Chapter 3 Methodology and Solution delves into the technical underpinnings, encompassing the environment setup, parser development, indexing metrics, mapping clusters, and sliding window techniques. Chapter 4 Decentralization Analysis segment rigorously examines Cardano's decentralization through both consensus and tokenomics lenses, scrutinizing stake pool decentralization, stake distribution, initial token allocation, and UTxO ownership. Chapter 5 Evaluation assesses the parser's performance and metric validity, including comparisons against alternative approaches. Finally, Chapter 6 Conclusion encapsulates research contributions, findings, and potential pathways for future research, while acknowledging inherent limitations.

Chapter 2

Background

2.1 Decentralization Analysis

Decentralization within blockchain systems serves as a powerful means to enhance security, although its importance surpasses being a sole objective. Instead, it acts as a foundational strategy to bolster several critical security attributes [23]. These attributes encompass safety, ensuring a uniform view of the ledger among honest participants; liveness, guaranteeing timely updates to the ledger's view with new transactions; privacy, allowing actions to remain unlinked from real-world identities; and stability, maintaining predictability in digital asset supply and market price. While decentralization doesn't directly ensure these properties, its influence can significantly reinforce them. The dynamics between centralization and decentralization entail distinct vulnerabilities and strengths within specific threat models.

Low levels of decentralization expose networks to vulnerabilities, including safety risks arising from long-range attacks, liveness concerns due to transaction censorship and potential reordering, sybil attacks, single points of failure, instability stemming from the dominance of powerful entities, and indirect threats like honest miners abstaining due to unstable rewards.

Analyzing the consequences of decentralization involves a systematic approach that considers each layer of the blockchain network. This approach encompasses identifying pivotal resources, determining the entities that control these resources, and assessing the potential compromise of security properties if these resources were to become centralized [23]. For instance, within Bitcoin's consensus layer, miners control hashing power, thereby impacting safety, liveness, stability, and privacy. To analyze decentralization's implications, the process involves identifying entities and resources in each layer, quantifying metrics to evaluate decentralization, and assessing its impact on the security attributes of the network.

The Decentralization Index is quantitatively measured through metrics like node distribution, power allocation, and wealth dispersion, reflecting the distribution of resources among participants. Normally, a higher Index value signifies enhanced network decentralization. Metrics such as Shannon entropy, Gini coefficient, Theil, Atkinson, Herfindahl-Hirschman indices, and the Nakamoto coefficient are utilized to gauge decentralization. These metrics provide insights into resource distribution dynamics and the extent of network decentralization, contributing to a holistic understanding of blockchain network structures and their implications.

2.2 Cardano

2.2.1 The Evolution of Cardano

Cardano, initiated in 2015, addresses the challenges of earlier blockchain generations by striving for a balanced, sustainable ecosystem meeting user and system integration needs. While Bitcoin's first-generation blockchain emphasized secure cryptocurrency transfer, Ethereum's second-generation added smart contracts and tokens. Cardano stands as a third-generation blockchain, amalgamating strengths from prior generations to meet evolving requirements. Cardano prioritizes security, scalability, functionality, sustainability, and interoperability for holistic blockchain development. Cardano uses a proof-of-stake (PoS) consensus protocol to offer a secure and sustainable platform for decentralized applications (dApps).

Cardano's development journey can be understood through two main phases, Byron and Shelley. Byron's foundation facilitated ada transactions on a proof-of-stake blockchain. Initially, Cardano's network was federated, and managed by Input Output Global and Emurgo stake pools. Transitioning from Byron, Shelley marked the shift to a decentralized ledger system, driven by distributed stake pool operators. Enhancements in user experience, stake pool operations, delegation, and incentives were central.

2.2.2 The Ouroboros Protocol: PoS Consensus

Cardano, founded on the pioneering consensus protocol called Ouroboros, stands as the first blockchain platform developed through peer-reviewed research. Cardano employs

PoS, which offers several advantages over Proof of Work (PoW), including enhanced security, reduced centralization risks, energy efficiency, and cost-effectiveness.

Ouroboros [24] carries three key responsibilities: facilitating block production decisions, managing chain selection, and validating generated blocks. In the PoS paradigm, stake pools form the nucleus of the Ouroboros protocol, enabling ada holders to delegate their stake to dependable server nodes managed by stake pool operators (SPO). These stake pools emphasize maintenance and collectively hold the stake of various stakeholders, ensuring broader participation within the protocol. Notably, it selects stake pools instead of miners to generate new blocks, based on their stake in the network. The protocol divides time into epochs and slots, organizing the leader election process to accommodate dynamic changes in stake distribution. Each slot has an elected leader responsible for appending a block to the chain and passing it to the next leader.

As for the milestones of this protocol, the Ouroboros Classic achieved key implementations such as energy-efficient consensus, introducing a mathematical framework for PoS analysis, and implementing a unique incentive mechanism.

Ouroboros BFT, the protocol's second iteration, established synchronous communication among federated servers, enhancing ledger consensus in a more deterministic manner, while Ouroboros Praos [9] introduced substantial security and scalability improvements. Praos operates in a semi-synchronous setting, safeguarding against adaptive attackers through private leader selection and forward-secure, key-evolving signatures.

In summary, Cardano's foundation on the Ouroboros PoS protocol embodies advancements in blockchain security, sustainability, and scalability, highlighting its commitment to fostering a secure and decentralized ecosystem.

2.2.3 Consensus Layer: Stake Pools and Slot Leaders

The Cardano blockchain employs the Ouroboros Praos protocol for consensus, dividing time into epochs consisting of slots lasting one second each. An epoch encompasses 432,000 slots, approximately 5 days. Within each slot, nominated block-producing nodes contend for leadership, with one node expected to be nominated every 20 seconds, averaging 21,600 nominations per epoch. Randomly elected slot leaders generate and append blocks, while the settlement delay safeguards against adversarial activities.

Stake pools form the backbone of Cardano's consensus mechanism. A stake pool, a reliable server node, handles ledger maintenance and combines resources (stake) of multiple stakeholders. Stake pools process transactions and create new blocks, with operators running the protocol 24/7 and delegators actively participating. The reward for running Ouroboros arises from transaction fees and inflation.

Slot leaders are chosen at random from stake pools, determining who produces new blocks. The probability of election is proportional to the stake controlled by the pool. To maintain decentralization, an incentive system prevents excessive concentration of stake within a few large pools. Slot leader selection involves calculating the VRF value of each pool's block producer against a threshold, with successful matches resulting in leadership roles for specific slots.

To ensure fairness and security in the slot leader selection process, a reliable source of randomness is crucial. The Ouroboros protocol incorporates a Global Random Oracle feature that produces fresh randomness each epoch. This feature uses a Verifiable Random Function (VRF) [33] to generate randomness based on stakeholder keys. The blockchain's hashing of previous epoch values forms the random seed, creating a new source of randomness for the protocol.

2.2.4 Tokenomics Layer: EUTXO Model

Cardano operates on the Unspent Transaction Output (UTXO) model, similar to Bitcoin, distinguishing it from account-based blockchains like Ethereum. Through the Alonzo upgrade, Cardano introduces the Extended Unspent Transaction Output (EUTXO) model, enabling support for multi-assets and smart contracts.

In the UTXO model, transactions have inputs and outputs. Transactions spend unspent outputs from prior transactions and create new outputs. Unlike account-based systems, assets are stored in the form of unspent outputs on the ledger. Transactions effectively unlock previous outputs and generate new ones.

The EUTXO model expands upon the UTXO model in two significant ways:

- Generalized Address Logic: EUTXO broadens the concept of 'address' by allowing addresses to incorporate complex logic in the form of scripts. These scripts determine whether a transaction is authorized to use a specific output as an input. For instance, during transaction validation, the script associated with an output's address is executed, and based on the outcome, the transaction's utilization of the output is determined.

- Inclusion of Custom Data: EUTXO enhances outputs by enabling them to carry additional data along with the address and value. This augmentation empowers scripts by enabling them to carry state information. Scripts could also access supplementary information termed “redeemers”, which the transaction supplies for each input. This comprehensive contextual information empowers scripts to make informed determinations in complex scenarios and diverse use cases.

2.3 Related Work

Karakostas et al. introduce a systematization approach to assess decentralization in blockchain systems. By categorizing blockchain layers, the paper identifies properties at risk due to centralization. However, it only includes a brief case study on Bitcoin. The authors acknowledge the need for more research on each blockchain under this framework. This methodology can aid in evaluating Cardano’s decentralization across its layers and highlighting areas for this project.

Similar to the above work, Zhang et al. [35] offer a thorough examination of blockchain decentralization through a comprehensive analysis of existing research. The study contributes to the field of blockchain decentralization by providing a systematic review of existing literature across multiple dimensions. It focuses on five essential facets: consensus, network, governance, wealth, and transaction. The authors introduce a decentralization index that serves as a tool to explore the dynamics of DeFi token transfers. However, it’s important to note a limitation in the paper: the self-proposed decentralization index, which is adapted from Shannon entropy, lacks rigorous mathematical proof.

Jia et al. [22] undertake the task of quantifying decentralization across various public blockchains, including Cardano. The approach revolves around assessing the distribution of governance token balances on the chain, utilizing metrics such as information entropy and Gini coefficients. The study aims to shed light on the trade-off between efficiency and decentralization in emerging blockchain platforms. This paper serves as an inspiring pioneer in horizontally comparing the decentralization levels of emerging blockchains. It contributes valuable insights into the degree of decentralization in various blockchain systems, aligning with the objective of the current project. The work done in this paper offers a benchmark for evaluating the decentralization of Cardano through the parsing of full nodes. Notably, Cardano is identified as having a higher degree of decentralization when compared to other public blockchains like Binance Smart Chain and Elrond. However, a limitation of the paper lies in its focus primarily on the distribution of governance token balances as an indicator of decentralization.

This approach provides an important perspective on the tokenomics layer but may not offer a comprehensive overview of the entire complex ecosystem of a blockchain.

Lin et al. [27] introduce a methodological framework for assessing decentralization in blockchains through a variety of metrics and temporal granularities. This study focuses on Bitcoin and Ethereum, employing metrics like the Gini coefficient, Shannon entropy, and Nakamoto coefficient, and considering different time granularities such as days, weeks, and months. The paper aims to deepen the understanding of decentralization by analyzing mining power distribution trends over the course of 2019. The process in the current project draws inspiration from the sliding windows and multi-granularities techniques introduced in this paper to capture cross-interval information for deeper analysis. One of the limitations of this paper is its focus on Bitcoin and Ethereum, without an extension to other emerging blockchains like Cardano.

Gochhayat et al. [16] addresses the shift from decentralization to centralization observed in systems like Bitcoin. The framework proposes metrics for assessing decentralization in governance, network, and storage layers. An outstanding aspect of this paper is its empirical analysis using a stratified approach.

The paper "Decentralization Analysis of Pooling Behavior in Cardano Proof of Stake" [30] authored by Ovezik and Kiayias investigates the decentralization dynamics within Cardano's Proof of Stake system through an agent-based modeling approach. Specifically, it examines how pooling behavior unfolds in the context of Nash dynamics and explores whether Cardano's incentive mechanism effectively fosters decentralization. The paper employs various decentralization metrics, including the Nakamoto coefficient, which aligns with one of the metrics utilized in the current research endeavor. This work provides a deep exploration of decentralization in Cardano's specific context, and its result serves as a valuable reference.

Niya et al. [29] present an exploratory analysis of Cardano, with a focus on its activity, wealth distribution, and decentralization concerns. The study employs a novel approach by applying heuristics-based clustering to group addresses within the same wallet, shedding light on the relationship between pseudonymous on-chain entities and real-world individuals. Furthermore, an analysis of reward distribution in Cardano is conducted, providing valuable insights into the blockchain's dynamics, stake balance distribution, and wealth concentration. This research contributes to the understanding of Cardano's decentralized nature and has implications for similar blockchain systems.

Chapter 3

Methodology and Solution

3.1 Environment Preparation

Before delving into the decentralized analysis of Cardano using full-node data parsing, a robust environment must be established to support the required processes. This section outlines the essential components and repositories involved in the environment setup.

- Cardano-Node: Full Node Functionality

  The core of the environment is the cardano-node [21] repository, which serves as a multifaceted tool for joining the Cardano peer-to-peer network. This versatile repository encompasses several key functionalities, including the ability to connect to relay nodes, gather original on-chain data, extend the chain tip, and maintain the ledger state. The cardano-node aggregates elements from different packages, such as consensus, ledger, and networking, forming a cohesive node structure.

- Cardano DB Sync: Data Collection and Management

  Another pivotal component in the environment setup is the cardano-db-sync [20] repository. The cardano-node and cardano-db-sync repositories work hand-in-hand to create a comprehensive environment for the analysis. The cardano-db-sync component interfaces with a locally running cardano-node through a Unix domain socket. This integration enables the cardano-db-sync to retrieve blocks from the running node, update its internal ledger state, and store relevant data in the PostgreSQL database. This synchronized interaction ensures that the data repository remains up-to-date with the latest blockchain information.

- Hardware Configuration

  To facilitate the analysis process, a virtual machine on Google Cloud Platform has been created. This machine adheres to the system requirements specified by the cardano-db-sync repository, featuring an Intel Broadwell CPU, 4 cores, 32GB of RAM, and 1TB of storage. The purpose of this virtual machine is to serve as the host for cardano-node and cardano-db-sync modules and the remote data repository for the parser.

In the environment setup, the self-developed parser interacts with the data repository by executing SQL queries against the PostgreSQL database. By leveraging this approach, analysts can access a plethora of valuable information, including details about blocks, transactions, addresses, stake pools, and more. This interaction empowers the parser to access on-chain data efficiently, allowing for intricate analysis of Cardano's decentralization dynamics.

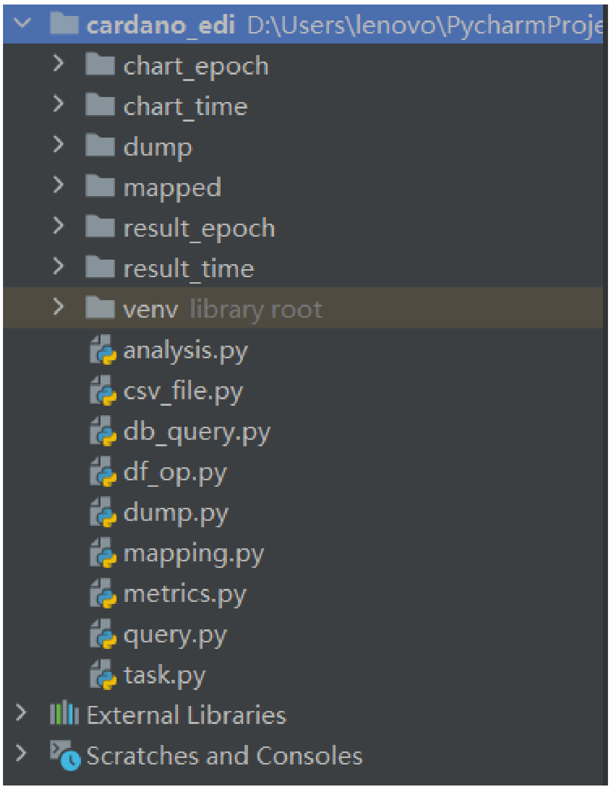
3.2 Parser Development

The development of the parser adopts a methodical approach, encompassing three integral modules: the parser module, mapping module, and metrics & analysis module. Each module serves a pivotal role in effectively retrieving, processing, and scrutinizing on-chain data, thereby contributing to the comprehensive analysis of Cardano's dynamic decentralization landscape. The code structure of the parser is illustrated in Figure 3.1.

Parser Module: Data Retrieval and Materialization

The parser module constitutes the foundational element of the development, orchestrating the extraction of on-chain data from the remote PostgreSQL database. Employing well-structured queries, this module targets relevant data crucial for the analysis. The retrieved data is subsequently materialized into local data files, ensuring accessibility for subsequent processing phases. This process of data retrieval and localization forms the bedrock for downstream analytical pursuits.

Within the parser module, several Python files collaborate to execute distinct functions. *query.py* contains structured SQL statements designed to query specific data from the PostgreSQL database managed by Cardano-db-sync. *db_query.py* establishes connections to the remote data repository, facilitating query operations using the *psycopg2* package. The file *csv_file.py* orchestrates local file I/O, efficiently transferring data between disk-stored CSV files and in-memory Pandas dataframes. The central



```
cardano_edi D:\Users\lenovo\PycharmProje
  chart_epoch
  chart_time
  dump
  mapped
  result_epoch
  result_time
  venv library root
  analysis.py
  csv_file.py
  db_query.py
  df_op.py
  dump.py
  mapping.py
  metrics.py
  query.py
  task.py
External Libraries
Scratches and Consoles
```

Figure 3.1: The code structure of parser developed in this study

orchestrator, *dump.py*, coordinates the functionalities of the aforementioned files, providing a unified interface for higher-level invocations. It equips specialized methods to retrieve desired data tables, thus facilitating a cohesive workflow.

Mapping Module: Data Transformation and Integration

The mapping module emerges as a pivotal intermediary, engaged in data cleansing, transformation, integration, and mapping. Raw data, procured from the parser module, undergoes meticulous refinement within this module. A suite of structured operations prepares the data for subsequent analysis, culminating in coherent and well-prepared datasets. Serving as a crucial bridge between raw data and comprehensive analysis, the mapping module guarantees the precision and trustworthiness of processed data.

Key components within the mapping module include *df_op.py*, encompassing utility functions for dataframe operations such as concatenation, joining, sorting, and null space filling. Additionally, *mapping.py* undertakes column mapping, clustering low-level entities into high-level entities based on identity relationships, and mapping epochs to corresponding dates.

Metrics & Analysis Module: Quantification and Visualization

The metrics & analysis module marks the culmination of parser development, concentrating on the computation of decentralization metrics and subsequent visualization of outcomes. Leveraging meticulously prepared data from the mapping module, this module quantifies pivotal decentralization indicators. These metrics provide discerning quantitative insights into Cardano's dynamic decentralization landscape. Additionally, the module facilitates visualization, yielding graphical representations that amplify comprehension of analysis findings.

The module's functionality resides within *metrics.py*, housing functions to calculate decentralization metrics such as the Nakamoto coefficient. *analysis.py* introduces a sliding window approach to dataframes, thereby generating a multi-granular index. It further provides visualization capabilities, generating insightful charts.

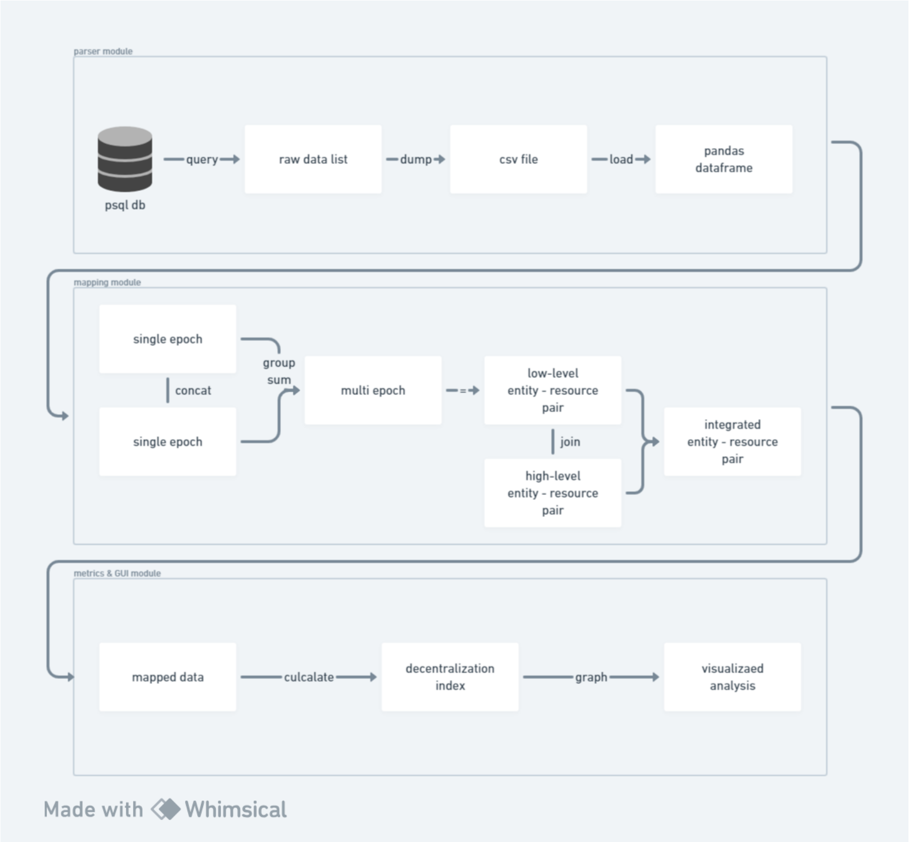


Figure 3.2: Parser Analysis Process

In harmony, these three modules synergistically enable the parser's efficacy. Their interwoven collaboration empowers the parser to fluidly traverse the stages of data retrieval, transformation, quantification, and visualization. The modular architecture engenders clarity in logic, elevating the parser's organizational coherence. This modular segregation simplifies invoking lower-level functions from higher-level calls, fostering streamlined development and augmenting the overall repository's scalability and adaptability. Figure 3.2 shows the logical demonstration of this parser's analysis process.

In summation, the parser's development adheres to meticulous labor division within these modules. Their harmonized interaction ensures the parser's coherent operation, thereby facilitating the exploration and analysis of Cardano's intricate decentralization panorama. Through this methodically structured approach, the parser emerges as an indispensable instrument, decoding intricate on-chain data and distilling meaningful insights about Cardano's decentralized ecosystem.

3.3 Cluster Mapping

The utilization of cluster mapping emerges as a pivotal technique in mitigating the inherent challenges posed by pseudonymous entities within the on-chain context. Particularly, the importance of accounting for off-chain relationships is emphasized, as entities often span multiple addresses [11]. This mapping methodology seeks to establish a holistic perspective by effectively consolidating scattered identifiers under a single overarching entity, thus enabling a more accurate and comprehensive assessment of an entity's holdings. The fundamental premise of the "mapping" technique lies in the consideration of off-chain connections that significantly influence the analysis. The technique serves as a countermeasure to the pseudonymous biases inherent in on-chain data, contributing to a more accurate and unbiased analysis.

In the context of this project, a simplified cluster mapping approach is adopted, specifically aimed at grouping stake pools under the same stake pool operator entity. This practice is particularly significant, as it aligns with the broader objective of mitigating biases stemming from pseudonymous on-chain entities. By clustering pools operated by the same entity, the analysis gains a clearer understanding of the stake pool operator's overall influence and holdings. This approach exemplifies the value of cluster mapping in fostering a more accurate, comprehensive, and unbiased analysis of Cardano's decentralized landscape, ultimately shedding light on the intricate relationships between on-chain entities and their real-world counterparts.

3.4 Sliding Window

The sliding window analysis emerges as a pivotal methodology aimed at illuminating nuanced trends and dynamics that might be obscured by conventional fixed window measurements. The measurement of decentralization, utilizing varying metrics and granularities, offers a comprehensive understanding of blockchain networks. However, a pertinent challenge lies in capturing dynamic trends and anomalies that manifest across different intervals. Fixed window measurements may inadvertently overlook critical cross-interval dynamics, thus limiting the depth of analysis [27].

To address this limitation, we introduce a sliding window-based measurement approach. The proposed approach is designed to bridge the gap left by fixed window measurements, enabling the detection of trends that span multiple intervals and enhancing the ability to identify abnormal situations. By embracing the sliding window technique, the analysis gains the capability to seamlessly traverse temporal boundaries, affording a continuous and uninterrupted assessment of decentralization dynamics.

The results obtained through the sliding window analysis are highly informative, as they provide a comprehensive picture of Cardano's decentralization dynamics over time. The approach effectively bridges the gaps between discrete intervals, offering a nuanced understanding of trends that span multiple timeframes. Additionally, the sliding window approach enriches the ability to identify and address abnormal situations that could impact the network's decentralization. Ultimately, this methodology contributes to a more holistic analysis by incorporating cross-interval insights and revealing the continuous trends that underpin Cardano's evolving decentralized landscape.

3.5 Decentralization Metrics

This section delves into the intricacies of the decentralization metrics employed to measure the relationship between stake pool operators/entities and the resource allocation of block production within the consensus layer. The metrics chosen for this analysis include the Herfindahl-Hirschman Index (HHI), Shannon entropy, Gini coefficient, and Nakamoto coefficient. Each metric provides distinct insights into the decentralization dynamics of stake pool operation and resource allocation.

3.5.1 Herfindahl-Hirschman Index (HHI)

The Herfindahl-Hirschman Index, denoted as HHI, serves as an insightful metric to evaluate the distribution of stake pool ownership and operation within the consensus layer of a proof-of-stake (PoS) blockchain like Cardano. This metric enables us to quantify the concentration or dispersion of stake ownership among different stake pool operators. In the context of PoS ecosystems, stake pools participate in block validation and accrue rewards based on their staked cryptocurrency. The decentralized nature of PoS networks mandates a diverse distribution of stake ownership to avert centralization risks.

The Herfindahl-Hirschman Index (HHI) formula is given by:

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

where n is the number of pools and s_i is the market share percentage of pool i expressed as a whole number.

Utilizing the HHI, we assess the extent of concentration or dispersion of stake ownership by considering both the number of stake pools and the magnitude of their stake holdings. A lower HHI value signifies a more decentralized distribution of stake ownership, indicative of a healthier and more resilient stake pool landscape. Conversely, a higher HHI value indicates concentrated stake ownership and a heightened risk of centralization. By continually monitoring the HHI, blockchain networks like Cardano can gauge the effectiveness of their strategies in maintaining a decentralized stake pool ecosystem.

3.5.2 Shannon Entropy

Shannon entropy, a staple of information theory, emerges as a valuable metric to quantify the degree of randomness and disorder in the allocation of resources—specifically, initial coin distribution within the tokenomics layer. Within the realm of blockchain, Shannon entropy allows us to assess the level of decentralization by evaluating the distribution of wealth among pseudonymous entities.

The Shannon entropy formula is given by:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$

where n is the number of entities, and $P(x_i)$ is the proportion of address x_i within the whole range.

Higher Shannon entropy values denote a more random distribution of wealth, which corresponds to a higher degree of decentralization. This randomness indicates that the allocation of resources is widely spread across addresses.

3.5.3 Gini Coefficient

The Gini coefficient, a renowned measure of economic inequality, finds applicability in assessing the distribution of balance (i.e. UTxO in Cardano) among addresses. Figure 3.3 shows an intuitive understanding of the Gini index [32].

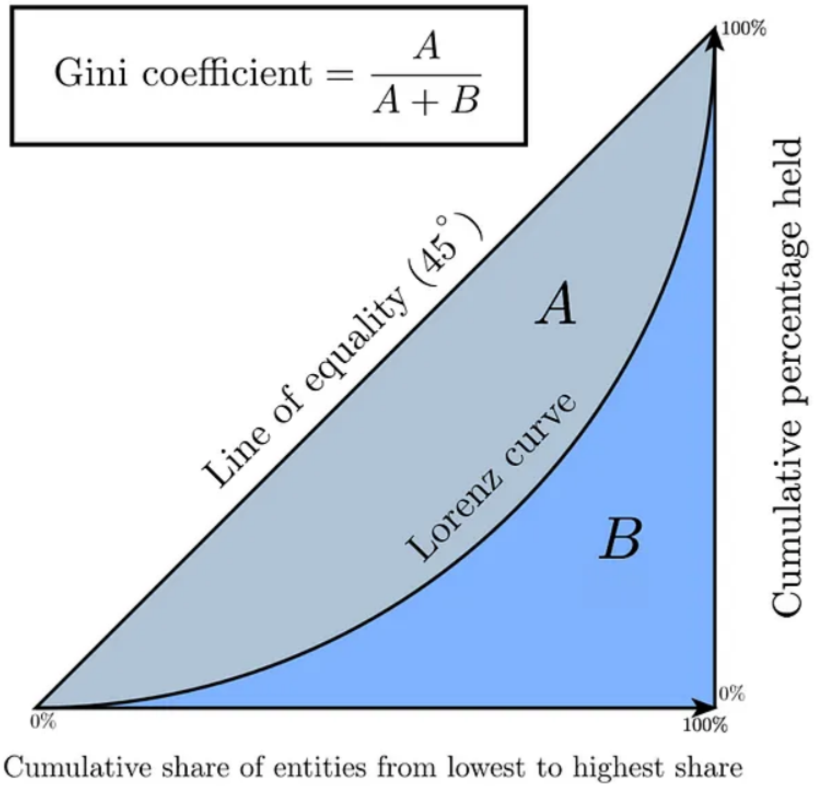


Figure 3.3: The Lorenz Curve and the Gini Coefficient.

A lower Gini coefficient signals a more decentralized distribution of wealth, implying that resources are equitably distributed among individuals. This equitable distribution minimizes the potential for centralization and promotes a balanced ecosystem.

3.5.4 Nakamoto Coefficient

The Nakamoto coefficient assumes center stage as a direct metric that quantifies the association between decentralization and security. This metric highlights the minimum number of entities required to collude in order to control over 51% of the overall block production within the consensus layer. The Nakamoto coefficient serves as a robust indicator of the resilience of the blockchain against malicious attacks and centralization risks [32].

The formula for the Nakamoto coefficient is given by:

$$N_s := min\{k \in [1,...,k] : \sum_{i=1}^{k} p_i \geq 0.51\}$$

, let $p_1 > ... > p_K$, and p_i is the proportion of resource controlled by entity i.

A higher Nakamoto coefficient signifies a more secure and decentralized blockchain, as a larger number of entities would need to collaborate to compromise the network's integrity. This metric provides a tangible link between decentralization and security, guiding blockchain networks in fortifying their systems against potential threats. Although the Nakamoto coefficient offers standardized comparability between networks, it is vital to acknowledge its limitations in capturing certain nuances of the data distribution.

Chapter 4

Decentralization Analysis

4.1 Consensus Layer

This section delves into pivotal metrics that bridges the connection between stake pool operators and the allocation of resources within the consensus layer of the Cardano blockchain. Notably, the epoch serves as a fundamental time unit in Cardano, acting as a key determinant for pool delegation and reward calculation. Within the consensus layer, the entity-resource relationship is examined and quantified using the epoch as the fundamental time period unit. Subsequently, this data can be transformed into a more relatable date format, enhancing its comprehensibility.

Additionally, a cluster mapping approach is applied within the consensus layer to establish connections among multiple stake pools controlled by the same Stake Pool Operator (SPO). Given the intricate landscape of approximately 3000 stake pools, this mapping endeavor ultimately consolidates them into nearly 2500 distinct SPOs. This strategic mapping strategy contributes to a clearer understanding of the underlying structure and relationships, aiding the overall assessment of decentralization in the Cardano blockchain ecosystem.

4.1.1 Pool Stake

The significance of stake pools and their distribution is pivotal as it serves as a reflection of power distribution and participation within the network. This analysis spans from epoch 300 to epoch 420, covering the time period from November 2021 to July 2023.

The Herfindahl-Hirschman Index (HHI) is utilized to measure the concentration of stake pool ownership, providing insights into the degree of decentralization and its

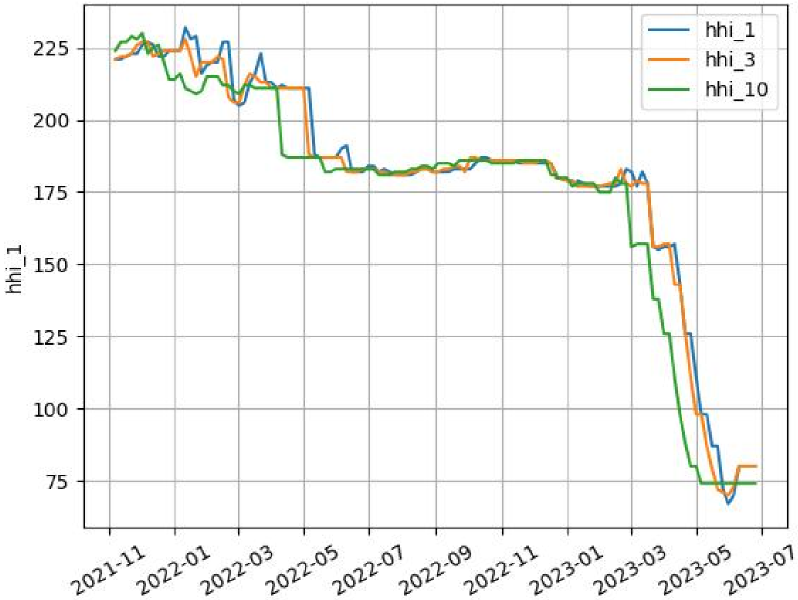


Figure 4.1: The trend of HHI over distribution on pool stake using sliding windows of multiple epochs

trends over time. According to figure 4.1, between November 2021 and May 2022, the HHI exhibited fluctuations within the interval of 210 to 225, suggesting a dynamic state of stake pool distribution. From July 2022 to March 2023, the HHI remained relatively stable at around 185. However, a significant phenomenon occurred after epoch 400, where the HHI experienced a rapid decline, plummeting to a mere 75 by epoch 420 (July 2023).

This intriguing phenomenon can be better understood through an examination of the top Stake Pool Operators (SPOs). Specifically, we analyze the stake proportions of these top SPOs, representing a substantial proportion of the total network stake. Notably, during epoch 300 and 400, an SPO like binance.com held approximately 10% of the market share shown in figure 4.2, with the remaining top SPOs maintaining stakes of 1-3% each. However, a notable trend emerged after epoch 400, as binance.com systematically reduced its stake holdings. By epoch 415, it had dropped out of the top 10 SPOs entirely. This decline in dominance led to the emergence of new top SPOs, each controlling less than 3% of the stake. As a result, the top ten SPOs collectively held a reduced share of only 20%.

Within the proof-of-stake (PoS) model, stake pool stake is closely linked to the power of block production and slot leader selection. A higher stake proportion grants

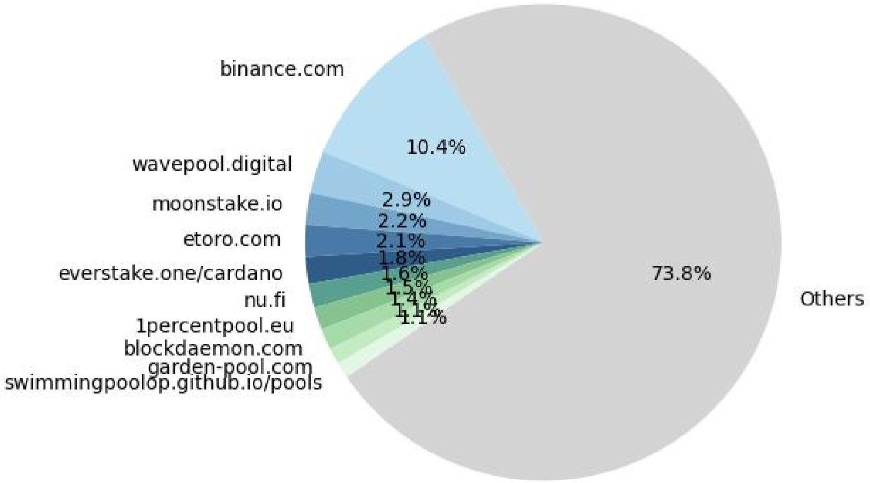


Figure 4.2: The stake share of top 10 SPO

increased opportunities for being selected as a slot leader. Consequently, the distribution of stake pool stake mirrors the distribution of power across the Cardano network. While Cardano's design, incorporating checkpoints in the Ouroboros protocol, provides resilience against long-range attacks, potential implications of single-point corruption arise when the network becomes too centralized. In comparison to delegators, SPOs possess significant influence and capabilities to convey their will to the network. Given that SPOs manage transaction queues in their local memory pools, the risk of censorship and transaction reordering prior to block inclusion becomes apparent, akin to a milder version of front-running as observed in Ethereum.

Comparing Cardano's decentralization metrics with Ethereum's [17], it is evident that Cardano's consensus layer HHI is significantly lower, measured at 75 compared to Ethereum's HHI exceeding 2500. Furthermore, when considering real-world domains like search engines, a prominent entity like Google contributes a substantial portion to the total HHI [13]. This comparative perspective reinforces the observation that Cardano's consensus layer enjoys a relatively high degree of decentralization, fostering a robust and balanced ecosystem. While emergent threats have yet to be identified, the continuous monitoring of decentralization trends becomes imperative to detect potential risks amid rapid market shifts.

4.1.2 Block Production

The significance of block production is paramount, as it reflects the power and influence of stakeholders in shaping the blockchain's operation. This analysis is conducted over the same time span, from epoch 300 to epoch 420, corresponding to the period between November 2021 and July 2023.

The Nakamoto coefficient, a key decentralization metric, demonstrates an upward trend over the two-year period. Interestingly, this trend stands in contrast to the analysis of pool stakes. This opposition in trends can be attributed to the positive correlation between pool stake and block production quantity, driven by the design of the proof-of-stake (PoS) protocol. As such, the contrasting trends in the Nakamoto coefficient and HHI suggest that the minimum number of entities required to collude and dominate the network has increased. Consequently, the cost associated with orchestrating coordinated attacks has risen, thus enhancing the robustness and security of the network. Similar to the previous section, the changes in these indices' trends and their underlying causes parallel the discussions presented earlier.

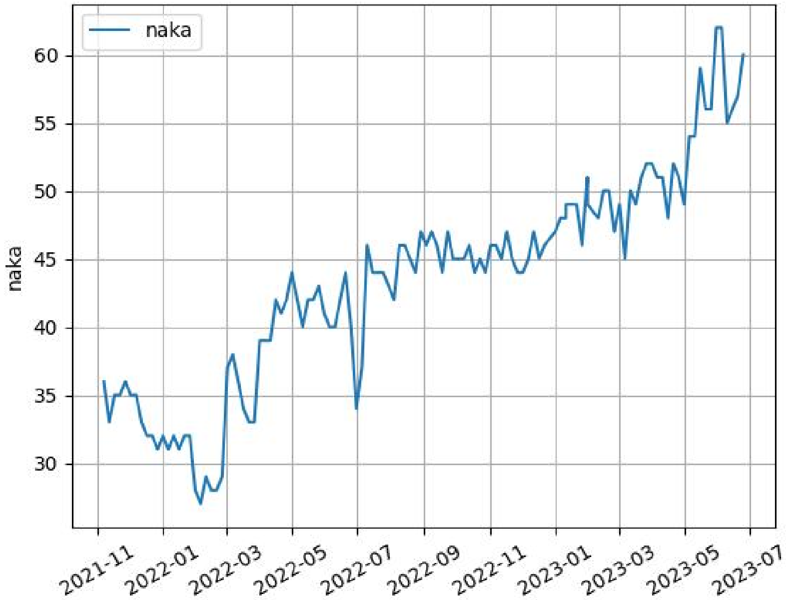


Figure 4.3: The trend of Nakamoto coefficient over distribution on block production using sliding windows of single epoch

However, upon segregating the line graphs based on different sliding window lengths, a noteworthy observation emerges. When analyzing Nakamoto coefficient values over individual epochs versus those encompassing ten epochs, a thought-provoking

Figure 4.4: The trend of Nakamoto coefficient over distribution on block production using sliding windows of ten epochs

phenomenon becomes apparent. Within shorter time windows, such as a single epoch in figure 4.3, Nakamoto coefficients exhibit more pronounced fluctuations compared to the relatively smoother trend observed when considering a ten-epoch span in figure 4.4. For instance, the lowest point of the Nakamoto coefficient occurred in January 2022 at a value of 31 within the ten-epoch sliding window. However, within a single epoch sliding window, the lowest Nakamoto coefficient was recorded at 27. This discrepancy suggests that over shorter time intervals, the network's decentralization experiences temporary dips, potentially lowering the costs of launching attacks like Sybil attacks. This realization underscores the importance of considering both smoothed and granular data to comprehend the ecosystem's vulnerability to rapid threats, while also appreciating its overall long-term trends.

A comparative analysis of Cardano's decentralization metrics with other blockchain networks [27] reveals its standout position. In the context of the Nakamoto coefficient, while Bitcoin's ranges from 4 to 5 and Ethereum's from 2 to 3, Cardano's Nakamoto coefficient in the consensus layer is remarkable. This achievement can be attributed to the design of stake pool reward distribution thresholds, which effectively curbs centralization tendencies in super-large mining pools, resulting in a more evenly distributed stake pool landscape. The Cardano network's resilience is further highlighted, as it remains less susceptible to single point failures caused by unexpected events such as power outages or communication disruptions.

4.2 Tokenomics Layer

4.2.1 Initial Coin Distribution

In this section, we delve into the assessment of initial coin distribution within the tokenomics layer of Cardano, aiming to uncover wealth distribution patterns and their implications for decentralization. According to official information from Cardano, the public sales distribution accounted for 25,927,070,538 ADA, contributing to a total available supply of 31,112,484,646 ADA at launch. Our analysis of the sum of outputs from the genesis blocks yielded a total of 31,112,484,745,000,000 Lovelace, a result closely aligned with the conversion rate of 1 ADA to 1,000,000 Lovelace.

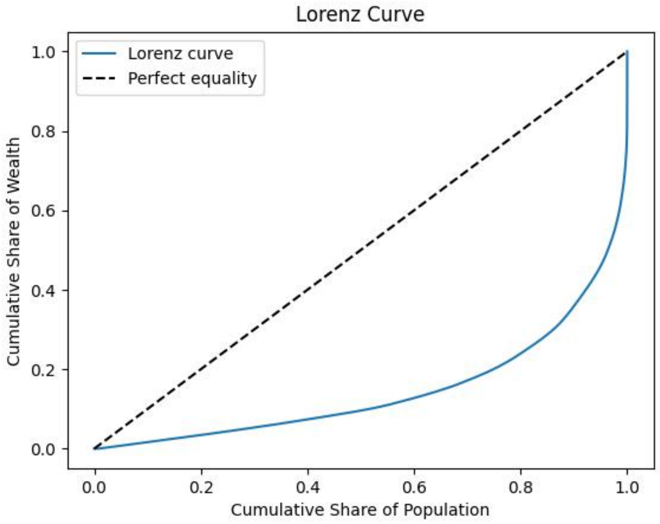


Figure 4.5: The Lorenz Curve of Initial Coin Offering

The computed Gini index of 0.77 and the corresponding Lorenz curve (shown in figure 4.5) depicting cumulative population and wealth distribution offer insights into wealth inequality. Additionally, the Shannon entropy value of 11.04 provides a quantification of randomness and disorder in the initial coin distribution.

Wealth inequality remains a significant concern within the cryptocurrency ecosystem, particularly when it contradicts the principle of decentralization, as evident in proof-of-stake based coins favoring larger holders. Remarkably, evidence from Roubini's testimony in the US Senate Committee on Banking, Housing and Community Affairs indicates that the Gini coefficient for wealth inequality in certain cryptocurrencies surpasses that of nations like North Korea and the United States. Criticism of pronounced wealth gaps has surfaced within the cryptocurrency community, with sentiments that such inequalities run counter to the ethos of decentralization and the fundamental principles of blockchain technology. The examination of ICO of most emerging blockchains has shown that the Gini index of token allocation frequently exceeds the notable threshold of 0.88 [18], underscoring the extent of wealth concentration in initial coin distribution processes.

High wealth distribution ensures that no single entity or a small group of entities can accumulate a disproportionate amount of wealth and influence within the network. This prevents the concentration of power and minimizes the risk of a small number of actors having undue control over the network's decision-making processes, transactions, and operations. Conversely, a highly centralized initial coin offering can result in a small number of entities holding a significant portion of the tokens. This concentration of ownership can lead to a situation where a few powerful holders have disproportionate influence over network decisions and governance, potentially sidelining the interests of smaller holders and undermining the principle of decentralization.

With a higher number of token holders participating in the network, there is a greater potential for diverse perspectives and ideas to be represented in governance decisions. Token holders have a say in proposals, upgrades, and changes to the protocol, contributing to a more inclusive and representative decision-making process. If the initial coin offering is highly centralized, it may discourage broader participation from the community. Smaller investors and participants might feel excluded or marginalized, leading to reduced engagement and collaboration. This can hinder the network's growth and adoption as well as limit the diversity of ideas and perspectives in its development.

4.2.2 Reward Distribution

By employing the Shannon entropy metric, we assess the distribution of rewards across the network's participants. Our analysis reveals that reward distribution over time typically maintains a stable interval between 12 and 13 on the Shannon entropy scale. This indicates a level of income or wealth distribution that is not excessively skewed towards any particular entity or group. However, our examination also exposes four significant troughs in reward distribution shown in figure 4.6, occurring during epochs

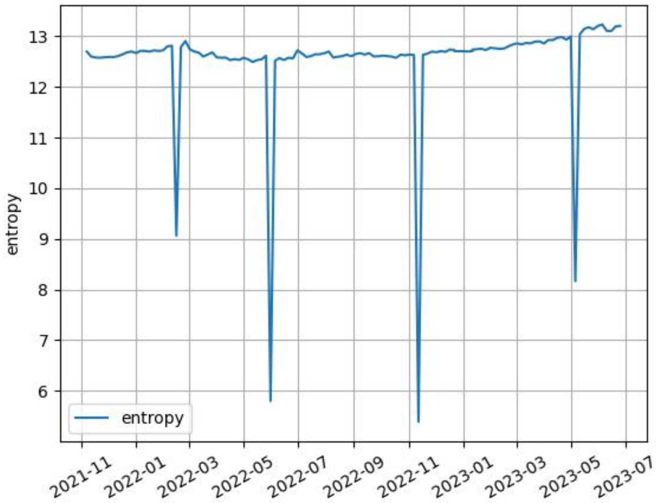


Figure 4.6: The trend of Shannon Entropy over reward distribution

320, 340, 374, and 409. These deviations from the stable interval signify periods of altered reward distribution, warranting further investigation.

The Shannon entropy value serves as a meaningful metric to gauge the equality of income or wealth distribution. As the value of Shannon entropy increases, it signifies a more equitable distribution of rewards. This metric encompasses the entire spectrum from 0, where only one outcome is certain, to its maximum value when all outcomes are equally likely. Countries such as China exhibited a high Shannon entropy of 5.1 in 2002, signifying a more even distribution of rewards within their economies. Notable countries, including India, the United States, and Argentina, fall within the intermediate spectrum of 4 to 5 on the Shannon entropy scale, implying moderate levels of income inequality in their respective economies. Comparatively, the reward distribution trends in Cardano exhibit a more even distribution of income.

High wealth distribution encourages broader participation from a larger number of stakeholders. When wealth is distributed across a diverse group of participants, each individual has a vested interest in the network's success and stability. This alignment of incentives enhances overall network participation, governance, and collaboration.

4.2.3 UTxO

Beginning with an overview, the circulating supply of 33,697,118,996.954581 ADA by epoch 400 serves as the backdrop for our analysis. Our exploration encompasses two distinct perspectives: first, considering the entirety of addresses and their corresponding balances; second, focusing on the top 1000 addresses by balance. Notably, in the former scenario, epoch 400 yields a Gini index of 14.41 and a Shannon entropy of 0.98 in figure 4.7. However, in the context of the top 100 addresses, we observe a shift in values across epochs, with epoch 300 showcasing a Gini index of 8.956 and Shannon entropy of 0.703, followed by epoch 400 with a Gini index of 9.032 and Shannon entropy of 0.688 (shown in figure 4.8). By epoch 420, the Gini index remains steady at 9.110, accompanied by a Shannon entropy of 0.686.

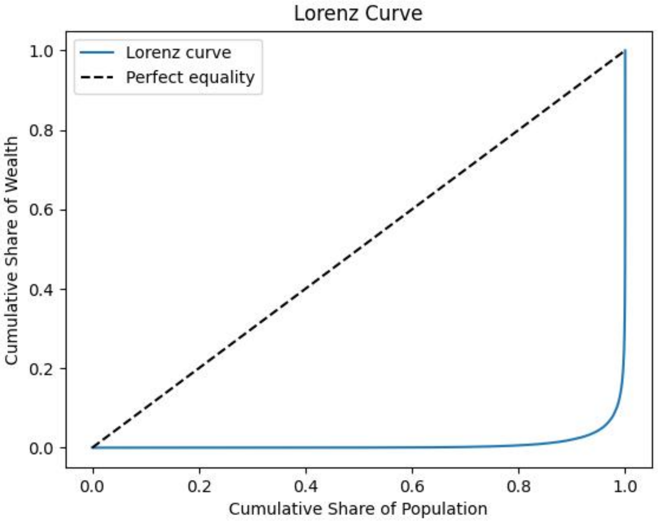


Figure 4.7: The Lorenz Curve of UTxO in Epoch 400

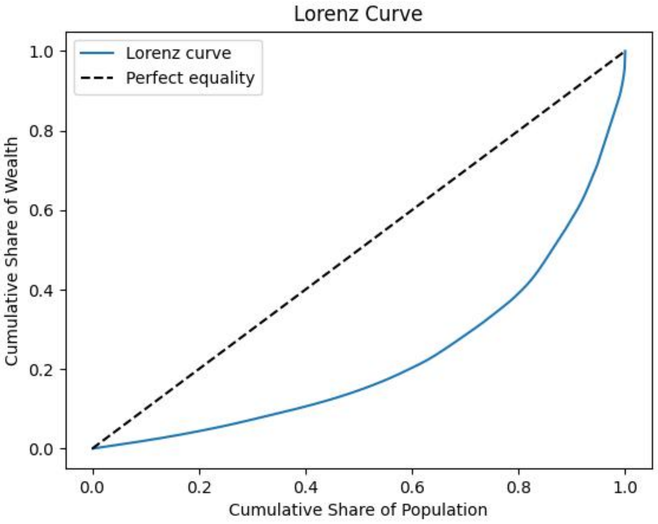


Figure 4.8: The Lorenz Curve of UTxO in Epoch 400 (top 1000 address)

Comparing our findings to real-world country-level Gini indices, where values range from 0.23 (Slovakia) to 0.63 (South Africa), we gain insights into the extent of wealth inequality. Blockchain shows a more severe wealth inequality than the real world.

Interestingly, when juxtaposed with the Gini indices and Shannon entropy of other prominent blockchain networks, such as Avalanche, Tron, and Polkadot, Cardano emerges as the most decentralized among the emerging public blockchains. However, it is important to acknowledge that the actual decentralization of a blockchain might be more intricate than what our address balance-based approach reveals. Address fragmentation, lost tokens, and exchange wallets can all contribute to deviations in our estimation.

The implications of high wealth distribution on the tokenomics layer in Cardano extend to several crucial dimensions:

- Long-Term Viability: A widespread distribution of wealth creates a more stable and sustainable ecosystem. It reduces the risk of sudden market shocks caused by large sell-offs from a small number of entities, which can destabilize the token's value. Additionally, it promotes long-term commitment and engagement from a larger user base.

- Reduced Centralization: A balanced distribution of wealth helps avoid the centralization of stake and decision-making power within a few entities or pools. This fosters a competitive environment where multiple stake pools can actively participate in block production, transaction validation, and other network activities. As a result, the network becomes more democratic and less reliant on a handful of entities.

- Market Resilience: A well-distributed wealth landscape helps the network remain resilient against external market forces and potential regulatory challenges. A concentration of wealth can make a network vulnerable to regulatory actions or market manipulation, whereas a decentralized distribution provides a more balanced and resilient response.

Chapter 5

Evaluation

5.1 Performance

In this section, we delve into a comprehensive evaluation of the performance metrics associated with our analysis framework, encompassing data retrieval, processing time, and associated costs. This assessment provides valuable insights into the feasibility and efficiency of our methodology.

5.1.1 Time Analysis

Our analysis reveals that the data retrieval process for typical data sizes exhibits commendable efficiency. For instance, parsers can extract over 100 epochs of data from the data repository within a minute for parameters like pool stake and pool block. The retrieval and processing of ICO and reward data through PostgreSQL typically require approximately 5 minutes for the parser to receive the results. However, the most time-intensive operation pertains to the UTxO query, which demands roughly half an hour for completion.

A detailed examination of a specific task, such as mapping from pool to stake pool operator (SPO), unfolds a time cost of approximately 6.53 seconds. Calculating metrics across three sliding windows necessitates an average time of about 16.27 seconds. Subsequently, generating line charts incurs a time expense of approximately 1.23 seconds. It is essential to acknowledge that configuring a remote repository on Google Cloud Platform (GCP) entails additional time considerations. The installation and restoration of snapshots from Cardano-db-sync images alone require a minimum of one day.

Taking into account the time needed for Cardano-node chain extension and Cardano-db-sync synchronization, the cumulative time cost expands to at least one week. The synchronization process aligns with Cardano's gradual chain extension approach and synchronizes with the most recent state.

5.1.2 Cost Analysis

When evaluating the associated costs, our chosen virtual machine configuration on GCP incurs a charge of \$0.41 hourly, translating to an estimated monthly expenditure of \$301.98. It is prudent to highlight, however, that there exists a degree of uncertainty regarding the potential risk of a virtual machine with blockchain-related modules being flagged and suspended due to suspicions of crypto mining activities by the GCP team.

5.2 Comparison

In this section, we engage in a comprehensive comparison of our developed parser-based framework with other prevalent solutions in the realm of decentralization analysis. This comparison serves to elucidate the distinct advantages and limitations inherent in each approach, providing valuable insights into the strengths and weaknesses of our methodology.

Online Explorers: e.g., CExplorer Online explorers offer an expedient solution characterized by cost-effective time and fund investment. These tools facilitate quick access to readily available indices, rendering them suitable for simple queries. However, their functionality is constrained by the inability to undertake customized tasks beyond the predefined set of operations. Furthermore, online explorers are limited to querying only the published indices, thereby restricting their utility for more specialized analysis.

Analysis Tools: e.g., BlockSci Advanced analysis tools like BlockSci provide automated processing capabilities and an integrated data analysis environment. These tools excel in their capacity to automate processes and streamline analytical workflows. However, a drawback arises in their inability to support the analysis of emerging blockchains, and their operational range may be limited in certain aspects. Despite their sophistication, they might not encompass all required operations for intricate decentralization analyses.

Online Data Repositories: Online data repositories feature distinct data schemes and incompatible formats that necessitate the development of specialized parsers. Moreover, the unverified nature of the data they host presents challenges to ensuring data accuracy and truthfulness. Despite their allure in providing a centralized repository of blockchain data, these repositories demand comprehensive data extraction and parsing efforts to attain the desired insights.

The parser-based approach we propose possesses its own set of advantages and limitations in the context of decentralization analysis.

This approach entails the highest cost and time input among the discussed solutions. A certain degree of familiarity with the blockchain platform and analytical procedures is requisite for successful implementation.

However, the advantages of our parser-based approach are pronounced. It empowers researchers to define and develop customized analyses, showcasing remarkable flexibility and compatibility. Building the parser from source code enables effortless integration of new features and upgrades, ensuring scalability over time. Moreover, our solution is characterized by its proprietary nature, diminishing reliance and trust on third-party tools or services. The ability to extract data directly from the blockchain network bolsters the authenticity and reliability of the obtained data, further enhancing the trustworthiness of analysis outcomes.

In summary, the comparison underscores the unique attributes of our parser-based framework, emphasizing its distinct flexibility, self-sufficiency, and potential for tailored analysis. While the investment of time and effort is notable, the framework's inherent advantages hold the potential to revolutionize the landscape of decentralization analysis within blockchain ecosystems.

Chapter 6

Conclustion

6.1 Contribution

In this dissertation, we provide a comprehensive perspective on the decentralization of the Cardano consensus and token layers. Our analysis uncovers Cardano's strong ecological environment within the blockchain domain, highlighting its remarkable attributes of safety, stability, and liveness from a decentralization standpoint.

This research delivers novel insights into the field by shedding light on Cardano's decentralized nature. We explore aspects that were previously unexplored or underrepresented in existing studies. This project addresses the existing gap in the academic literature regarding decentralized analysis of Cardano. It marks the first application of a layered methodology in this domain, establishing a foundation for future research standardization. The systematic and theoretical nature of our study renders its findings highly valuable. By combining sliding window and clustering mapping techniques, we bring the analysis closer to real-world dynamics, capturing the changing trends and realities of Cardano's decentralization.

Compared with Previous Work, this work draws on a multitude of existing studies in the field, incorporating valuable research methods and technologies. This enriches the academic value of our paper. Unlike prior studies that often focused on limited aspects, our research encompasses five distinct entity-resource pairs: pool stake, pool block, ICO, reward, and UTxO. Our analysis delves into potential centralization risks and attack vectors that may threaten blockchain networks. This study produces authentic raw data and analytical metrics from five different perspectives within the Cardano blockchain. This includes original data, visualized line charts capturing dynamic changes, and insightful analysis results.

The analysis extends beyond Cardano itself, exploring the practical significance of decentralization indices and their real-world applications. By comparing the results with real-world data and response data from other blockchain platforms, we enhance the practical relevance of our findings. The research findings hold relevance for scholars in the field of decentralization. Additionally, this work has contributed to the advancement of the Edinburgh Decentralization Index project. The data and analytical results presented in this paper also offer valuable guidance for developers and investors in blockchain platforms like Cardano.

This study has the potential to stimulate further research in the realm of Cardano decentralization. It serves as a catalyst for continued investigation, exploration, and innovation in this area.

6.2 Limitation

While this study has provided valuable insights into Cardano's decentralization, it is important to acknowledge certain limitations that provide opportunities for further research and areas of improvement in future analyses.

Firstly, the methodologies employed in this study are based on specific assumptions and models. The accuracy of the results is contingent upon the correctness of these assumptions and the applicability of the models to real-world scenarios. Further refinement of these methodologies, coupled with sensitivity analysis, can bolster the robustness of the findings.

Secondly, the availability and accuracy of data play a pivotal role in shaping the outcomes of our analysis. While diligent efforts were made to collect and validate data from reliable sources, the presence of data discrepancies or inaccuracies could potentially impact the conclusions drawn. For instance, the absence of clustering mapping in the tokenomics layer—such as merging multiple addresses controlled by a solo entity or differentiating addresses owned by cryptocurrency exchanges—could introduce bias into the analysis results. Collaborative efforts with blockchain platforms and regulatory bodies to access verified and current data can help mitigate this limitation.

Furthermore, our analysis predominantly concentrates on Cardano and its decentralized features within the consensus and tokenomics layers. Essential components of decentralization, including governance mechanisms, community engagement, and network security, merit further exploration. Comparative analyses spanning various ecosystems can furnish a more comprehensive comprehension of decentralization trends.

Lastly, the rapid evolution of blockchain technology and the dynamic nature of the cryptocurrency landscape imply that the findings of this study remain subject to change over time. Future research should contemplate conducting longitudinal analyses to capture the progression of decentralization properties and their corresponding implications.

In spite of these limitations, this study furnishes valuable insights to the realm of blockchain decentralization. The elucidation of these limitations serves as a guidepost for future research endeavors, fostering a more profound understanding of the intricacies and dynamics inherent to decentralized networks.

6.3 Further Work

As this study delves into the decentralization analysis of Cardano, it unveils numerous opportunities for further exploration and investigation. These avenues for future research can expand the depth of understanding in the realm of blockchain decentralization and its implications.

One prospective avenue involves refining and extending the methodologies employed in this study. Conducting sensitivity analyses to assess the impact of varying assumptions and parameters on the results can enhance the credibility of the findings. Moreover, the incorporation of advanced data analytics techniques, such as machine learning and network analysis, could provide deeper insights into the dynamics of decentralization.

Exploring the broader landscape of blockchain ecosystems represents another promising direction. Comparative analyses across multiple blockchain platforms can shed light on shared trends and distinctive characteristics of decentralization. Investigating how varying consensus mechanisms, governance structures, and economic models influence decentralization properties can contribute to a comprehensive understanding of the field.

The integration of qualitative research methods can complement quantitative analyses by capturing the perspectives and experiences of stakeholders within the blockchain community. In-depth interviews, surveys, and case studies can provide valuable context and insights into the practical implications of decentralization.

Furthermore, considering the multidimensional nature of decentralization, future research could delve into its social, economic, and political implications. The impact of decentralization on network security, user behavior, economic growth, and regulatory frameworks remains a fertile ground for exploration.

To ensure the relevance and accuracy of future analyses, ongoing collaboration with blockchain platforms, regulatory bodies, and academic institutions is paramount. Access to timely and verified data sources can mitigate potential biases and inaccuracies.

In conclusion, the findings and insights presented in this study lay the groundwork for an array of future research endeavors. By embracing these opportunities, scholars and practitioners can collectively advance the understanding of blockchain decentralization and its multifaceted implications.

Bibliography

[1] Nick Arnosti and S. Matthew Weinberg. Bitcoin: A natural oligopoly. *CoRR*, abs/1811.08572, 2018.

[2] Balaji S. Srinivasan. Quantifying decentralization. https://news.earn.com/quantifying-decentralization-e39db233c28e, 2017.

[3] Sven Barac, Ivica Botički, Gabrijela Perković, Vjekoslav Radošević, and Ivan Terzić. Cardano - what is it and how to start working with it. In *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, pages 1727–1732, 2023.

[4] Carlo Campajola, Raffaele Cristodaro, Francesco Maria De Collibus, Tao Yan, Nicolo' Vallarano, and Claudio J. Tessone. The evolution of centralisation on cryptocurrency platforms, 2022.

[5] Ren Chen, I-Ping Tu, Kai-Er Chuang, Qin-Xue Lin, Shih-Wei Liao, and Wanjiun Liao. Endex: Degree of mining power decentralization for proof-of-work based blockchain systems. *IEEE Network*, PP:1–6, 08 2020.

[6] Ling Cheng, Feida Zhu, Huiwen Liu, and Chunyan Miao. On decentralization of bitcoin: An asset perspective. *CoRR*, abs/2105.07646, 2021.

[7] Lin William Cong, Zhiguo He, and Jiasun Li. Decentralized Mining in Centralized Pools. *The Review of Financial Studies*, 34(3):1191–1235, 04 2020.

[8] Lin William Cong, Ke Tang, Yanxin Wang, and Xi Zhao. Inclusion and Democratization Through Web3 and DeFi? Initial Evidence from the Ethereum Ecosystem. NBER Working Papers 30949, National Bureau of Economic Research, Inc, February 2023.

[9] Bernardo David, Peter Gaži, Aggelos Kiayias, and Alexander Russell. Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake blockchain. In

Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2018*, pages 66–98, Cham, 2018. Springer International Publishing.

[10] Francesco Maria De Collibus, Alberto Partida, Matija Piškorec, and Claudio J. Tessone. Heterogeneous preferential attachment in key ethereum-based cryptoassets. *Frontiers in Physics*, 9, 2021.

[11] Dmitry Ermilov, Maxim Panov, and Yury Yanovich. Automatic bitcoin address clustering. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 461–466, 2017.

[12] Ittay Eyal and Emin Gün Sirer. Majority is not enough: Bitcoin mining is vulnerable. *CoRR*, abs/1311.0243, 2013.

[13] Christian Fuchs. The google and facebook online advertising duopoly. In *Online Advertising Tax as the Foundation of a Public Service Internet, The*, pages 11–19. University of Westminster Press London, 2018.

[14] Adem Efe Gencer, Soumya Basu, Ittay Eyal, Robbert van Renesse, and Emin Gün Sirer. Decentralization in bitcoin and ethereum networks. *CoRR*, abs/1801.03998, 2018.

[15] A. Gervais, G. O. Karame, V. Capkun, and S. Capkun. Is bitcoin a decentralized currency? *IEEE Security amp; Privacy*, 12(03):54–60, may 2014.

[16] Sarada Prasad Gochhayat, Sachin Shetty, Ravi Mukkamala, Peter Foytik, Georges A. Kamhoua, and Laurent Njilla. Measuring decentrality in blockchain based systems. *IEEE Access*, 8:178372–178390, 2020.

[17] Dominic Grandjean, Lioba Heimbach, and Roger Wattenhofer. Ethereum proof-of-stake consensus layer: Participation and decentralization. *arXiv preprint arXiv:2306.10777*, 2023.

[18] Mingyu Guo, Zhenghui Wang, and Yuko Sakurai. Gini index based initial coin offering mechanism. *Autonomous Agents and Multi-Agent Systems*, 36:1–20, 2022.

[19] Manas Gupta and Parth Gupta. Gini coefficient based wealth distribution in the bitcoin network: A case study. In Rajnish Sharma, Archana Mantri, and Sumeet Dua, editors, *Computing, Analytics and Networks*, pages 192–202, Singapore, 2018. Springer Singapore.

[20] input-output hk. cardano-db-sync. https://github.com/input-output-hk/cardano-db-sync. Accessed: August, 2023.

[21] input-output hk. cardano-node. https://github.com/input-output-hk/cardano-node. Accessed: August, 2023.

[22] Yongpu Jia, Changqiao Xu, Zhonghui Wu, Zichen Feng, Yaxin Chen, and Shujie Yang. Measuring decentralization in emerging public blockchains. In *2022 International Wireless Communications and Mobile Computing (IWCMC)*, pages 137–141, 2022.

[23] Dimitris Karakostas, Aggelos Kiayias, and Christina Ovezik. Sok: A stratified approach to blockchain decentralization, 2022.

[24] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov. Ouroboros: A provably secure proof-of-stake blockchain protocol. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology – CRYPTO 2017*, pages 357–388, Cham, 2017. Springer International Publishing.

[25] Yujin Kwon, Jian Liu, Minjeong Kim, Dawn Song, and Yongdae Kim. Impossibility of full decentralization in permissionless blockchains. *CoRR*, abs/1905.05158, 2019.

[26] Chao Li and Balaji Palanisamy. Comparison of decentralization in dpos and pow blockchains. In Zhixiong Chen, Laizhong Cui, Balaji Palanisamy, and Liang-Jie Zhang, editors, *Blockchain – ICBC 2020*, pages 18–32, Cham, 2020. Springer International Publishing.

[27] Qinwei Lin, Chao Li, Xifeng Zhao, and Xianhai Chen. Measuring decentralization in bitcoin and ethereum using multiple metrics and granularities, 2021.

[28] CoinMarketCap OpCo LLC. Today's cryptocurrency prices by market cap. https://coinmarketcap.com/, 2023. April 22.

[29] Sina Rafati Niya, Ivana Mesić, Georgios Anagnostou, Gabriele Brunini, and Claudio J. Tessone. A first analytics approach to cardano. In *2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pages 1–5, 2023.

[30] Christina Ovezik and Aggelos Kiayias. Decentralization analysis of pooling behavior in cardano proof of stake. In *Proceedings of the Third ACM International*

Conference on AI in Finance, ICAIF '22, page 18–26, New York, NY, USA, 2022. Association for Computing Machinery.

[31] Ashish Rajendra Sai, Jim Buckley, and Andrew Le Gear. Characterizing wealth inequality in cryptocurrencies. *Frontiers in Blockchain*, 4, 2021.

[32] Balaji S. Srinivasan. Quantifying Decentralization. https://news.earn.com/quantifying-decentralization-e39db233c28e. Accessed: August, 2023.

[33] Hui Wang and Wenan Tan. Block proposer election method based on verifiable random function in consensus mechanism. In *2020 IEEE International Conference on Progress in Informatics and Computing (PIC)*, pages 304–308, 2020.

[34] Keke Wu, Bo Peng, Hua Xie, and Zhen Huang. An information entropy method to quantify the degrees of decentralization for blockchain systems. pages 1–6, 07 2019.

[35] Luyao Zhang, Xinshi Ma, and Yulin Liu. Sok: Blockchain decentralization, 2023.