NTNU

Kunnskap for en bedre verden

Department of Computer Science

MACS490 - Masters Thesis

Decentralized Exchanges: A qualitative comparison against centralized exchanges

Author:
Dipesh Pandey

June, 2022

Acknowledgements

This thesis was made possible through the resources provided by the Norwegian University of Science and Technology(NTNU)'s Decentralised Systems Engineering(DSE) Lab. I want to thank my supervisor **Mariusz Nowostawski** for constant guidance and support through regular code reviews as well as meetings. Despite his busy schedule, he was the mediator in helping manage the resources from the lab for this project.

Additionally, I would like to thank Mr. **Michael Herbert Ziegler**, a Ph.D. student at NTNU in blockchain technologies, for his valuable help in code debugs. He was always swift to reply to my emails and help solve the issues when I was stuck.

During my short stint at E-Krone, I could meet some incredible researchers and developers working to make the Norwegian blockchain ecosystem a better place. Specifically, I would like to thank the Founder and CEO of E-Krone, **Leander Etemad** as well as **Jonatan Blum**, founder and CTO of E-Krone. The work I did with E-Krone and my colleagues' discussions helped me understand a few vital concepts about my thesis more clearly.

Dipesh Pandey

Abstract

Innovations in cryptocurrency exchanges in recent times have made it easier for people to trade cryptocurrency assets. The increasing number of exchanges has made it clear that this domain will keep evolving in the coming years. Decentralized exchanges revolutionized the way we trade cryptocurrencies by eliminating third parties and allowing transparent and open access to all the activities occurring on the platform.

However, the open nature of such platforms and low barrier to entry have increased clutter and noise in the space. In order to understand such anomalous patterns with granular details, it becomes crucial to build tailor-made datasets from the exchanges. Most relevant researches have opted for third-party tools and even paid proprietary resources to obtain data for their analyses. In this regard, an empirical method that explores the possibilities of obtaining data from exchanges and using the collected data to perform analyses would be a valuable contribution to the field.

Considering the points above, a study was carried out exploring the options of obtaining data from the cryptocurrency exchanges like Uniswap, PancakeSwap, and Binance. Initially, a systematic literature review was conducted that sought out existing research focusing on obtaining data from the exchanges and using it to perform the analyses. The review resulted in very few papers that explored the nuances of data collection. Thus, both the on-premise resources and third-party tools were investigated for data collection to understand the merits of each of them. The collected data was then used to inspect how these exchanges manage their markets regarding pricing, transaction fees, and trading pairs.

The results show that using in-premise resources is difficult and tedious to manage but offers incredible speed and flexibility, while third-party resources, despite being slower and expensive, work well when the required data is available. Additionally, the analyses performed show certain areas like usability and fees where Binance does better than Uniswap and PancakeSwap but is comparatively less robust in aspects like transparency and availability.

Contents

Figures

Tables

Code Listings

List of Algorithms

Glossary

atomic transaction A transaction in which either all the operations are successfully complete, or neither. If one of the operations fail, the entire transaction is rolled back. 23

coin The native token of a blockchain, for eg: ETH is the native coin of the Ethereum network, and ADA is the native coin of the Cardano network. 15

composability composability in blockchainis is the property of dApps and DAOs to integrate with each other by easily cloning each other, and for software components such as tokens and messages to be interoperable between them. 16

doublespendingproblem The double-spending problem is the issue with digital currencies that cash does not have where it is possible to spend an already spent coin. For example: if you pay for a coffee with a $10 bill once, it's not possible to use the $10 bill elsewhere. A transaction using a digital currency like bitcoin, however, occurs entirely digitally and might allow such a flaw if not designed properly.. 5

fungible token token whose type and value is the same as another token. 14

genesis block the first block of the blockchain. 12

KYC Know Your Customer (KYC) is process one has to go through while opening new accounts in banks that allows them to keep a record of the user's identities. 17, 21

mempool A mempool or memory pool is a mechanism used by cryptocurrency nodes for storing information on unconfirmed transactions. It works like a queue or waiting room for transactions that have not yet been included in a block. 26

PoSA Proof of Staked Authority (PoSA) is a variation of PoS where instead of stake with the monetary value, the identity of a validator performs the role of stake. 7, 25

token The assets built on upper layers of the blockchain, for eg: SHIB is a token built on Binance Smart Chain (BSC) and UNI is a token built on the Ethereum network. 15

web3 a catch-all word for next-generation internet services that allow for more decentralization and engagement of individuals at various stages of the value chain.. 36

Acronyms

AMM Automated Market Maker. iii, 21

BNB Binance Coin. 15

BSC Binance Smart Chain. xi, 7, 8, 11, 15, 25

CeFi Centralized Finance. v, 17

CEX Centralized Exchange. 18

CPMM Constant Product Market Maker. 21

DAO Decentralized Autonomous Organization. x

dApp Decentralized App. iii, x, 15

DeFi Decentralized Finance. v, 4, 16, 17, 28

DEX Decentralized Exchange. 18

ETH Ether. 8, 13, 15

EVM Ethereum Virtual Machine. 7, 8, 12

PoS Proof of Stake. x, 7, 8

PoW Proof of Work. 7, 8, 25

Chapter 1

Introduction

This chapter introduces main research motivations behind the study tied with the research questions that were a recurrent subject of focus during the entire process. Furthermore, the main contributions of this research in the field of blockchain are included, followed by a brief summary of the thesis structure.

1.1 Motivation

The field of research in the blockchain[1] space, especially Decentralized Finance(DeFi)[2], is still very new[3]. As of May 29, 2022, the trading volume generated by Decentralized Exchanges(DEX)[4] is about 67% of the total trading volume across all DeFi projects[5]. Thus, DEXes play a central role in the DeFi space. However, there seems to be a scarcity of easy-to-follow academic research when performing qualitative analyses of such exchanges. Specifically, although decentralized exchanges advocate free and open data, many hurdles make the life of a researcher difficult. The issues in data collection from such exchanges is usually caused due to one or more of the following reasons:

- inavailability of existing APIs and servers as blockchain based systems need regular maintenance
- scarcity of free and open-source tools, requiring researchers to pay to get the data

[1] A blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain. More detail is provided on section 2.1.

[2] Decentralized Finance is a financial technology where unlike traditional finance, the goal is to get rid of the third parties that are involved in making financial transactions. More detail is provided on section 2.4.

[3] It was only on 2018 that the term DeFi was coined.(`https://www.trality.com/blog/decentralized-finance`)

[4] Decentralized exchanges help users exchange cryptocurrencies without the involvement of third parties. More detail is provided on section 2.4.3.

[5] This data was taken from coingecko(`https://www.coingecko.com/en/dex`) on May 29, 2022, 12:30 PM UTC)

- lack of easy to follow standards that define the relevant terms
- at present, blockchains have grown to be massive in size, usually in the range of multiple terabytes[6], so collecting all the data on-premise requires a lot of resources and time

In order to mitigate these issues, there are a few third parties tools aimed at easing the process of data collection from DEXes. Therefore, most researchers have relied on third-party tools to perform their analyses. However, such tools are costly and limited to only the data structures they provide.

Additionally, there has not been enough research using empirical methods to scrutinize cryptocurrency exchanges. So, exploring the nuances of on-premise and third-party resources to produce tailored data with an aim to compare different types of exchanges is the primary motivation of this research.

1.2 Research Questions

As mentioned in section 1.1, the field of cryptocurrency exchanges is still new and evolving. Such exchanges produce an enormous amount of data every day, and because they are based on public blockchains, it is possible to extract them to understand the nature of the data. With the motivation that data is an essential aspect of producing valuable analytical research in the cryptocurrency exchange space, the following research questions were asked at the beginning of the study and later addressed during the entire process:

Research Question 1 What is the current state of academic research in the empirical analysis of decentralized exchanges?

Research Question 2 What are the trade-offs of using own copies of blockchains and own resources vs. using third-party resources for collecting exchange data?

Research Question 3 What metrics are the most important to evaluate the performance of cryptocurrency exchanges?

Research Question 4 How can we use free and openly available data to compare decentralized and centralized exchanges?

1.3 Planned Contributions

We believe the experiments and their results produced from this thesis will have several contributions to the field of study. Outlined below are the planned novel contributions of this thesis:

[6] Ethereum's entire archived blockchain is of size 10.7TB as of May 30, 2022 (`https://etherscan.io/chartsync/chainarchive`

- producing research consisting of a systematic literature review that identifies the options available for data acquisition from cryptocurrency exchanges.

- outlining a method to increase the time-frequency of the price data from blockchain-based smart contracts to about 10 seconds[7], while the current frequency from third-party tools like the Graph[8] is 1 hour.

- building generic metrics to compare both centralized and decentralized exchanges in an empirical way.

1.4 Thesis Structure

Chapter 1: Introduction provides a brief introduction about the motivation of this research. This chapter also points out the research questions addressed by this study and the contributions made by this work.

Chapter 2: Background makes the reader familiar with all the cryptocurrency exchange topics. It builds the topics in a funnel way, starting from high-level general topics initially and then narrowing down to the topics most relevant to this study as the chapter progresses.

Chapter 3: Literature Review presents the summary of researches performed by other researchers that involve empirical analysis of cryptocurrency exchanges. It presents the current state of research in this field and shows where this study fits in the domain.

Chapter 4: Methodology includes a detailed description of the processes used to perform the literature review and the technical implementation of the study. It shows the data collection process, and then the analyses performed using the data.

Chapter 5: Results and Discussion shows the results of the experiments performed during the study. This chapter also discusses how the results obtained from the study answer the research questions asked in Chapter 1 while trying to relate to the works done by other researchers.

Chapter 6: Conclusion concludes the research by evaluating at what level research questions were answered by this study. This chapter also outlines the study's limitations and presents the possible future work along the line of this study.

[7] The block time of a blockchain can vary based on aspects like the computing power of the network and the complexity of the problem being solved, however, in Ethereum, the avearage block time is about 12 to 14 seconds (between 12 to 14 seconds).

[8] The Graph is an open-source tool used to index and query blockchain data (`https://thegraph.com/en/`

Chapter 2

Background

In order to understand how cryptocurrency exchanges work under the hood, it is vital to introduce the topics surrounding the domain. This section presents a brief overview of how blockchains came to be, how they gave rise to Decentralized Finance (DeFi), and finally focuses on how exchanges are the most crucial building blocks of DeFi.

2.1 Blockchain

A blockchain is a data structure used for redundant recording, syncing, and sharing of data across multiple data stores in a distributed network[1]. They store and transmit data in the form of packages called ’blocks’ linked together in the form of a ’chain.’ Each node in the network can always view all the network transactions. In 2008, Satoshi Nakamoto envisioned a technology that is "a purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution"[1]. By sending 10 BTC to Hal Finney on January 12, 2009, Nakamoto launched the blockchain revolution through Bitcoin.

The core of a blockchain lies in the use of cryptography. Mainly, asymmetric key cryptography and hash functions are the primary building blocks. Let us take the most common use case of blockchain, i.e., cryptocurrency. Imagine Alice sends 1 BTC[2] to Bob in a blockchain. In order to do this, each of them first maintains a digital wallet that holds their public/private key pair. The public key is shared with others during the transaction. On the other hand, the private key is used to authorize and assign the cryptocurrency to be spent or sent to someone else in the network. In simplest terms, the flow of the transaction between Alice and Bob happens in the following steps, as shown in figure 2.1:

1. The transaction, including the amount sent and the address of Bob, along

[1] On the network of a blockchain lie the vast web of computers called ’nodes’.

[2] BTC is a native coin of the Bitcoin network

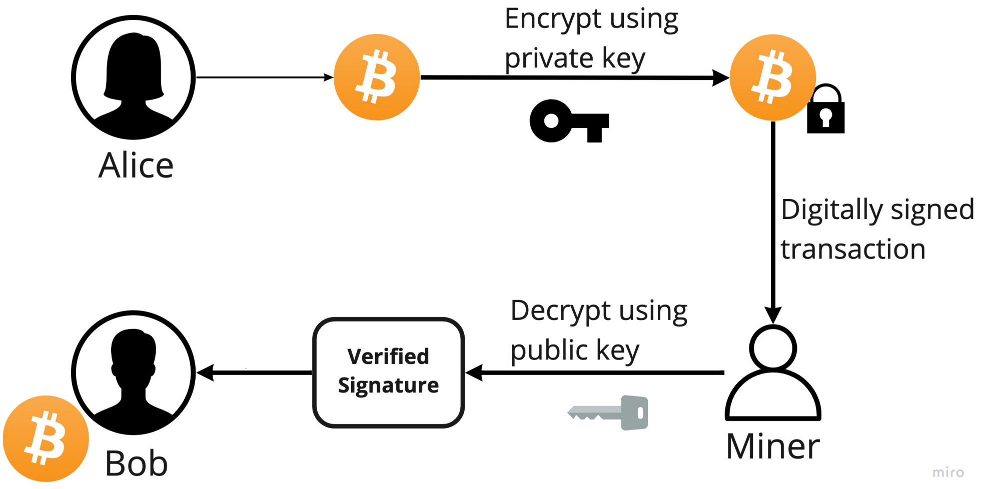


Figure 2.1: Process of digitally signing transaction in a blockchain

with other information, is hashed and then signed using Alice's private key. This signed transaction is broadcasted across all the nodes on the network.

2. Special nodes in the blockchain (usually called miners or validators) check the transaction and validate the sender's authenticity. These validator nodes use the public keys of Alice stored in the blockchain, which allows them to check the authenticity. More on this is covered in section 2.1.2.

3. If the transaction is valid, the transaction is added to the block by the validator, and Bob now has access to 1 BTC which he got from Alice. Otherwise, the transaction gets discarded and is not added to the block.

Blockchain aims to solve the centralization problem by allowing users to interact with each other in a secured peer-to-peer manner without trusting a central authority. Apart from decentralization and security, blockchains guarantee immutability which means that once data has been written, it cannot be modified later. It makes blockchain a valuable tool in use cases like record management, identity management, financial transactions, and others. Another long-standing problem solved by the blockchain is the double spending problem, i.e., it removes the possibility of reproducing a single asset infinitely by confirming that a single unit of transaction happened only once.

2.1.1 Data structure

The blocks in a blockchain record the transactions across many computers(also called nodes) so that it is impossible to alter the backdated content of a single block without having to alter all the subsequent blocks. It allows the nodes to

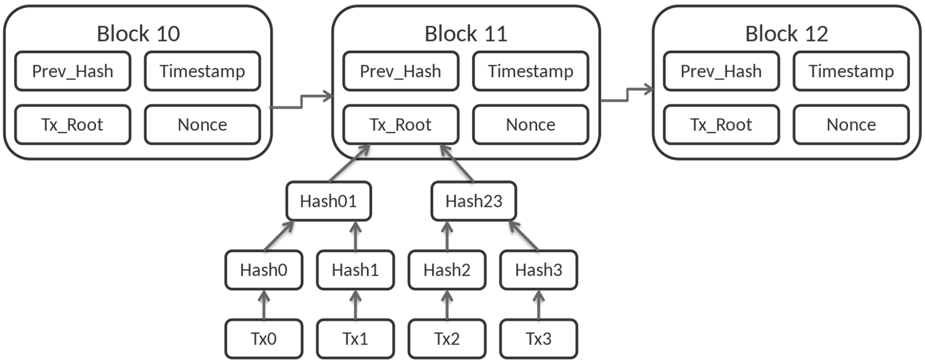


Figure 2.2: Merkle tree structure for storing block data in Bitcoin[3]

verify and validate the transactions independently.

Each block contains a batch of valid hashed transactions and is encoded and stored in a Merkle tree(also known as binary hash trees) for efficient and secure storage of data[2]. Figure 2.2 shows the tree representation of transactions into blocks in Bitcoin. It consists of "leaves" at the bottom that hold the hashes of the raw transactions, intermediate "branches," and the "root" that holds the hash at the top. Block 11 in this example contains the hashed data of 4 transactions: Tx0, Tx1, Tx2, and Tx3(usually, a block in Bitcoin contains about 1700 transactions[4]). The root or the block header is combined with metadata information of the block like the timestamp, the previous block's hash, and the nonce, and run through a hash function producing the block's unique hash. This hash is stored in the next block, as shown in the figure.

Using a Merkle tree provides the following benefits:

1. They provide integrity and validity of data
2. They require little memory or disk space to store
3. They are tamper-proof by nature, i.e., it is easy to track the changes and inconsistencies efficiently in the tree

2.1.2 Consensus in Blockchain

In order to make sure that there exists only one single valid copy of a record in the blockchain, consensus mechanisms are used. They ensure that the records are kept consistently in the decentralized network where all the nodes are brought in agreement by trusting each other, while in general, there is no trust among each other.

[3] Image taken from: `https://en.wikipedia.org/wiki/Blockchain`

[4] The average number of transactions per block in Bitcoin was obtained from: `https://ycharts.com/indicators/bitcoin_average_transactions_per_block` at May 30, 2022.

Proof of Work (PoW)

PoW is a cryptographic proof used to prove that one party proves to another party that a certain computational effort has been spent. Bitcoin popularized it as a consensus mechanism in which network miners compete to append blocks and mint new currency. The work or the computation involved in PoW should be difficult to find but easy to verify. Generally, the energy and hardware-control requirements to manipulate the data in PoW are too big. Thus, it is a perfect system to deter fraudulent transactions and actors in blockchains. PoW is the consensus mechanism used in most blockchains, including Bitcoin and Ethereum.

In PoW, whenever a new transaction has to be added to the blockchain, the miner nodes have to compete to find the right solution to a cryptographic puzzle. This process consumes much energy as all the miners have to use their resources until the solution is found. When a miner finds the right solution, it broadcasts it to the whole network. In return, the miner gets a block reward by the PoW protocol. This reward adds to the circulation of the currency of the blockchain. Another issue related to PoW is that if more than 50% of miners have control over the nodes of the blockchain, they can take control over the entire network. At the time of writing, the block reward in Bitcoin is 6.25 BTC. Ethereum uses a similar approach for consensus and provides a block reward of 3.0 ETH for the winning miner.

Proof of Stake (PoS)

PoS started as an alternative for PoW in order to mitigate PoW's energy consumption concerns. Instead of having miners compete to solve a cryptographic puzzle and mint new coins, in PoS, there are validators chosen in proportion to their quantity of holdings of the cryptocurrency. These validators are responsible for adding the transactions to the blockchain without performing any computation-intensive work, in return for which they usually earn a reward. For the validators to take control over the network, they need to stake(hold) a majority of the blockchain tokens.

In PoS, all the tokens that are to ever exist are already in circulation from day 1. This process might create some centralization of power to a small group of initial founders in the initial days. So, it is not usually considered as fair and reliable as the PoW systems.

Others

Apart from PoW and PoS systems, there have been several other variations of these models that are used for consensus. Binance Smart Chain (BSC) uses Proof of Staked Authority (PoSA) which is a variant of PoS and is EVM compatible[5]. Other

[5] An EVM compatible blockchain establishes an environment that allows it's code to be executed in ways similar to Ethereum Virtual Machine

consensus mechanisms include Proof of History in Solana and Proof of Space in Chia.

A study by Zhang and Kin[3] shows that PoW based mechanisms are helpful for high reliability and fairness systems but are not energy efficient. PoS systems, on the other hand, significantly reduce the energy consumption but are less reliable and fair. Thus, they suggest that a mixed mechanism that combines both PoW and PoS tends to solve the reliability and fairness problem while still reducing energy consumption.

2.1.3 The Ethereum Ecosystem

This research focuses mainly on the decentralized exchanges that run on two blockchains: Ethereum and Binance Smart Chain (BSC). BSC is a repository-fork[6] of Ethereum, so the basic concepts of Ethereum are also applicable to BSC."There is a single, canonical computer called the Ethereum Virtual Machine (EVM) whose state everyone on the network agrees on," according to the Ethereum yellow paper[4]. Every participant keeps an exact duplicate of the machine's current state. They can send a request to this machine to do blockchain computations. Other parties in this situation verify, validate, and execute the computation. As mentioned in section 2.1.2, PoW enables the miners to verify and validate the transactions before they are included into the block. Ethereum has its native cryptocurrency called Ether (ETH). Ether allows for a market of computation in the network, i.e., whenever someone in the network broadcasts a transaction request, they have to offer some amount of ether to the network as a bounty. This amount is awarded to the miners to verify transactions and keep the network running. In particular, the following entities frequently occur in the discussion surrounding these Ethereum like blockchains:

Accounts

An Ethereum account is an identifier entity of the blockchain with an Ether (ETH) balance that can send transactions in the Ethereum network[7]. There are two types of Ethereum accounts:

1. **Externally owned(EOA)**: These accounts can be controlled by anyone with private keys. Creating such accounts costs nothing, and the transactions between such accounts can only be ETH/token transfers.
2. **Contract**: These accounts are created after the deployment of smart contracts to the network and are controlled by code. It has a cost associated with it. External accounts can trigger the code in such accounts to perform many different actions, like token transfer or the creation of new contracts.

[6] A repository fork means when a project forks an existing repository from another project and builds on top of it independently as a different project

[7] `https://ethereum.org/en/developers/docs/accounts/`

Transactions

In Ethereum, transactions are cryptographically signed instructions from the accounts. They are the requests made by accounts to update the state of the Ethereum network[8]. Transferring ETH from one account to another is the simplest transaction in the Ethereum network. An Ethereum transaction consists of the following information:

1. **recipient**: The receiver's address(if EOA, the value will be transferred, if a contract, contract code will be executed).
2. **signature**: The identifier of the sender
3. **value**: amount of ETH to transfer from sender to recipient(in WEI[9]).
4. **data**: optional field to include arbitrary data
5. **gasLimit**: maximum amount of gas units that the transaction can consume
6. **maxPriorityFeePerGas**: maximum amount of gas to be included as a tip to the miner
7. **maxFeePerGas**: maximum amount of gas willing to be paid for the transaction

Gas

Gas is the unit that measures the computational effort required to execute operations on the Ethereum network. In other words, gas denotes the fee required to conduct a transaction successfully in the network[10]. Ether(ETH) is the currency used to denote gas fees. The gas prices are denoted in gwei, where $1 gwei = 10^{-9}$ ETH. One of the reasons why gas fees exist is to help keep the Ethereum network secure. Requiring fees helps filter out bad actors from spamming the network.

[8] `https://ethereum.org/en/developers/docs/transactions`

[9] Wei is the smallest denomination of ether—the cryptocurrency coin used on the Ethereum network. One ether = 1,000,000,000,000,000,000 wei (10^{18}).

[10] `https://ethereum.org/en/developers/docs/gas`

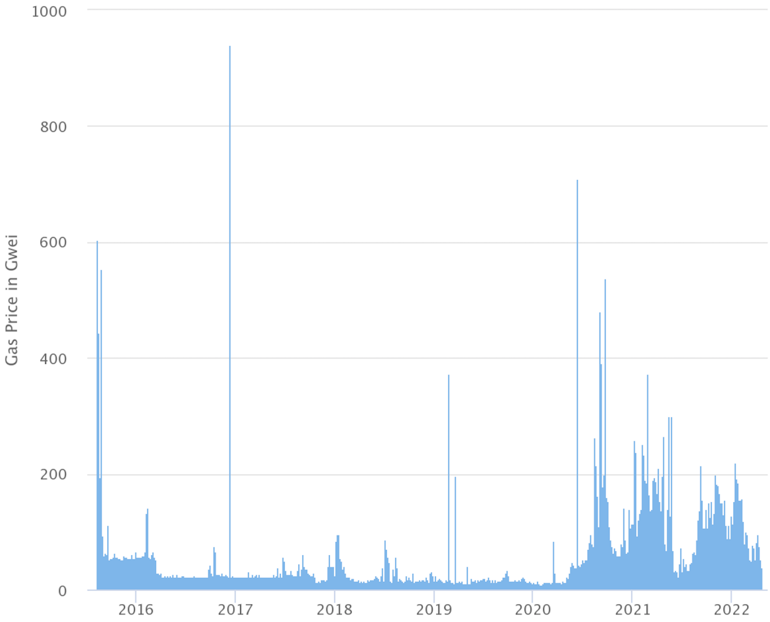


Figure 2.3: Gas price in Ethereum over the years[11]

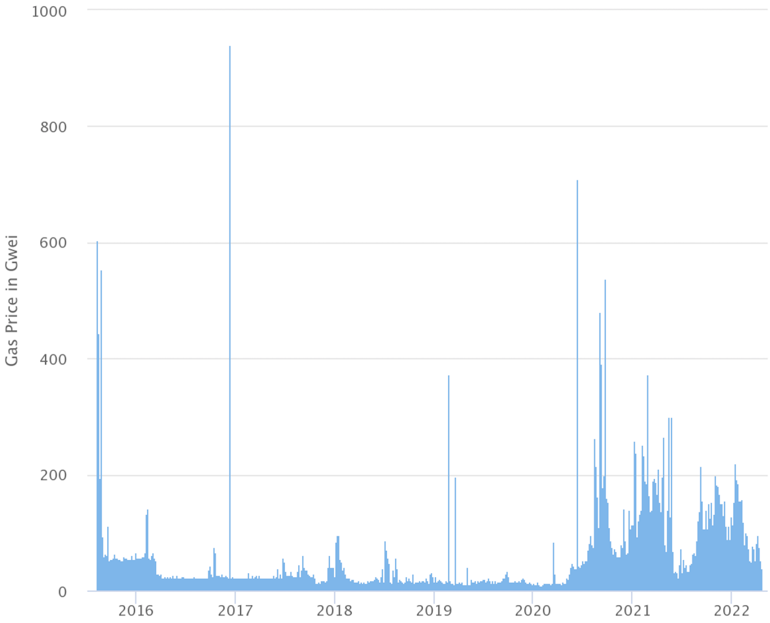


Figure 2.4: Gas price in BSC over the years[12]

[11] The gas price data for Ethereum and BSC were obtained on April 30, 2022 from `https://etherscan.io/chart/gasprice`

[12] The gas price data for BSC were obtained on April 30, 2022 from `https://bscscan.com/chart/gasprice`

Figure 2.3 and 2.4 show the fluctuation of Ethereum and BSC gas fees over the years. Some researches have looked into the reasons behind fluctuating gas prices in blockchain networks. Donmez and Karaivanov show that changes in service demand in the network significantly affect the gas price[5].

Networks

The Ethereum blockchain provides a set of protocols followed by multiple independent networks that do not interact with each other[13]. Usually, these networks are used for testing, development, or production purposes, similar to general software engineering patterns. Ethereum provides two major types of networks for production and testing purposes.

1. **Mainnet**: The Ethereum Mainnet is the major public Ethereum blockchain network. It holds the actual value of transactions that occur on the network. The mainstream Ethereum(ETH) prices and news circulated is the one of the Ethereum Mainnet.
2. **Testnet**: Ethereum Testnets are the test environments provided to developers and testers to experiment with their smart contracts before they deploy them in the mainnet. The protocol developers also use it to test out new features in the blockchain. The ETH on testnets has no value, although they work the same way as the protocol coin in the mainnet. Testnet ETH can be acquired using faucets which are web apps that send ETH to the desired address. Sepolia[14], Görli[15], Ropsten[16], etc. are some of the popular Ethereum testnets.

Nodes and clients

In order to run a node to verify blocks and transaction data in Ethereum(or BSC), we need a client on our computer. Ethereum clients exist in different programming languages like Go, Rust, Javascript, and Python[17]. These clients allow syncing of the data with the latest Ethereum state, which is one of the processes used for this study. The clients can run three different types of nodes:

1. **Full node**: A full node stores the entire blockchain data. It participates in block validation and verifies all blocks and states. It also serves the network and provides data on request.
2. **Light node**: A light node stores only the header chain and requests everything else. It can verify data validity against the state roots in the block headers and is helpful for low-capacity devices that cannot store gigabytes of data.

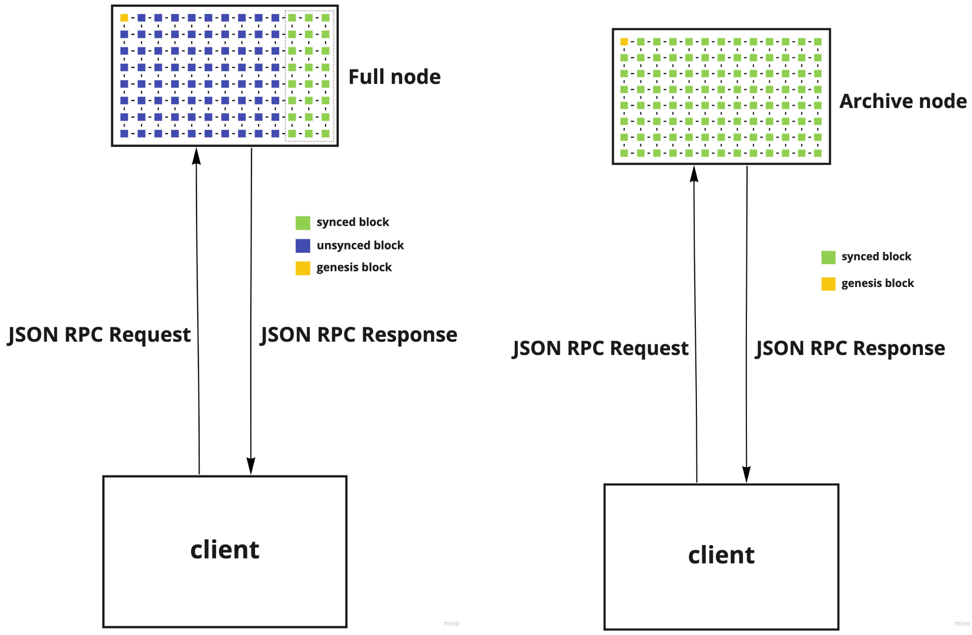
[13] https://ethereum.org/en/developers/docs/networks/

[14] https://sepolia.dev/

[15] https://goerli.net/

[16] https://ropsten.etherscan.io/

[17] https://ethereum.org/en/developers/docs/nodes-and-clients/



(a) Ethereum Full Node

(b) Ethereum Archive Node

Figure 2.5: Full node vs Archive node

3. **Archive node**: An archive node stores all the data of a blockchain from its inception. These data are usually in the order of terabytes and thus are more useful for services like chain analytics and block explorers.

In most cases, either a full node or an archive node is used to perform data collection activities. Figures 2.5a and 2.5b show the basic difference between the organization of blocks in full node and archive node. A full node stores the state of the most recent 128 blocks. These recent blocks are represented with the green blocks in figure 2.5a. In addition to the states, it also stores about one week of trace data. An archive node, on the other hand, stores all the states of each block since the genesis block.

2.2 Smart Contracts

Introduced by Nick Szabo in the 1990s as "a set of promises, specified in digital form, including protocols within which the parties perform on these promises"[6], smart contracts are now the building blocks of all the projects built in the blockchain. It is a computer program that executes automatically and performs actions according to the specified set of terms in the digital contract. While Bitcoin allows simple scripts to execute in its blockchain, Ethereum took this to the next level by providing a Turing-complete programming language where developers could write code that gets executed in the EVM. In Ethereum, a smart contract is a type of an Ethereum account. Below is a simple smart contract implementation of a digital vending machine as depicted from the Official Ethereum Documentation Website[18].

```
pragma solidity 0.8.7;

contract VendingMachine {

    // Declare state variables of the contract
    address public owner;
    mapping (address => uint) public cupcakeBalances;

    // When 'VendingMachine' contract is deployed:
    // 1. set the deploying address as the owner of the contract
    // 2. set the deployed smart contract's cupcake balance to 100
    constructor() {
        owner = msg.sender;
        cupcakeBalances[address(this)] = 100;
    }

    // Allow the owner to increase the smart contract's cupcake balance
    function refill(uint amount) public {
        require(msg.sender == owner, "Only the owner can refill.");
        cupcakeBalances[address(this)] += amount;
    }

    // Allow anyone to purchase cupcakes
    function purchase(uint amount) public payable {
        require(msg.value >= amount * 1 ether, "You must pay at least 1 ETH per
             cupcake");
        require(cupcakeBalances[address(this)] >= amount, "Not enough cupcakes in
             stock to complete this purchase");
        cupcakeBalances[address(this)] -= amount;
        cupcakeBalances[msg.sender] += amount;
    }
}
```

Code listing 2.1: Solidity implementation of a simple vending machine

Like in general programming languages, Solidity has programming constructs like variables, constructors, and functions, which are enclosed within a contract. In the code listing 2.2, the smart contract allows anyone to purchase cupcakes from the digital vending machine. It also has another function that refills the cupcake balance when needed. There is no restriction to anyone to write a smart contract and deploy it in the blockchain network. One needs to know the language and have enough ETH to pay gas fees to deploy the contract. Once the contract gets deployed, it acts as a public API that can be called from other smart contracts or client tools to interact with the functionalities of the deployed contract.

In addition to working as code blocks that run on the blockchain, smart contracts also act as endpoints for data. Usually, important information of dApps(2.3.2)

[18] The famous smart contract of a digital vending machine was taken from the Ethereum docs website: `https://ethereum.org/en/developers/docs/smart-contracts/`.

is stored as the states of the smart contract in the blockchain. So, for example, if someone wants to know the price of ETH at block number 12474190, he can easily query the blockchain to acquire the data as long as the smart contract defines the price of ETH as its state. This study uses this strategy to obtain data like prices and token information from the smart contracts of decentralized exchanges.

2.2.1 ERC-20 Tokens

Ethereum provides a set of rules to be followed by developers to build tokens(discussed more in section 2.3). A token in Ethereum can range from fiat currency to an event ticket to shares of a company. In order to be able to build such mechanisms, developers need to follow a standard that allows them to become interoperable with each other.

The ERC-20 is a standard for fungible tokens to be followed while developing the smart contracts for such tokens. This standard requires that the body of the smart contract implement the following methods as mentioned in the Ethereum official website[19]:

```
//methods
function name() public view returns (string)
function symbol() public view returns (string)
function decimals() public view returns (uint8)
function totalSupply() public view returns (uint256)
function balanceOf(address _owner) public view returns (uint256 balance)
function transfer(address _to, uint256 _value) public returns (bool success)
function transferFrom(address _from, address _to, uint256 _value) public returns (
    bool success)
function approve(address _spender, uint256 _value) public returns (bool success)
function allowance(address _owner, address _spender) public view returns (uint256
    remaining)

//events
event Transfer(address indexed _from, address indexed _to, uint256 _value)
event Approval(address indexed _owner, address indexed _spender, uint256 _value)
```

Code listing 2.2: Mandatory methods and events for an ERC-20 contract

2.3 Tokenization in Blockchain

Bitcoin started with the aim of decentralizing the traditional finance model. Bitcoin is still the most prominent cryptocurrency and is being used as a legal tender in many countries already. However, with the inception of Ethereum, we have started to see more innovation in space. With smart contracts, the possibilities of different types of projects that can exist are endless. Most of these projects have

[19] The required methods and events to be implemented for an ERC-20 token were taken from the Ethereum official docs: `https://ethereum.org/en/developers/docs/standards/tokens/erc-20/`.

one thing in common: they have their token or coin that represents some value in the real world which can be designated in several layers of the blockchain.

In general, the tokenization occurs in three layers of the blockchain, as depicted in figure 2.6. These three tokenization methods in the blockchain represent specific assets or rights in the real world. An attempt has been made in this section to explain the concepts of different types of tokens circulated in the blockchain space by providing their real-life analogies.

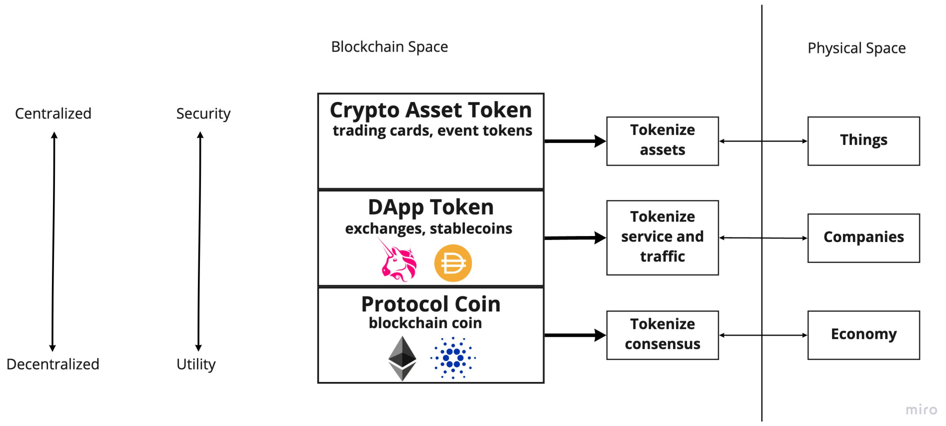


Figure 2.6: Tokenization in Blockchain layers

2.3.1 Protocol Coins

Blockchains can be different based on the consensus mechanisms and the tokenomics they adopt. The core of each of them has a native currency, also called a protocol coin. Consensus is the basic activity performed at the base layer of the blockchain, so it is safe to say that protocol coins tokenize the consensus. Ether (ETH) is the protocol coin of the Ethereum network whereas Binance Coin (BNB) is the protocol coin of BSC.

In the real-world, such coins correspond to the taxation system employed by the government. The government sets rules for the citizens based on their actions. For instance, when a citizen buys a car, some portion of the cost goes to the running of the government as tax. It is similar to gas fees to be paid by someone when she interacts with the blockchain to use certain services.

2.3.2 Decentralized App (dApp) Tokens

Decentralized apps or DApps are the latest innovation in the blockchain space. With the advent of smart contracts, we have started to see many projects built on top of the base layer of the blockchain. This layer is also often called the application layer. These DApps run autonomously and can work as service providers on top of the base consensus layer. An example of such a service provider is Uniswap.

It allows the exchange of tokens in a decentralized manner, all of which are managed by a smart contract. Similarly, Aave is another project that allows the lending and borrowing of cryptocurrency assets. These DApps have to interface with the core blockchain layer to provide their services. They will pay the gas fees when they make transactions, and in addition, they can charge additional charges to their users.

In the real world, an example that is similar to DApps is a stock exchange. It is a separate service provider that helps users exchange their fiat currencies by charging them some fees. In hindsight, however, the stock exchange does so based on the taxation laws set by the government and pays some portion of the tax from its earnings to the government.

2.3.3 Crypto Asset Tokens

Finally, above the application layer lies the crypto asset token layer, which usually represents the natural world's physical objects. A few examples of usage of such tokens could be event tickets and trading cards. So, a company that runs an event management company can issue such tokens to its customers. The customers can buy these tokens and use them as their identity to enter an event. These coins are more flexible and can be programmed according to need. They can be programmed to be valid up to a specific period and expire automatically after that time. Similarly, one company might decide to make the tokens non-transferable, while others can make them transferable based on their demands. In similar ways, one can either use the token to attend an event or transfer it to a friend who can then use the token to attend the event, given that it is used only once as an entry ticket.

2.4 Decentralized Finance(DeFi)

While Bitcoin showed that blockchain could be used very effectively as a cryptocurrency, it had several limitations. One is that it is only confined to crypto. Ethereum solves this issue by providing a general-purpose blockchain that anyone can program, and once the program is deployed, it can be accessed by anyone without any special rights. This powerful idea of smart contracts saw a lot of innovative projects built on top of the core Ethereum blockchain. In particular, the most popular projects have come out to be touching finance one way or the other. They have shown that it is possible to have a decentralized version of the traditional finance, also called DeFi, which runs autonomously without the involvement of third parties like banks, brokers, and other intermediaries. All of this is made possible with the help of smart contracts, as shown in figure 2.7.

Just like in traditional finance, several components make DeFi. These components are often called DeFi legos, which can be stacked on top of each other to build new functionalities. It has been made possible because of the easy interfacing of one component's smart contract with another. This composability feature

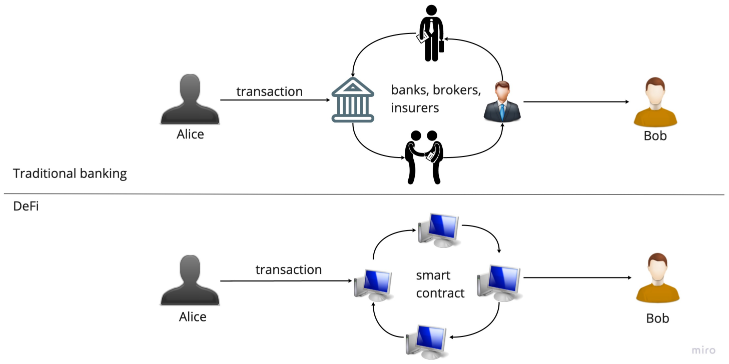


Figure 2.7: DeFi vs CeFi

has led to over 500 DeFi projects currently in circulation in Ethereum alone[20].

Lending and borrowing

DeFi platforms like Aave and MakerDao allow users to lend and borrow cryptocurrency tokens without going through the Know Your Customer (KYC) process. Lenders can deposit their coins and tokesn into the DeFi protocol-based smart contracts. In return, they get newly minted tokens native to the protocol as well as interests in their deposits, e.g., Aave lenders get AAVE tokens, Compound lenders get COMP token, and MakerDao lenders get Dai tokens. These tokens can be redeemed by the users at any time.

Borrowers, on the other hand, can choose to borrow from these protocols by putting certain amount of tokens(like ETH) as a collateral. In such platforms, the loans are over-collateralized, usually to account for the volatile nature of the cryptocurrencies. For instance, in Aave, lenders can deposit cryptocurrencies like Dai(which is a stablecoin). The borrowers can then use their collateral to withdraw crypto tokens listed in Aave. If 10 ETH is deposited by a borrower in Aave, because the loan-to-value ratio is 80%, he can only borrow upto 80% of his collateralized value. If the value of 10 ETH is $20,000, he can only withdraw $16,000 worth of crypto tokens.

At the time of writing, there are 46 lending and borrowing DeFi protocols in circulation.[21].

[20] The data for the number of DeFi was taken on May 30, 2022 from coinmarketcap: `https://coinmarketcap.com/view/lending-borowing/`

[21] The data for number of lending and borrowing DeFi protocols was taken on May 30, 2022 from coinmarketcap: `https://coinmarketcap.com/view/lending-borowing/`

Stablecoins

Stablecoins are cryptocurrencies whose price is pegged to the value of another asset, usually another cryptocurrency, a fiat currency, or commodities like precious metals. The use of stablecoins in DeFi is vital because they help deter the extreme volatility of prices in most other cryptocurrencies. Pegging the price of a coin to, for example, the USD allows for correlation with the economics related to the USD, which is more often than not considered stable. Additionally, stablecoins allow to open up financial services in places that did not have access to USD before. With a stablecoin, a person sitting in remote Africa can access a cryptocurrency that works similar to the USD by just having access to a smartphone and the internet. Tether (USDT), USDC, and DAI are among the most popular stablecoins in the Ethereum network. As of the time of writing, there are 100 stablecoins in circulation, including USD, EUR, GBP, and others.[22].

Decentralized Exchange

Having a market to exchange currencies is significant in any economy. While we have the traditional centralized exchanges with services like Coinbase and Binance, they do not fully utilize the features of the blockchain. That is where Decentralized Exchange (DEX) comes into play. Using decentralized exchanges, people have complete control over their crypto assets. So, instead of trusting a massive company like Coinbase(a Centralized Exchange (CEX)) to trust my assets and my private keys, I will instead opt to use a that provides complete control over my private keys and my assets. I can then use this wallet to interact with the decentralized exchanges, which are dApps built on top of the blockchain and allow exchanging assets using smart contracts. Uniswap, Compound, Curve, and Balancer are among the prominent decentralized exchanges in the Ethereum network. As of the time of writing, there are 216 decentralized exchanges in circulation.[23].

Derivatives

In traditional finance, derivatives are the assets that derive their value from another asset, such as a stock, bond, commodity, or currency. While derivatives in traditional finance are regulated heavily and have taken many decades to evolve, in DeFi, because of the permissionless and open nature, derivatives have been innovating at a rapid pace.

Popular DeFi derivative protocols include Synthetix, UMA, DyDx, etc. Synthetix, for instance, allows users to create synthetic assets that track the value of several tradeable things. These things can be fiat currencies, cryptocurrencies,

[22] The data for number of stablecoin DeFi protocols was taken on May 30, 2022 from coinmarketcap: `https://coinmarketcap.com/view/stablecoin/`

[23] The data for number of DEXes DeFi protocols was taken on May 30, 2022 from coinmarketcap: `https://coinmarketcap.com/view/decentralized-exchange/`

and commodities like oil, gold, silver, etc. Similar to overcollateralization used in lending and borrowing protocols like Aave, Synthetix allows traders to trade synthetics (“synths”) which are overcollateralized derivatives on cryptocurrencies, fiat currencies, commodities, and the stock market. Users can create synthetic tokens that track the price of underlying assets using smart contracts and oracles without buying them.

As of May 30, 2022, there are 96 derivatives platforms in circulation.[24].

2.5 Cryptocurrency Exchanges

Cryptocurrency exchanges provide a way to exchange cryptocurrencies just like the forex market in traditional finance, which provide a way to trade fiat currencies. These exchanges allow the exchange of one crypto to another, e.g., bitcoin to Ethereum, buy crypto using fiat currency, or convert crypto to fiat currency. These exchanges reflect the current market price of the currency in their platform, meaning the price of the same coin can be different(but close) in different crypto exchanges. Customers usually look for the following aspects in a crypto exchange before deciding on using them[25]:

1. **Accessibility**: It should be readily available, e.g., in an easy to use web app or an app
2. **Security**: It should be able to secure the funds and private keys of its users
3. Fees: It should have affordable transaction fees
4. **Liquidity**: There should be enough liquidity in the platform so that when trades do not move, the prices too much
5. **Trade pairs offered**: The exchange should offer a variety of crypto token pairs and do it rapidly because there are tons of new tokens are added to the blockchain every day
6. **Tax information**: The exchange should either handle taxes for the users based on their geographical location or at least be able to guide them to the rules

In a broader sense, these cryptocurrency exchanges can be categorized into two sections: the ones that follow the traditional order book models that also happen to be usually centralized, and the other ones that are usually automated market model-based and usually happen to be decentralized.

2.5.1 Order book based exchanges

An order book-based exchange uses a list of orders that records the interest of buyers and sellers in a particular asset. A central authority maintains the pool of

[24] The data for number of derivative DeFi protocols was taken on May 30, 2022 from coinmarketcap:(`https://coinmarketcap.com/view/derivatives/`).

[25] The criteria for users deciding to use a certain crypto exchange was taken from: `https://time.com/nextadvisor/investing/cryptocurrency/what-are-cryptocurrency-exchanges/`

assets available in the system where the buyer wants to buy cheaper while the seller wants to sell dearer. If the wishes of both parties can be met, the order gets executed.

Volume	Price
300	1.27
650	1.26
1200	1.25
750	1.24
400	1.23
200	1.21

(a) Buy Side

Volume	Price
240	1.20
400	1.19
800	1.18
650	1.17
450	1.16
200	1.15

(b) Sell Side

Table 2.1: Illustration of order book model

Consider the scenario in table 2.1 where buyers are buying and selling a token for USD. On the left side is the buy-side of the order book, and on the right is the sell-side. E.g., a buyer is willing to buy 240 tokens at the rate of $1.20 per token, but the only seller willing to sell at a price closer to 1.20 will sell 200 tokens at the rate of $1.22 per token. It is where the buyer and seller have to wait until they get the desired rates. There also exist other types of buyers who do not wish to wait and want to buy at the current available price in the market. The exchange keeps track of these orders, and after execution, the prices go up or down on the table.

Binance

Binance is an example of an order book-based cryptocurrency with the highest trading volume among all exchanges as of May 30, 2022[26]. Started in June 2017, it helps users trade variety of cryptocurrencies including ERC-20 tokens, protocol coins, and even fiat currencies. Although it is the biggest exchange at present, there are a few disadvantages of Binance:

1. It is centralized, i.e., a single authority has full control over the private keys and assets of its customers
2. The buyers and sellers have to wait until they find the right price on the platform, so it creates an idle time situation
3. Companies like Binance have a non-transparent corporate structure, so people find it difficult to trust them
4. They have issues with being hacked[27] and having weak standards in security compliance[28].

[26] The ranking data of exchanges was taken from Coinmarketcap (`https://coinmarketcap.com/rankings/exchanges/`) on May 30, 2022.

[27] `https://www.wired.com/story/hack-binance-cryptocurrency-exchange/`

[28] `https://www.reuters.com/investigates/special-report/finance-crypto-currency-binance/`

2.5.2 AMM based exchanges

In order to mitigate the idle time issue of the order book model, automated market makers(AMM) are employed. They lie in the core of decentralized exchange by allowing automatic trading of crypto without the need for an intermediary. Instead of using an order book, in AMMs, the prices of the tokens are determined by a mathematical formula. The section below uses Uniswap as an example of AMM to explain how such exchanges work.

Uniswap

Uniswap is the first decentralized exchange based on AMM. It uses a constant product mathematical formula[7] to determine the prices of the assets. Unlike centralized exchanges, Uniswap uses liquidity pools to facilitate more efficient markets than the order book market makers[8]. Individuals act as "liquidity providers" and provide liquidity to the platform by adding a pair of tokens to the smart contract. People can then interact with the platform to buy and sell the tokens. Whenever a trade(buy and sell) happens, the amount of tokens in the pool also changes, which results in a change in the price. Uniswap does not require individuals to provide any KYC or additional fees to create an account. Additionally, no fees are required for the liquidity providers to list tokens on the platform.

In order to manage the prices and assets, Uniswap uses a Constant Product Market Maker formula. Figure 2.8 shows how the CPMM can be used to represent assets of a token pair in a liquidity pool. The quantity of assets/tokens(A and B) follow a mathematical formula:

$$x * y = k$$

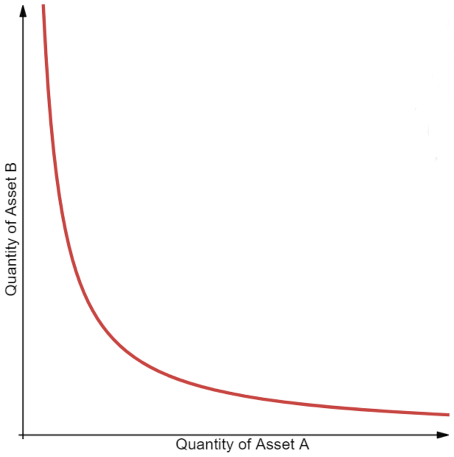


Figure 2.8: Constant Product Market Maker in Uniswap

where x is the quantity of asset A and y is the quantity of asset B in the pool. Uniswap pairs are represented as smart contracts, which are composed of the liquidity pool of the reserves of two ERC-20 tokens2.2.1. Figure 2.9 shows how this graph changes during the process of making a trade in the platform.

An example

Consider the initial state of the pool where the number of Token A is 1200, and that of Token B is 400. In this state, the price of Token B is $1200/400 = 3$. Now, a trader wants to trade his Token A to receive Token B using Uniswap. He inputs 3 Token A's and pays the 0.3% swap fee to the platform to receive 1 Token B based on the current price. In doing so, now, the total number of Token A in the pool will be $1200 + 3 + 0.03 = 1203.03$ whereas the total number of Token B will be $400 - 1 = 399$. It is represented in the lower section of the graph in figure 2.9. The new price of Token B in terms of Token A will now be $1203.03/399 = 3.015$. As we can notice, based on the constant product formula, the product of the number of tokens in the liquidity pool does not change, i.e.

$$1200 * 400 = 1203.03 * 399$$

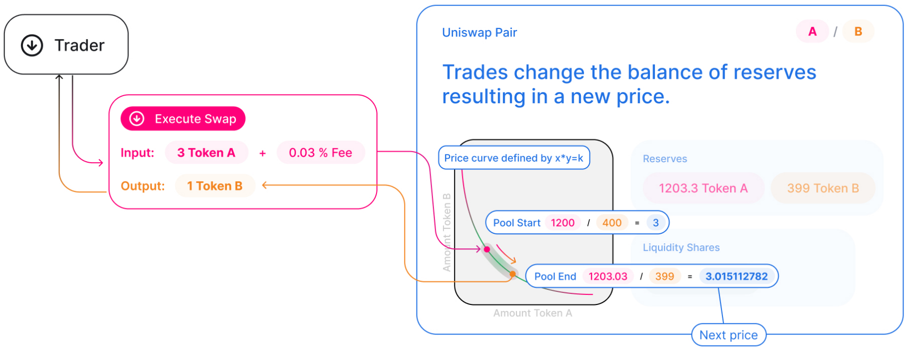

Trader
Execute Swap
Input: 3 Token A + 0.03 % Fee
Output: 1 Token B
Uniswap Pair
A / B
Trades change the balance of reserves resulting in a new price.
Price curve defined by x*y=k
Amount Token B
Amount Token A
Pool Start 1200 / 400 = 3
Pool End 1203.03 / 399 = 3.015112782
Reserves
1203.3 Token A
399 Token B
Liquidity Shares
Next price

Figure 2.9: Uniswap trading process[29]

Uniswap has changed its structure since its start with Uniswap V1 in 2018. Uniswap V2[9] launched in 2020 and solved several problems with V1, and now with Uniswap V3[10], a lot more flexibility has been added to the platform. A brief comparison between these versions is listed in table 2.2:

Decentralized vs Centralized Exchanges

Special Features of DEXes

DEXes bring some unique features that bring better utility in crypto markets compared to the CEX counterparts. Being a fully digital and autonomous mechanism certainly benefits the economy. Listed below are a few of such unique features.

[29] `https://docs.uniswap.org/protocol/V2/concepts/core-concepts/swaps`

Criteria	Uniswap V1	Uniswap V2	Uniswap V3
Liquidity	ETH-ERC20 possible, ERC20-ERC20 requires 2 swaps	ERC20-ERC20 swap possible	ERC20-ERC20 swap possible
Order types supported	Trades	Trades and Flash swaps	Trades, Flash swaps and range orders
Fees	0.3% of transaction value	0.3%(with a switch of sending 0.05% as a protocol fee)	Three tieres: 0.05%, 0.30%, 1.00%
Language	Vyper	Solidity	Solidity
Number of pairs	N/A^1	$75,736^2$	$6,891^3$
Total value locked	N/A^1	\$ $1.7B^2$	\$ $13B^3$
Total trading volume	N/A^1	\$ $484B^2$	\$ $587B^3$

1: Unable to find data for Uniswap V1
2: As of May 30, 2022, data was extracted from Uniswap V2 Subgraph(https://api.thegraph.com/subgraphs/name/ianlapham/uniswapv2)
3: As of May 30, 2022, data was extracted from Uniswap V3 Subgraph(https://api.thegraph.com/subgraphs/name/uniswap/uniswap-v3)

Table 2.2: Comparison between different versions of Uniswap

1. **Flash Swap**: Flash swaps are useful because they allow withdrawing the liquidity of any ERC-20 token from the pool without any cost, given that either of the following two conditions is met[30]:

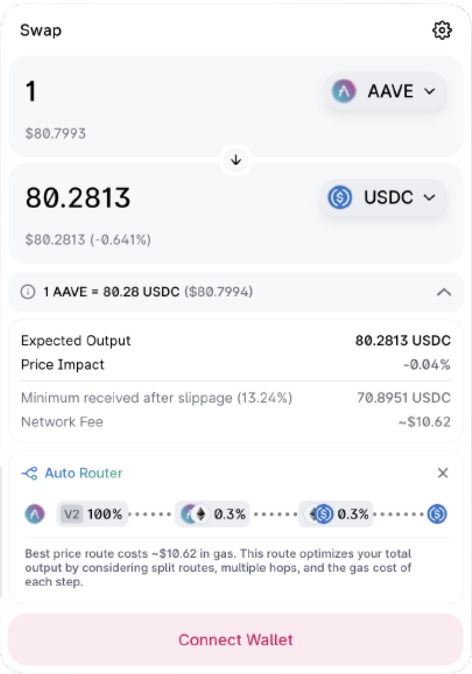
   a. pay for the withdrawn ERC20 tokens with the corresponding pair tokens
   b. return the withdrawn ERC20 tokens

   Failing to adhere to either of the two above conditions means that the transaction will fail and rollback. Flash swaps are atomic transactions, allowing for only complete transactions to occur. One use case of flash swaps is that it allows arbitrageurs to profit while also balancing the on-chain price with the outside market.
   Consider an example where ETH's price is \$1,200 on Kyber protocol and the price is \$1,000 on Uniswap. Now, using a flash swap, one can withdraw ETH on Uniswap, sell it to Kyber, return the original amount to Uniswap, all in one transaction making a profit of \$200 without investing any capital.

[30] https://docs.uniswap.org/protocol/V2/concepts/core-concepts/flash-swaps

2. **Price Oracles**: Price oracles are tools used to look up the price information of a given asset. Because multiple smart contracts talk to each other at a given time and the price of an asset is essential in such DeFi protocols, the problem of "what is the best way to retrieve the price of an asset on-chain?" is very relevant in this space.
   In order to solve this problem, several oracle designs have been implemented on Ethereum. It has enabled hackers to attack the oracle implementation[11]. Uniswap V2 added several improvements to deter such manipulations of public price feeds.

3. **Routers**: Exchanges like Uniswap use routers to determine the most efficient path of swaps in order to get the lowest slippage when there is no direct trading pair available. Additionally, Uniswap also uses an Auto Router that can optimize price for any swap by considering split routes, using gas cost awareness, and using a more robust algorithm that considers a more extensive data set for larger trades and better prices[31].
   An illustration of price while using an auto-router for swapping AAVE to USDC is shown in figure 2.10. At the given time, the router identifies that AAVE/ETH and ETH/USDC pools are the optimal intermediate pools used to convert AAVE to USDC.



Swap

1 AAVE

$80.7993

80.2813 USDC

$80.2813 (-0.641%)

1 AAVE = 80.28 USDC ($80.7994)

Expected Output	80.2813 USDC
Price Impact	-0.04%
Minimum received after slippage (13.24%)	70.8951 USDC
Network Fee	~$10.62

Auto Router

V2 100% 0.3% 0.3%

Best price route costs ~$10.62 in gas. This route optimizes your total output by considering split routes, multiple hops, and the gas cost of each step.

Connect Wallet

Figure 2.10: Using autorouter in Uniswap

[31] https://uniswap.org/blog/auto-router

DEX and DeFi terms

This section defines some of the major terms used when describing DEXes(and sometimes other DeFi projects).

- **Market cap**: Market cap of an asset is the total dollar value of all the coins in circulation. It is the product of the number of coins in circulation by the current market price of the coin. For example, at the point of writing, the amount of

- **Liquidity pool**: In a DEX, A liquidity pool is a store where providers deposit their assets to create trading pairs in the market and make it liquid for others who want to trade it. One of the liquidity pool in Uniswap is the ETH-USDT Pool, which is made of the ETH and USDT tokens.

- **Total volume locked(TVL)**: TVL is the total value of assets deposited in the decentralized exchange. It is the amount of funds(usually represented in USD) deposited by the liquidity providers to the exchange.

- **Trading volume**: Trading volume, usually expressed in USD, is the value of transactions that occured on the exchange over a period of time, e.g., 24 hours.

Pancakeswap

PancakeSwap started as a dApp of the BSC blockchain in September 2020. Its version 2, or PancakeSwap V2, is a fork of Uniswap V2, just like BSC is a fork from Ethereum, but it uses the PoSA consensus mechanism, unlike PoW in Ethereum. However, the previous topics covering Uniswap will still apply to Pancakeswap. Pancakeswap has already gained much popularity in a very small time duration[32]. The number of pairs it supports is very high compared to Uniswap, and because of its low trading fees, the trading volume has also already surpassed that of Uniswap. However, it is a bit difficult to get access to the data from Pancakeswap and BSC in general, although they advocate themselves as open source. Binance, a centralized crypto exchange, is behind BSC and Pancakeswap.

Problems with DEXes

While there are many positives of DEXes, they have their challenges. Millions of dollars have been compromised in these systems over the past few years[34][35]

[32] https://coinfomania.com/pancakeswap-overtakes-uniswap-bsc-explodes/

[33] This graph was taken from CoinmarketCap (https://coinmarketcap.com/currencies/ethereum/) at May 30, 2022: 11:07 UTC.

[34] https://www.theblockcrypto.com/linked/144491/stablecoin-dex-saddle-finance-hacked-for-10-million

[35] https://securityaffairs.co/wordpress/101895/cyber-crime/uniswap-lendf-me-hacked.html

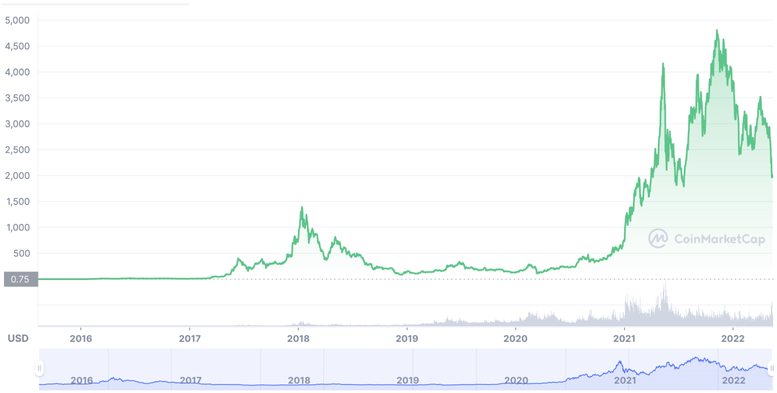


Figure 2.11: Price of 1 ETH in USD over time[33]

indicating that they are far from perfect. Most of these issues compromise the technical vulnerabilities of the blockchain or the smart contract itself. Some of such issues are described below.

- **Frontrunning**: Frontrunning is the technique of placing a specific transaction in the mempool with the inside knowledge of a transaction possibly happening in the future that is about to affect the price of an asset substantially. In Ethereum, the transactions in a block are ordered based on gas price and time by default. Exploiters can use bots to quote higher gas fees than a possibly higher value pending trade(that happens in the future), ensuring their transaction takes place earlier. Full node providers, usually miners, keep an eye on the network activities, so they have insider knowledge about the pending transactions.

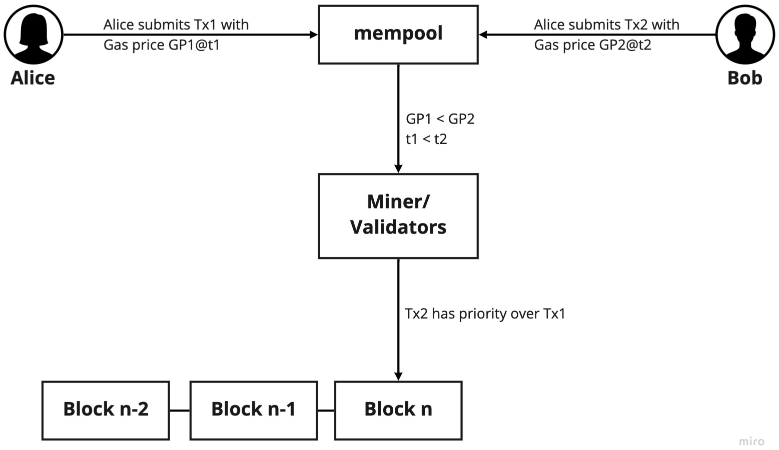


Figure 2.12: Illustration of frontrunning in blockchain

Figure 2.12 shows a simple illustration of how it works in a blockchain.

Here, an attacker(Bob) front-runs Alice and buys an asset at a lower price, although his transaction's gas fees are higher, thus getting to be in the queue earlier than Alice. Immediately after buying, Bob can then sell to Alice at a higher price. Thus, this negatively affects both the exchange and the user, Alice in this case.

- **Price Impact**[36]: Price impact is the affect a trade has on the market price of an asset after the trade is made in the underlying asset pair. It is the difference between the current market price and the expected fill price(or the price when the trade gets executed in real). Price impact is dependent on the amount of assets available in the pool, and can be very high for illiquid markets, resulting in heavy loss for the trader.
  Let us take a simple example of calculation of price impact in an ETH/USDC pool. Consider the pool has 2,000,000 USDC tokens and 1,000 ETH tokens. So, the initial market price of 1 ETH is 2,000 USDT. Using constant product formula, we get $k = 2,000,000,000$. Now, let's consider someone swaps 10,000 USDT of his tokens for ETH. Now, the number or USDT in the pool becomes 2,010,000, and using constant product formula, we get the number of ETHs as 995.024$(2,000,000,000/2,010,000)$. The trader thus receives 4.976 ETH$(1,000-995.024)$ at a price of 2009.64$(10,000/4.976)$ USDC per ETH. Thus, the price impact is 0.48%(9.64/2000).

- **Liquidity risks**: In any financial market, liquidity plays an important role. The liquidity in any exchange is usually tied to the trading volume. It is evident in DEXes, which have a comparatively low share of trading volume compared to the centralized exchange. The reason behind this is that DEXes are relatively new, and they are still comparatively less user-friendly.
  Given that the price and volume of transactions of a specific coin occurring in a DEX can not remain independent of the CEXes, it is usually a struggle for DEXes like Uniswap to keep up with CEXes like Binance. Although Uniswap has liquidity providers who operate in a decentralized manner, there are still high risks of losing the value of the assets because of impermanent loss. Impermanent loss means the change in the price of a pool's token at a later time compared to when they were deposited in the beginning. It is a common issue with crypto tokens because their prices are volatile, leading to a higher chance of impermanent loss. Figure 2.11 shows the price volatility of ETH since the beginning.

After providing an overview of how several components fit together in the crypto exchange space in this chapter, the upcoming chapter outlines the past works relevant to this study.

[36] `https://research.paradigm.xyz/amm-price-impact`

Chapter 3

Literature Review

With decentralized exchanges being a critical component of DeFi, several studies that scrutinize the quality of such exchanges have been done in the past. Most of these studies either compare them with the traditional exchanges or dive deep into the technical novelty and the vulnerabilities that come with it. This chapter gives a brief overview of the papers that cover the topics closely related to both decentralized exchanges and traditional exchanges while also trying to look into the data collection strategies used by them.

Over the last few years, some studies have closely observed Uniswap V1 till V3, critiquing the price volatility aspects of crypto in general, which opens up arbitrage opportunities. A study done by Berg et al. suggests that there were price inaccuracies in the DEXes, especially during the DeFi Summer of 2020[12]. They observe that although DEXes Uniswap and Sushiswap quickly adapt to such inaccuracies, they still struggle to track the reference prices of the market, resulting in chances for cyclic arbitrage. Similarly, Han et al. imply that the decentralized infrastructure built on blockchain and smart contracts can provide an alternative solution to cases where a consensus underwritten by a credible central party is not feasible or too costly to obtain. They do so using data from Binance and Uniswap and checking if investors on Binance and Uniswap trade in response to prices on the two exchanges. They find that it is indeed the cases[13]. Similarly, Wang et al. perform a systematic investigation on cyclic arbitrage in DEXes. Using the transaction-level data of Uniswap V2, they analyze the profitability conditions and optimal trading strategies for traders. They show that "traders have executed 292,606 cyclic arbitrages over eleven months and exploited more than 138 million USD in revenue." They question that the markets of DEXes may not be efficient enough because they allow for such massive arbitrage opportunities[14].

Another research category focuses on the vulnerabilities and possible attacks in decentralized exchanges. While such DEXes are somewhat less prone to being hacked in ways similar to the CEXes, several incidences have been noticed that cleverly compromise the technical weaknesses of DEXes. One of the significant vulnerabilities is the innate nature of blockchain: open and no-barrier-to-entry. A study done by Xia et al. has identified over 10K scam tokens listed on Uniswap, suggesting that roughly 50% of the tokens listed on Uniswap are scam tokens. They further estimate that scammers have gained a profit of at least $16 million from 39,762 potential victims. In their study, they believe that their approach can be used to identify and stop scam tokens from a DEX in their early stages[15]. Mazorra et al. use Uniswap V2 dataset of 20K tokens and propose a machine learning model to classify a token as a scam. They used the data from an archive Infura node and interacted with it using smart contract calls to get the relevant data for their analysis. They suggest that their methods can be used to categorize tokens as a scam or not scam not only after it performs malicious actions but also before. Several other studies have focused on detecting scams in the decentralized exchanges and found results that show that it has been one of the biggest struggles in this decentralized market[16]. Similarly, recent research by Tjiam et al. shows that smart contract vulnerabilities like transaction-ordering dependency and oracle manipulation have been exploited to extract hundreds of millions of dollars from those smart contracts. They do so by primarily focusing on the Uniswap smart contracts[17]. A very comprehensive data-driven study by Daian et al. shows that DEX arbitrage bots are employed hugely to make profits using techniques like frontrunning and transaction reordering. They claim that DEX design flaws threaten the underlying blockchain security, which might cause consensus-layer security threats in the blockchain[18]. In addition, there is more research available that delves into the technical vulnerability aspects of the blockchain facilitating bad actors to exploit them[19–21].

Taking a different approach from these empirical studies, some studies have also tried to compare DEXes with CEXes in terms of usability. An interesting study done by Zhou and Shen explores the user experience aspect of cryptocurrency exchanges. They observe a lack of understanding in this aspect among users, especially novice users. They argue that CEXes provide better usability and lower fees, due to which most of the end-users still go with centralized exchanges[22].

While most of the studies mentioned above are skeptical towards DEXes, there have been several studies that show the areas where DEXes are better suited to the actors in the market. Angeris et al. use formal analysis methods to constant product market makers to see how these markets must closely track the reference market prices. They also numerically demonstrate via large-scale agent-based simulation that Uniswap is "stable under a wide range of market conditions"[23]. Another study done by Lo and Medda concludes that Uniswap's simplicity enables liquidity providers and arbitrageurs to ensure the ratio of reserves matches the trading pair price[24]. Danos et al. formalize routing and arbitrage on DEX networks as convex optimization problems, empirically showing that such options can indeed solve the routing and arbitrage problems in Uniswap[25]. Similarly, Krishnamachari et al. show that by using dynamic AMM models, arbitrage opportunities can be mitigated, maintaining the pool price to be identical to the market price[26].

More closely related to our study was done by Lehar and Parlour, where they collected and analyzed all transactions of Uniswap V2(19M at the time) since its start. They compare the liquidity pool model used in Uniswap with the data collected from Binance and conclude that constant product models are more or less stable and, in some cases, more effective than order book-based models[27]. Similarly, the study done by Barbon and Randall investigates the quality of decentralized exchanges and compares them with their centralized counterparts based on two aspects: price efficiency and market liquidity. Using their comprehensive dataset, they conclude that while CEX provides better overall market quality, DEX is more competitive for high-volume transactions. They also propose a model that identifies quantitative conditions for DEX to overtake CEX in the future[28]. For price stability, both of them use the price data of ETH in USD over a 24H period. While Barbon and Randall claim that Binance is more stable than Uniswap based on their data, Lehar and Parlour prove that the claim is incomplete because the periodicity of the data was only one hour. Lehar and Parlour used data with about 9 seconds periodicity instead and show with this more comprehensive data that Uniswap's pricing is more stable than Binance's.

Additionally, the study by Aspris et al. also compares centralized and decentralized exchanges where they show that end-users tend to show a strong preference for deeper and more liquid markets provided by centralized exchanges compared to their decentralized counterparts[29].

Chapter 4

Methodology

This chapter describes in detail how the research process was carried out. At first, to validate the research questions, a literature survey was carried out. Based on the survey's outcome, the focus area of this research was determined. The sections below explain how the literature survey was carried out and then delineate the technical methods used to perform the analysis. The technical methods are divided into two parts: data and market analysis.

4.1 Literature Survey

This study delved into cryptocurrency exchanges with a focus on two aspects:

1. **Data collection**: It involved investigating how relevant data can be extracted from different blockchain as well as non-blockchain based exchanges
2. **Market**: It involved looking into how the exchanges manage their markets and co-ordinate with the actors involved in the platforms

While the second aspect has been probed in a few research papers as covered in detail in chapter 3, there was an evident lack of research resources available in the first aspect, i.e., methods available to extract quality data from the cryptocurrency exchanges. To support this claim, a systematic literature review was conducted that went through research papers published in the major publications to find detailed methods of obtaining data and inquire about their quality.

In order to look into the data aspect of crypto exchanges, first top academic publishers were queried. Specifically, ACM, IEEE, Springer, and ScienceDirect were the databases used, and the following query was used to get the relevant results until April 15, 2022:

"*uniswap*" *AND* "*data*"

The reason behind using "uniswap" as one of the keywords is that Uniswap is the first decentralized exchange, and most of the research investigating decentralized exchanges has to mention Uniswap one way or the other. It is also the right exchange candidate in terms of data collection because it ticks all the methods possible to carry out the task, as shown in table 5.8.

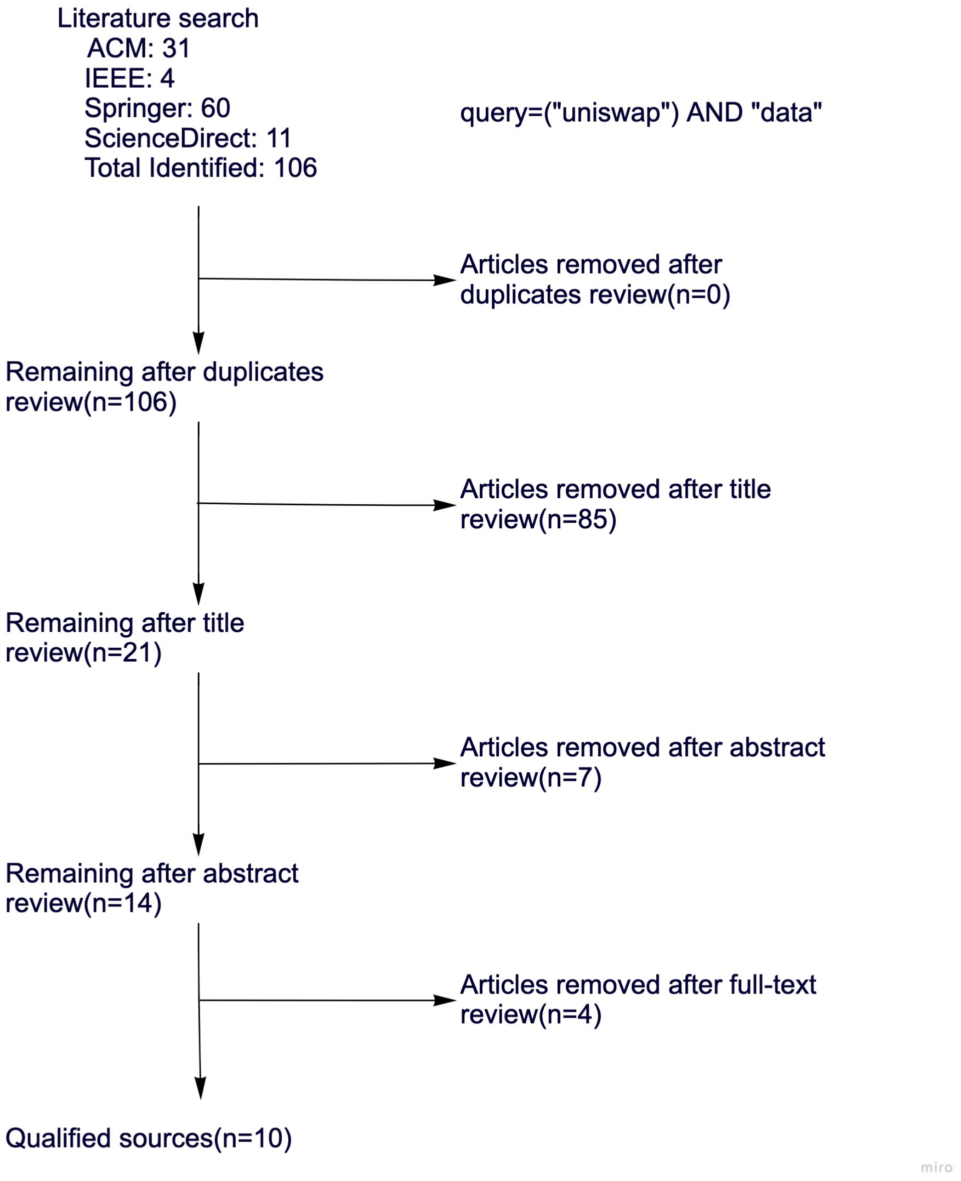

Literature search
ACM: 31
IEEE: 4
Springer: 60
ScienceDirect: 11
Total Identified: 106
query=("uniswap") AND "data"
Articles removed after duplicates review(n=0)
Remaining after duplicates review(n=106)
Articles removed after title review(n=85)
Remaining after title review(n=21)
Articles removed after abstract review(n=7)
Remaining after abstract review(n=14)
Articles removed after full-text review(n=4)
Qualified sources(n=10)
miro

Figure 4.1: Systematic Literature Review in top academic publications

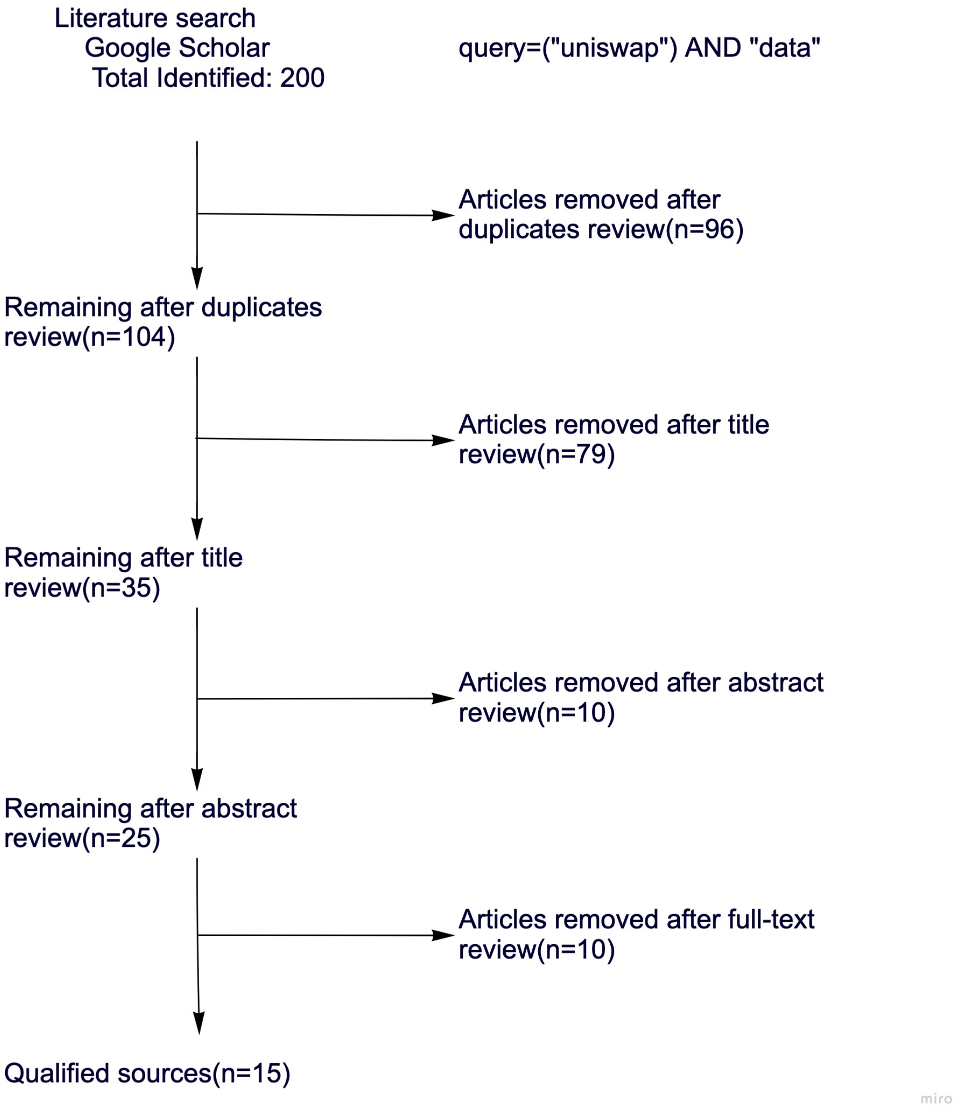

Literature search
Google Scholar
Total Identified: 200
query=("uniswap") AND "data"
Articles removed after duplicates review(n=96)
Remaining after duplicates review(n=104)
Articles removed after title review(n=79)
Remaining after title review(n=35)
Articles removed after abstract review(n=10)
Remaining after abstract review(n=25)
Articles removed after full-text review(n=10)
Qualified sources(n=15)
miro

Figure 4.2: Systematic Literature Review in Google Scholar

In order to be more aligned with the current state of research in this space, another literature search was done with Google Scholar. Google Scholar is a better fit for this study because it covers non-peer-reviewed and non-academic publications like blog posts and whitepapers. It would give an idea about how the field of blockchain is still in the process of being more systematic. In Google Scholar, the search was limited to the first 200 results because the results later started to significantly diverge from the expected field of study.

In total, 200 search results from Google Scholar and 106 results from the combined four publications were used to perform a systematic literature review as shown in figures 4.1 and 4.2. In Google Scholar, most of the relevant results were from non-peer-reviewed archives like arxiv.org[1]. Another crucial point is that these publications are relatively new, with more than 90% published after 2021. It proves that most of the research in this field is still evolving.

The search results were further filtered based on the title and abstract review, and the qualified papers were studied to check their relevance to this study. In the last filter, which required going through the full text of the articles, the same two aspects as mentioned in 4.1 were sought out. Ultimately, the total corpus of articles, including Google Scholar and the other four publications, and the number of qualified sources in terms of data were 25. Most of these identified relevant studies relied on third-party data sources to perform their analyses. A handful of papers(3) used their node to obtain the blockchain data, and some of them even used proprietary data to perform their analyses(1).

So, in light of the conditions above, a comprehensive review of the possibilities and limitations of obtaining data from these exchanges was performed. Additionally, with the obtained data, some analyses were performed that cover the exchanges' market/economy aspects. The detailed findings obtained from the literature survey are discussed in chapter 3.

4.2 Technical Implementation

4.2.1 Data Collection

On the AMM side, Ethereum and bsc were the blockchains of interest, and both of them provide ways to interact with the blockchain. On the order book side, the official API of Binance was used to obtain data from the exchange. On a broader level, the options to collect data fell under two main categories: using in-house resources or third-party resources.

Using in-house resources

The required resources to collect data from Ethereum were provided by the Decentralized and Systems Engineering (DSE) Lab at NTNU[2]. These were the specifications of the resources used during the process:

The initial plan was to use these resources to capture Uniswap(using Ethereum node) and PancakeSwap(using BSC node) data. However, the documentation and links provided for BSC to run a full(or archive) node did not work. So, the data collection for PancakeSwap was limited to only a few free third-party resources.

[1] arXiv(`https://arxiv.org`) is an open-access repository of electronic preprints and postprints (known as e-prints) approved for posting after moderation, but not peer review.

[2] Decentralised Systems Engineering Lab(DSE) is a part of Norwegian University of Science and Technology(NTNU) that focuses on research in fields like security, privacy, decentralization, artificial intelligence (AI), machine learning (ML), and big data analytics and combination and interaction of those various technologies `https://www.ntnu.edu/idi/dse`

Item	Specs
RAM	1.5TB
CPU	64-core AMD EPYC 7713 with 64 hardware threads each
Disk	4 x 3.5TB RAID-0 nvme drives

Table 4.1: Specification of resources used

Consequently, the DSE lab resources were used to set up an archive node for Ethereum, and clients were used to interact with it to get the data for Uniswap V2.

The Ethereum full-node blockchain was of size 800GB, while the archive node was about 10TB at writing. Because an archive node stores all the states of all smart contracts from the beginning (including their historical states and logs), it was better suited than the full node in this study. On the other hand, the full node stores the states and logs of the smart contracts of only a few recent blocks in the Merkle trees. However, it stores all the transactions that occurred in the blockchain, providing a way to replay all the transactions through a smart contract and storing the relevant data separately in our storage.

The storage issue was the main challenge while using our node to get the data. Initially, we started syncing with the Ethereum full node. The complete sync took about 800GB and was done within a week, which was manageable with available resources. However, it was later realized that we needed the archive node instead to get all the historical transactions in Ethereum, because of which the process had to be restarted from scratch. An archive node takes much time to sync to the latest block of Ethereum because of its massive size. So, because of time limitations, we could only sync up to June 2021. So, the data collection was limited up to this date.

Replaying historical transactions using full node

In order to obtain data from historical transactions via smart contracts in the Ethereum blockchain, the following resources are needed:

1. Storage space of about 800GB
2. Additional storage space to store the transactions(and other data) of the specific smart contract(This could be a database depending on where the data is to be stored)
3. A client to interact with the full node
4. The block number(N) in which the first transaction of a smart contract A occurred (as shown by the purple block in figure 4.3.

With the resources mentioned above, the steps needed to get relevant transactions and state data from the smart contracts are outlined in algorithm 1. The algorithm gets the transactions and state outputs from the Uniswap V2 router

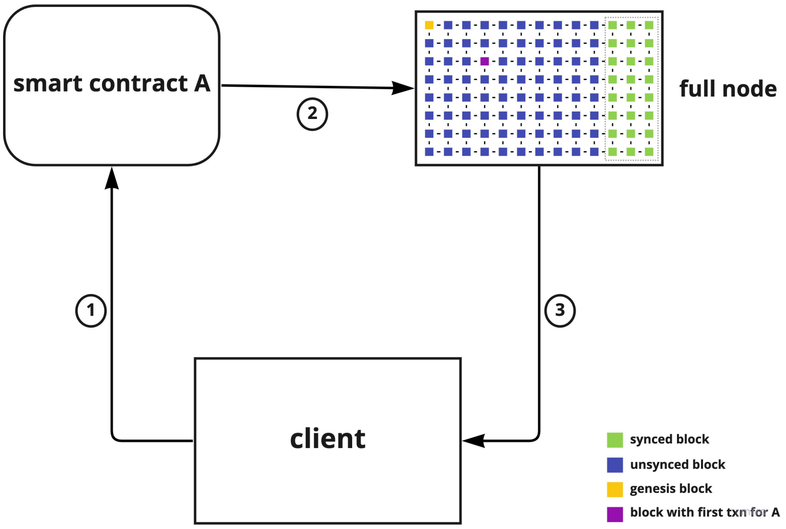


Figure 4.3: Collection of historical data for smart contract A using full node

smart contract, which is the one responsible for making all the vital exchange transactions in Uniswap V2. At first, the first block containing the transactions of Uniswap V2 Router is identified. And from there till the latest block in Ethereum, the transactions of the router are filtered. On identifying the transaction of a block that belongs to the router, the details of the transaction like outgoing/incoming addresses, amount sent, etc. are passed to the relevant smart contract method of the router. The method can then produce the relevant output based on the inputs provided. This output is then validated with other peers connected to the full-node. Once we get the same result from the peers, we know that the output is valid, and we can store the output in our own storage.

Using third party resources

For Binance, the official Binance API was used to collect their historical and real-time data. BSC has several JSON RPC URLs[3][4] ready to use to interact with the blockchain. However, because none of them were reliable and using an in-house setup to archive the full node from the beginning was out of the scope of this research because of resource constraints, it was decided that the following tools were used to collect the data until the resources in the DSE lab were available:

- **Infura**[5]: Infura is a web3 backend and an Infrastructure-as-a-service(IAAS) that provides services and tools for blockchain developers. The main service provided by Infura is API access to the Ethereum network, both mainnet, and testnet. In other words, it provides access to enterprise-ready cloud-

[3] `https://docs.binance.org/smart-chain/developer/rpc.html`

[4] `https://bsc.streamingfast.io/subgraphs/name/pancakeswap/exchange-v2/graphql`

[5] `https://infura.io`

hosted Ethereum nodes. In this study, Infura was used as a node provider for Ethereum because it was straightforward to set up and test. The figure compares throughput using an Infura Ethereum node vs. a local Ethereum node. As expected, because Infura is cloud-hosted, it is comparatively slower than having the resources on-premise.

- **The Graph**[6]: The Graph is an open-source indexing protocol used to collect, store and process data from various blockchains. It allows anyone to host easily accessible subgraphs, which are open-source APIs.
  Figure 4.4 shows the basic architecture of how data flows within several components of the subgraph of the Graph. The dApp adds data to Ethereum or other blockchains[7] through a transaction on a smart contract. This smart contract then emits events during the processing of the transaction. The Graph Node, which is at the center, scans Ethereum continuously for new blocks and queries for new data that the subgraph might contain. When the Graph node finds Ethereum events for the subgraph, they are mapped into data entities using the WebAssembly(WASM)[8] module and then stored in the Graph Nodes. Finally, a GraphQL endpoint is provided for the dApp to query the indexed data from the Graph Node. One of the examples dApp which uses such a subgraph under the hood is the official Uniswap Info website[9].
  Using The Graph helped to get the essential aggregate data for Uniswap, like its daily trading volume, pairs, and tokens. The Graph also added inspiration to the design of the schema for the data collected in this study. The Graph is very comprehensive and stores almost all necessary fields needed in the blockchain data.
  Some of the relevant schema for this study for Uniswap V2 are presented in tables 4.2, 4.4, 4.5 and 4.3[11].

[6] `https://thegraph.com`

[7] The Graph subgraphs can be created cross-chain, meaning multiple blockchains are supported, including mainnets and testnets

[8] `https://webassembly.org/`

[9] `https://v2.info.uniswap.org/`

[10] Image taken from: `https://thegraph.com/docs/en/about/introduction/`

[11] `https://thegraph.com/hosted-service/subgraph/uniswap/uniswap-v2`

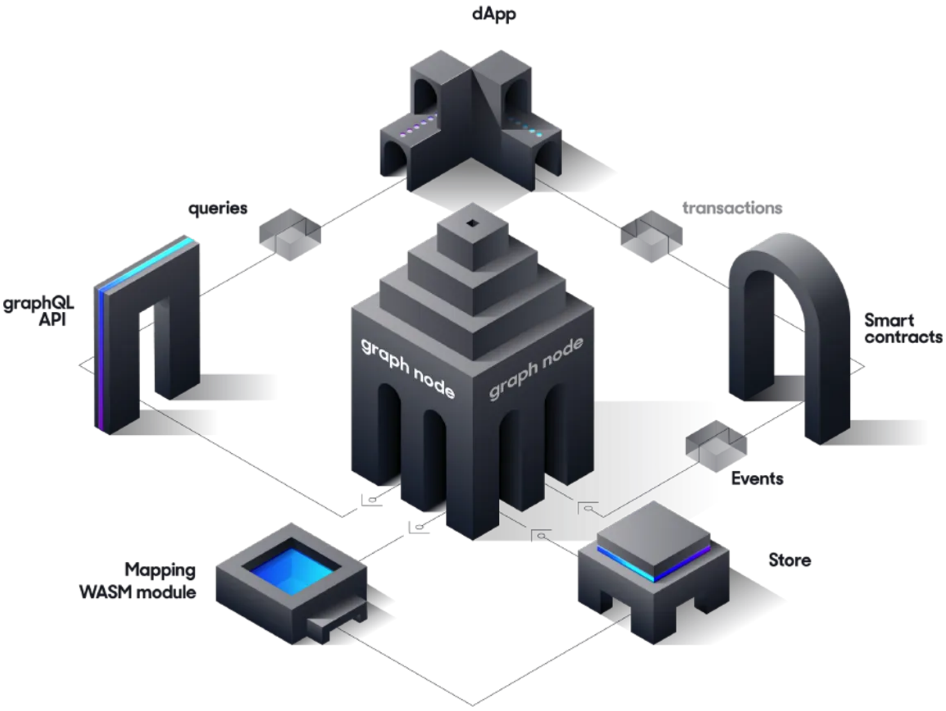


Figure 4.4: Architecture of The Graph[10]

Field	Data type	Mandatory
id	ID	Yes
symbol	String	Yes
name	String	Yes
decimals	BigInt	Yes
totalSupply	BigInt	Yes
tradeVolume	BigDecimal	Yes
tradeVolumeUSD	BigDecimal	Yes
untrackedVolumeUSD	BigDecimal	Yes
txCount	BigInt	Yes
totalLiquidity	BigDecimal	Yes
derivedETH	BigDecimal	No

Table 4.2: The Graph API Uniswap Token Entity

Field	Data type	Mandatory
id	ID	Yes
date	Int	Yes
dailyVolumeETH	BigDecimal	Yes
dailyVolumeUSD	BigDecimal	Yes
dailyVolumeUntracked	BigDecimal	Yes
totalVolumeETH	BigDecimal	Yes
totalLiquidityETH	BigDecimal	Yes
totalVolumeUSD	BigDecimal	Yes
totalLiquidityUSD	BigDecimal	Yes
maxStored	Int	No
txCount	BigInt	Yes

Table 4.3: The Graph API UniswapDayData Entity

Field	Data type	Mandatory
id	ID	Yes
token0	Token	Yes
token1	Token	Yes
reserve0	BigDecimal	Yes
reserve1	BigDecimal	Yes
totalSupply	BigDecimal	Yes
reserveETH	BigDecimal	Yes
reserveUSD	BigDecimal	Yes
trackedReserveETH	BigDecimal	Yes
token0Price	BigDecimal	Yes
token1Price	BigDecimal	Yes
volumeToken0	BigDecimal	Yes
volumeToken1	BigDecimal	Yes
volumeUSD	BigDecimal	Yes
untrackedVolumeUSD	BigDecimal	Yes
txCount	BigInt	Yes
createdAtTimestamp	BigInt	Yes
createdAtBlockNumber	BigInt	Yes
liquidityProviderCount	BigInt	Yes

Table 4.4: The Graph API Pair Entity

Field	Data type	Mandatory
id	ID	Yes
blockNumber	BigInt	Yes
timestamp	BigInt	Yes
mints	[Mint]	Yes
burns	[Burn]	Yes
swaps	[Swap]	Yes

Table 4.5: The Graph API Transaction Entity

- **Moralis**[12]: Moralis is an alternative to Infura, but it not only provides APIs for Ethereum but also operates cross-chain providing support for blockchains like BSC and Polygon/Matic. In this study, it was a good choice for accessing PancakeSwap because other options tried out for BSC were not reliable. Apart from serving as Blockchain-as-a-Service(BaaS), Moralis provides a complete set of tools like Software Development Kits(SDK) and user authentication tools that make it easy to build dApps.

- **Dune Analytics**[13]: Dune Analytics is an on-chain analytics platform that allows users to turn blockchain data into actionable charts and metrics. Dune supports multiple blockchains, including Ethereum, Polygon, Binance Smart Chain, Optimism, and Gnosis Chain. It picks up the internal calls and events from these blockchains but does not have the state/storage data[14]. Figure 4.5 shows how the entire system works. The smart contract events and calls are indexed into a PostgreSQL database, which can then be queried either by using Dune's dashboard[15] to get the results.
  Using Dune Analytics helped quickly test some hypotheses about the decentralized exchanges.

[12] `https://moralis.com`
[13] `https://dune.com`
[14] `https://docs.dune.com/`
[15] `https://dune.com/browse/dashboards`

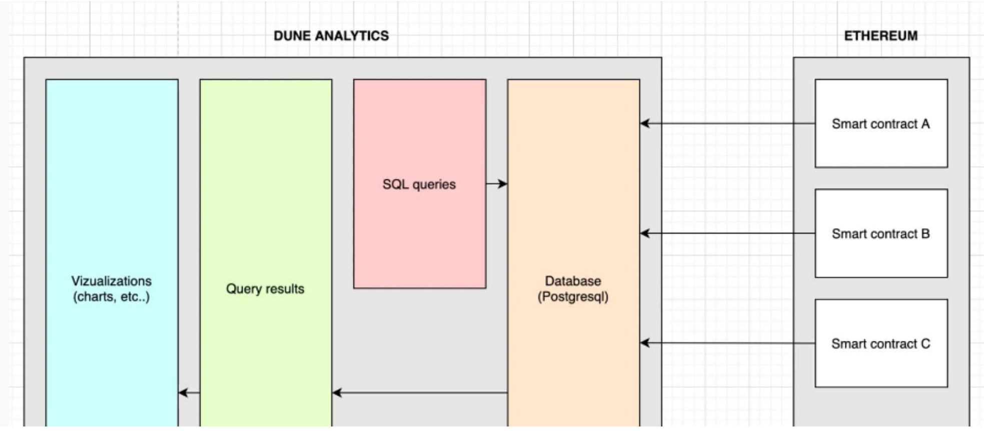


Figure 4.5: Architecture of Dune Analytics[16]

- **Uniswap[17] & PancakeSwap[18] web interface**: Apart from the third-party tools mentioned above, some data was collected from the official web interface of Uniswap and PancakeSwap. It was done to test whether the real-time prices shown in the web interface match the ones produced from the smart contracts. The following data was collected:
  - Price of ETH in terms of USDT
  - Possible gas fees incurred during the price calculation

A summary of how the first three tools mentioned above differ from each other is included in section 5.3.

4.2.2 Choice of tools

After experimenting with all the methods described above, the following choices were made to collect data for this study. Some of these data were stored in a local database depending on the nature of the analyses to be performed.

1. **Transaction data**: Smart contract calls, The Graph
2. **Pricing data**: Smart contract calls, Scraping
3. **Pair data**: Smart contract calls, The Graph
4. **Data Storage**: PostgreSQL
5. **Exchange Aggregate data**: The Graph API and Pancakeswap Graph API
6. **Binance data**: Binance official API[19]
7. **Visualizations**: Matplotlib, Dune Analytics
8. **Block explorers**: Etherescan[20] for Ethereum and BSCScan[21] for BSC

[16] Image taken from: https://academy.moralis.io/blog/defi-deep-dive-exploring-dune-analytics

[17] https://app.uniswap.org/?use=v2#/swap?chain=mainnet

[18] https://pancakeswap.finance/swap

[19] https://github.com/binance-exchange/binance-official-api-docs/blob/master/rest-api.md

[20] https://etherscan.io

[21] https://bscscan.com

4.2.3 Data Description

The data extracted from blockchain-based exchanges and Binance was stored in a relational database for this study. It was done because it was convenient for the study as only a few entities had to be stored. This model is similar to what Dune analytics has under the hood. However, the tables and their fields were inspired from the schema of The Graph which is included in sections 4.2, 4.3, 4.4, and 4.5. Below is the description of tables used for the database during this study:

1. **Token**: This table stores the data related to a token listed in the blockchain or Binance.
2. **Pair**: This table stores the trading pairs listed in the exchanges
3. **Price**: This includes the price of a specific token in terms of another in several exchanges. For this study, the pair of ETH wrt. USDT was extracted and stored.
4. **Transaction**: The transaction table includes the details of a transaction that happened in the blockchain.

One vital point to notice is that the fields in these tables are very sparse compared to The Graph. We had the option to model the fields and tables according to what we needed and discard other information from the exchanges. Tables and fields of exchange like Uniswap V2 in The Graph, on the other hand, are very comprehensive and try to include almost all of the possible data.

Field	Required	Type	Description
address	Yes	String	
symbol	Yes	String	
blockchain	Yes	String	
decimals	Yes	Integer	

Table 4.6: Schema of Token

Field	Required	Type	Description
token0	Yes	Token	
token1	Yes	Token	
pair_address	Yes	String	
created_at	Yes	String	

Table 4.7: Schema of Pair

4.3 Market Analysis

Using the tools and techniques from the experiments above, the data collected was used to analyze the exchanges' market quantitatively. While the central focus

Field	Required	Type	Description
price	Yes	String	
source	Yes	String	
ts	Yes	String	
token0	Yes	Token	
token1	Yes	Token	

Table 4.8: Schema of Price

Field	Required	Type	Description
blockchain	Yes	String	
blockNumber	Yes	Integer	
from	Yes	String	
to	Yes	String	
gas	No	Integer	
gasPrice	No	Integer	
hash	Yes	String	
value	Yes	String	

Table 4.9: Schema of Transaction

of this study is on exploring data quality from the exchanges, some experiments were performed with the data collected from several sources to validate the results and compare them with past research. Specifically, the analysis focused on four aspects of exchanges. Data collection for all these aspects used the Binance Official API for Binance's data, while for Ethereum and BSC, several options were needed, which are discussed below.

4.3.1 Transactions

A transaction in a crypto exchange usually includes the information about the buy and sell activity occuring through the exchange. By just looking at the transaction details over a certain time range, it is possible to understand the pattern of influx and outflux in the platform.

In order to collect transaction data, the Infura node and a local node were used for Uniswap V2 data(in Ethereum Mainnet). In contrast, the Moralis node was used for PancakeSwap data(in BSC). While the initial plan was to collect all the Uniswap V2 and Pancakeswap to date via a locally hosted archive node, the time and space requirements were underestimated. Because of time constraints, it was later decided that transactions that occurred during some specific periods would be stored. Specifically, transactions during the DeFi Summer of May 2021 were extracted. Picking specific time ranges in the past is helpful because it validates the results with the actual events that occurred during that time range. E.g., let us say someone converts a massive amount of ETH to USDT using Uniswap. This activity can easily be traced to the account who did it and what volumes of assets were moved.

Below is the algorithm used to extract transaction data from the blockchain node:

4.3.2 Pricing

Another aspect of a cryptocurrency exchange is how it manages the prices of its assets. While with CEXes, we are limited with the type and quality of data they provide through their APIs, in DEXes, we can trace back and collect all the pricing data they ever published during their entire existence. Blockchain-based DEXes usually store the end-of-block price in the blockchain, making it easy to extract them back using gasless smart contract calls. It was the option we opted for in this study. Uniswap V2 has the Router02 smart contract with the method ‘getAmountsOut‘ with a signature shown in listing 4.1.

’getAmountsOut’ takes two parameters(amountIn is the amount of tokens of tokenA to be exchanged, and address is the address of the exchange pair tokenA-tokenB in Uniswap V2) and returns the maximum amount of tokenB possible during a trade. The valuable point to note about this smart contract method call is that it can be made to get the output of the amount of tokenB both at the given instant and in the past. These smart contract outputs are stored for all blocks from the beginning only by an archive node.

Apart from getting the prices from smart contract calls, they were also extracted from the official web interface of Uniswap and PancakeSwap. It was done to check the discrepancies in prices from the two sources, if any. This price data was extracted for a 1H period in Uniswap V2 and PancakeSwap V2 and compared with Binance. Doing so gives a sense of how such an exchange reacts to certain events when a token price fluctuates. The results are discussed in section 5.2.1.

```
function getAmountsOut(uint amountIn, address[] memory path) internal view returns
    (uint[] memory amounts);
```

Code listing 4.1: Signature of the getAmountsOut method in Uniswap Router02

Outlined in pseudocodes 3 and 4 are the steps followed to store pricing data with the frequency of the blockchain’s block time. In both the algorithms, smart contract calls are made to the GetPriceOfPair() method of smart contract that takes the pair address, amount of input tokens, and the block number. The specialty of this method is that it can trace back to the old blocks so that any block number can be passed to it, and it returns the price of the pair at that instant.

4.3.3 Transaction fees

People choose to go with a particular exchange based on its transaction fees. While there are several factors to a transaction fee ranging from gas fees(if the blockchain is PoW based), to additional fees for the exchange, to protocol fees, people usually care about how much they receive when they want to exchange. Keeping this aspect in mind, transaction fees over the three exchanges for different volumes of trades were compared.

In order to get the transaction fees, for Binance, Uniswap V2, and PancakeSwap V2, their respective web portal were used. The exchange was made with USDT-ETH pair in all these exchanges for different volumes of USDT. For Uniswap, the gas fees were also noted, along with their respective trading fees(0.3% for Uniswap V2 and 0.25% for Pancakeswap V2, 0.1% for Binance). However, in PancakeSwap, although gas fees incur, the web interface doesn't show the gas fees for the trades happening. So, only the final worth of ETH was noted.

4.3.4 Trading Pairs

Finally, this study also examines how the exchanges manage the trading pairs in their platforms. Uniswap V2 and PancakeSwap were tracked over 3 hours to see how many new trading pairs were added to the platforms. Doing something similar in the Binance exchange was impossible because the API endpoints subscribe to new trading pairs and get their data like creation time, total transactions, and trading volume. These numbers were not possible to find for Binance, so the study was limited to only Uniswap and Pancakeswap in this aspect.

The collected pairs data was then used to compare the frequency of a new pair generated into the platform. It is an essential factor to consider because PancakeSwap, which started quite later than Uniswap, already has about twice as many trading pairs as Uniswap.

In order to analyze the pairs in the system, the tokens involved in the pairs were studied. It was done using the primary DeFi metrics like Trading Volume, Number of Transactions, and Total Volume Locked in these pairs. These metrics were checked about a week later after the pairs were added to the platforms because it is usually expected that there should be some activity involving these trading pairs by then. Additionally, there is a growing trend of shitcoins[22] among the cryptocurrency communities. Using the names of the tokens involved in these trading pairs, an attempt was made to check if it is possible to spot test tokens or shitcoins in the platform of PancakeSwap. The result obtained from the word cloud generated from it is included in figure 5.10.

[22] Shitcoins are cryptocurrencies with no apparent function or meme tokens. They are potentially undervalued projects or those with a low market cap of fewer than 1 billion dollars

Algorithm 1 An algorithm to replay the all historical smart contract states using a full node

```
procedure GETTXNSINBLOCK($N$)    ▷ Method to get all transactions from a block
    return txns
end procedure

procedure GETLATESETBLOCK    ▷ Method to the latest block number from the blockchain
    return $N$
end procedure

procedure SCMETHOD($txnDetails$)    ▷ Method of a deployed smart contract which is executed in the blockchain
    return output
end procedure

$txns \leftarrow []$
$scoutputs \leftarrow []$
$s \leftarrow' 0x7a250d5630B4cF539739dF2C5dAcb4c659F2488D'$    ▷ Address of Uniswap V2 router
$N1 \leftarrow 12474190$    ▷ Beginning block number of smart contract (This can be found from tools like Etherescan)
$N2 \leftarrow$ GETLATESTBLOCK()
while $N1 \neq N2$ do
    $ts \leftarrow$ GETTXNSINBLOCK($N1$)
    for t in ts do
        $to \leftarrow t.to$
        $from \leftarrow t.from$
        if $from$ =s or $to = s$ then
            $output \leftarrow$ SCMETHOD($t.details$)    ▷ Run the Smart contract method locally
            $txns.insert(t)$
            $scoutputs.insert(output)$
        end if
    end for
    $N1 \leftarrow N1 + 1$
end while
```

Algorithm 2 An algorithm to capture and store transactions from a smart contract

```
procedure GetTxnsInBlock($N$)    ▷ Method to get all transactions from a block
    return txns
end procedure
$txns \leftarrow []$
$s \leftarrow' 0x7a250d5630B4cF539739dF2C5dAcb4c659F2488D'$    ▷ Address of Uniswap V2 router
$N1 \leftarrow 12474190$    ▷ Beginning block number
$N2 \leftarrow 12484190$    ▷ End block number
while $N1 \neq N2$ do
    $ts \leftarrow$ GetTxnsInBlock($N1$)
    for t in ts do
        $to \leftarrow t.to$
        $from \leftarrow t.from$
        if $from = sorto = s$ then
            $txns.insert$(t)
        end if
    end for
    $N1 \leftarrow N1 + 1$
end while
```

Algorithm 3 An algorithm to extract old price data from the blockchain based exchanges for a range of blocks

```
procedure GetPriceOfPair($p$, $a$, $block$)    ▷ Smart contract method that returns transactions from a block
    return outTokens
end procedure
$prices \leftarrow []$
$ETHUSDTPAIR \leftarrow' 0xa478c2975ab1ea89e8196811f51a7b7ade33eb11'$
$N1 \leftarrow 12474190$    ▷ Beginning block number
$N2 \leftarrow 12484190$    ▷ End block number
while $N1 \neq N2$ do
    $price \leftarrow$ GetPriceOfPair($ETHUSDTPAIR, 1, N1$)
    $prices.insert$(price)
    $N1 \leftarrow N1 + 1$
end while
```

Algorithm 4 An algorithm to extract real-time prices from blockchain based exchanges

```
procedure GetPriceOfPair($p$, $a$, $block$)    ▷ Smart contract method that returns price of a pair
    return $price$
end procedure
$prices \leftarrow [\,]$
$ETHUSDTPAIR \leftarrow' 0xa478c2975ab1ea89e8196811f51a7b7ade33eb11'$
while $True$ do
    subscribe for new blocks in the blockchain
    if new_block then
        $block \leftarrow$ block_from_subscription
        $price \leftarrow$ GetPriceOfPair($ETHUSDTPAIR, 1, block.number$)
    end if
    $prices.insert$(price)
end while
```

Chapter 5

Results and Discussion

As mentioned in section 4.1, there is a dearth of research on cryptocurrency exchanges in terms of collecting data to perform the analysis. Apart from the literature mentioned in this study, another point to note is that most of the other studied resources fall under blogs or non-peer-reviewed publications. Majority of researchers relied on third-party data sources like The Graph and Etherscan[1] which allow developers to focus on the analysis without having to spend time and resources for data collection. It, however, has led to drawbacks in some research. Barbon et al., for instance, used The Graph as the source of Uniswap V2 data but had limitations with the frequency of time interval of how they collected the price of a token. Lehar et al., on the other hand, used their own resources to collect data from Uniswap V2 and were able to bring the frequency down to a median of 15 seconds. In this context, our research tends to be the missing link between these two: act as an easily followable methodology to extract data from blockchains using smart contracts. Algorithm 4 shows how using our resources; we can leverage this and store pricing data more frequently in our database. It is a definite improvement in the capability of The Graph.

Using our resources is also not the ultimate solution. While Uniswap is comparatively open and easy to obtain data, the Ethereum archive node takes a massive amount of storage(10 TB at present) and computation resources to obtain all the data. It is why most researchers do not opt to go with this approach. Additionally, in BSC, although they mention that they are open and transparent, their graph or node URLs do not usually work. Researchers have to rely on paid resources to access their blockchain data. This study also ended up using the free version of Moralis for the same reason. The same goes for the order book exchange Binance. They have their API, but it is very limited in what can be obtained from it. Barbon et al. mention in their study that they use proprietary level data from Binance for their analysis[28].

Going back to Research Question 1, it can therefore be concluded that data

[1] https://etherscan.io/

collection from cryptocurrency exchanges is still very new. Given the amount of research we see happening recently, it can be expected that the data aspect of such exchanges will get better in the coming years.

5.1 Data

This section shows how the different experiments performed for data collection compared with each other. Table 5.1 shows the latency comparison when using a self-hosted local node vs. an Infura node for downloading 10,000 transactions from the Ethereum full node. Infura does not provide free access to an archive node, so only the full node was tested. Additionally, table 5.2 shows the latency comparison for the extraction of prices from archive nodes for Binance Smart Chain and Ethereum. Because we were able to set up an archive node locally for Ethereum but not for Binance because some of their official URLs did not work, we had to use third-party service Moralis for BSC's archive node. The tables clarify that having the node on-premise is faster and eases the data collection process. However, there is a trade-off between these two approaches, and one should decide which one to use based on the requirements. Self-hosted on-premise nodes allow great flexibility and are very quick but require the programmer to do the heavy lifting with all the setups before getting the data. Third-party nodes, on the other hand, make the data collection process more straightforward by handling the infrastructure themselves. However, in addition to being slower, they might sometimes not be able to provide the data required. In such a situation, the system of such tools needs to be modified internally, which is not in the user's control.

Using the on-premise resources for the archive node,

This aspect was also brought up in Research Question 2. As mentioned above, the tradeoff of using on-premise and third-party tools is lies in flexibility vs overhead work. It is, therefore, not surprising to see that only 3 out of 25 research papers mentioned using their self-hosted nodes to collect data.

Criteria	Self-hosted node	Infura node
Number of transactions	10,000	10,000
Time taken	4min 45s	20m

Table 5.1: Throughput comparison between self-hosted node vs third-party node for transaction data

Table 5.3 shows the summarized comparison between four third-party tools tried out during this study. The results show that each has its limitations, although they serve very well for the required tasks in most cases. One crucial point to notice is that only The Graph is open source, so developers have the option to modify the codebase and make it work according to their needs. On the other hand, Dune Analytics is relatively inflexible in this regard, i.e., its tables and entities are fixed,

Criteria	Self-hosted node	Moralis node
Number of prices	10,000	10,000
Time taken	5 min	55min

Table 5.2: Throughput comparison between self-hosted node vs third-party node for pricing data

and if something new is to be done, it is not possible without the maintainers making changes to their system. Additionally, Dune Analytics also doesn't offer an API to call from external services, requiring that the queries be input only in it's web dashboard. This is a big limitation for someone who might need to run scripts to get the data from multiple sources of Dune at the same time.

Criteria	**The Graph**	**Moralis**	**Dune Analytics**	**Infura**
Pricing per month	N/A^*	$49	$390	$50
Open source	✓	x	x	x
Blockchains supported	$multiple^1$	$multiple^2$	$multiple^3$	$Ethereum^4$
Database	Graph	NoSQL (MongoDB)	PostgreSQL	-
Historical data	✓	✓	✓	✓
APIs Provided	GraphQL HTTP and Websocket	SDK[5]	Dashboard[6]	HTTP and Websocket endpoints

*: Data about pricing in TheGraph was not found
1: Ethereum(testnets and mainnet), Polygon, BSC, Avalanche
2: Ethereum(testnets and mainnet), Polygon, BSC, Avalanche, Fantom
3: Ethereum(testnets and mainnet), Polygon, Optimism and Binance Smart Chain
4: Both Mainnet and Testnets
5: Moralis Software Development Kit(SDK) includes fullstack workflow for building web3 apps
6: Dune doesn't provide any API but provides a dashboard to enter queries

Table 5.3: Comparison between third-party tools to access blockchain data

5.2 Market

The second component of this thesis is using the collected data to analyze how cryptocurrency exchanges manage their markets. In this section, initially, the data collected for the most prominent metrics of DeFi projects are analyzed for different exchanges, and then additional metrics are proposed and discussed.

The most recurring metrics in DeFi applications are the total volume locked, 24H trading volume, and the number of transactions that happened in the network over time. The figures below show the comparison of these metrics over the past year between Uniswap V2, Uniswap V3, and PancakeSwap V2. As we can see from figures 5.2 and 5.4, although PancakeSwap V2 started on 23 April 2021, it has caught up with Uniswap in terms of 24H trading volume and the volume of USD locked in the exchanges. Figure 5.1 shows that Pancakeswap has surpassed both versions of Uniswap in terms of daily transactions over the year.

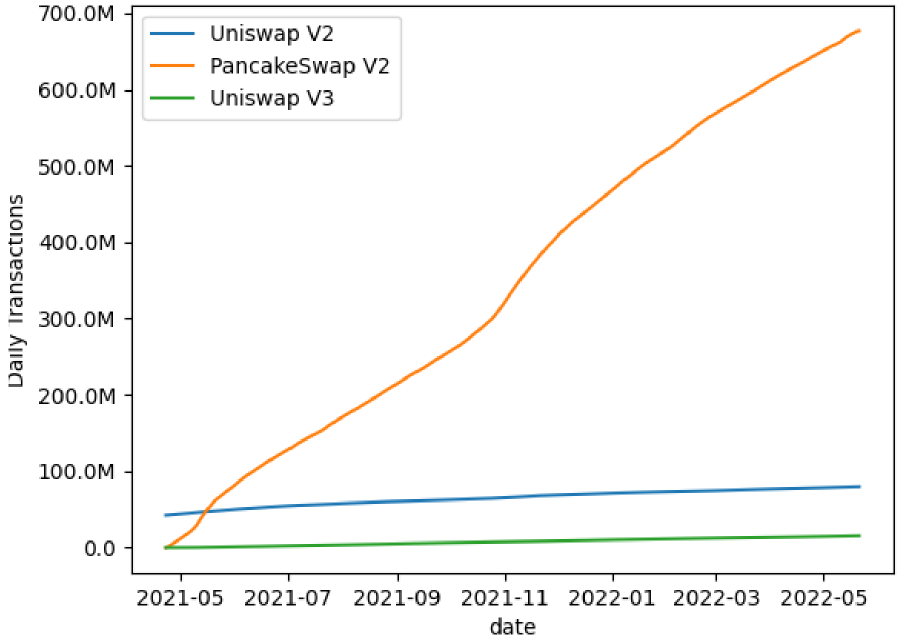


Figure 5.1: Daily transactions comparison between Uniswap V2, Uniswap V3 and PancakeSwap V3

However, it is not easy to decide the quality of an exchange with just one metric. Likewise, an unusual spike is seen around mid-May 2022 in Uniswap V2 and Binance in figure 5.2. There have been several huge crashes in Binance's trading volume during the past year. These crashes can be explained by understanding what happened at those specific points back in time. For instance, the crash of May 2021[2] and September 2021[3] happened when China banned cryptocurrency

[2] https://www.reuters.com/technology/chinese-financial-payment-bodies-barred-cryptocurrency-business-2021-05-18/

[3] https://www.bbc.com/news/business-57169726

trading all over the country. Such regulations on countries like China, which accounted for the majority of crypto trades globally, is that the prices of major coins like Bitcoin and Ethereum crash with it, causing panic selling among people and inducing clogging in the centralized exchanges.

A more recent crash happened in May 2022, which affected both Binance and Uniswap. A zoomed-in version of May 2022 of this graph is shown in figure 5.3. It can be explained using the infamous stablecoin crash of the Luna currency[4]. As a result of the panic among traders, there was a massive spike in Binance trading around May 13. However, later Binance had to shut down the trading of Luna, because of which traders shifted to Uniswap V2 at around May 15 until Luna completely shut down its blockchain.

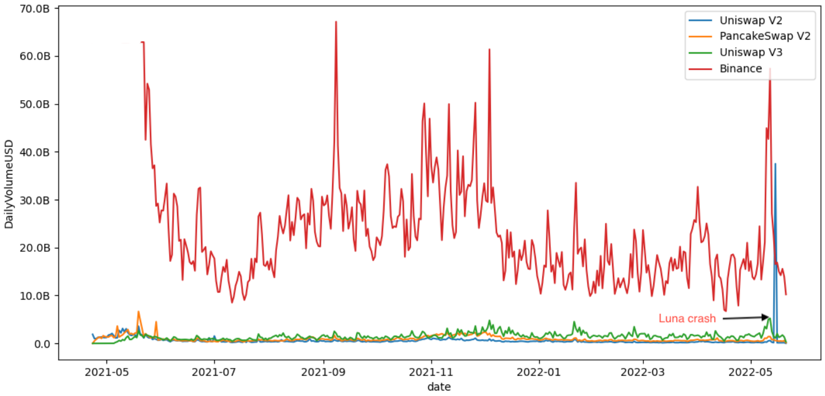


Figure 5.2: 24H Trading volume comparison between Uniswap V2, Uniswap V3, PancakeSwap V2, and Binance

There is an important point to note between the nature of these two types of trends during the crashes. In the first few Binance crashes, they did not have many effects on the decentralized exchanges. However, a significant spike on May 15 in Uniswap V2 shows that people seek alternatives to centralized exchanges when they fail to work in crunch situations. Despite having comparatively significantly higher trading volume, centralized cryptocurrencies like Binance and Coinbase are prone to usage limits because of their centralized nature during crunch times. On the other hand, the decentralized ones can usually deal with high traffic situations very well.

[4] https://www.euronews.com/next/2022/05/12/terra-luna-stablecoin-collapse-is-this-the-2008-financial-crash-moment-of-cryptocurrency

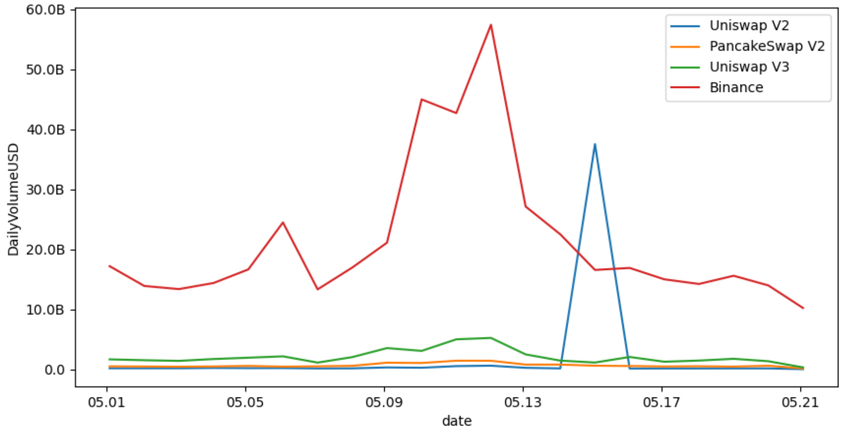


Figure 5.3: 24H Trading volume comparison between different exchanges for the month of May 2022

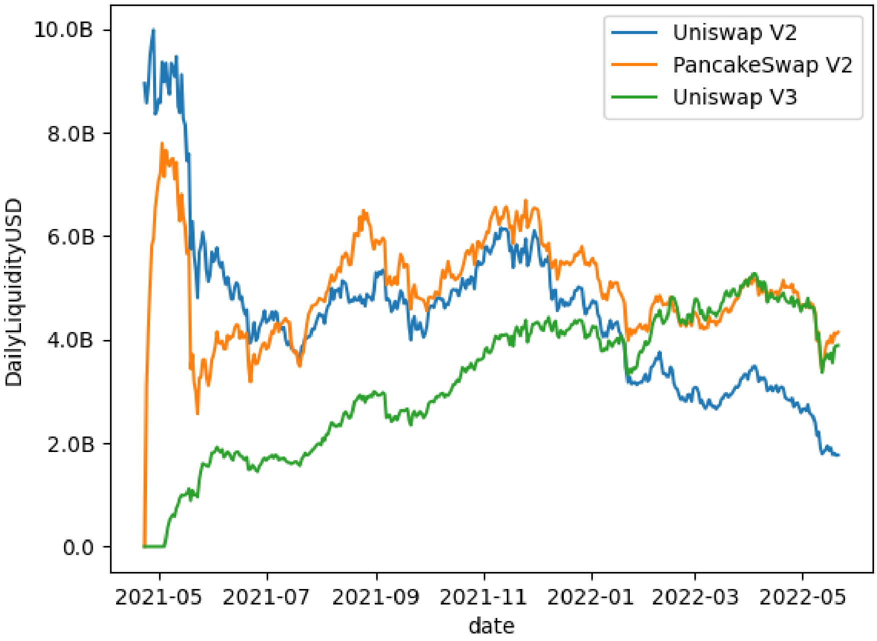


Figure 5.4: Total Volume Locked(USD) comparison between Uniswap V2, Uniswap V3 and Pancakeswap V2

Additionally, looking from the liquidity perspective, as shown in figure 5.4, it can be observed that the ride for decentralized exchanges has not been smooth.

The amount of USD volume locked[5] in the smart contracts of these exchanges looks dependent on the price of ETH in the same period, which shows that in the process of trying to correct the market, liquidity has to be moved in or out of the system.

One of the vital research questions was related to the metrics for evaluating the performance of exchanges: Research Question 3. In addition to the metrics proposed and studied by past researchers like Barbon and Randall who study price efficiency and market liquidity[28], Wang et al. who investigate arbitrage opportunities in DEXes[14], in this study, the nature of tokens in the trading pairs of the exchanges are also studied.

5.2.1 Pricing

A few experiments were performed to measure how the exchanges maintain their price over a period of time. Specifically, the price stability of the ETH token across three exchanges: Binance, Uniswap V2, and PancakeSwap V2, was tested. Figure 5.5 shows the price of the token over 24 hours in different exchanges on May 21, 2021. This experiment tries to build on top of the works done by Barbon Randalo[28] and Lehar and Parlour[27] who did the same work with Binance and Uniswap V2. Their strategies characterize stability in an exchange based on how much fluctuation there is in the price of a coin over a period of time. While their strategy indicates how the prices change over time, I think the more helpful factor to consider when measuring the efficacy of exchange is how well it reacts to external market prices.

At first glance, the price of ETH seems to be almost the same in all three exchanges in figure 5.5. However, the zoomed-in version of the exact figure for the first few minutes in figure 5.6 shows that Uniswap's price is more stable than the others. For the first increasing trend at around 02:00, Uniswap seems to gain stability much quicker than Pancakeswap and Binance. Pancakeswap V2 started around April 23, 2021. So, this could be why its prices seem to be fluctuating a bit more than the other two, as it might not have as much liquidity of the assets as Uniswap.

Figure 5.7 shows the difference in prices obtained from the web interface of Uniswap V2 and the prices obtained from smart contract calls for 1 hour. A notable result is that these scraped prices are not the same as those obtained from the smart contract calls. One explanation for this could be that the web interface uses a different autorouter to optimize prices, but this feature may not be available in smart contract calls. Additionally, the web interface could be using different settings for multihop swap compared to the price obtained from the smart contract. Thus, there need to be better standards to be used in such context when defining the prices and how they are different. This is tied back to the Research Question

[5] https://medium.com/multi-io/defi-explained-the-tvl-metric-99187587f8f0

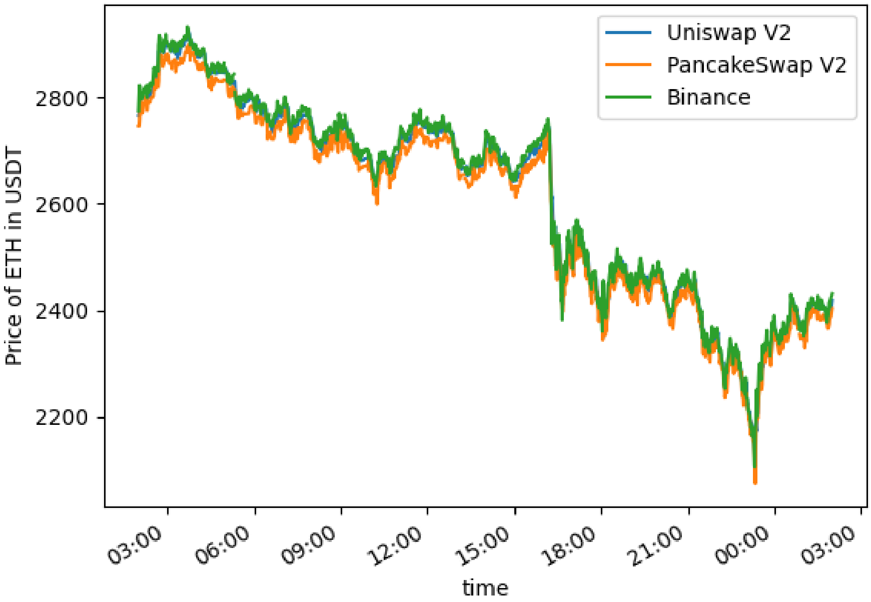


Figure 5.5: Checking price stability in different exchanges over 24H

3 where the idea was to define how pricing work in exchanges. It is clear that providing a detailed explanation of the price nuances can help the researchers.

5.2.2 Fees

Another important factor that comes into play with cryptocurrency exchanges is how they handle the transaction fees. The fee mechanism is different in DEXes like Uniswap or PancakeSwap compared to CEX like Binance. However, the customers' overall idea remains the same: they want to get the maximum amount of Output Token when they exchange their Input Token. Tables 5.4, 5.5, and 5.6 show the comparison of transaction fees for different volumes of transactions in the three exchanges.

As we can see from the tables, the transaction fees in decentralized exchanges are higher than in Binance. While the DEXes like Uniswap and Pancakeswap write every transaction in the blockchain, they have to incur gas fees for each of them. It is where centralized exchanges have a benefit. They usually store the transaction history of each user in their database and handle the outgoings and incomings internally. Later, they batch process these transactions and write the record in blockchains in bulk, saving gas fees. It is one of the reasons why the transaction fees are meager in exchanges like Binance.

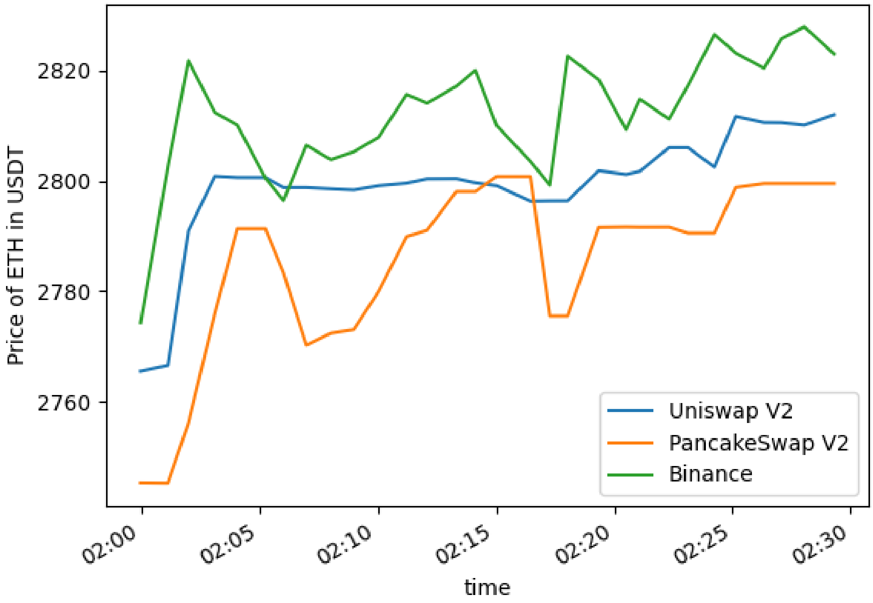


Figure 5.6: Checking price stability in different exchanges for 30 minutes

Another roadblock faced during this research is that the gas fees for the trades from PancakeSwap Web interface are not visible like in Uniswap. So, the row 2 in the tables are the prices before gas fees. Refering back to the trend of gas price on the BSC network compared to Ethereum network, it can be assumed that the gas fees are significantly lower.

Exchange	Taker Fee	$20 trade	$50 trade	$100 trade
Uniswap $V2^*$	0.3%	19.93(-8.27)	49.83(-7.97)	100.67(-8.32)
PancakeSwap V2	0.25%	19.95	49.32	99.98
Binance	0.1%	19.98	49.95	99.9

Table 5.4: Transaction fees on small size trades across several exchanges

*: Prices for Uniswap V2 inside the brackets are gas fees. These gas fees have been deducted from the net amount to left of it

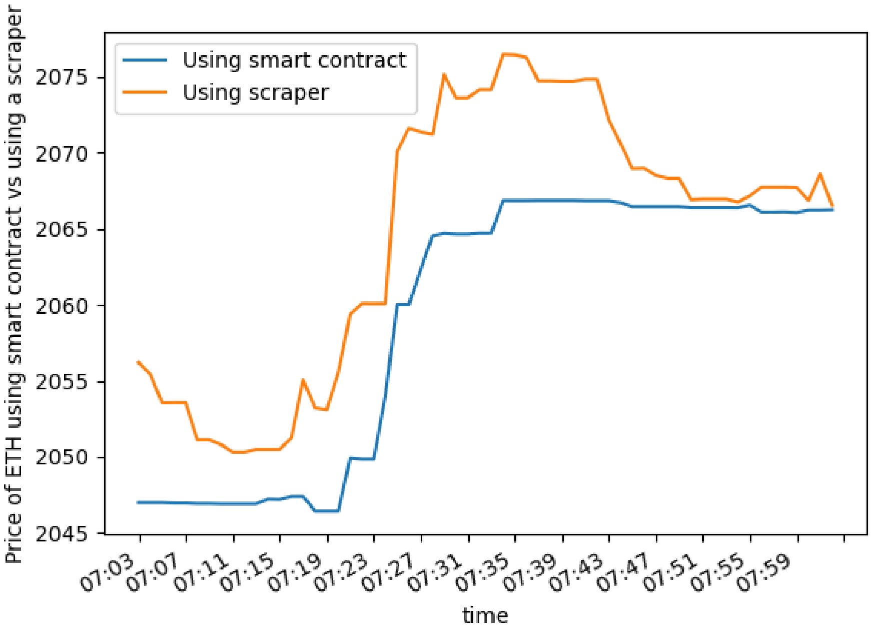


Figure 5.7: Comparing scraped price with smart contract price

Exchange	Taker Fee	$200 trade	$500 trade	$1000 trade
Uniswap V2*	0.3%	199.21(-7.39)	498.03(-6.20)	999.66(-5.71)
PancakeSwap V2	0.25%	199.43	499.01	999.72
Binance	0.1%	199.8	499.5	999

Table 5.5: Transaction fees on medium size trades across several exchanges

*: Prices for Uniswap V2 inside the brackets are gas fees. These gas fees have been deducted from the net amount to left of it

Exchange	Taker Fee	$10,000 trade	$100,000 trade	$1,000,000 trade
Uniswap V2*	0.3%	9,993.42(-6.45)	99,774.79(-21.24)	995,069.86(53.20)
PancakeSwap V2	0.25%	9,995.31	99,802.23	995,071.21
Binance	0.1%	9,990	99,990	999,000

Table 5.6: Transaction fees on large size trades across several exchanges

*: Prices for Uniswap V2 inside the brackets are gas fees. These gas fees have been deducted from the net amount to left of it

5.2.3 Pairs

The analysis of trading pairs available in exchange helps develop a top-view picture of the activities happening in the exchange. Traditionally, in centralized systems, adding or delisting a pair from the exchange lies in the hand of the central authorities, with the end-users having not much role to play in it. However, the same thing in DEXes can be handled using the voting of the stakeholders by making the decisions transparent to the general public. Such a situation has arisen multiple times in Binance, which has delisted many tokens in the past[6]. It, however, does not mean that DEXes cannot delist tokens in a decentralized manner with voting. A similar thing has happened in the past with Uniswap, which delisted several tokens from its platform without much information about how they did it[7][8]. This is usually considered to be problematic for the ethos of decentralization. In this light, an analysis was performed with the most active trading pairs in the past of Uniswap V2 and Pancakeswap V2.

Additionally, to check whether the results obtained from historical pairs hold for the new pairs, an experiment was also performed to track new pairs in these two exchanges. In a period of 3 hours, it can be seen from figure 5.10 that new pairs in PancakeSwap get added at a rate of more than 13 times the pairs in Uniswap. To be more specific, in 3 hours, there were 118 new pairs added in PancakeSwap V2, whereas the number was only 9 in Uniswap V2 in the same period. It also correlates with the number of transactions made in the PancakeSwap network, as shown in figure 5.1. To see whether PancakeSwap is better than Uniswap, we need to analyze one more aspect: i.e., the quality of a few top tokens added in pairs during the period. Table 5.7 shows the nature of the tokens involved in the new pairs in the exchanges.

Token	Total Liquidity	Number of Transactions
CASS	4.9M	922
AIElon	962M	106
BEARLON	848M	196
Huri Inu	7.18*10¯11	6
ElonaGate	944K	36
Baby Akihiko Inu	950,000,659T	23
Adam And Eve	20M	60
Sanada Inu	1.118*10¯12	2
Alpha	8884373033	290

Table 5.7: Concentration of new tokens in Uniswap V2 by key metrics

[6] https://www.binance.com/en/support/announcement/92eea3f48aa04210a454511151f3d362

[7] https://www.coindesk.com/markets/2021/04/01/uniswaps-token-issue/

[8] https://www.coindesk.com/podcasts/mapping-out-eth-2-0/3-reasons-why-uniswaps-token-delisting-sparked-controversy/

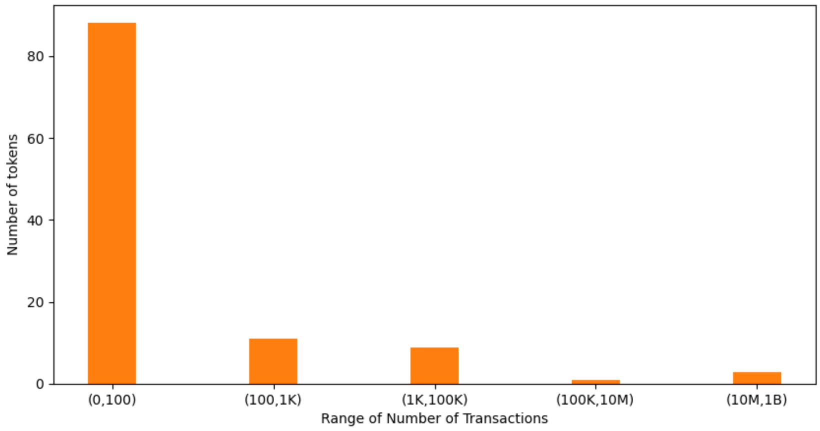


Figure 5.8: Number of new tokens distribution in Pancakeswap V2 based on total number of transactions

Table 5.7 shows the liquidity and number of transactions made by the new tokens in Uniswap V2 over the period of 3 hours. There were only 9 new tokens added, compared to 112 new tokens added in Pancakeswap V2 over the same period. All of these 9 tokens had no trading volume whatsoever when checked after a week. Also, barring from the CASS token, all of the others have made less than 200 transactions over the period. Similarly, among the 112 unique tokens constituting the 118 token pairs, more than 90 tokens had less than 100 transactions a week after being introduced on the platform. In addition, almost 50% of these tokens had a total trading volume of less than $100 during the same period. It shows how although the growth of Pancakeswap is exponential, the tokens that appear on the platform might not necessarily be of high quality. To further evaluate this, the name of these symbols were visualised in the form of a wordcloud in figure 5.1. It can be clearly seen that the occurence of words like 'test', 'Doge', 'Elon', 'Crazy', 'Luna', etc. in the majority of them shows that these tokens are either built for test purposes or just to follow the crypto hype.

It shows that the decentralized nature of these exchanges has also added a lot of noise in their platforms. In centralized options like Binance, the process is a bit less haphazard, and only the tokens that have been tested and filtered out in decentralized exchanges make the cut[9]. It is a sign that there should possibly be some sort of regulations in the process of adding new pairs in the decentralized exchanges. However, this strategy might go against the core idea of decentralization and openness. So, exchanges need to find the right balance between openness and regulation when it comes to adding new tokens. Coming back to Research Question 3, we can thus conclude that while the basic metrics like trading volume, transactions, pricing, etc. matter, we should also keep a close eye on what kind of tokens are being added to the exchanges. Given that there is a lot of noise in the

[9] https://www.binance.com/en/support/announcement/c-48

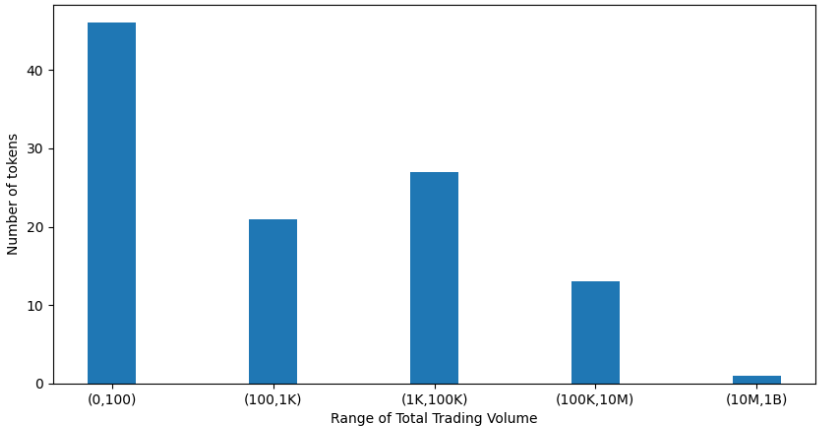


Figure 5.9: Number of new tokens distribution in Pancakeswap V2 based on total trading volume

market because of their decentralized nature, looking at the nature of token pairs could give an idea of signal to noise ratio of the quality of tokens in the platform.

An brief summary of the exchanges analyzed in this study is shown in table 5.8. It can be concluded that Uniswap V2 is better available in terms of data collection compared to the other exchanges. Similarly, in terms of metrics like trading volume and fees, Binance leads the other two. One interesting result seen in the top 5 trading pairs by volume row is that in Binance and PancakeSwap, the top 5 tokens are mostly the mix of ETH or BNB based tokens along with stablecoins like USDT, USDC, DAI, etc. However, in Uniswap, the fifth biggest pair by trading volume is the SCAMMY-WETH. The SCAMMY token released a controversial smart contract in March 2021 that allowed nearly $11 billion in trading volume for a single token to be registered over the last 24 hours on Uniswap despite having less than 1 cent in actual liquidity[10]. This is a huge number because until this point in time, Uniswap had not done any day with more than $2.22 billion in a day[11].

Such vulnerabilities are expected when in decentralized technologies where there are no specific set of rules like we see in the centralized exchanges. These mishaps and hiccups in Uniswap are similar to what were obtained by Xia et al[15, 17]. While their results show that a massive number of the tokens listed in Uniswap are scam, we can still expect such decentralized markets to get better with time because as seen from the results in this chapter in terms of transparency of data and availability of trading pairs, decentralized exchanges are better than

[10] Delta.financial was behind the massive spike in trading volume shown in Uniswap(`https://decrypt.co/63458/11-billion-fake-uniswap-volume-defi-project-dex-clash`

[11] This article explains how using Flash Swap, Uniswap and Curve were hacked resulting in unprecendented trading volume in the systems(`https://decrypt.co/46339/dex-volumes-5-billion-harvest-25-million-defi-hack`).

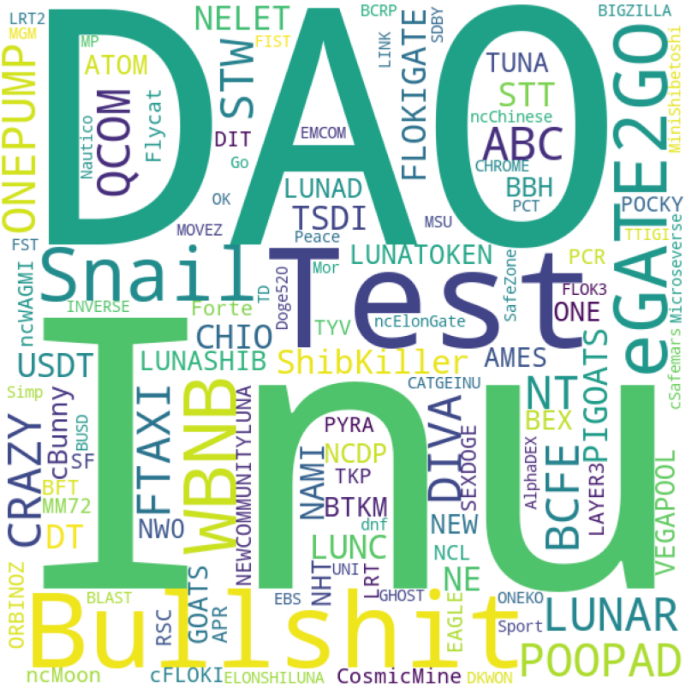


Figure 5.10: Wordcloud of new tokens added in PancakeSwap V2

centralized ones.

Thus, in this research, Research Question 4 remains partially answered. While we were able to compare some metrics across decentralized and centralized exchanges, because of lack of data available in Binance, a comprehensive answer was not found. Based on the data available, it can be concluded that there are some aspects like usability, first mover advantages, etc. that make centralized cryptocurrency exchanges better, but in terms of factors like transparency, availability, and price efficiency, decentralized exchange have the edge.

Criteria	Sub-criteria	Binance[1]	Uniswap V2[2]	PancakeSwap V2[3]
data collection	Local node	N/A [4]	✓	✓*
	Hosted node		✓	✓
	External APIs	✓	✓	✓
	Historical data	✓[5]	✓	✓
	Best options	Binance API	SC[6]$Calls$& $TheGraph$	SC[6]$Calls$& $PancakeSwap$ $Subgraph$ $GraphQL$ $Interface$
Statistics	24H Trading Volume	$12.6B	$99.4M	$459.5M
	Total Transactions	N/A	80.1M	688.06M
Price	Time resolution	1 min	Ethereum block time	BSC block time
Pairs	total trading pairs	1,458	75,732	962,489
	Top 5 Trading Pairs by volume	ETH-USDT BTC-USDT BTC-BUSD ETH-BUSD GMT-USDT	USDC-WETH WETH-USDT UST-mNFLX DAI-WETH SCAMMY-WETH	WBNB-BUSD USDT-WBNB Cake-WBNB USDT-BUSD ETH-WBNB
Fees		0.1%	0.3% + Gas Fees	0.25% + Gas fees

*: The URLs available to run a local node in BSC don't work all the time

1: This data for Binance was obtained on May, 30, 2022 from
`https://coinmarketcap.com/exchanges/binance/`

2: This data for Uniswap V2 was obtained on May, 30, 2022 from
`https://api.thegraph.com/subgraphs/name/ianlapham/uniswapv2`

3: This data for PancakeSwap V2 was obtained on May, 30, 2022 from
`https://bsc.streamingfast.io/subgraphs/name/pancakeswap/exchange-v2/graphql`

4: N/A: Not available

5: Limited availability

6: SC: Smart contract

Table 5.8: Comparison between Binance, Uniswap, and PancakeSwap

Chapter 6

Conclusion and Future Work

This chapter aims to summarise the overall study regarding how well the research questions asked in the beginning were addressed. After the summary, the limitations of this study are discussed, finally wrapping up with the future work possible on top of this work.

6.1 Conclusion

To summarise, the research questions of this study have been answered with a few hiccups. The study started with an attempt to probe the status of empirical research in cryptocurrency exchanges. As a result, the systematic literature research shows that the field is still very new but exploding at a very high rate with time. Thus, to answer Research Question 1: although there does not seem to be enough systematic research right now, there are comparatively a lot of non-peer-reviewed resources available, and it can therefore be expected that the status will change in the coming few years.

After the literature research was carried out, it was felt necessary that the data collection strategies used by the researchers need to be replicated and tried out on our own to understand the various options used by them. It was the objective of Research Question 2. It was found that there were some trade-offs in using on-premise resources vs. using third-party resources after trying both of these strategies. While on-premise resources provide great flexibility and control, it is not for someone who can not meet the blockchain nodes' expensive and high resource requirements. On the other hand, using third-party resources makes it very easy to interact with the blockchain but offers less flexibility and is usually very expensive when there is a need to perform extensive work.

Apart from the quality of data and the options to access them, this study also tried to critically evaluate the cryptocurrency exchanges by using already existing metrics and proposing new ones. To answer Research Question 3, the metrics most common in existing research were explored with Binance, Uniswap V2, and PancakeSwap V2. However, it was observed that getting granular from Binance was usually not possible in most cases, limiting the comparison to Unsiwap and PancakeSwap. It was also noticed that some existing metrics like prices and fees need to be more detailed. Additionally, a proposal was made to scrutinize the quality of exchange based on the evaluation of tokens available in the system.

Finally, the purpose of Research Question 4 was to understand what data collection is necessary to compare decentralized and centralized exchanges empirically. This aspect of research faced some roadblocks because of time limits and the limitations of the APIs from the exchanges, especially Binance. Nevertheless, the collected data was used to critique the characteristics of several exchanges based on how they manage their markets in terms of pricing, fees, and trading pairs.

6.2 Future Work

While at the beginning of this study, the aim was to come up with a comprehensive framework to collect all the data required to compare decentralized and centralized exchanges, there were several challenges during the process that limited this study. As a result, the scope of the study has to be decreased given the limited time at hand. In the future, in collaboration with DSE at NTNU, the following work would be handy for research purposes:

1. Explore the possibilities of making the local Ethereum node as well as other blockchain nodes node open for research purposes
2. Index and store other types of data from the blockchains to build our price oracles as well as data explorers
3. Build tools to perform exploratory as well as anomaly analyses from the collected data

In addition to these extended works, it is also necessary to track how blockchains change in the future. Ethereum, for instance, is about to go through a significant upgrade with Ethereum 2.0, which is also going to use Proof of Stake. So, it will automatically decrease the gas fees tremendously. With Ethereum 2.0, inevitably, Uniswap will also undergo changes. However, the basics of how an exchange operates will stay the same. Such cryptocurrency exchanges have had their lows multiple times in the past, but only those who were quick to remedy their faults and adapt are still surviving. So, in the future, we can expect the exchanges that are resilient and can adapt quickly will continue to survive and dominate the cryptocurrency markets.

Bibliography

[1] S. Nakamoto, *Bitcoin: A peer-to-peer electronic cash system*, Accessed: 2015-07-01, Dec. 2008. [Online]. Available: `https://bitcoin.org/bitcoin.pdf`.

[2] R. C. Merkle, 'A digital signature based on a conventional encryption function,' in *Advances in Cryptology — CRYPTO '87*, C. Pomerance, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 1988, pp. 369–378, ISBN: 978-3-540-48184-3.

[3] R. Zhang and W. K. (Chan, 'Evaluation of energy consumption in blockchains with proof of work and proof of stake,' *Journal of Physics: Conference Series*, vol. 1584, no. 1, p. 012 023, Jul. 2020. DOI: `10.1088/1742-6596/1584/1/012023`. [Online]. Available: `https://doi.org/10.1088/1742-6596/1584/1/012023`.

[4] G. Wood *et al.*, 'Ethereum: A secure decentralised generalised transaction ledger,' *Ethereum project yellow paper*, vol. 151, no. 2014, pp. 1–32, 2014.

[5] A. Donmez and A. Karaivanov, 'Transaction fee economics in the ethereum blockchain,' *Economic Inquiry*, vol. 60, no. 1, pp. 265–292, 2022.

[6] N. Szabo, 1996. [Online]. Available: `https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html`.

[7] X. C. Yi Zhang and D. Park, 'Formal specification of constant product (x × y = k) market maker model and implementation,' *GitHub*, Dec. 2018. [Online]. Available: `https://github.com/runtimeverification/verified-smart-contracts/blob/uniswap/uniswap/x-y-k.pdf`.

[8] Y. Lo and F. Medda, *Uniswap and the rise of the decentralized exchange munich personal repec archive*, `https://mpra.ub.uni-muenchen.de/103925/`, (Accessed on 04/28/2022), 2020.

[9] H. Adams, N. Zinsmeister and D. Robinson, 'Uniswap v2 core whitepaper,' vol. 12, 2020.

[10] H. Adams, N. Zinsmeister, M. Salem, R. Keefer and D. Robinson, *Uniswap v3 core*, 2021.

[11] *Taking undercollateralized loans for fun and for profit*, `https://samczsun.com/taking-undercollateralized-loans-for-fun-and-for-profit/`, (Accessed on 05/14/2022).

[12] J. A. Berg, R. Fritsch, L. Heimbach and R. Wattenhofer, *An empirical study of market inefficiencies in uniswap and sushiswap*, 2022. DOI: `10.48550/ARXIV.2203.07774`. [Online]. Available: `https://arxiv.org/abs/2203.07774`.

[13] J. Han, S. Huang and Z. Zhong, 'Trust in defi: An empirical study of the decentralized exchange,' *Available at SSRN 3896461*, 2021.

[14] Y. Wang, Y. Chen, H. Wu, L. Zhou, S. Deng and R. Wattenhofer, *Cyclic arbitrage in decentralized exchanges*, 2021. DOI: `10.48550/ARXIV.2105.02784`. [Online]. Available: `https://arxiv.org/abs/2105.02784`.

[15] P. Xia, H. wang, B. Gao, W. Su, Z. Yu, X. Luo, C. Zhang, X. Xiao and G. Xu, *Trade or trick? detecting and characterizing scam tokens on uniswap decentralized exchange*, 2021. arXiv: `2109.00229 [cs.CR]`.

[16] B. Mazorra, V. Adan and V. Daza, *Do not rug on me: Zero-dimensional scam detection*, 2022. DOI: `10.48550/ARXIV.2201.07220`. [Online]. Available: `https://arxiv.org/abs/2201.07220`.

[17] K. Tjiam, R. Wang, H. Chen and K. Liang, 'Your smart contracts are not secure: Investigating arbitrageurs and oracle manipulators in ethereum,' in *Proceedings of the 3rd Workshop on Cyber-Security Arms Race*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 25–35, ISBN: 9781450386616. [Online]. Available: `https://doi.org/10.1145/3474374.3486916`.

[18] P. Daian, S. Goldfeder, T. Kell, Y. Li, X. Zhao, I. Bentov, L. Breidenbach and A. Juels, *Flash boys 2.0: Frontrunning, transaction reordering, and consensus instability in decentralized exchanges*, 2019. arXiv: `1904.05234 [cs.CR]`.

[19] C. Ferreira Torres, A. K. Iannillo, A. Gervais and R. State, 'The eye of horus: Spotting and analyzing attacks on ethereum smart contracts,' in *Financial Cryptography and Data Security*, N. Borisov and C. Diaz, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2021, pp. 33–52, ISBN: 978-3-662-64322-8.

[20] K. Qin, L. Zhou, B. Livshits and A. Gervais, 'Attacking the defi ecosystem with flash loans for fun and profit,' in *Financial Cryptography and Data Security*, N. Borisov and C. Diaz, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2021, pp. 3–32, ISBN: 978-3-662-64322-8.

[21] L. Zhou, K. Qin, C. F. Torres, D. V. Le and A. Gervais, *High-frequency trading on decentralized on-chain exchanges*, 2020. DOI: `10.48550/ARXIV.2009.14021`. [Online]. Available: `https://arxiv.org/abs/2009.14021`.

[22] Z. Zhou and B. Shen, *Toward understanding the use of centralized exchanges for decentralized cryptocurrency*, 2022. DOI: 10.48550/ARXIV.2204.08664. [Online]. Available: https://arxiv.org/abs/2204.08664.

[23] G. Angeris, H.-T. Kao, R. Chiang, C. Noyes and T. Chitra, *An analysis of uniswap markets*, 2019. DOI: 10.48550/ARXIV.1911.03380. [Online]. Available: https://arxiv.org/abs/1911.03380.

[24] Y. Lo and F. Medda, 'Uniswap and the rise of the decentralized exchange,' 2020.

[25] V. Danos, H. E. Khalloufi and J. Prat, 'Global order routing on exchange networks,' in *Financial Cryptography and Data Security. FC 2021 International Workshops: CoDecFin, DeFi, VOTING, and WTSC, Virtual Event, March 5, 2021, Revised Selected Papers*, Berlin, Heidelberg: Springer-Verlag, 2021, pp. 207–226, ISBN: 978-3-662-63957-3. DOI: 10.1007/978-3-662-63958-0_19. [Online]. Available: https://doi.org/10.1007/978-3-662-63958-0_19.

[26] B. Krishnamachari, Q. Feng and E. Grippo, 'Dynamic automated market makers for decentralized cryptocurrency exchange,' in *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2021, pp. 1–2. DOI: 10.1109/ICBC51069.2021.9461100.

[27] A. Lehar and C. A. Parlour, 'Decentralized exchanges,' working paper, University of Calgary and University of California, Berkeley, Tech. Rep., 2021.

[28] A. Barbon and A. Ranaldo, *On the quality of cryptocurrency markets: Centralized versus decentralized exchanges*, 2021. arXiv: 2112.07386 [q-fin.TR].

[29] A. Aspris, S. Foley, J. Svec and L. Wang, 'Decentralized exchanges: The "wild west" of cryptocurrency trading,' *International Review of Financial Analysis*, vol. 77, p. 101 845, 2021, ISSN: 1057-5219. DOI: https://doi.org/10.1016/j.irfa.2021.101845. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1057521921001782.

Appendix A

Additional Material

A.1 Code and datasets used for the study

All the code used used for this study is open-source and available for anyone who wants to reproduce the work. The link to the Github repository is: `https://github.com/dipespandey/blockchain-analysis`. The repository also contains some sample datasets produced during the process.

A.2 Some GraphQL Queries Used with The Graph

```

// get the current price of Ethereum in Uniswap V2
query{
    bundle(id:"1"){
    ethPrice
  }
}

// get the pairs in Uniswap V2 added between a time range
query{
      pairs(first:120, where:{timestamp_gt:1653040552, timestamp_lt:1653046706, volumeUSD:"0"}){
      block
      token0{symbol tradeVolumeUSD totalTransactions totalLiquidity }
      token1{symbol tradeVolumeUSD totalTransactions totalLiquidity}
      totalTransactions
      volumeUSD
    }
}

// get the top tokens in Uniswap V2 by transaction count
query{
  tokens(first:20, orderBy: txCount, orderDirection:desc){
    symbol
    tradeVolumeUSD
    txCount
    totalLiquidity
  }
}

//get daily token data for Tether (USDT)
query{
  tokenDayDatas(where: {token: "0xdac17f958d2ee523a2206206994597c13d831ec7"}, first:365, orderBy:date, orderDirection:desc){
    token{symbol}
    date
    dailyVolumeUSD
    dailyTxns
    priceUSD
  }
}

// get the stats about Uniswap V2
query{
    uniswapFactories(first: 1) {
    pairCount
    totalVolumeUSD
    totalVolumeETH
    totalLiquidityUSD
    }
}

// get the daily stats of the Uniswap V2 network for the last 365 days
query{
    uniswapDayDatas(first: 365, orderBy:date, orderDirection:desc) {
    dailyVolumeUSD
    date
    txCount
    totalLiquidityUSD
  }
}
```