BACHELOR

Tracing the Transactions and Connections of Binance on the Bitcoin Network
An Exploratory Network Analysis

Lee, Calvin C.E.

Award date:
2023

Technische Universiteit Eindhoven
University of Technology
Department of Mathematics and Computer Science

Tracing the Transactions and Connections of Binance on the Bitcoin Network: An Exploratory Network Analysis

Bachelor's Thesis

Calvin Lee

Supervisor:
Dr. Claudia Zucca

Eindhoven, Sunday 18th June, 2023

Abstract

In an age characterized by the rapid digitization of monetary and financial systems, this thesis focuses on Binance. As one of the largest cryptocurrency exchanges, we investigate it within the context of the Bitcoin network. Through a network analysis approach, we specifically focus on three research aspects: the visual representations of communities, influential nodes, and the overall structural composition. We found that the Louvain method for community detection on an undirected interpretation of the network results in more cohesive and higher-quality communities. In addition, we found the most influential node within network *1NDyJtN*, using degree, eigenvector, and PageRank centrality. Finally, the results of the descriptive network statistics suggest that most network participants only transact with a select few users or services of high influence. However, our analyses are mostly limited by the static interpretation of the dynamic Binance network and the limited timeframe of a single day of transaction data. With our approach, we look to provide a foundation for further analyses facilitating more comprehensive studies of the roles, behaviors, and classification of network participants. Network analysis on the blockchain has potential applications for the forensic tracking of illicit transactions, conducting investment research, or analyzing financial data that is typically hard to obtain due to security and privacy concerns. As the monetary system continues to evolve, this research contributes to our understanding of sophisticated financial networks. It offers valuable insights that are beneficial for academics, policymakers, and industry professionals.
Keywords: Network analysis, blockchain, transaction network

Chapter 1

Introduction

The world is producing an ever-increasing amount of data over time. This has led to the emergence of increasingly complex relations between various entities resulting in sophisticated systems that can be analyzed through network analysis techniques. We have seen many applications of network analysis in various research fields including transportation, biology, and financial markets [1]. However, gaining a thorough understanding of the underlying networks in complex systems is crucial to gain insights into these increasingly complex systems [2]. Network analysis in the context of cryptocurrency transactions has become a popular research area in recent years [3]. Binance is one of the world's largest cryptocurrency exchanges, which are brokers that facilitate the trading of digital currencies and other financial services. Its transaction network on the Bitcoin blockchain is a good example of a complex system that can be analyzed using network analysis techniques. Moreover, The transparent nature of the blockchain provides promising data mining opportunities as it contains publicly available rich information on economic activity. In contrast to many traditional financial scenarios, this kind of information is often obfuscated or inaccessible to the general public due to security and privacy concerns. By interpreting the blockchain as a network considering addresses as nodes and the flow of funds as edges, network analysis could provide knowledge discovery in other financial systems as this field is still relatively unexplored due to the constraints of traditional financial systems [4, 5]. In addition, much of the existing literature focuses on the Bitcoin network without specifically considering cryptocurrency exchanges. Analysis of protocols that extend functionalities of the blockchain like the Lightning Network [6] are even less frequently encountered.

Bitcoin is a decentralized digital currency, often referred to in the literature as a peer-to-peer payment system [7]. Since the inception of Bitcoin as the first and most popular real-world application of blockchain technology, we have seen explosive growth in the popularity of cryptocurrencies and other applications of blockchain technology. This surge has been driven by a combination of factors, including the increasing awareness by governments [8] and the adoption of Bitcoin as legal tender in countries such as El Salvador and the Central African Republic [9]. One other significant factor is the emergence of cryptocurrency exchanges, which play a crucial role in the cryptocurrency ecosystem by providing services that enable the exchange of digital assets with both government-issued currencies and other digital assets [10]. These exchanges operate through platforms that prioritize user-friendliness and ease of use which lowers the threshold for interaction with the blockchain significantly. They allow users to interact, trade, and benefit from the properties of the blockchain such as privacy practically without any knowledge of the underlying mechanics of the Bitcoin blockchain. At the time of writing, Binance is the largest cryptocurrency exchange by a large margin [11] despite being a relatively young exchange founded in July 2017. As the most dominant exchange in the market, the role of Binance is crucial in the industry as its behavior could impact a significant portion of the network. Therefore, studying the actions of Binance on the Bitcoin blockchain using an ego-centric data collection approach can yield valuable insights into behavioral patterns and financial strategies. This information might be useful for businesses or financial institutions that provide similar services to cryptocurrency exchanges such as banks or financial brokers where such behaviors and strategies are highly valued. In addition, financial institutions could utilize this research to identify illicit activity without compromising the privacy of their users.

In this thesis, our objective is to investigate the Binance network around the halving date. The halving date refers to a predetermined event in the Bitcoin system where the incentive for Bitcoin miners is reduced by half [12, 7]. Therefore, with this thesis, we aim to answer the following research question: "What are the key network characteristics and properties of the Binance network around the halving date, and how can the communities within the network be described?".

The Bitcoin blockchain can be regarded as a distributed and public ledger [13] that contains all historical Bitcoin transactions in the form of a list of blocks, where a block is defined as a data structure that contains a set of transactions on the Bitcoin blockchain. From a modeling perspective, the Bitcoin blockchain can be interpreted as a network. As such, we will refer to the Bitcoin blockchain and the Bitcoin network interchangeably in the rest of this thesis. There have been different network interpretations of the Bitcoin blockchain which we will discuss in Chapter 3. Our primary focus will be to explore how different network measures such as centrality measures, number of nodes and edges, and degree distribution capture different aspects and characteristics of the Binance network. This thesis will provide an overview of these network measures and their intuition, which we'll explore in further detail in Chapter 3 To our knowledge, research on network analysis of the Bitcoin blockchain with regard to a single entity such as Binance remains limited. Much of the existing research investigates subsets of the Bitcoin blockchain prior to the founding of Binance. In this thesis, we will conduct an exploratory network analysis of the Binance network on the Bitcoin blockchain. We aim to provide a more detailed perspective of the Bitcoin network by exploring several network measures to gain a deeper understanding of its characteristics and behaviors. As such, the main contributions of this thesis are listed as follows:

- Identified influential nodes in the Binance network.
- Calculated descriptive network statistics of the Binance network.
- Visualized communities from both a directed and an undirected perspective
- Developed a methodology to conduct network analysis on the Binance blockchain utilizing publicly available data sources.

The rest of this thesis is organized as follows: Chapter 2 provides a brief summary of the related works. In Chapter 3 we will discuss our methodology which includes an overview of data acquisition, data validation, and network construction methods. In addition, we will provide a detailed overview of the network measures used, their calculations, and interpretations in the context of network analysis on the Binance network. In Chapter 4 we will present our findings followed by a discussion of results, potential future work, and conclusions in Chapter 5.

Chapter 2

Literature review

Network analysis in blockchain-based systems has been conducted from various perspectives, often focusing on specific or multiple time intervals of the Bitcoin blockchain. Specifically, previous studies investigated network properties during the early stages of the Bitcoin network from 2009 until 2013. In this Chapter, we will discuss several works in which the authors analyze network properties and additional observations chronologically ordered by publishing date.

Kondor et. al [14] investigated the growth of the Bitcoin network by analyzing the change in network properties over time during the early years of the Bitcoin network. The authors also related external features with daily snapshots of the Bitcoin network using principal component analysis [15]. In [14] Kondor et. al observed significant fluctuations in the degree distributions in the early stages of the Bitcoin network (2009-2010). In comparison, the more recent evolution of the network was observed to be more stable as characterized by the disassortative degree correlations and the degree distributions. Furthermore, the authors investigated the accumulation of wealth and preferential attachment [16], finding that the growth of the Bitcoin network was mainly explained by linear preferential attachment and wealth accumulation driven by sublinear preferential attachment.

Baumann et. al [17] investigated network measures over a period from March 2009 to October 2013. The authors observed that the majority of nodes had a low degree value, while a small number had high degree values. They also observed a high global clustering coefficient where the coefficient would decrease with increasing network activity. Furthermore, the authors observed that the number of edges (transactions) was highly correlated with the exchange price of Bitcoin at the time.

Maesa et. al [18] focused on utilizing network analysis for the Bitcoin blockchain. Since there is no limit on how many addresses a user can create on the Bitcoin network, it is necessary to find heuristics to facilitate the clustering of addresses when performing analysis on the user level. Maesa et. al defined a scalable address clustering algorithm using all Bitcoin transactions from 2009 until 2015. They characterized the evolution of the Bitcoin network by examining the network measures such as densification and small-world properties finding support for observations made by Baumann et. al. The authors also found that there are central nodes that act as a hub between different groups of users. This phenomenon was analyzed in another study where Maesa et. al proved the existence of artificial transaction chains. The authors characterized and presented theories for the transaction structures found [19]. Finally, Maesa et. al concluded that when compared to social networks, the diameter of the Bitcoin network is significantly larger. Moreover, it was also supposed that the degree distribution shows sharp increases at specific values, unlike other more evenly dispersed distributions found in social networks for example.

In their analysis of the Bitcoin network between January 2009 and September 2014, Alqassem et al. [20] identified different node types, including checking accounts, saving accounts, and intermediate accounts. Similar to Maesa et. al, the authors also observed a non-linear increase in the diameter of the Bitcoin network. A proposed theory for this behavior is the increased adoption of Bitcoin over time which naturally results in more usage of the network. Alqassem et. al also analyzed the degree assortativity which is a measure of the tendency of interaction between nodes with similar in and out degrees. They concluded that the Bitcoin transaction graph is negatively assortative over time, meaning that transactions are likely to occur between high-degree and low-degree checking and saving addresses. Finally, the authors investigate the time-evolving community structure of the Bitcoin network which are groups of nodes that are densely connected within themselves but have relatively fewer connections to nodes outside the group. The authors concluded using the modularity of the graph that the total number of communities did not increase. Instead, it was observed that smaller communities were merging into larger ones over time.

Xiang et. al [21] investigated different types of Bitcoin address subgraphs by classifying subgraphs with different characteristics using four different machine learning models. The input variables of these models consisted of 10 network measures which include average degree, betweenness centrality, average shortest path length, and more. Xiang et al then identified eight different categories for these subgraphs or patterns which all interact uniquely. For instance, exchanges typically own multiple addresses which can broadly be classified into two roles. Internal addresses are used for internal transfers, and exchange addresses are used to interact with clients. Naturally, cryptocurrency exchanges and their clients on the network exhibit behaviors that are different from gambling service providers or dark web markets which each have their own transactional patterns. The results showed that the classification models were quite effective as observed in their study with F1-scores up to 0.91.

There have been many different interpretations of the Bitcoin transaction network, such as the transaction network, address network, and user network where each variation uses a different node representation [3]. The transaction network is constructed with transactions as nodes, the address network utilizes addresses, and the user network represents individual users as nodes. The primary drawback of the user network model is that it is impossible to conclusively associate users with addresses unless the user verifies their addresses. Various proposals have been made for heuristic clustering methods to approximate the actual users graph. For example, Reid and Harrigan [22] introduced the user network where all input addresses for a multi-input transaction are regarded as a single user. This heuristic is based on the fact that every input address needs to sign the transaction with its private key to facilitate a transaction [23, 7]. Therefore, if a transaction contains multiple inputs they should be controlled by the same owner since the input transactions are signed simultaneously. This method of clustering has become widely used in subsequent analyses [24, 25, 26].

However, despite the amount of research on the topic of analysis of the Bitcoin network, the number of comprehensive studies specifically focusing on exchanges utilizing recent data is still limited [5]. In this thesis, our objective is to fill this gap in the literature by characterizing the Binance network, one of the largest cryptocurrency exchanges at the time of writing [27]. We will be using data from May 14, 2020, by considering the proposed methods and representing the network in its simplest form as a network of addresses without considering clustering as introduced by Kondor et. al [14]. The research questions we will answer in this thesis are:

- What insights can we gain into the visual representations and distribution of communities within the Binance network?
- What are the most influential nodes or addresses within the Binance network, as evaluated by degree, eigenvector, and PageRank centrality measures?
- How can we describe the structural composition of the Binance network through the use of network statistics?

We aim to provide an analysis of the network properties and describe the communities exhibited within this network. We will investigate various network measures and properties of the network such as degree distributions, clustering coefficients, centrality measures, and community structure of which we will give a more detailed description in Chapter 3. Additionally, we will identify the most active nodes and analyze the triadic patterns in the network.

Chapter 3

Methodology

In this chapter, we aim to describe the methodology of how we will answer the research questions presented in Chapter 2. We will introduce basic network-related concepts and examine several network property measures of the Binance network around the halving date. In addition, we will describe and explain why these measures are suitable for our analysis. The Binance network, which we model as a transaction graph, can formally be defined as follows: "$BitN(A, T)$, where A represents the finite set of unique addresses (vertices) on the network which are identifiers to send and receive Bitcoins. T represents the set of all Bitcoin transactions (edges) that have occurred between addresses." A transaction between two distinct addresses, denoted as $a_i, a_j \in A$, is defined as a directed edge $t_{ij} \in T$ between a_i, a_j

Local Network Properties

- *Number of Nodes and Edges.* The number of nodes and edges in a network are metrics to measure the size and activity of a network. However, it has been observed by multiple researchers that there is a large presence of intermediate transactions. A plausible theory for this behavior is that users are looking to obfuscate the final destination of the funds. In addition, there is no limit on how many addresses a single user can create. Therefore, it should be noted that the number of nodes and edges is subject to manipulation and may not necessarily reflect user adoption, but instead indicate only the level of network activity and size, irrespective of the number of users. This limitation can be mitigated by using methods such as implementing clustering heuristics [14], or by considering the transaction frequencies or degree distributions. In this thesis, our approach will utilize ego data to construct the Binance network. Given our scope, we will implement the network in its simplest form without considering clustering. We will provide a detailed description of our methodology in Section 3.1.

- *Degree Distribution.* In an undirected network the degree of a node is equal to the number of edges. Where N_k denotes the number of nodes with degree k and N denotes the total number of nodes in the network. The degree distribution is calculated as follows:

  $$P(k) = \frac{N_k}{N} \tag{1}$$

  In a directed network, we will distinguish between incoming edges (in-degree) and outgoing edges (out-degree). As a result, we will have two distributions calculated using the same formula with one only considering the in-degree and the other considering the out-degree.

- *Node Importance.* The process of identifying important nodes in a network can be achieved through a variety of methods. Lü et al. [28] provide a comprehensive review of node importance measures. The authors discuss the use of centrality measures to assign a value representing each node's importance. These measures are derived from the structural information of nodes within the network. In this thesis, we will consider both neighborhood-based centrality measures and path-based centrality measures. Neighborhood-based measures take into account the immediate neighbors of a node while path-based centrality is based on the number of shortest paths that go through the node. We will evaluate three centrality measures.

Degree centrality [29], denoted by:

$$C_{D(a_i)} = deg(a_i) \tag{2}$$

Degree centrality is one of the simplest measures of a node's importance. Here, a_i is a node in the network, $C_{D(a_i)}$ represents the degree centrality of node a_i, and $deg(a_i)$ is the number of edges or connections it has.

Eigenvector centrality [30], expressed as:

$$C_E(a_i) = \frac{1}{\lambda} \sum_j A_{ij} C_E(a_j) \tag{3}$$

Here, $C_E(a_i)$ is the eigenvector centrality of node a_i, and A_{ij} represents the entries in the adjacency matrix A. λ is the largest eigenvalue of the adjacency matrix, and $C_E(a_j)$ represents the eigenvector centrality of nodes connected to a_i. This measure accounts for both the quantity and quality of connections, assigning higher scores to nodes connected to other high-scoring nodes. The eigenvector is calculated as $Ax = \lambda x$, the basic equation for an eigenvalue (λ) and eigenvector (x) of a matrix (A).

PageRank centrality [31]:

$$PR(a_i) = (1 - d) + d \sum \frac{PR(a_j)}{N(a_j)} \tag{4}$$

In this equation, $PR(a_i)$ is the PageRank of node a_i, d is a damping factor often set to 0.85 by default, $PR(a_j)$ represents the PageRanks of nodes a_j linking to node a_i, and $N(a_j)$ is the number of links outgoing from each a_j. This reflects the concept that a node's importance is determined by the importance of the nodes linking to it.

Global Network Properties

- *Clustering Coefficient.* The clustering coefficient is a measure of the tendency for nodes to form a triangle structure and can be measured in two different ways. The local clustering coefficient correlates with the probability that the neighbors of a node will also be connected to each other in a triangle structure. The local clustering coefficient C_a of node a can be calculated as follows:

$$C_a = \frac{e_a}{(d_a(d_a - 1))/2} = \frac{2e_a}{d_a(d_a - 1)} \tag{2}$$

  Where node a's participation in directed triangle structures is denoted by e_a, while its out-degree is represented by d_a. The global clustering coefficient is measured as the proportion of closed length-two paths in a network [32]. The global clustering coefficient is calculated as follows:

$$C = \frac{1}{N} \sum_{i=1}^{N} C_a \tag{5}$$

  Where N is the total number of nodes.

- *Community Detection.* Community detection involves identifying and analyzing groups of entities or communities within a network. It aims to represent real-world scenarios of more densely connected clusters or communities of nodes relative to the rest of the network. There are different methods to detect and partition communities within a network. In this thesis we will implement the Louvain method [33, 34, 35] to detect communities within the Binance network. This method is often chosen due to its efficient use of computational resources and it does not require a predefined number of communities. The Louvain algorithm starts by classifying each node as its own community. The algorithm then looks to improve the modularity of the communities by assigning nodes to neighboring communities. If there are no improvements in modularity, the node will not change communities. Modularity gain, ΔQ is calculated as:

$$\Delta Q = \frac{k_{a,in}}{2m} - \gamma \frac{\Sigma_{tot} \cdot k_a}{2m^2} \tag{6}$$

where m represents the graph's size, $k_{a,in}$ is the sum of link weights from node a to the community nodes, Σ_{tot} is the total weight of links incident to community nodes, k_i is the sum of the weights of links incident to node a, and γ is the resolution parameter.

For directed networks, modularity gain is calculated as:

$$\Delta Q = \frac{k_{a,in}}{m} \gamma \frac{k_a^{out} \cdot \Sigma_{tot}^{in} + k_a^{in} \cdot \Sigma_{tot}^{out}}{m^2} \tag{7}$$

where k_a^{out} and k_a^{in} are the outer and inner weighted degrees of node a respectively, and Σ_{tot}^{in} and Σ_{tot}^{out} are the sums of weights of in-going and out-going links to community nodes respectively. This process continues until there are no more modularity improvements. In the next phase, a new network is constructed using the identified communities as nodes iteratively repeating the first step until no modularity improvements are observed.

- *Triadic Census.* In our research we utilized the triadic census [36] to understand the local structure and patterns of the Binance network. The triadic census is a method that classifies every combination of three vertices (triads) in the network into one of the sixteen types as shown in Figure 4.2. These insights provide additional insight into the local structure of the network. For the directed network, we will calculate the number of each of the sixteen possible triadic relations within the Binance network $BitN(A, T)$. Triadic configurations will be computed following Batagelj and Mrvar's methods [36]. This measure provides valuable insights into the reciprocity and transitivity patterns within the network which is crucial to understanding the local structures of the Binance network.

3.1 Data Collection

The fact that all information on the blockchain is publicly available does not mean that the data is suited for research. Instead of manually interacting with the Bitcoin blockchain and parsing the data into an adequate format, in this thesis, we take advantage of the transparent nature of the blockchain and utilize publicly available analytical tools for our analysis. Blockchair data dumps are pre-compiled datasets [37] that encapsulate all Bitcoin network activities in a structured format sorted by date. Instead of directly querying the Blockchair API, we will utilize the data dumps obtained from blockchair.com [37] as the primary data source to gather the transactional data of the Bitcoin blockchain. As opposed to using the live API, the data dumps allow us to obtain the complete network data for an entire day in two files. This approach eliminates the necessity to work with rate limits imposed by the API, thus simplifying the data-gathering process and enhancing efficiency. Even though various researchers have used Blockchair as a data source for cryptocurrency transactions [38, 39, 40], we plan to manually cross-test the collected data with the blockchain.info API to ensure the validity of our data. This is done by iterating through all unique transactions using their unique transaction hash [7] and comparing the inputs and outputs of the corresponding transaction with the returned information from Blockchain.com. In light of the collapse of FTX [41], Binance undertook measures to enhance transparency. The most evident action taken is the introduction of the proof of reserves system [42] which allows Binance users to verify their own assets by querying the Bitcoin blockchain. The public disclosure of Binance's wallets facilitates easier data collection. In addition, numerous public tools like blockchain.com [43] have already labeled previously known wallets for Binance. We will combine these sources to extract all transactions involving these addresses. We will consider all transactions on the 14th of May 2020.

3.1.1 Preprocessing

The raw data obtained from the data dumps consists of compressed input and output files, which are separated by day. Given our scope, we will limit our analysis to the data for a single day (May 14th, 2020). Each input and output file is stored in a comma-separated format that contains a set of attributes such as the time of the transaction, the transaction hash, the value of the transaction, and other related information. Given that a single transaction hash can correspond to multiple inputs and outputs, it is possible for a transaction hash to have multiple entries in the table. In order to build our graph, we require only the transaction hash and either the recipients or senders of transactions based on whether an input or output file is being processed. For pre-processing, we made use of Pyspark to read the compressed data and subsequently encoded it to the utf-8 format. To optimize data management, we implemented filtering to exclude transactions involving the transfer of bitcoins from and to the same address (self-loops), as they are not considered for our analysis. This is due to the UTXO system where the change of a transaction is incorporated as a subtransaction within a multi-transaction that is sent back to the same address [7]. To construct a graph from the inputs and outputs table we performed a cross-join operation while keeping only the rows that shared the same transaction hash across the two tables. This procedure resulted in a new table containing all sub-transactions corresponding to each unique transaction hash.

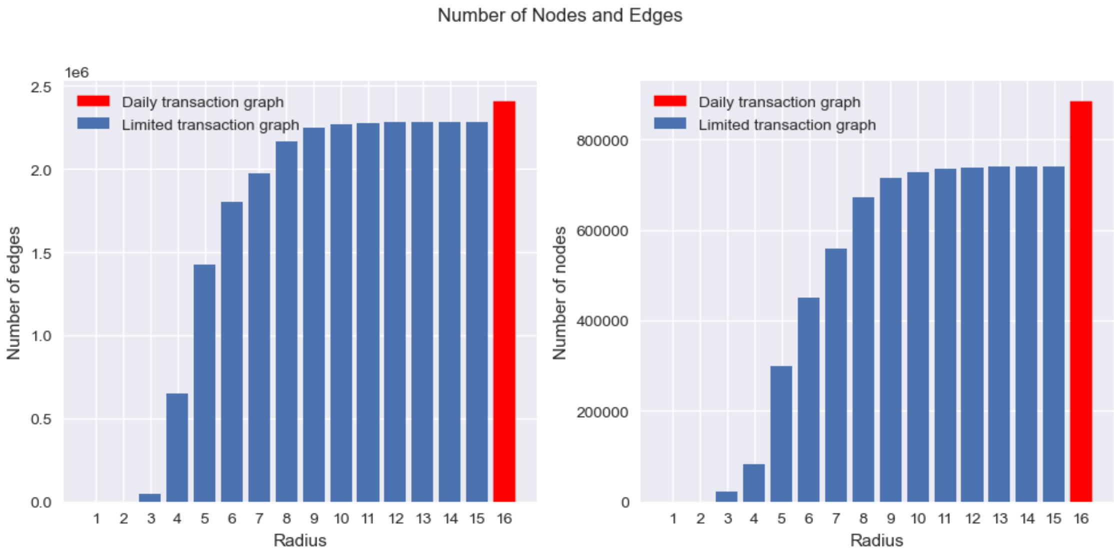

Figure 3.1: Nodes and edges in the network

3.2 Implementation

Following our preprocessing steps, the dataset becomes small enough to be loaded into memory. We use the toPandas() function available in Pyspark to construct a graph object using the NetworkX library in Python as the table resulting from our pre-processing is essentially an edge list and can be directly converted into a NetworkX object. To extract the Binance network, we utilize the ego_graph() function. This limits the graph by setting a maximum geodesic distance from the center node to control the number of nodes included in the induced subgraph. We will refer to this distance as the radius of the induced subgraph. We can see from Figure 3.1 that the number of edges and nodes approaches the daily transaction graph of May 14th, 2020 after a radius of 10. To clarify, the daily transaction graph is a snapshot of the network including all transactions that occurred within a day. The address of interest was selected based on two primary criteria: Bitcoin balance corresponding to the address and its transactional activity around the halving date on May 11, 2020. Our address of interest (*3M219KR5vEneNb47ewrPfWyb5jQ2DjxRP6*) is labeled as Binance 3 by blockchain.com. Known addresses like the ones associated with cryptocurrency exchanges are often already labeled by external parties such as blockchain.com. Binance 3, is the third largest public Bitcoin wallet address associated with Binance. The two largest addresses had no transactions on the halving date. Even though our chosen address also did not engage in transactions specifically on the halving date, it was the most active address close to the halving date. All network measurements were executed using the NetworkX library in Python.

3.2.1 Data validation

Bitcoin clients such as Bitcoin Core [44] provide access to all historical transactions on the Bitcoin blockchain. Since transactions are considered final once confirmed on the blockchain [7], we do not expect there to be noticeable differences between the data included in the Blockchair data dump compared to the data retrieved from the Blockchain.com API assuming the original source of the data is extracted using an authentic client like Bitcoin Core. However, to ensure thoroughness, we will still compare both data sources to identify any differences. After the transformations described in the previous section, we iterated through each unique transaction hash while comparing the corresponding inputs and outputs obtained from the Blockchair dump with those retrieved using the Blockchain.com API.

Chapter 4

Results and findings

4.1 Network description

The extracted Binance network has a total of 2,283,377 edges and 741,170 nodes. As we have seen before, this constitutes a large portion of all transactions on May 14th, 2020. The entire daily transaction represents all transactions that occurred on that day and consists of 884,759 nodes and 2,408,816 Million edges. In Chapter 3 we described our methodology for data validation, and the results showed that there was no significant discrepancy between the inputs and outputs obtained from our two data sources.

The total degree distribution as well as the distribution of in-degree and out-degree are depicted in Figure 4.1. We observe a large number of nodes with low degree values and only very few nodes in the network have high degree values. This is in line with earlier observations by researchers mentioned in Section 2. This suggests that there are a select few nodes that are more connected than other nodes in the network. The average clustering coefficient equates to 0.0136. This is relatively low and suggests that the majority of nodes in the Binance network have only a few connections rather than being connected to a large number of nodes. This is in line with Figure 4.1 as we can see a large number of nodes having low in and out degrees, with the frequency decreasing sharply as the degree values increase.

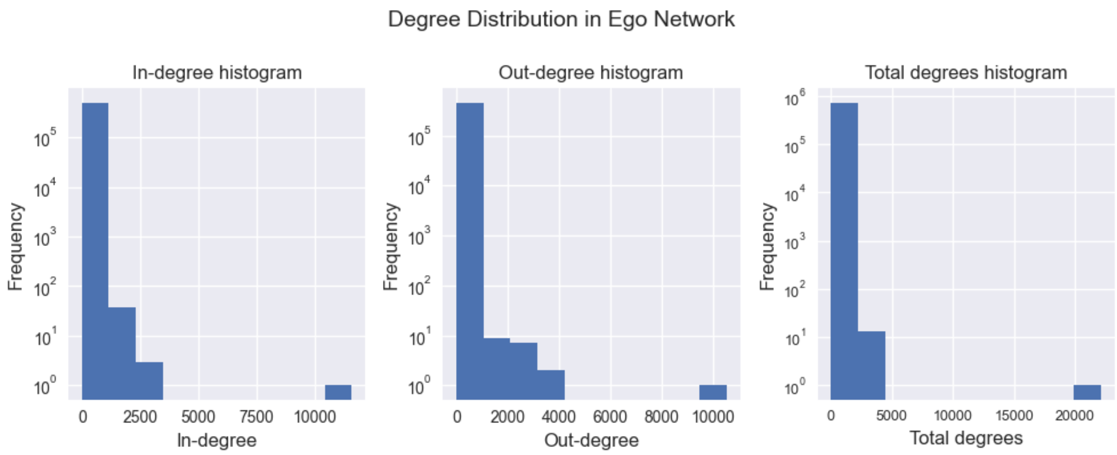

Figure 4.1: Degree distributions.

4.1.1 Triadic Census

The triadic census [36] results for the Binance network are presented in Table 4.3 and a visual representation of the different triad types is presented in Figure 4.2. We observe a decreasing trend in the number of triads per type as the number of connections in the triad increases. The most frequent triad type is 003 with a count of 65836 trillion. This indicates that many potential connections between nodes in the network are not formed, implying that the network is quite sparse. The second most frequent triad type is 012, with a count of 1.69 trillion. This type of triad represents a structure in which only a single node is connected to one other node in a group of three nodes. This finding suggests that many nodes act as intermediaries or bridges. In network theory, a bridge [45] is a node or edge that, if removed, increases the number of disconnected parts of the network. Bridges have the potential to enable information or resource flow from one part of the network to another. We will analyze the presence of bridges in Chapter 5. The third most frequent structure, type 102 describes two nodes that are connected to each other in both directions but not to the third. It suggests that many pairs of nodes have a strong relationship with each other but are not well-connected to the rest of the network. This finding suggests that the network may consist of communities as there are groups of nodes that are more closely connected to each other than the rest of the network. Other significant triad types include 021D, 021U, and 021C, which all involve two nodes that are connected to a third node in different ways. These types of triads can represent different patterns of fund flow or influence in the network. The least frequent triad types are 300 and 210. This suggests that closely connected triples are not as common in the network. This finding corresponds with the low clustering coefficient we measured as complex structures such as 300 and 210 are rare. This could imply that the network is not organized or clustered, or that communities are small and not tightly connected. However, the triadic census is just one aspect of the overall structure of the network.

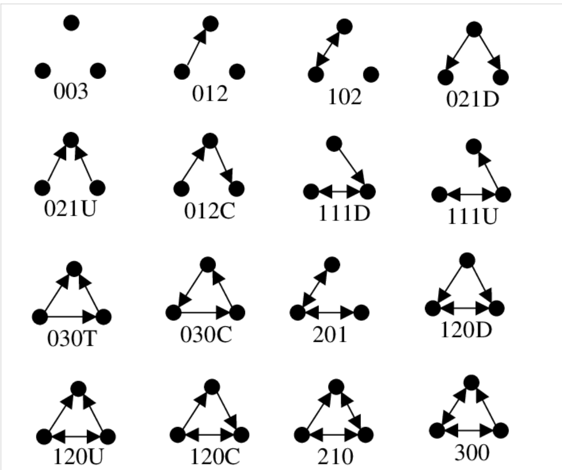

Figure 4.2: Overview of types of triads.

Figure 4.3: Triad census results.

Node	Triad census.
003	67856 Trillion
012	1.69 Trillion
102	559 Million
021D	197 Million
021U	208 Million
021C	137 Million
111D	71 Thousand
111U	59 Thousand
030T	73 Thousand
030C	1750
201	1086
120D	384
120U	1323
120C	274
210	204
300	204

4.1.2 Centrality

The three centrality measures, PageRank [31], degree [29], and eigenvector centrality [30], have been computed for the Binance network, and are presented in Tables 4.1, 4.2, and 4.3. The addresses are truncated to show only the first 7 characters to improve readability. Each of these centrality measures provides a different way of ranking nodes in a network based on their importance as we discussed in Section 3.

The PageRank centrality table shows that node 1NDyJtN has a significantly higher score compared to the other nodes (in full textit1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s). PageRank was originally developed by Google to rank web pages based on the number of links to a webpage where the algorithm assigns importance scores to a node based on the number and quality of incoming links to that node. The individual centrality scores are calculated based on the PageRank algorithm. This means that nodes with higher scores tend to be linked to other nodes that also have high PageRank scores. Therefore, nodes with higher PageRank scores are more likely to be involved in transactions with other nodes that are considered important in the network according to the PageRank centrality measure.

Similar to the PageRank results, the degree centrality table also shows that node 1NDyJtN has a much higher degree centrality score than the rest of the nodes. With node bc1qzu7 as the second most important, albeit with a considerable difference. Degree centrality simply evaluates the importance of a node based on its number of connections in the network as nodes with high scores are highly linked to other nodes. In other words, the highest-scoring nodes are the ones engaged in the most transactions on the Binance network and thus the most active and well-connected. It is possible that these nodes serve as intermediaries for transactions between nodes. One disadvantage of degree centrality is that the measure does not take into account the influence or importance of neighboring nodes [46].

Finally, the eigenvector centrality results show that the node with the highest eigenvector centrality value is also 1NDyJtN, followed by 1AozLV7 and 3LtpFPp. Here the first two addresses are significantly more important when compared to the other centrality measures where there was a more significant difference between the first and the second most important nodes. The eigenvector centrality measure can be used to identify nodes that have a high level of influence or importance in terms of transaction activity. Nodes with high eigenvector centrality are likely to have connections to other highly connected nodes, indicating their significance in the network.

There are many ways to assess the importance of nodes in a network and each measure provides a different perspective of the importance of the nodes. This is evident as the rankings of the three measures we investigated are not the same. Depending on the objective some centrality measures might be more suitable as they can reveal different aspects of a node's significance. It is important to consider multiple centralities since it is important to investigate the properties and roles of the nodes with high centrality values to gain a deeper understanding of the network's structure and dynamics.

Table 4.1: Pagerank centrality results.

Node	**Pagerank**
1NDyJtN	0.00631
bc1qzu7	0.00234
bc1qx57	0.00181
3Qf4e6M	0.00176
1AHkzeT	0.00170
1AozLV7	0.00141
bc1qtga	0.00110
bc1qqk6	0.00108
bc1q8cv	0.00103
bc1q0tq	0.00101

Table 4.2: Degree centrality results.

Node	**Degree**
1NDyJtN	0.02987
1GX28yL	0.00550
3FxUA8g	0.00498
bc1qx57	0.00435
3Qf4e6M	0.00423
3JXRVxh	0.00410
1AHkzeT	0.00401
34epuEj	0.00388
1HckjUp	0.00384
3LrytkU	0.00371

Table 4.3: Eigenvector centrality results.

Node	**Eigenvector**
1NDyJtN	0.12247
1AozLV7	0.09429
3LtpFPp	0.04658
12oMTi6	0.04076
1GNy7Eu	0.03579
36hLx7H	0.03543
12xjTvg	0.03532
3Qf4e6M	0.03452
1JsfNPN	0.03307
19Kedre	0.03126

4.1.3 Community detection

The calculated modularity for the directed network is 0.693. Suggesting that the Louvain method has successfully identified communities. The modularity score implies that the cohesiveness and quality of these communities could be classified as slightly above average.

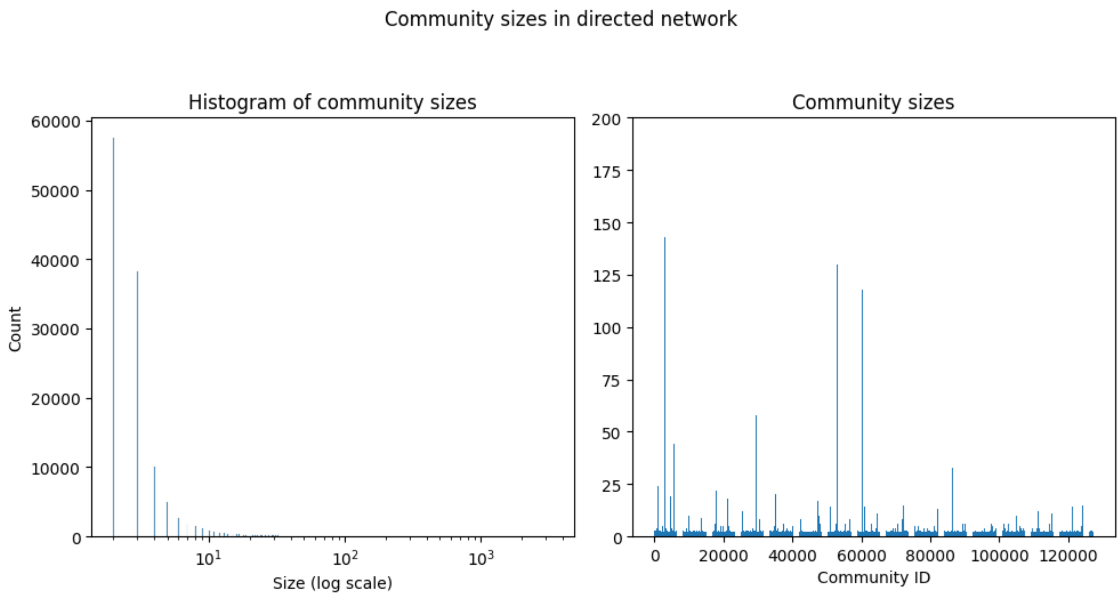

Figure 4.4: Community sizes in the directed network.

In Figures 4.5 we can see visualizations of 12 randomly selected communities. The size of each node in the visualization is determined by its degree centrality, while their node positions are calculated using the Kamada-Kawai layout algorithm. We note the presence of central nodes having edges directed towards many other nodes or vice versa. These highly connected nodes could be exchanges or gambling websites receiving or sending a large number of transactions. Overall, the community detection algorithm on the Binance network results in structures with a single node having a large degree value. This pattern is mainly a consequence of the directional information being incorporated into the community clustering

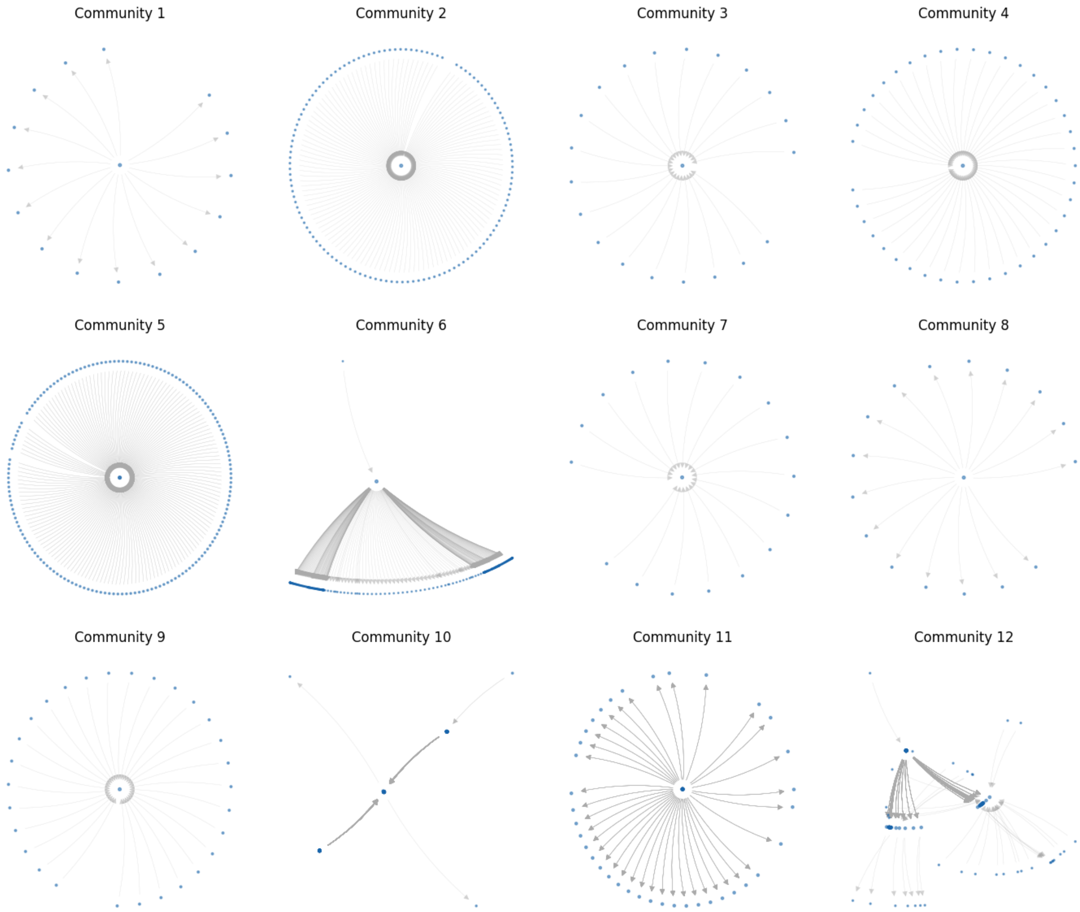

Figure 4.5: Communities in the directed network.

Chapter 5

Discussion and conclusion

In this Chapter, we will discuss the findings and their implications focusing on answering our three core research questions:

- What insights can we gain into the visual representations and distribution of communities within the Binance network?
- What are the most influential nodes or addresses within the Binance network, as evaluated by degree, eigenvector, and PageRank centrality measures?
- How can we describe the structural composition of the Binance network through the use of network statistics?

In our study, we used centrality measures and other network measures, to provide insights into transaction patterns and key entities within the network, to answer these questions. We will first discuss and present our findings, followed by addressing limitations and suggesting future work. Finally, we conclude the thesis with a summary of our key findings.

5.1 Interpretation of results

In Chapter 4 we observed that the Louvain method identified many smaller-sized communities. These communities mainly exhibited a structure where a single, highly connected node with a high degree value was linked to the other nodes in the community. Nonetheless, this characterization of the communities within the Binance network might not be indicative of the more intricate relationships and complex structures found in a transaction network of this size. Therefore, given the exploratory nature of our study, we will repeat our community detection methodology on an undirected version of the network. The directed interpretation allows us to analyze the flow of Bitcoin transactions between nodes, offering insights into the directionality of transactions. In contrast, the undirected interpretation provides an overview of the overall structure and relationships between nodes, independent of transaction directions. A visualization of 12 randomly selected communities is depicted in Figure 5.1.

The calculated modularity score for the undirected network is 0.874. We observe fewer but larger communities when compared to the directed network. In Figure 5.2 we can clearly observe two different types of communities, smaller communities where the first type mostly consists of 10 to 100 nodes. While the second type consists mostly of communities that are not larger than 15 nodes. This is due to the difference in the calculation of modularity between the directed and the undirected graphs. While in an undirected graph, the modularity considers only the weight (in this case amount of edges since all weights are equal). In the directed case, modularity also includes direction to be included in the community calculation. This optimizes the modularity function to identify more densely connected groups of nodes with stronger intra-community edges than inter-community edges. Using the Louvain method we observe more complex structures when we interpret the Binance network as an undirected network with sub-communities that could represent groups of nodes that are involved in similar types of transactions. This could for example be the group of customers of a certain exchange.

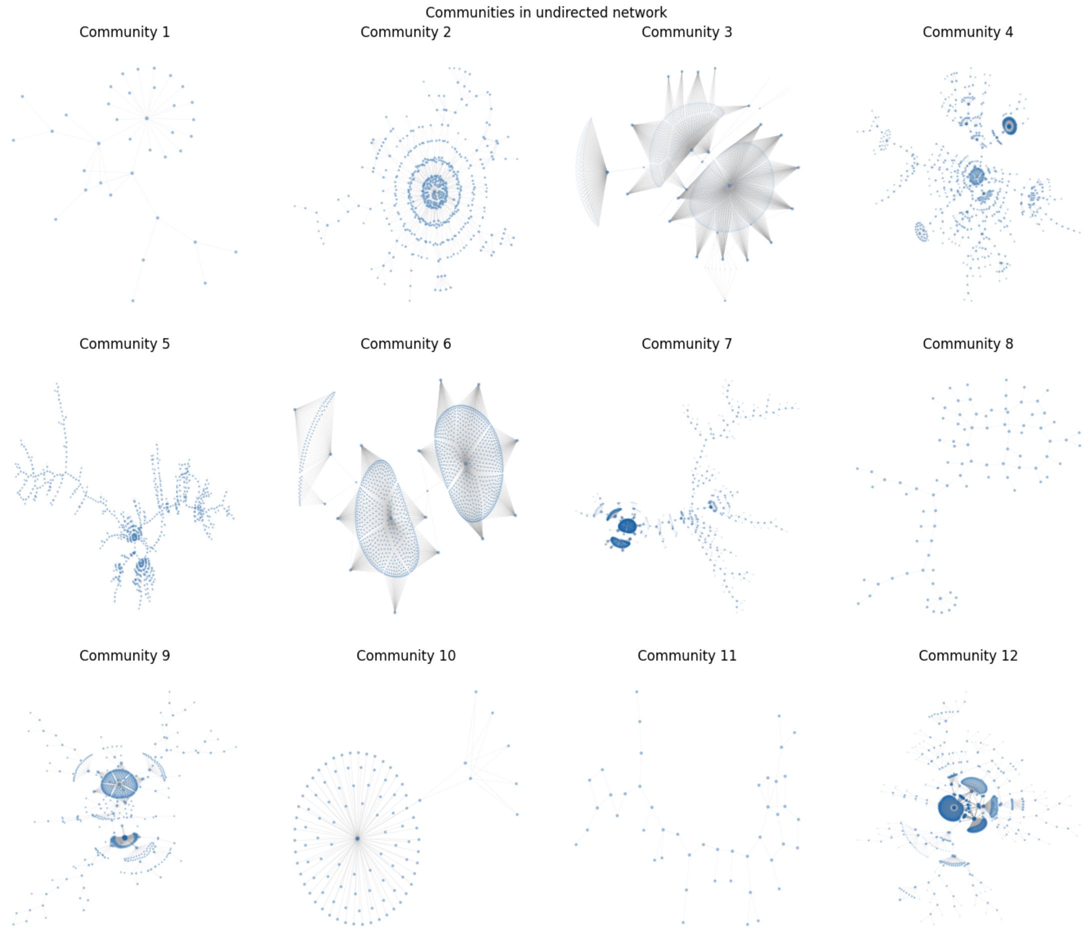

Figure 5.1: Communities in the undirected network.

In Figure 5.3, communities 7, 10, and 9 are visualized with node sizes based on three different centrality measures. We observe that PageRank and degree centrality visualizations are more similar when compared to eigenvector centrality. In the context of the Binance network, degree centrality measures the direct connections of a user, in other words, how many counterparties a user has transacted with on the 14th of May 2020. On the other hand, PageRank, a variant of Eigenvector centrality calculates scores using a more complex calculation as the measure accounts for the global network structure. PageRank assesses the significance of a user based on both the quantity and quality of their transactions. This means that if an address frequently transacts with nodes with a high PageRank centrality, that address will also have a higher PageRank centrality score. Thus, nodes with high PageRank scores are typically influential, well-connected entities within the Binance network, possibly serving as transaction hubs or hotspots in the network. However, more research is needed to confirm the existence and analyze the influence of hubs. Eigenvector centrality is calculated using an iterative process using the dominant eigenvector of the adjacency matrix of the network. In the context of the Binance network, much like PageRank centrality, a node with high eigenvector centrality is not only involved in many transactions but is also transacting with other nodes who themselves are highly influential or well-connected. As depicted in Figure 5.3, the eigenvector centrality seems to assign lower scores to nodes that are

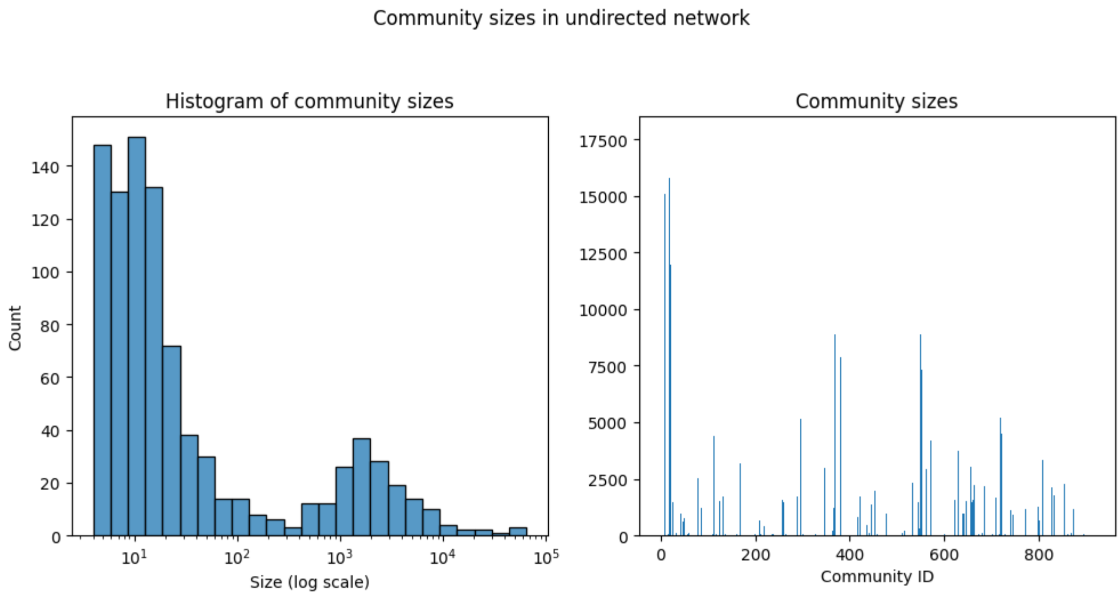

Figure 5.2: Community sizes in the undirected network.

sparsely connected nodes or belong to smaller sub-communities. In contrast, the eigenvector centrality assigns higher scores to nodes that are more densely connected and are part of larger sub-communities. This result differs from the other two centrality measures. Therefore the eigenvector centrality offers a more pronounced separation between nodes of high influence and nodes of low influence within the network.

One of the objectives of our research was to identify the most influential nodes or addresses within the Binance network, as evaluated by degree, eigenvector, and PageRank centrality measures. Based on the results presented in Table 4.1, 4.2, and 4.3, it becomes evident that according to all three centrality measures, the address *1NDyJtN* is the most influential node within the network. Not only was this address involved in the most transactions compared to any other node, but it also ranked the highest in terms of the quality of its connections according to the PageRank and Eigenvector centrality. Meaning that the *1NDyJtN* is well-connected and influential across the entire network on the 14th of May 2020, frequently transacting with other prominent nodes. In the context of the Binance network, it means that the address has been involved in many transactions with influential counterparties. These nodes could potentially be key players in the network, such as cryptocurrency exchanges or gambling services. These services are crucial as they facilitate the flow of government-issued currencies into and out of the Binance network. Future research could, for instance, focus on analyzing illicit activities [47] by tracing the movement of suspicious funds to and from an exchange. Such investigations and other further research will be discussed in detail in Section 5.2.

To address our research question regarding the structural composition of the network, we will take into account the insights provided by descriptive network statistics. The average clustering is observed to be 0.0137 which is relatively low and indicates a generally low connectivity between nodes in the network. In the Binance network, this means that users often do not transact with many different counterparties but rather with a select few users or services. We have seen that this is due to a few reasons. Many of the nodes have low degrees with only a select few nodes having higher degrees. This means that there is less connectivity between the majority of the nodes and only a select few nodes have high degree values. This finding might indicate that there are key players in the network who have more influence and transactions in the network. This is evident as we observe that most of the communities found in the directed Binance network exhibit a structure characterized by single nodes with connections to many other nodes or many nodes

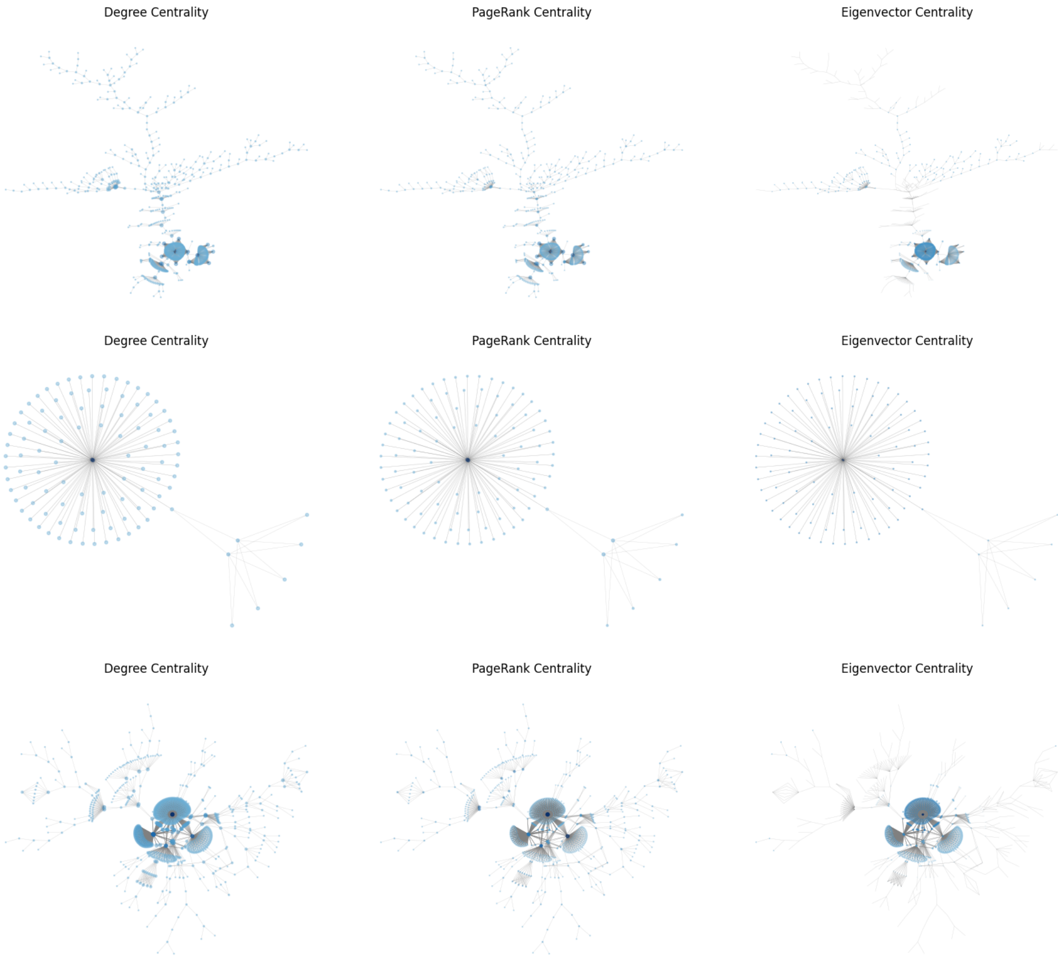

Figure 5.3: Centrality measures.

connecting to a single one.

We also observe long paths in the undirected network where many nodes are only connected by one other node which also contributes to the low average clustering score. In addition, our findings presented in Chapter 4 suggested the possibility of bridge nodes existing within the network due to the high occurrence of type 021 in the triadic census, a pattern indicating a single directed edge between two nodes in a triad. To further analyze the long paths and the triadic structure we use the methods presented by Tarjan [45] to identify the bridges in the network, and found that as much as 39% of nodes within the network qualify as bridges. This finding aligns with the long paths and the low clustering score and may indicate potential inefficiencies in the network as direct transactions could potentially be more efficient. The presence of long paths is interesting and warrants more research as it could be interpreted in multiple ways.

One explanation might involve users on the blockchain that are actively obfuscating their transaction trail for illicit activities by repeatedly sending transactions from one address to another creating long paths. There can be multiple explanations for these structures as suggested by previous studies, for instance, Maessa et. al [48] have characterized the long paths in the Bitcoin network. The authors also suggested that some transactions might not reflect traditional payments and instead, serve as the primary method of updating the state of the blockchain. These might be for instance fund management tasks such as internal transfers or payments from network participants such as exchanges or gambling services. However, the authors note that these explanations only remain speculative, and further research is needed to establish definitive conclusions.

5.2 Limitations and Future Work

The most significant limitation of this study compared to previous studies is the time horizon of the analysis. Given the limitations of resources, we only studied a single day of transactions. Therefore the results and insights we generated might be prone to sampling bias and will be limited by the lack of generalizability of our findings. This might also explain the differences in our results when compared to for instance the results obtained by Baumann et al. [17] where the authors observed a high global clustering coefficient. It is likely that the results generated from a single day might not be generalizable over longer periods of time. Further research is needed to confirm and validate our findings, which could involve a more comprehensive analysis of the behavior and characteristics of the Binance network over different periods of time.

While we identified influential nodes by assessing the quality and quantity of transactions the node has participated in, further research could look into the interpretation of an important or influential node. This could include taking into account the transaction size, frequency, or examining the role of a node to determine if it acts as a broker or a hub within the network. Such an analysis would contribute to our understanding of different roles and their behavior within the network. This is particularly interesting as it allows us to study the behavior of different participants and their responses to particular events such as the halving date as their influence and actions could affect the entire network.

Although our analysis focused on a limited and simplified subset of the network spanning only one day while utilizing ego data, the Bitcoin network is much more complex. Our depth of research may be too limited to make definitive conclusions. Further research could include a more exhaustive analysis of the communities on a larger scale such as performing analysis on individual communities and comparing the characteristics and dynamics between communities and classifying them into different types. The investigation of the long paths we observed in combination with the analysis of different roles certain nodes have within the network potentially lead to a significant study. This is due to the possible correlation between these two elements, as transactions need to pass through nodes associated with an exchange or other service to convert Bitcoin into fiat currencies. This could prove useful in the context of forensic analysis where the interest lies in tracking the movement of illicit funds within the network.

In addition, it may be interesting to compare the Binance network a set of similar-sized random networks as a baseline to study and benchmark the unique properties that the Binance network exhibits. We interpreted the Binance network as a static network, however since transactions are dynamic, this might not accurately reflect the behavior and characteristics of the network. It would be interesting to study the evolving nature of the network by examining the dynamic relationships such as the change in transaction patterns over time, and the overall network growth over time due to the time-dependent characteristics of the network.

5.3 Conclusion

In conclusion, we studied the daily transaction graph of the Binance network on the 14th of May 2020. Our main objective was to explore several fundamental properties and key characteristics of the Binance network during our timeframe of interest. Furthermore, we conducted an exploratory study to analyze the overall structure of the communities within the network. We developed an approach utilizing publically available data and employed distributed computing to extract and transform blockchain data to ensure its suitability for analysis. In our analysis, we investigated the degree distributions in the network and observed two peaks in Figure 5.2. This finding indicates that while most nodes exhibit a low degree value, there are specific nodes with high degree values. We then investigated the triad census of the Binance network, centrality measures, and explored the communities of both the directed and the undirected representations of the network. The triad census results showed that most of the nodes in the network are not interconnected. Most nodes exhibit either no connections or are connected to only a single other node resulting in long paths. These findings contradict previous research as the observed clustering coefficient of the Binance network is relatively low. The three centrality measures identified *1NDyJtN* as the most influential node in the network, considering both the number and the quality of its connections.

When analyzing the communities within the network we observed a large number of small communities. These communities exhibited a structure where one node was involved in many transactions, while the remaining nodes participated in only a few transactions. However, when performing community detection on an undirected interpretation of the network, we find larger, more cohesive structures. Some long paths we have seen serve as bridges in the Binance network. This approach differs from the directed communities as we no longer consider directional information within the network. Instead, the communities are formed based on their connections, regardless of the direction of transactions.

It is important to acknowledge the limitations of this study. Our analysis only focused on the transaction data of a single day while utilizing ego data and interpreting the Binance network as a static network. The dynamic nature of transactions within the network needs further investigation to understand its complex and dynamic behavior. For instance, exploring time-dependent relationships, changes in transaction patterns over time, and overall network growth. Future research should account for these limitations and perform more comprehensive investigations into the complexities of the Binance network. In summary, this study provides an exploratory overview of the properties and characteristics of the Binance network on 14th May 2020. However, further research is necessary to enhance our understanding, particularly with regard to time-dependent analysis, given the exploratory nature of this investigation.

Bibliography

[1] Sergey N Dorogovtsev and Jose FF Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002. 1

[2] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013. 1

[3] Natkamon Tovanich, Nicolas Heulot, Jean-Daniel Fekete, and Petra Isenberg. Visualization of blockchain data: A systematic review. *IEEE Transactions on Visualization and Computer Graphics*, 27:3135–3152, 2019. 1, 4

[4] Ayana T Aspembitova, Ling Feng, and Lock Yue Chew. Behavioral structure of users in cryptocurrency market. *Plos one*, 16(1):e0242600, 2021. 1

[5] Jiajing Wu, Jieli Liu, Yijing Zhao, and Zibin Zheng. Analysis of cryptocurrency transactions from a network perspective: An overview. *Journal of Network and Computer Applications*, 190:103139, 2021. 1, 4

[6] Joseph Poon and Thaddeus Dryja. The bitcoin lightning network: Scalable off-chain instant payments, 2016. 1

[7] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008. 1, 2, 4, 7, 8, 9

[8] Charles Guandaru Kamau. The cryptocurrency market in kenya: A review of awareness and participation by the youth. *Journal of Asian Business Strategy*, 2022. 1

[9] Rfi. Central african republic adopts bitcoin as legal currency, Apr 2022. 1

[10] Pengcheng Xia, Haoyu Wang, Bowen Zhang, Ru Ji, Bingyu Gao, Lei Wu, Xiapu Luo, and Guoai Xu. Characterizing cryptocurrency exchange scams. *Computers Security*, 98:101993, 2020. 1

[11] The block. Spot archives: https://www.theblock.co/data/crypto-markets/spot. 1

[12] Peter D DeVries. An analysis of cryptocurrency, bitcoin, and the future. *International Journal of Business Management and Commerce*, 1(2):1–9, 2016. 2

[13] Alex Hughes, Andrew Park, Jan Kietzmann, and Chris Archer-Brown. Beyond bitcoin: What blockchain and distributed ledger technologies mean for firms. *Business Horizons*, 62(3):273–281, 2019. 2

[14] Dániel Kondor, Márton Pósfai, István Csabai, and Gábor Vattay. Do the rich get richer? an empirical analysis of the bitcoin transaction network. *PloS one*, 9(2):e86197, 2014. 3, 4, 5

[15] Dániel Kondor, István Csabai, János Szüle, Márton Pósfai, and Gábor Vattay. Inferring the interplay between network structure and market effects in bitcoin. *New Journal of Physics*, 16(12):125003, 2014. 3

[16] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002. 3

[17] Annika Baumann, Benjamin Fabian, and Matthias Lischke. Exploring the bitcoin network. *WEBIST (1)*, 2014:369–374, 2014. 3, 19

[18] Damiano Di Francesco Maesa, Andrea Marino, and Laura Ricci. Uncovering the bitcoin blockchain: an analysis of the full users graph. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 537–546. IEEE, 2016. 3

[19] Damiano Di Francesco Maesa, Andrea Marino, and Laura Ricci. Detecting artificial behaviours in the bitcoin users graph. *Online Social Networks and Media*, 3-4:63–74, 2017. 3

[20] Israa Alqassem, Iyad Rahwan, and Davor Svetinovic. The anti-social system properties: Bitcoin network data analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(1):21–31, 2018. 3

[21] Yuexin Xiang, Tiantian Li, and Yuquan Li. Leveraging subgraph structure for exploration and analysis of bitcoin address. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1957–1962, 2022. 4

[22] Fergal Reid and Martin Harrigan. *An analysis of anonymity in the bitcoin system.* Springer, 2013. 4

[23] Tim Ruffing, Pedro Moreno-Sanchez, and Aniket Kate. Coinshuffle: Practical decentralized coin mixing for bitcoin. In *Computer Security-ESORICS 2014: 19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11, 2014. Proceedings, Part II 19*, pages 345–364. Springer, 2014. 4

[24] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography and Data Security: 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers 17*, pages 6–24. Springer, 2013. 4

[25] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140, 2013. 4

[26] Elli Androulaki, Ghassan O Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. Evaluating user privacy in bitcoin. In *Financial Cryptography and Data Security: 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers 17*, pages 34–51. Springer, 2013. 4

[27] May 2023. 4

[28] Linyuan Lü, Duanbing Chen, Xiao-Long Ren, Qian-Ming Zhang, Yi-Cheng Zhang, and Tao Zhou. Vital nodes identification in complex networks. *Physics Reports*, 650:1–63, 2016. Vital nodes identification in complex networks. 5

[29] Linton C Freeman et al. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge*, 1:238–263, 2002. 6, 12

[30] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182, 1987. 6, 12

[31] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford infolab, 1999. 6, 12

[32] Mark Newman. *Networks.* Oxford university press, 2018. 6

[33] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008. 7

[34] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):5233, 2019. 7

[35] Nicolas Dugué and Anthony Perez. *Directed Louvain: maximizing modularity in directed networks*. PhD thesis, Université d'Orléans, 2015. 7

[36] Vladimir Batagelj and Andrej Mrvar. A subquadratic triad census algorithm for large sparse networks with small maximum degree. *Social networks*, 23(3):237–243, 2001. 7, 11

[37] Blockchain api documentation: https://blockchair.com/api/docs. 7

[38] Luis P de la Horra, Gabriel de la Fuente, and Javier Perote. The drivers of bitcoin demand: A short and long-run analysis. *International Review of Financial Analysis*, 62:21–34, 2019. 7

[39] Alex De Vries. Renewable energy will not solve bitcoin's sustainability problem. *Joule*, 3(4):893–898, 2019. 7

[40] Malte Möser, Rainer Böhme, and Dominic Breuker. An inquiry into money laundering tools in the bitcoin ecosystem. In *2013 APWG eCrime researchers summit*, pages 1–14. Ieee, 2013. 7

[41] Alex Hern. What do we know so far about collapse of crypto exchange ftx?, Nov 2022. 8

[42] Binance proof of reserves (por): Merkle tree. https://www.binance.com/en/proof-of-reserves. 8

[43] Live blockchain transactions: https://www.blockchain.com/explorer/mempool/btc. 8

[44] Rostislav Skudnov. Bitcoin clients. 2012. 9

[45] R.Endre Tarjan. A note on finding the bridges of a graph. *Information Processing Letters*, 2(6):160–161, 1974. 11, 18

[46] Francis Bloch, Matthew O Jackson, and Pietro Tebaldi. Centrality measures in networks. *Social Choice and Welfare*, pages 1–41, 2023. 12

[47] Yan Wu, Fang Tao, Lu Liu, Jiayan Gu, John Panneerselvam, Rongbo Zhu, and Mohammad Nasir Shahzad. A bitcoin transaction network analytic method for future blockchain forensic investigation. *IEEE Transactions on Network Science and Engineering*, 8(2):1230–1241, 2020. 17

[48] Damiano Di Francesco Maesa, Andrea Marino, and Laura Ricci. An analysis of the bitcoin users graph: inferring unusual behaviours. In *Complex Networks & Their Applications V: Proceedings of the 5th International Workshop on Complex Networks and their Applications (COMPLEX NETWORKS 2016)*, pages 749–760. Springer, 2017. 18