Imperial College London

Department of Computing

Community-based identification of important entities on the Ethereum blockchain.

Author:
Ivan Gorshkov

Supervisor:
Dr. Arthur Gervais

Submitted in partial fulfillment of the requirements for the MSc degree in MSc Computing of Imperial College London

September 2021

Abstract

This project focuses on detecting influential entities on the Ethereum blockchain. In particular, it models USD Tether Ethereum ERC-20 token transactions as a complex network and uses degree, closeness, betweenness and PageRank centralities to identify the most influential addresses in the network. Furthermore, the study combines centrality algorithms with the Louvain community detection algorithm to determine the extent to which the most influential addresses tend to cluster with each other. It also uses the Louvain algorithm in conjunction with the centrality metrics to assess how the relative importance of the most influential addresses changes on the community level. The experimental results show that the USDT transaction network exhibits a strong centralisation tendency. However, a significant clustering tendency among the influential entities was not identified. Furthermore, it was found that the relative importance of addresses with high PageRank and Betweenness centrality tends to reduce on the community level.

Acknowledgments

I would like to thank the following people for helping me throughout this project:

- Dr. Arthur Gervais, for his advice and for introducing me to people that helped me throughout the project.
- My second marker, Professor William Knottenbelt, for his support, as well as his course, Introduction to C++ Programming, which taught me a lot fundamental programming principles.

Contents

Chapter 1

Introduction

1.1 Motivation

The first cryptocurrency, Bitcoin was introduced in 2008 by Nakamoto [1]. In contrast to traditional financial instruments, blockchain-based currencies do not require a trusted third party to execute transactions.

Decentralisation is one of the key ideas behind cryptocurrencies. However, previous studies have shown that they might not exhibit this property. A small number of addresses accounts for a large proportion of the transactions. Sui et al. [2] used PageRank and Hubs and Authorities algorithms to identify the most influential addresses on Ethereum blockchain. They found that out of the top 50 influential addresses 58% are controlled by centralised exchanges, and 12% are related to the mining pools.

These findings highlight the importance of identifying and studying the most important entities performing transactions on the blockchain. Identifying such entities can help us to develop a better understanding of the behaviour of blockchain users and the activity of companies providing cryptocurrency-related services.

In contrast to the traditional firms operating in the financial sector, companies that use blockchain to deliver their services are often more transparent. This is due to the fact that data stored on blockchains is often publicly available. Therefore, it presents a unique opportunity to study the phenomenon of interconnectedness, centralisation and influence in the financial markets.

For instance, token transaction data from the Ethereum blockchain introduced by Buterin [3] is publicly available. Identifying the most influential users of Ethereum-based tokens is of particular interest. Ethereum is the second-largest cryptocurrency by market capitalisation and it allows for the creation of smart contracts. Some of them can potentially be considered to be influential entities as well.

Moreover, identifying the most influential entities performing transactions on the blockchain can have real-world applications. For instance, cyberattacks affecting the key players in the blockchain ecosystem can have an impact on less influential entities that are related to it and the whole ecosystem. Therefore, identifying such entities may allow us to introduce preventive measures where possible. Furthermore, identifying the key players can be useful for market research purposes and targeted on-chain advertising. Wu et al., [4] proposed a methodology that employs Bitcoin and Ethereum transaction data for targeting ads. Identifying the most popular services combined with this methodology can be used to understand user preferences. Finally, studying the behaviour of the most influential entities and their relationship to each other can be used to detect collaborations between them.

The project has the following objectives:

1. Describe a set of tools that can be used to quantify and to compare the influence of entities operating on the Ethereum blockchain.

2. Test and compare influence metrics.

3. Use Ethereum blockchain data to rank the companies and blockchain protocols by their influence.

4. Determine the extent to which influential addresses tend to cluster with other influential addresses.

5. Identify communities of users that cluster around the most influential addresses.

6. Propose a way to identify addresses that are not important in the network as a whole, but are important inside the smaller address groups.

1.2 Project Overview

This study employs the tools developed in the network science field to identify the most influential addresses performing transactions on the Ethereum blockchain. In particular, it uses the transaction data of USD Tether (USDT), the second biggest ERC-20 token by market capitalisation. This section briefly describes the approach employed in this project.

1.2.1 Modelling a network of transactions

Ethereum blockchain can be regarded as a complex system of interacting entities. Identifying the key actors in the system of interacting entities is a common problem in many academic disciplines including ecology, economics and medicine. Such systems can be modelled as complex networks where network nodes represent entities and edges represent the relationship between entities. Therefore, this study models USDT transactions as a complex network. To satisfy the requirements of the algorithms employed in this study, three different versions of the transaction network were used: directed weighted, directed unweighted and undirected unweighted.

1.2.2 Measuring influence

To identify the most influential addresses engaged in USDT token transactions the study uses four different network centrality metrics. In particular, degree centrality, closeness centrality, betweenness centrality and PageRank. The study calculates the centrality metrics for every address in the dataset and presents the influence score distribution for every centrality metric. The results are used to assess the extent to which the USDT transaction network exhibits the centralisation tendency. Furthermore, the study compares the centralisation metrics based on their Pearson correlation coefficient and uses the Spearman rank correlation coefficient to compare the rankings produced by different centrality metrics. Analysing the relationship between the centrality metrics allows us to derive insights into the overall structure of the transaction network and the behaviour of entities using USDT.

1.2.3 Detecting Communities

The study attempts to analyse the behaviour of the most influential entities in the USDT transaction network and to identify entities that rely on it to perform their operations. To achieve this objective, we cluster all addresses in our dataset into different groups based on their relationship with each other. The study employs the Louvain community detection algorithm for this purpose.

After detecting the communities of addresses, we identify the communities that contain the top 100 most influential addresses according to each centrality metric. This allows us to analyse the extent to which the most influential addresses tend to cluster with each other.

1.2.4 Detecting Locally Important Entities

After detecting the most influential addresses and identifying the communities that they belong to, we determine whether the most influential addresses remain influential on the community level. This is done by creating a subgraph for each community and calculating centrality scores for every address in each community based on its relationships with other entities within its community. These locally important nodes are then compared to nodes that are the most influential on the global scale.

Then, based on the findings of this section, we discuss the usefulness of combining centrality and community detection algorithms to detect the most influential addresses on the Ethereum blockchain. Furthermore, the results are used to assess whether the most influential addresses in the network tend to create clusters of less influential around them.

The rest of the project is organised as follows. At first, a technical implementation of the project will be described. This will be followed by the section that outlines the relevant academic literature. Then, project contributions will be listed. The contribution section will be followed by the background section which outlines the key definitions and concepts used in this study. Afterwards, data and analysis sections will describe data sources and present the analysis. Finally, a discussion section will contain a description of the project results and limitations. The report will conclude with a short summary of findings.

Chapter 2

Implementation

Figure 2.1 presents a schematic representation of the data collection and analysis process. Golang was used to download transaction data. Go-Ethereum library was employed for this purpose. The data was downloaded from the Ethereum archive node provided by the Imperial College London Department of Computing. An algorithm that searches for the required data on blockchain was developed for this purpose.

To download address label data, a Python script that collects data from Etherscan.com was developed. It uses the BeautifulSoup package. In addition, some of the address labels were provided by the Imperial College London Department of Computing. These addresses were stored in the PostgreSQL database and SQL combined with Python was used to extract this data and load it into the Neo4j database.

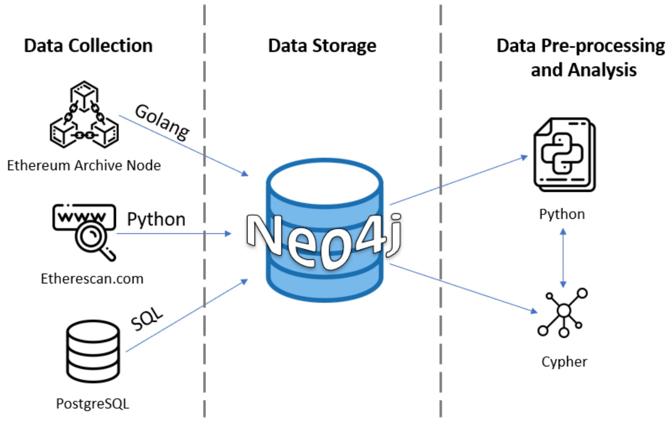

Figure 2.1: Implementation.

Data from the Ethereum archive node and Etherscan.com were stored in the Neo4j database. Neo4j is a graph database that stores data as nodes and edges. A graph database was chosen, since storing social network data this way makes it easier to perform pre-processing and analysis on graphs.

The data were analysed with Python and Cypher. Cypher is a querying language used in Neo4j. Cypher language combined with built-in Neo4j algorithms was used to calculate degree, closeness, betweenness and PageRank centralities. Neo4j built-in algorithms were also used to identify weakly connected components and to partition the transaction network into communities with the Louvain algorithm. Python was used to clean the data, calculate correlation coefficients between the centrality metrics and visualise the results.

Chapter 3

Literature Review

This section describes the academic literature that deals with the applications of network science to the transactions on blockchains. At first, it describes research papers that attempted to model blockchain transactions as a network and analysed its general properties. Secondly, it covers the literature that measured address centrality on blockchain transaction graphs. Then, it lists the studies that used community detection algorithms to identify communities on blockchain transaction graphs. Lastly, it describes the literature that focuses on combining centrality and community detection algorithms.

3.1 Graph Representation of Blockchain Transactions

Tasca, Liu and Hayes [5] clustered Bitcoin addresses into different business categories and used them to model a network of payments between them. This enabled them to analyse how the Bitcoin network evolved with time. They have identified three key stages in the development of Bitcoin: early prototype, ”sin” enterprise and a progression from ”sin” to a legitimate enterprise. Alqassem, Rahwan and Svetinovic [6] also analysed the Bitcoin transaction graph from its inception. Their findings show that the graph exhibits many dynamics and properties that are common for other social networks. However, the anonymity-seeking behaviour of Bitcoin users significantly affects the network structure. In particular, it increases its diameter, creates tree-like communities and leads to the creation of many accounts with zero cumulative balances.

Chen et al.[7] studied Ethereum blockchain via graph analysis. They have constructed three different graphs: a money flow graph, a smart contract creation graph and a smart contract invocation graph. They have used their analysis to propose attack forensics, anomaly detection and deanonymisation methods. Furthermore, Guo, Dong and Wang [8] modelled Ethereum blockchain as a graph and studied its statistical properties. In particular, they found that Ethereum transaction volume and node degree distributions can be approximated by the power law.

3.2 Network Centrality

Networks of interacting entities often contain actors that are relatively more important than other entities in the network. Algorithms used to identify influential entities have applications in a lot of different academic disciplines. For example, Ashtiani et al., [9] used network centrality metrics to study protein-proten interaction networks. Also, Ahern [10] used eigenvector centrality to study stock returns. He found that stocks in more central industries have greater market risk.

Network centrality algorithms were used to study blockchain transaction networks as well. Kumar [11] used degree centrality, betweenness centrality, eigen vector centrality and PageRank to study the Ethereum transaction graph. He analysed how these techniques can be used to breach transaction privacy. Sui et al. [2] used PageRank and Hubs and Authorities algorithms to study the transactions of the Ethereum blockchain native currency. They found that out of the top 50 influential addresses 58% are controlled by centralised exchanges, and 12% are related to the mining pools.

3.3 Community Detection

Real-world systems of interacting entities often contain community structures. According to Fortunato [12], graphs can be used to model these structures. Graphs were first employed to model communities in the early 1970s. Algorithms used to partition graph vertices into groups are called community detection algorithms.

A wide research study used community detection algorithms to understand the properties of the complex systems of interacting entities. For instance, Girvan and Newman [13] used the betweenness centrality metric to identify nodes that serve as bridges between the different parts of the network and determined community boundaries based on them. Their algorithm was tested on the collaboration network of scientists and a food web of marine organisms. It was shown that the algorithm is good at capturing the known community structures.

Community detection algorithms were also used to identify blockchain addresses sharing similar properties. For instance, Prado-Romero et al., [14] used the Louvain algorithm to develop a tool that identifies communities on the Bitcoin transaction network. Then, they searched for anomalies within those communities to identify coin mixers. Wu et al., [15] used community detection to cluster Bitcoin and Ethereum users and proposed a methodology that can be used to deliver targeted on-chain advertising.

3.4 Community Detection and Centrality

A number of studies have attempted to combine community detection algorithms with centrality metrics to enhance the quality of community detection and to identify influential entities on the local scale. For instance, Ahajjam et al., [16] used eigenvector centrality to identify neighbouring nodes with a similar centrality rank and grouped them into communities. They have developed two Leader-Community Detection Algorithms. At first, these algorithms identify the most influential nodes in the network and then search for communities by finding nodes that a similar to the leader nodes.

Furthermore, Lie, Zheng and Liao [17] combined the Infomap community detection algorithm developed by Rovsal [18] with Influence Maximization via Martingales (IMM) algorithm to detect influential addresses on the Ethereum blockchain. The study detected communities with the Infomap algorithm and calculated the importance of addresses inside those communities with IMM. It also compared the results produced by IMM with the results produced by the PageRank algorithm.

Perrin [18], used the Louvain algorithm in conjunction with PageRank to identify communities in biological networks. In this study, PageRank and Louvain algorithms were applied recursively to detect communities with a predefined size.

Yaowen and Yanchang [19] combined the Louvain algorithm with PageRank to detect Zombie accounts on social media. They used the Louvain algorithm to decompose a social media network into communities. Then, PageRank was used to identify zombie accounts in these communities.

Chapter 4

Contribution

The study focuses on finding the most influential addresses on the USDT transaction network and identifying the communities that these addresses belong to. The findings and methodologies presented in this project contribute to several areas of academic research.

To begin with, the project employed degree, betweenness, closeness and PageRank centralities to identify the most influential addresses in the USDT transaction network. The results of this part of the project contribute to the existing literature focused on studying the most influential entities operating on the Ethereum blockchain.

Furthermore, the study analysed the relationship between the centrality metrics by calculating their Pearson and Spearman rank correlation coefficients. Findings in this section contribute to the literature focused on analysing the relationship between the centrality metrics.

Then, the Louvain algorithm was used to identify communities of addresses in the USDT transaction network. The results were used to identify communities that the influential addresses belong to. This part of the study introduced a new methodology that can be used to determine the extent to which the most influential nodes in the network tend to cluster with each other. Furthermore, a methodology that can be used to identify addresses that are important on the community level was proposed. In addition, the study compared addresses' relative influence on the global level to their relative influence on a community level. Findings of this part of the study contribute to the literature that attempts to combine community detection algorithms with centrality metrics to study complex networks.

Chapter 5

Background

This section covers the key concepts and definitions used in this project. It starts with the explanation of the blockchain technology focusing on Ethereum and Smart Contracts. Then, it defines a social network and explains how Ethereum transactions can be modelled as a network. Lastly, it describes centrality and community detection algorithms that will be used in this study.

5.1 Blockchain And Cryptocurrencies

Blockchain is a decentralised and distributed digital ledger. It consists of blocks of transaction data. Each block is linked to the previous block via a cryptographic hash thereby creating a chain.

Cryptocurrencies can be defined as decentralised peer-to-peer payment systems empowered by blockchain technology. The first and the most famous cryptocurrency-Bitcoin was introduced by Nakamoto [1], its main purpose was to substitute central banks by creating a deflationary currency. Later, other types of cryptocurrencies were developed. Some of them attempt to change the way that people trade and exchange goods and services with each other by employing smart contract technology. For instance, Ethereum introduced by Buterin [3] gave a rise to the Decentralised Finance (DeFi) Industry which uses smart contracts to create new financial instruments.

5.2 Ethereum

Ethereum is an open-source blockchain introduced by Buterin [3]. The key difference between Bitcoin and Ethereum is that the latter has a smart contract functionality. Smart Contracts will be described later in this section.

5.2.1 Address

Ethereum address is a unique string of 42 characters that can be used to receive Ethereum and other Ethereum blockchain-based entities like tokens and collectables. Addresses can be of two different types. They can be externally and internally owned. External addresses can be accessed by users via a private key. Internally owned addresses, on the other hand, can not be accessed directly, they belong to smart contracts and therefore, can only be accessed via a function call.

5.2.2 Transactions

A transaction is an event that changes the state of a blockchain ledger. Our study will mainly focus on transactions that transfer tokens from one Ethereum address to another. However, Ethereum transactions can also alter the state of the ledger in more complicated ways.

5.2.3 Smart Contracts

The concept of a Smart Contract was first introduced by Nick Szabo [20] in the 1990s. Later, this idea was implemented on the Ethereum blockchain. A smart contract can be defined as a computer program that operates on blockchain. In other words, blockchain keeps track of the commands and sequence in which they should be executed. The key difference between smart contracts and other computer programs is their autonomous nature which allows them to execute automatically when a certain event occurs.

5.2.4 Decentralised Finance

The invention of smart contracts gave a rise to Decentralised Finance (DeFi). Companies in this industry use smart contracts to create new financial instruments and products that utilise blockchain technology. DeFi protocols have recently gained popularity and widespread adoption among cryptocurrency users.

5.2.5 ERC-20

ERC-20 is a smart contract standard that is commonly used to create tokens on Ethereum blockchain. Technically, ERC-20 tokens are similar to other cryptocurrencies except the fact that instead of their own blockchain they use Ethereum. These tokens can serve different purposes. For instance, they can be used to raise money for project financing and to represent ownership and voting rights in DeFi protocols. Furthermore, they can be used as a store of value. Some tokens are pegged to fiat currencies (stablecoins).

ERC-20 specifies a set of functions that should be included in every token. Every time a user calls a function, the contract stores an event that specifies the details of the executed transaction. The fact that this data is standardised, makes it easy to collect and compare data on the transactions of those tokens.

5.3 Social Networks and Graphs

Conceptually, a graph is a set of objects where some of those objects are connected. Formally, a graph can be denoted as $G = (V, E)$. V denotes a set of objects called vertices or nodes. E is a set of pairs of vertices u,v also called edges. A graph that represents a real-world system is often called a network. Therefore, in this project, these two terms will be used interchangeably.

Systems, where multiple entities interact with each other, can be modelled as a network. Every node in this network represents an entity and every edge shows how two entities interact with each other. Representing a complex system in this fashion enables us to analyse its structure and to get insights about the behaviour patterns of its entities.

5.4 Modelling ERC-20 transactions as a network.

Transactions of ERC-20 tokens on Ethereum Blockchain can be represented as a network. One way to model it is to represent addresses as nodes in the network and the flow of funds between the two addresses as an edge between them. This approach is depicted in figure 5.1. It shows three addresses *a1, a2* and *a3* as network nodes and depicts a situation where *a1* transfers 1 and 2 tokens to *a2* in two transactions. *a2* transfers 5 tokens to *a3* and *a3* transfers 4 tokens to *a2*. To analyse the resulting graph, all edges connecting a pair of nodes and pointing in the same direction can be aggregated into one edge. The value of this edge will be equal to the sum of the transfer values of all edges aggregated into it. The aggregated graph is presented in figure 5.2. Two edges representing transactions from *a1* to *a2* are aggregated into one edge with value 3.

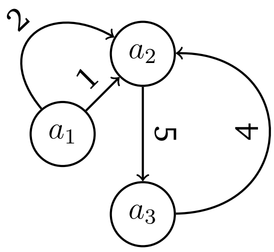

Figure 5.1: ERC-20 transactions as a social network

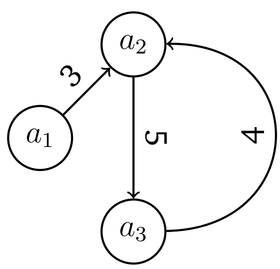

Figure 5.2: Aggregated ERC-20 transactions

5.5 Centrality and Influence

One way to measure an entity's importance is to model its relationships with other entities as a network and then assess the importance of every node in this network. Characteristics that determine the influence of a single entity in a system of interacting entities depend on the nature of the system. There is no universal metric that can be used to quantify influence or centrality. Furthermore, the nature of influence may vary within the network as well. For instance, some nodes can be considered to be important because they are connected to a lot of different entities. This might imply that a large number of actors in the network use a service provided by this node. On the other hand, some nodes might have a low number of connections but serve as a "bridge" from one part of the network to another. To account for different influence types this study employs four different metrics to identify the most influential address.

The relative importance of a node in a social network can be measured by its centrality index. Algorithms used to calculate centrality use the information about the network's nodes, edges and their attributes to rank nodes based on their importance. The rest of this section will focus on describing those centrality metrics.

5.5.1 Degree Centrality

Degree centrality was introduced by Freeman [21]. It accounts for the number of in-coming and out-coming edges of a node. According to this metric, the most central nodes have more connections than others. It can be used to identify the most "popular" entities. If an address interacts with many other different addresses, this may imply that these addresses rely on it to perform some kind of service or to store their funds.

One can distinguish between in and out degree centrality. In-degree centrality calculates the total number of addresses that sent tokens to a given address and out-degree centrality indicates the number of addresses that received funds from a given address. Our study uses the undirected version of degree centrality. Therefore, before calculating the score, all directed edges get converted into undirected edges. If two nodes have two directed edges between them, these two edges are aggregated into a single edge.

5.5.2 Closeness Centrality

Closeness centrality was also proposed by Freeman[21]. It measures how far on average a node locates to all other nodes. Closeness centrality is often used to identify entities that can influence the whole network most quickly. The algorithm has the following steps:

1. For each pair of nodes, find the shortest path between these two nodes.

2. For every node, calculate the sum of the shortest paths from this node to all other nodes.

3. Invert the resulting sum.

4. Normalise the score by multiplying it by the number of nodes in the network minus one.

In this study, closeness centrality was calculated on an undirected version of the network, since in directed graphs there is a possibility that there is no path between two nodes.

5.5.3 Betweenness Centrality

Similarly to degree and closeness centralities, Betweenness centrality was first proposed by Freeman [21]. It uses the concept of the shortest path to calculate the nodes' centrality scores. The shortest path between two vertices in the graph is the minimum possible set of edges that connects them. Betweenness centrality quantifies the extent to which a node lies in between all other nodes. Nodes with high betweenness scores might serve as "bridges" from one part of the network to another. It is often used to find entities that influence the flow from one part of the network to another.

The algorithm has the following steps:

1. Find shortest paths between all nodes in the network.

2. For every node, calculate the sum of the shortest paths from this node to all other nodes.

3. For every node, calculate the number of times that this node lies on the shortest path between all pairs of nodes.

An unweighted version of the algorithm was used.

5.5.4 PageRank

PageRank is an algorithm that was developed by Larry Page and Sergey Brin [22] to rank web pages. PageRank measures the importance of the web page based on the number of other pages that contain a link to it. It also takes into account the importance of the web pages that contain a link to the page of interest. According to this algorithm, a page is important if other important pages link to it.

$$PR(A) = (1-d) + d\left(\frac{PR(N_1)}{C(N_1)} + ... + \frac{PR(N_k)}{C(N_k)}\right) \tag{5.1}$$

To apply PageRank to the ERC-20 transaction network, one needs to use Ethereum addresses instead of the webpages and transactions instead of the web links. PageRank of node *A* is determined according to equation 5.1. N_1 ... N_k are the nodes representing the addresses that sent tokens to *A*. *d* is a damping factor. Originally, Larry Page and Sergey Brin described it as a probability that a user is going to stop clicking pages while surfing the web. *d* is usually set to 0.85. The algorithm is applied iteratively to update the PageRank of all nodes based on the PageRank of all nodes that send funds to it.

Since PageRank takes into account the edge direction and the connection weight, it can be used to identify addresses with the highest authority. PageRank was run on the directed weighted version of the graph.

5.6 Community Detection

Community detection is a technique that can be used to discover clusters of entities in the network that share similar properties or frequently interact with each other. A community can also be defined as a group of nodes within the network that is more densely connected within that group than to the other nodes.

5.6.1 Louvain Algorithm

Louvain community detection algorithm developed by Blondel et al., [23] is one of the most widely used community detection tools for large networks. It is an unsupervised learning algorithm that compares the node connection density of subgraphs in the network to the average node connection density of subgraphs generated by a random network.

The Louvain algorithm starts from a partition where each node is in its own community. The algorithm has two stages. In the first stage, it goes through all nodes in the graph one by one and puts them into the community with one of their neighbours. The neighbour is chosen based on the community modularity score.

Modularity is measured with the following formula:

$$Q = \frac{1}{2m}[A_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j) \tag{5.2}$$

- A_{ij} is the weight of an edge between nodes *i* and *j*.
- k_i, k_j represent the sum of the weights of all edges attached to nodes *i* and *j*.
- *m* ids the sum of all of the edge weights in the graph.
- c_i and c_j are the communities of the nodes.
- δ is a Kronecker delta function where $\delta(x,y) = 1$ if $x = y$ and 0 otherwise.

Modularity takes a value between -1 and 1. A high modularity score means that the nodes within the community are relatively more densely connected to nodes in this community, than to nodes outside the community.

Stage two consists of merging every community gathered in the first stage into a single node. Then, the Louvain algorithm repeats the first stage on a newly created graph. The algorithm runs recursively until it is not possible to increase the modularity any further.

5.6.2 Weakly Connected Components

A weakly connected component is a set of nodes in a network that are connected with each other by paths. In other words, any vertex in a weakly connected component can reach any other node by traversing edges.

A network may contain components that are not connected. Community detection and centrality algorithms produce more intuitive results when the input network consists of a single connected component. Therefore, before applying those algorithms, one needs to identify weakly connected components.

Chapter 6

Data

6.1 ERC-20 transactions data

The study uses transaction data for the second biggest ERC-20 token by market capitalisation on 12.08.2021. This token is Tether USD (USDT). USDT is a stablecoin. It is claimed to be backed by US dollar reserves and other assets. The second biggest token was chosen instead of the first biggest token, because the largest token (BNB) had significantly less transactions during the time period under analysis.

Transaction data was collected from block 12979784 to block 13011784. The time period between these blocks is roughly equal to 5 days.

Not all USDT transaction data was collected. The study focuses solely on the transfer events. Transfer event gets recorded on Ethereum blockchain when one address transfers a token to another address.

Transaction data was collected from the archive Ethereum node. The dataset contains data on transactions performed by 265006 unique addresses. There are 631649 transactions in the dataset. After edge aggregation, the number of transactions reduced to 479647.

6.2 Address label data

Some address label data was provided by the Imperial College London Department of Computing and some of it was collected from Etherscan.com. There are 1066 labelled addresses in the dataset Therefore, 0.4% of the addresses in the dataset have labels.

The dataset contains addresses that mainly belong to centralised exchanges, decentralised exchanges, cross-chain bridges and various decentralised finance protocols.

Chapter 7

Analysis

7.1 Pre-processing

Centrality and community detection algorithms require inputs to have a graph-like structure. Therefore, the USDT transaction network was modelled as a graph. Furthermore, weakly connected components in this graph should be identified to ensure that algorithms run correctly. Therefore, this section describes the process of graph creation and weakly connected components detection.

7.1.1 Graph Creation

Three different types of graphs were used to model the USDT transaction network. In particular, directed weighted, directed unweighted and undirected unweighted versions were used for different algorithms.

Edges in a directed weighted graph have a direction. This direction identifies source and target addresses for the transferred funds. Edges weights represent the amount of USDT transferred from one address to another. A directed unweighted graph is similar, but it does not have edge weights.

The undirected unweighted graph does not have directions and weights. In order to create an undirected graph from ERC-20 transaction data, all pairs of nodes that had at least one transaction between them were identified and an undirected unweighted link between them was created.

7.1.2 Weakly Connected Components

The USDT transaction network contains 6392 connected components. The largest connected component contains 250738 addresses. The second-largest connected component contains 74 nodes. The largest connected component contains 94,62% of all addresses in the transaction network. Addresses in this component will be used for further analysis. Data from the rest of the found components will be ignored.

7.2 Centrality

This section presents the analysis of the scores generated by the centrality algorithms. To begin with, it presents the distributions of the degree, closeness, betweenness and PageRank centrality scores. Then, the centrality scores are compared with Pearson and Spearman rank correlation coefficients. In addition, top 100 key players in USDT transaction network according to each centrality metric can be found in tables 10.1, 10.2, 10.3 and 10.4 in the Appendix.

7.2.1 Degree

Degree centrality was calculated on the directed unweighted version of the network. The in and out degrees for every node in the network were calculated and the scores were summed up to get the degree centrality.

The average degree centrality was found to be equal to 3.295. It means that one address transacted with around three other addresses on average. Address with the highest degree centrality transacted with 25930 addresses. The lowest score is equal to 1, which means that some addresses transacted only with one address.

Table 7.1 shows that 99.9% of all addresses transacted with less than 119 addresses. Figure 7.2 presents the cumulative kernel density estimation for the degree centrality score of nodes in the largest weakly connected component. It shows that most of the addresses have a low degree centrality and that less than 1% of addresses have extremely high scores. Figure 7.1 presents a more detailed breakdown of the degree centrality score distribution. 250728 addresses transacted with 1 to 5000 entities. 9 addresses transacted with between 5000 and 20000 addresses. Finally, one addresses transacted with 25930 addresses. The top-performing address belongs to Binance exchange and is labelled as Binance 14. All the above indicated that the distribution of the degree centrality scores is strongly right-skewed.

Right-skewed degree centrality distribution implies that there is a subset of nodes that transact with many more addresses than other nodes. This might indicate that a large proportion of USDT users uses the services provided by the owners of these addresses.

Table 7.2 lists 10 most influential addresses according to degree centrality. The dataset contains labels for nine of ten listed addresses. All labelled addresses belong to centralised exchanges. A half of those addresses belong to Binance Exchange.

7.2.2 Closeness Centrality Results

The closeness centrality algorithm searches for the shortest paths between all pairs of nodes. In directed networks, there is a possibility that there is no path between a pair of nodes. Therefore, closeness centrality was calculated on the undirected unweighted version of the network.

Average closeness centrality was found to be equal to approximately 0.228. Figures

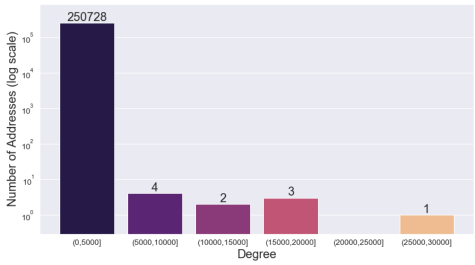

Figure 7.1: Degree Distribution across addresses

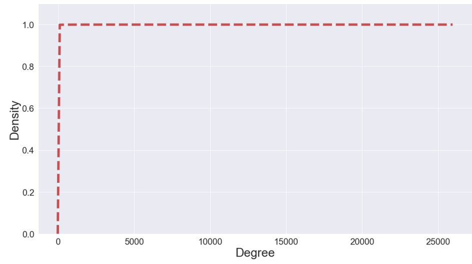

Figure 7.2: Degree Cumulative Kernel Density Estimation

Statistic	Degree Centrality	Closeness Centrality	Betweenness Centrality	PageRank
p99	13.0	0.28024	3642111.0	0.0000235
min	1.0	0.08520	0.0	0.0000011
max	25930.0	0.37078	16748380160.0	0.0561945
mean	3.29508	0.22833	504743.0	0.0000039
p90	3.0	0.27049	442061.0	0.0000026
p50	2.0	0.23109	14.0	0.0000011
p999	119.0	0.29307	28955647.0	0.0003134
p95	4.0	0.27103	922871.0	0.0000046
p75	2.0	0.25147	151660.0	0.0000015

Table 7.1: Centrality Metrics Summary Statistics

7.3 and 7.4 show that closeness centrality is more evenly distributed across addresses than the degree centrality. 137424 addresses scored between 0.2 and 0.25 according to this metric. 45450 addresses scored less than 0.2 and 67864 addresses scored more than 0.25. The best performing address scored 0.37. It belongs to the Binance exchange and is labelled as Binance 14 in the dataset.

Closeness centrality can take values from 0 to 1. The highest centrality score in the USDT transaction network is equal to 0.37 and the average is 0.228. These scores are lower than 0.5. A low average closeness centrality score indicates that the average number of transactions or links that lie in between two addresses is high. This, in turn, implies that nodes locate far from each other in the network.

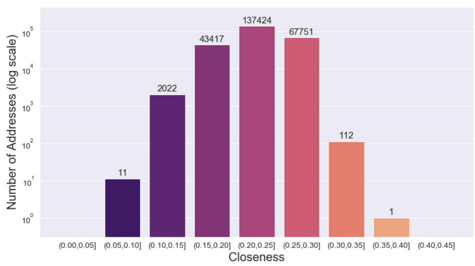

Figure 7.3: Closeness Distribution across addresses

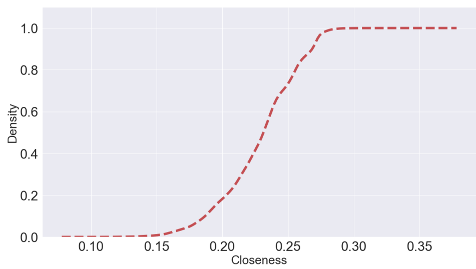

Figure 7.4: Closeness Cumulative Kernel Density Estimation

Table 7.3 lists ten most influential addresses according to closeness centrality. Six out of ten addresses are labelled. All labelled addresses belong to centralised exchanges. The most common address owner is Binance exchange, it owns five addresses.

7.2.3 Betweenness Centrality Results

Betweenness centrality was calculated on the directed unweighted version of the graph. The betweenness centrality algorithm can be used with directed weighted networks. However, a classic implementation of the algorithm takes into account edge weights while it searches for the shortest paths between all pairs of nodes. In this implementation, the shortest path chosen by the algorithm is the path between two nodes whose sum of edge weights is the lowest. This shortest-path definition is not suitable for the problem at hand. Therefore, an unweighted network was used.

As one can see in figure 7.6, betweenness centrality is distributed similarly to degree centrality. Less than 0.001% of addresses have a score higher than 28955647. Figure 7.5 shows that 250734 addresses lie on between 0 and 2 billion shortest paths, three addresses lie on between 2 and 4 billion paths and lastly, one address is between 16 and 18 billion paths. This indicates that the betweenness centrality score distribution is strongly right-skewed. The average score was found to be equal to 504743. The best performing address scored 16748380160 which is roughly 33182 times higher than the average. Address with the highest score belongs to Binance and is labelled as Binance 14.

The fact that there is a small number of addresses with an extremely high betweenness centrality indicates that there is a small subset of addresses that lie in-between all other addresses. In other words, large amounts of tokens flow through this subset addresses and they serve as bridges between addresses that don't interact with each other directly.

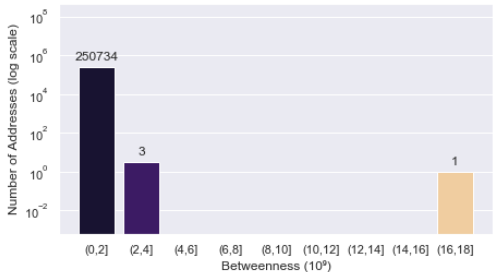

Figure 7.5: Betweenness Distribution across addresses

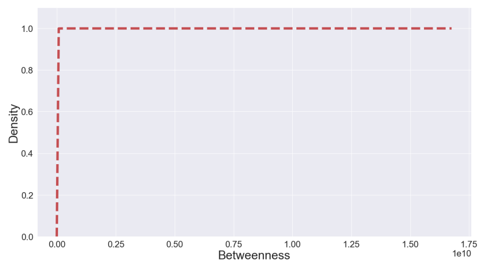

Figure 7.6: Betweenness Cumulative Kernel Density Estimation

Table 7.2 lists 10 most influential addresses according to betweenness centrality. The dataset contains labels for nine of ten listed addresses. All labelled addresses belong to centralised exchanges. Four out of ten addresses belong to Binance Exchange.

7.2.4 PageRank Results

PageRank algorithm was run on the weighted directed version of the network. Having edge directions is crucial for the PageRank algorithm and accounting for edge weights can potentially improve its performance. PageRank score was normalised to have values between 0 and 1 with L1 Norm.

Figure 7.8 presents a PageRank score cumulative kernel-density estimation (KDE). The KDE curve has a strong positive slope, in the beginning, indicating that most of the addresses have a low PageRank and that a few addresses have an anomalously high score. Figure 7.7 presents a more detailed background of the distribution of PageRank scores across addresses. As one can see, 250734 addresses scored between 0 and 0.01. There is one address that scored between 0.01 and 0.02 and one address that scored between 0.05 and 0.06. Therefore, the distribution is strongly right-skewed. Address with the highest PageRank scored approximately 0.056. This score is approximately 3.7 times higher than the score of the second most central address. Therefore, one can conclude that this score is anomalously high. The address with the highest score belongs to the Binance exchange and is labelled as Binance 14 on Etherscan.com.

Strongly right-skewed PageRank distribution indicates that there is a small number of addresses that interact with other influential addresses much more frequently than other addresses.

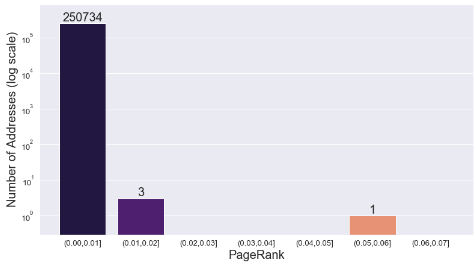

Figure 7.7: PageRank Distribution across addresses

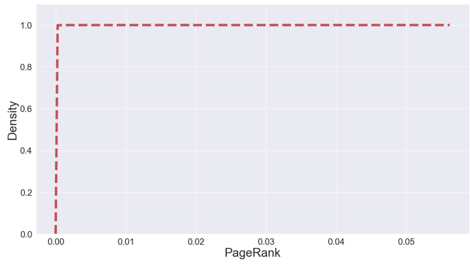

Figure 7.8: PageRank Cumulative Kernel Density Estimation

Table 7.3 lists ten most influential addresses according to PageRank. Seven out of ten addresses are labelled. All labelled addresses belong to centralised exchanges. The most common address owner is Binance exchange, it owns four addresses.

7.2.5 Comparing Centrality Metrics

To sum up, the distributions of degree centrality, betweenness centrality and PageRank scores are strongly right-skewed. This implies that according to these three metrics there are subsets of addresses in the network that interact with other addresses more frequently, act as bridges between different parts of the network and transact with other influential addresses more frequently. According to these metrics, less

	Top degree centrality addresses	Degree score	Top betweenness centrality addresses	Betweenness score
1	Binance 14 binance exchange	25930.0	Binance 14 binance exchange	1.674833e+10
2	0xe59cd29be3be4461d79c0881d238cbe87d64595a	18190.0	0xe59cd29be3be4461d79c0881d238cbe87d64595a	3.767116e+09
3	Coinbase 5 coinbase	16350.0	Coinbase 5 coinbase	3.020358e+09
4	Coinbase 4 coinbase	15372.0	Coinbase 4 coinbase	2.665609e+09
5	Binance 17 binance exchange	12262.0	Binance 17 binance exchange	1.905957e+09
6	Binance 18 binance exchange	11811.0	Binance 18 binance exchange	1.865442e+09
7	Binance 15 binance exchange	8833.0	Crypto.com 2 crypto-com exchange	1.686126e+09
8	Binance 16 binance exchange	8181.0	FTX Exchange	1.454656e+09
9	Crypto.com 2 crypto-com exchange	7123.0	Binance 15 binance exchange	1.372418e+09
10	FTX Exchange	5326.0	Kraken 7 kraken exchange	1.340465e+09

Table 7.2: Ten most influential addresses according to Degree and Betweenness centralities.

	Top PageRank addresses	PageRank score	Top closeness centrality addresses	Closeness score
1	Binance 14 binance exchange	0.056194	Binance 14 binance exchange	0.370783
2	0xe59cd29be3be4461d79c0881d238cbe87d64595a	0.015660	Binance 17 binance exchange	0.344749
3	Binance 18 binance exchange	0.011791	Binance 18 binance exchange	0.343566
4	Binance 17 binance exchange	0.011034	Binance 15 binance exchange	0.340063
5	Binance 15 binance exchange	0.009832	Binance 16 binance exchange	0.335967
6	Kraken 7 kraken exchange	0.008545	Crypto.com 2 crypto-com exchange	0.335231
7	Coinbase 5 coinbase	0.008375	0xb8b53751f76492cc32c73553764e07ccd6a98517	0.332728
8	FTX Exchange	0.008043	0x18877aeddef4fde5be9b818d713ee39a7ca79761	0.331670
9	0xfa103c21ea2df71dfb92b0652f8b1d795e51cdef	0.007529	0x95b564f3b3bae3f206aa418667ba000afafacc8a	0.330233
10	0xc176761d388caf2f56cf03329d82e1e7c48ae09c	0.007351	0xfa103c21ea2df71dfb92b0652f8b1d795e51cdef	0.324924

Table 7.3: Ten most influential addresses according to PageRank and Closeness centrality.

than 0.1% of addresses have extremely high values in comparison to others. This indicates that the transaction activity on the USDT network is concentrated around a small number of addresses. In contrast, the closeness centrality score is more evenly distributed. According to this metric, centralisation is less severe.

Binance exchange is the largest owner of the highest-scoring addresses according to these four metrics. In particular, there is one dominant address. It is labelled as Binance 14. It has the highest score according to degree, betweenness, PageRank and closeness centralities. Its score much higher than the score of the second-best addresses. Therefore, one can conclude that a large proportion of the economic activity performed with USDT is centred around an address that belongs to the Binance exchange.

7.2.6 Relationship between Centrality Metrics

To compare the results produced by the centrality metrics, the study employs Pearson and Spearman rank correlation coefficients. Pearson correlation coefficient was found by correlating the scores produced by one centrality metric with every other centrality metric. The results are presented on figure 7.10. All correlations are positive. PageRank and degree centralities have a strong positive correlation coefficient equal to 0.844. Betweenness and degree centralities also have a positive correlation equal to 0.831. The strongest positive correlation was found between PageRank and Betweenness centrality, it is equal to 0.946.

In contrast, a correlation coefficient between the closeness centrality and all other metrics is close to zero. Pearson correlation coefficient between closeness centrality and degree, betweenness and PageRank centralities is equal to 0.035, 0.026 and 0.031 respectively.

The fact that the degree centrality is not correlated with closeness centrality might imply that entities that transact with many other addresses tend to be embedded in clusters that locate far away from other nodes. Furthermore, low correlation between betweenness and closeness centralities indicates that a small number of addresses with high betweenness scores serve as bridges from a small number of entities to a large set of entities. Lastly, the low correlation between the PageRank score and closeness centrality does not have an intuitive explanation. However, it might imply that influential addresses tend to cluster together and separate from less influential addresses.

Low correlation between closeness centrality and other centrality metrics might imply that addresses that have a high number of connections, serve as bridges connecting other addresses and connect to other influential entities are embedded in clusters that are not well-connected to the rest of the network. Furthermore, a correlation coefficient between degree centrality, PageRank and betweenness centrality is positive and close to 1. Therefore, the most influential addresses, tend to transact with many entities, serve as bridges and transact with other influential addresses at the same time.

To sum up, correlation analysis of the centrality metrics shows that there might exist extremely influential addresses and these addresses tend to form clusters that locate far away from the rest of the network.

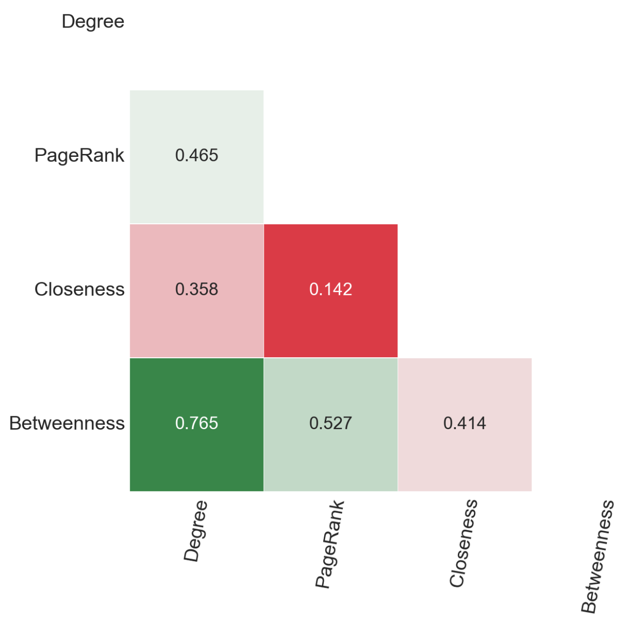

Figure 7.9: Spearman correlation coefficient between centrality Rankings

The Spearman rank correlation between two centrality metrics is equal to the Pearson correlation between the address ranks produced by those centrality metrics. It is equal to 1 if both metrics assign each address a similar rank when it is compared to other addresses. Spearman rank correlation is equal to -1 if the ranks produced by one of the metrics are exactly opposite to the ranks produced by the second metric. Spearman rank correlation assesses the monotonic relationship between two centralities, while Pearson's correlation deals with a linear relationship. In other words, Pearson correlation does not tell us how the relative centrality scores change between two metrics. Therefore, one needs to use the Spearman correlation in order to assess this.

Figure 7.9 presents Spearman rank correlations between 6 pairs of centralities. All correlations are positive. In comparison to the Pearson correlation coefficient, Spearman rank correlation is lower for all pairs despite the pairs that include closeness centrality as one of the variables. Furthermore, the strongest correlation is now between the betweenness and degree centralities. Correlations between degree and betweenness and PageRank and betweenness are the only ones that exceed 0.5. Spearman rank correlation between closeness centrality and all other centralities is well above 0. However, closeness centrality still has the lowest correlation with all pairs.

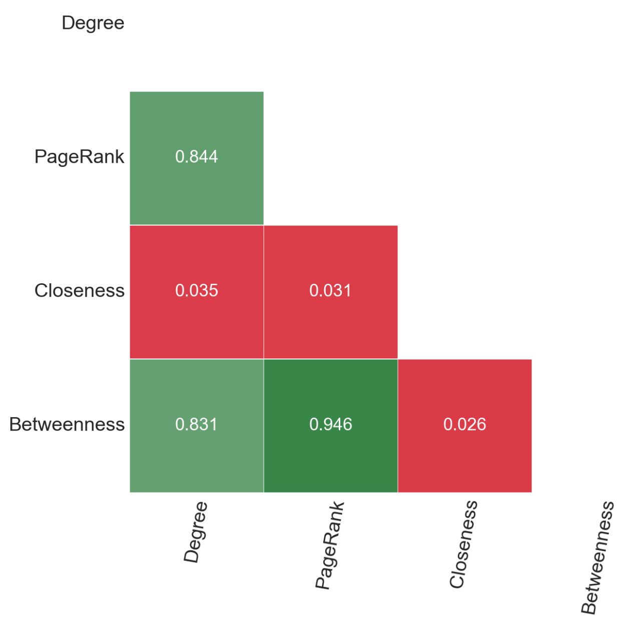

Figure 7.10: Pearson correlation coefficient between centrality Rankings

Pearson and Spearman rank correlation coefficients produce different results. According to Pearson correlation, for all centrality metrics despite the closeness, the high value of one centrality metric usually implies high values for all other metrics and vice-versa. Ranks produced by different metrics are correlated as well. However, this correlation is much weaker. This finding implies that to compare the influence or importance of a single address in comparison to others, one needs to use all four metrics together.

7.3 Community detection

This section is focused on the process of identifying communities on the USDT transaction network. At first, it describes the process of choosing the optimal configuration for the Louvain algorithm. Secondly, it presents the analysis of the communities identified by the algorithm. Then, it shows how one hundred most influential addresses according to each centrality metric are distributed across the communities. Lastly, it concludes with a discussion on the degree of clustering among the most influential addresses.

7.3.1 Methodology

In order to identify and to analyse communities that contain influential addresses, the following steps were performed:

1. Apply the Louvain community detection algorithm to the largest weakly connected component of the transaction network.
2. Perform optimal parameter tuning for the Louvain algorithm and choose the best model configuration.
3. Identify one hundred addresses with the highest score for each centrality metric.
4. Identify communities that the most influential addresses belong to.
5. Perform analysis of the communities that the most influential addresses belong to.

7.3.2 Louvain Algorithm Parameter Tuning

The Louvain algorithm was run on the undirected weighted version of the network. The key parameter that affects the result produced by the Louvain algorithm is the tolerance value. The tolerance value is the minimum possible change in modularity between the algorithm iterations. If the tolerance value is larger than the change in modularity between iterations, the algorithm stops and outputs the results.

Ten different tolerance values were tested for the Louvain algorithm in order to identify the best possible configuration. As one can see from table 7.4, every tolerance level results in partitioning with modularity equal to a value that is extremely close to 0. The number of communities produced by different values does not vary significantly as well. Therefore, it was decided to use the 0.1 value for the tolerance parameter, since the algorithm runs faster with a smaller tolerance value. Picking any other tolerances level with the same modularity should not have a significant impact on the analysis.

Tolerance	Modularity	Number of Communities	Mean	Minimum Size	Maximum Size
0.1	0.0	15393	16.289	1	43220
0.01	0.0	15406	16.275	1	43231
0.001	0.0	15409	16.272	1	43208
0.0001	0.0	15403	16.278	1	43173
0.00001	0.0	15374	16.309	1	43189
0.000001	0.0	15401	16.280	1	43228
0.0000001	0.0	15398	16.283	1	43281
0.00000001	0.0	15415	16.265	1	43206
0.000000001	0.0	15400	16.281	1	43193
0.0000000001	0.0	15406	16.275	1	43212
0.00000000001	0.0	15403	16.278	1	43295

Table 7.4: Louvain Algorithm Tolerance Value Tuning

7.3.3 USDT Community Analysis

Table 7.4 contains the summary statistics for the partitioning generated by the Louvain algorithm with the tolerance value equal to 0.1. The average number of addresses in the community is equal to approximately 16.3. The largest community contains 43220 addresses. The smallest community consists of one single address. 90% of all communities contain less than 9 addresses.

Figure 7.11 shows that Louvain partitioning produces a lot of small communities and that large communities are less frequent. There are 15390 communities that contain between 1 and 1000 addresses. Two communities have between 1000 and 2000 addresses. Only 1 community has more than 40000 addresses. In addition, 519 communities contain a single address. Single-node communities might be created due to the limitations of the Louvain algorithm. These limitations will be discussed in the limitations and the future work section.

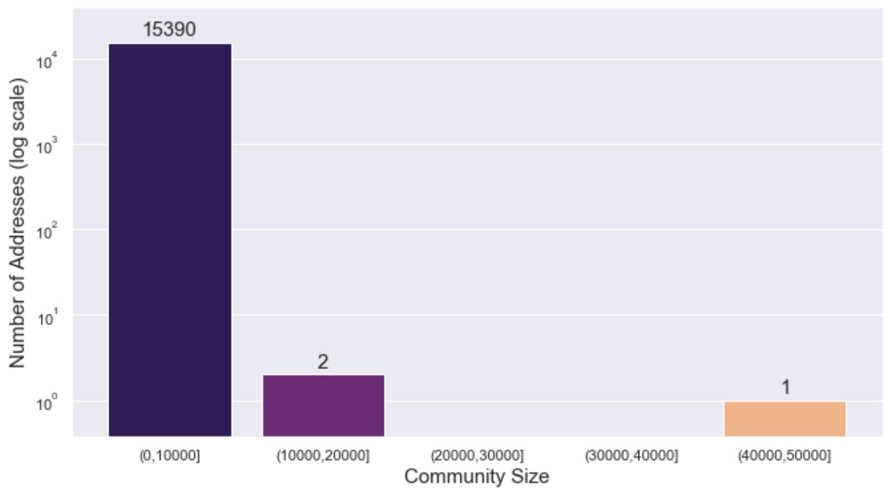

Figure 7.11: Community Sizes

7.3.4 Communities of the most influential nodes

This section describes how the most influential addresses are distributed across the communities produced by the Louvain algorithm. In particular, it focuses on one hundred most influential addresses according to degree, closeness, betweenness and PageRank centralities.

Degree

One hundred addresses with the highest degree centrality score are distributed across 79 communities. As one can see in figure 7.12, 72 of those communities contain one influential address, four communities contain two influential addresses and there are three communities containing contain three, five and twelve influential addresses.

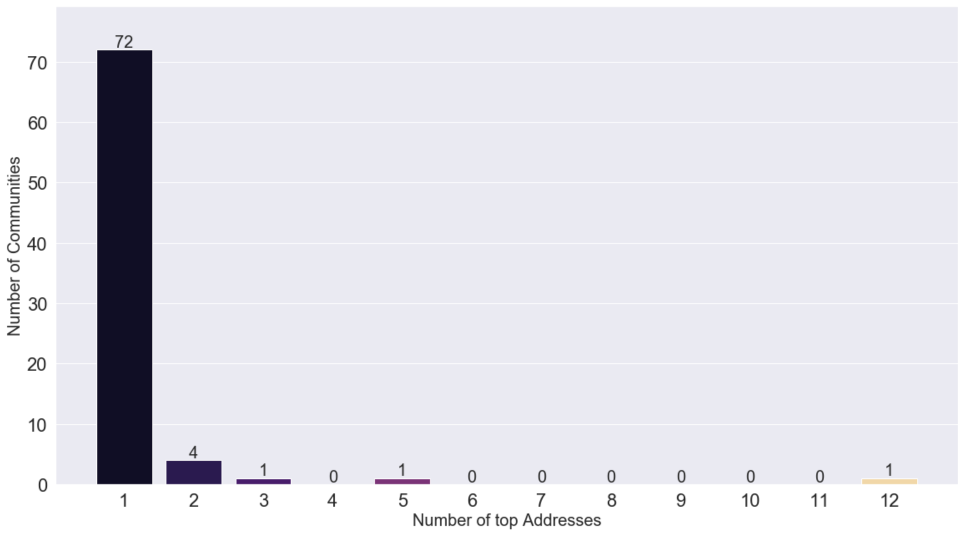

Figure 7.12: Distribution of top 100 entities according to Degree centrality across communities.

Closeness

One hundred addresses with the highest closeness centrality score are distributed across 53 communities. As one can see in figure 7.13, 40 of those communities contain one influential address, six communities contain two influential addresses, four communities contain 3 influential addresses and there are three communities containing 7, 11 and 18 influential addresses.

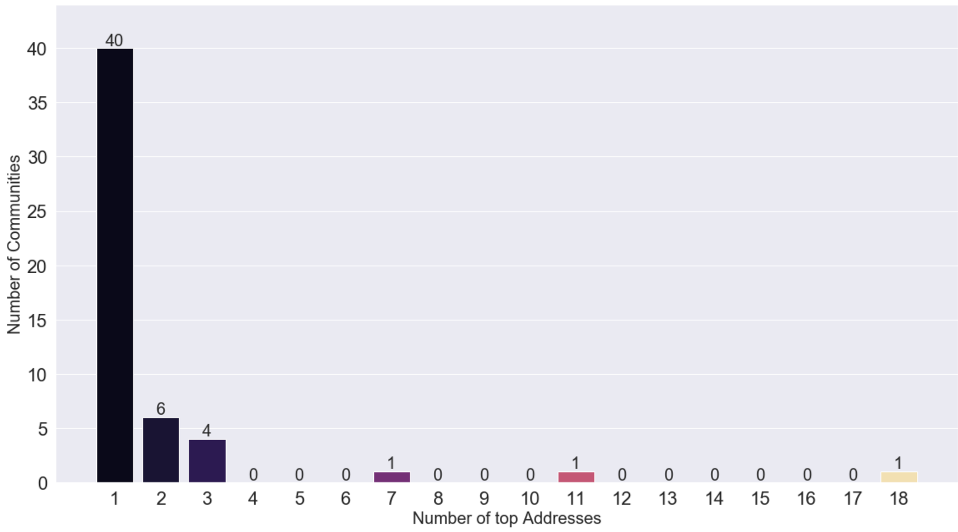

Figure 7.13: Distribution of top 100 entities according to Closeness centrality across communities.

Betweenness

One hundred addresses with the highest betweenness centrality score are distributed across 69 communities. As one can see in figure 7.14, 55 of those communities contain one influential address, 8 communities contain 2 influential addresses, 4 communities contain 3 influential addresses and there are two communities containing 4 and 13 influential addresses.

PageRank

One hundred addresses with the highest PageRank score are distributed across 79 communities. As one can see in figure 7.15, 36 of those communities contain one influential address, 11 communities contain 2 influential addresses, 2 communities contain 3 influential addresses, 2 communities contain 4 influential addresses and there are three communities containing 7 and 16 influential addresses.

Clustering of the most influential addresses

The most influential addresses according to the closeness centrality score display the highest clustering tendency since they are distributed across 53 communities. The highest scoring addresses according to PageRank and degree centrality show the lowest clustering tendency.

Overall, the analysis shows that there is no significant clustering of influential addresses. In contrast, some communities contain only a single influential entity. 72%

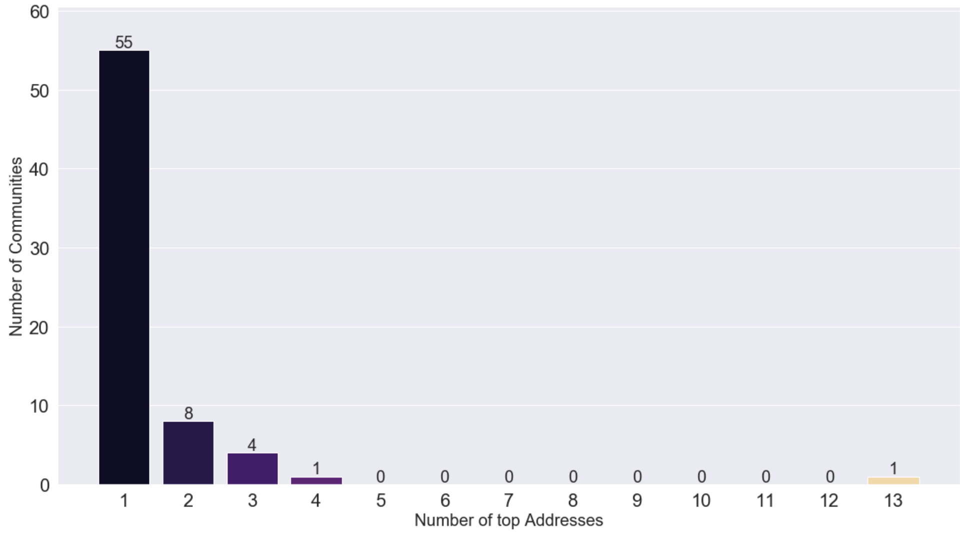

Figure 7.14: Distribution of top 100 entities according to Betweenness centrality across communities.

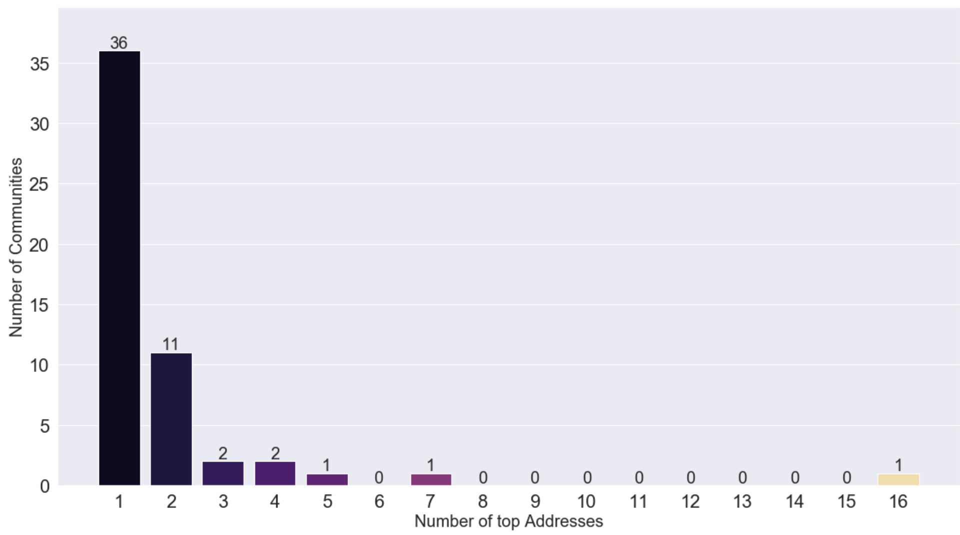

Figure 7.15: Distribution of top 100 entities according to PageRank across communities.

of the most influential addresses according to degree centrality are the only addresses from the top 100 list in their community. This proportion is equal to 40%, 55% and 36% for closeness, betweenness and PageRank centralities. However, there is always a community that contains more than 10% of all influential addresses. This community is the same for all metrics and is also the largest community overall.

7.4 Locally Important Nodes

This section is focused on analysing the influence of the most important addresses inside their communities. It starts with describing the process of graph creation for communities. Then, it compares the centrality rankings of the most influential addresses on the global scale to their centrality ranking inside the communities that they belong to.

7.4.1 Methodology

Locally important addresses detection process has the following steps:

1. Apply Louvain community detection algorithm to the largest connected component of the USDT transaction network.

2. For all communities that contain at least one influential node from the top 100 influential nodes creates a separate network.

3. For all community networks, calculate degree, closeness, betweenness and PageRank centrality scores for all nodes.

4. Determine the most influential address in every community.

5. Compare locally important nodes with the nodes that are important on the global scale.

7.4.2 Graph Creation

To identify addresses that are central within communities, a separate network was created for every community that contains at least one influential node according to any of the four metrics. A community network contains all addresses that are included in the community and all transactions between them. The Community network does not contain any edges that connect addresses within the community to addresses outside the community. The graph creation approach is similar to the one used for the whole network. In particular, three versions of the community graph were created: directed weighed, directed unweighted and undirected unweighted.

7.4.3 Comparing Global and Community-level centralities

After creating three versions of the network for each community graph, Four centrality algorithms described in the previous section were applied to each community graph to identify the most influential nodes within the communities. The most influential addresses within the communities were then compared to the most influential addresses for the whole network for every centrality metric.

The top hundred best-scoring nodes according to degree centrality are distributed across 79 communities. Degree centrality for all addresses inside these communities was calculated. Then, one addresses with the highest score from each community was identified. It was found that all most influential nodes on the community scale are included in the list of the top hundred most degree-central addresses on the global scale. In other words, in all 79 communities, the highest-scoring address is one of the most influential addresses on the global scale. This means that addresses that are the most influential according to degree centrality in the whole network, remain influential when they get separated into the communities.

Out of the most influential addresses on the global scale according to PageRank, 53 addresses have the highest PageRank in their community. This implies that relative PageRank centrality does not always remain the same on the local scale. The most influential nodes according to PageRank dominate 67.1% of the communities that they are in.

Closeness centrality results are similar to those of the PageRank, 28 out of 100 most influential addresses are the highest scoring addresses in their communities. They dominate 48.7% of all communities. Therefore, less than half of the addresses remain influential on the local scale according to this metric.

Lastly, 67 of the most influential betweenness centrality addresses on the global scale remain the most influential on the local scale. This is in 97.1% of all of the communities.

To sum up, addresses with a high degree and betweenness centrality scores remain important according to these metrics on the community scale. However, relative PageRank and closeness centrality scores of the highest-scoring nodes on the global scale deteriorate on the local scale.

Chapter 8

Discussion

8.1 Results

The study described and tested a set of tools that can be used to quantify and compare the influence of addresses on the Ethereum blockchain. Degree, betweenness, closeness and PageRank centralities were employed for this purpose. Several interesting insights about the structure of the USDT transaction network was discovered throughout the research.

To begin with, the distributions of degree, betweenness and PageRank centrality scores are strongly right-skewed. This implies that there is a small number of anomalously influential addresses in the USDT transaction network. Out of the top 10 highest scoring addresses, all labelled addresses belong to centralised exchanges. This result holds for all centrality metrics. None of the top 10 influential addresses belongs to decentralised finance protocols. This is a surprising finding, considering the growing popularity of such protocols. It confirms the previous findings of Sui et. al. [2] that most of the influential addresses on the Ethereum blockchain belong to centralised exchanges.

In addition, centrality analysis showed that the Binance exchange is the most common owner of the influential addresses. Furthermore, one address in the network is the most influential according to all four metrics. This address belongs to Binance exchange is labelled as Binance 14 on Etherscan.com.

After calculating centrality scores for every address in the network, the relationship between the centrality metrics was assessed by calculating Pearson and Spearman rank correlation coefficients between the centrality scores. It was found that degree, betweenness and PageRank centralities have a strong positive Pearson correlation coefficient. This implies that if an address is influential according to one of those metrics it is also likely to be influential according to the rest of them and vice-versa. This result, combined with the finding that a small number of addresses has anomalously high centrality scores indicates that there exists a set of addresses that are dominant in the network. In addition, the correlation between closeness centrality and the rest of the metrics was found to be close to 0. This is an unusual finding that does not have a clear intuitive explanation and requires further research.

Spearman rank correlation coefficient was found to be lower than the Pearson correlation coefficient for all pairs of centralities except the ones that include closeness centrality. This implies that although centrality scores are correlated, they produce different ranking tables. Therefore, they can complement each other in the analysis of the most influential entities.

Then, the Louvain algorithm was used to identify communities of addresses. Louvain algorithm has identified 15393 communities. The average community contains 16.2 addresses and the largest community contains 43220 addresses. Analysing the communities that the most influential addresses belong to, showed that these addresses do not exhibit a strong clustering tendency. A lot of communities contain a single influential address. This brings us to the conclusion that communities in the USDT transaction network tend to originate around single influential entities.

After identifying communities of the most influential addresses, the most influential addresses on the community level were detected. It was found that the addresses with high degree and betweenness centrality scores on the global scale tend to remain the most influential on the community scale. However, The relative influence of addresses with high PageRank and closeness centrality deteriorates on the community level.

These findings highlight the fact that using the traditional centrality metrics on their own might not be sufficient to identify the most influential entities. Addresses that are influential on the global scale might not remain influential on the community level. Therefore, it is important to analyse community-level centrality to correctly identify the most influential entities.

Furthermore, the results outlined above complement the study conducted by Li et al., [17] that proposed to combine Infomap community detection algorithm with IMM centrality algorithm to identify the most influential addresses on the Ethereum blockchain. They have used PageRank as a baseline metric to determine the relative performance of their algorithm. They found that the IMM algorithm outperforms PageRank in identifying important nodes in communities. Our findings demonstrate that the relative influence of addresses with high PageRank scores reduces on the community level. Therefore, PageRank might not be an ideal baseline metric to determine the effectiveness of such algorithms. Betweenness and degree centrality can potentially be more suitable for this purpose.

In addition, the methodologies outlined in this study can have practical applications. In particular, they can be used to improve the existing on-chain advertisement delivery strategies. Wu et al., [15] proposed a community detection algorithm that can be used to identify groups of users that use similar services and used this information to deliver targeted advertisements. Their algorithm matches Ethereum users that use the same smart contracts and partitions them into groups based on this. This algorithm can be improved by implementing the methodology described in this study. Instead of matching addresses based on their smart contract usage, one can determine whether they belong to the community with the same influential address.

8.2 Limitations and Future Work

The approach to identifying the key entities participating in ERC-20 token transactions and analysing the degree of decentralisation used in this study has some drawbacks. This section describes some of those drawbacks and explains how they can be solved in future work.

To begin with, the final modularity score produced by the Louvain algorithm while partitioning the USDT transaction network was equal to 0. The modularity score can take values between -1 and 1, therefore 0 is not a terrible result. However, other community detection algorithms can potentially produce better results. For instance, the Leiden algorithm introduced by Traag, Waltman and van Eck[24] can potentially help to overcome some of the drawbacks of the Louvain Algorithm. Also, a label propagation algorithm introduced by Raghavan, Albert and Kumara[25] can be used as well.

Furthermore, the approach described in this study does not take into account address balances. Addresses with high balance can potentially be more influential. Therefore, this factor should be accounted for in future work.

Finally, the study focuses solely on the properties of the transaction network while determining address' influences. Therefore, using some external data like market capitalisation of the companies performing ERC-20 transactions and the information about their business operations can potentially be combined with the network centrality metrics to determine the most influential entities.

Chapter 9

Conclusion

The study proposed a set of algorithms that can be used to identify the most influential ERC-20 users and services operating on the blockchain. The algorithms were tested on the transfer transaction data of the USD Tether. USDT transactions were modelled as a complex network and the most influential addresses according to degree, closeness, betweenness and PageRank centralities were identified. According to degree, betweenness and PageRank centralities, USDT transaction network exhibits a strong centralisation tendency. Influential addresses mainly belong to centralised cryptocurrency exchanges. These results confirm the previous findings on the centralisation of economic activity on the Ethereum blockchain.

Furthermore, the study used the Louvain community detection algorithm to identify communities of addresses that the most influential entities belong to. A clustering tendency among the most influential addresses was not identified. On the other hand, there tends to be one influential address per community. This implies that influential addresses tend to create communities of less influential addresses around them.

Lastly, it was found that addresses that have a high degree and betweenness centrality on the global scale remain relatively influential on the community scale. However, the influence of addresses with high PageRank and closeness centrality deteriorates on the community level. This finding highlights the importance of combining network centrality metrics with community detection algorithms to identify influential entities in the complex networks.

To conclude, the methodology proposed in this study can potentially be used to study other complex networks in different fields including medicine, biology and sociology. However, to account for the network-specific properties different community detection algorithms can be used. For instance, Infomap, Leiden and influence propagation algorithms can be employed for this purpose.

Bibliography

[1] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008. URL `https://git.dhimmel.com/bitcoin-whitepaper/`. pages 1, 11

[2] Danning Sui, Gezim Sejdiu, Damien Graux, and Jens Lehmann. The hubs and authorities transaction network analysis using the sansa framework, 2021. URL `http://ceur-ws.org/Vol-2451/paper-25.pdf`. pages 1, 8, 34

[3] Buterin. Ethereum whitepaper — ethereum.org, 2013. URL `https://ethereum.org/en/whitepaper/`. pages 1, 11

[4] Jiajing Wu, Jieli Liu, Weili Chen, Huawei Huang, Zibin Zheng, and Yan Zhang. Detecting mixing services via mining bitcoin transaction network with hybrid motifs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–13, 2021. doi: 10.1109/tsmc.2021.3049278. pages 2

[5] Paolo Tasca and Shaowen Liu. The evolution of the bitcoin economy: Extracting and analyzing the network of payment relationships. *SSRN Electronic Journal*, 2016. doi: 10.2139/ssrn.2808762. pages 7

[6] 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, 2020. doi: 10.1109/tsmc.2018.2883678. pages 7

[7] Ting Chen, Zihao Li, Yuxiao Zhu, Jiachi Chen, Xiapu Luo, John Chi-Shing Lui, Xiaodong Lin, and Xiaosong Zhang. Understanding ethereum via graph analysis. *ACM Transactions on Internet Technology*, 20(2):1–32, 2020. doi: 10.1145/3381036. pages 7

[8] Dongchao Guo, Jiaqing Dong, and Kai Wang. Graph structure and statistical properties of ethereum transaction relationships. *Information Sciences*, 492: 58–71, 2019. doi: 10.1016/j.ins.2019.04.013. pages 7

[9] Minoo Ashtiani, Ali Salehzadeh-Yazdi, Zahra Razaghi-Moghadam, Holger Hennig, Olaf Wolkenhauer, Mehdi Mirzaie, and Mohieddin Jafari. A systematic survey of centrality measures for protein-protein interaction networks. *BMC Systems Biology*, 12(1), 2018. doi: 10.1186/s12918-018-0598-2. pages 8

[10] Kenneth R. Ahern. Network centrality and the cross section of stock returns. *SSRN Electronic Journal*, 2012. doi: 10.2139/ssrn.2197370. pages 8

[11] E. Sandeep Kumar. Preserving privacy in ethereum blockchain. *Annals of Data Science*, 2020. doi: 10.1007/s40745-020-00279-9. pages 8

[12] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5): 75–174, 2010. doi: 10.1016/j.physrep.2009.11.002. pages 8

[13] Michelle Girvan and Aaron Newman. Community structure in social and biological networks, 2021. URL `https://www.pnas.org/content/99/12/7821.full`. pages 8

[14] Prado-Romero. (pdf) discovering bitcoin mixing using anomaly detection, 2018. URL `https://www.researchgate.net/publication/322914722`. pages 8

[15] Sissi Xiaoxiao Wu. Community detection in blockchain social networks, 2021. URL `https://ieeexplore.ieee.org/abstract/document/9387705`. pages 8, 35

[16] Sara Ahajjam, Mohamed El Haddad, and Hassan Badir. A new scalable leader-community detection approach for community detection in social networks. *Social Networks*, 54:41–49, 2018. doi: 10.1016/j.socnet.2017.11.004. pages 9

[17] Xin Li, Weidong Zheng, and Hao Liao. Community-based propagation of important nodes in the blockchain network, 2021. URL `https://link.springer.com/chapter/10.1007/978-981-16-2540-4_51`. pages 9, 35

[18] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008. doi: 10.1073/pnas.0706851105. pages 9

[19] Qiu Yaowen, Li Yin, and Lu Yanchang. Zombie account detection based on community detection and uneven assignation pagerank, 2021. URL `https://arxiv.org/abs/2101.00922`. pages 9

[20] Nick Szabo. Smart contracts, 2021. URL `https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html`. pages 12

[21] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978. doi: 10.1016/0378-8733(78)90021-7. pages 14, 15

[22] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. URL `http://ilpubs.stanford.edu:8090/422/`. Previous number = SIDL-WP-1999-0120. pages 15

[23] 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, oct 2008. doi: 10.1088/1742-5468/2008/10/p10008. URL `https://doi.org/10.1088/1742-5468/2008/10/p10008`. pages 16

[24] V. A. Traag, L. Waltman, and N. J. van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1), 2019. doi: 10.1038/s41598-019-41695-z. pages 36

[25] Nandini Raghavan, Reka Albert, and Soundar Kumara. Detecting user's spreading influence using community structure and label propagation, 2021. URL `https://arxiv.org/pdf/0709.2938.pdf`. pages 36

Chapter 10

Appendix

Rank	Name/Address	Degree	Rank	Name/Address	Degree
1	Binance 14	25930	**51**	Coinex	1023
2	0xe59cd29be3be4461d79c0881d238cbe87d64595a	18190	**52**	0x6b71dcaa3fb9a4901491b748074a314dad9e980b	1006
3	Coinbase 5	16350	**53**	Hotbit 3	945
4	Coinbase 4	15372	**54**	Polygon (Matic): ERC20 Bridge	943
5	Binance 17	12262	**55**	0x74dec05e5b894b0efec69cdf6316971802a2f9a1	939
6	Binance 18	11811	**56**	AMMWrapperWithPath	935
7	Binance 15	8833	**57**	Poloniex 4	923
8	Binance 16	8181	**58**	0x6753818a769f2024a454fd471d663fac4ddaeefb	911
9	Crypto.com 2	7123	**59**	0x8c8d7c46219d9205f056f28fee5950ad564d7465	858
10	FTX Exchange	5326	**60**	0x477b8d5ef7c2c42db84deb555419cd817c336b6f	828
11	Uniswap V2: USDT 2	4892	**61**	0xa25979e9150a8ab187b7db0fccd3d0ea0e7aa03a	824
12	Kraken 7	4479	**62**	HitBTC 3	811
13	0x5041ed759dd4afc3a72b8192c143f72f4724081a	4223	**63**	0xb02f1329d6a6acef07a763258f8509c2847a0a3e	797
14	BitMart 2	4061	**64**	0xeec0ed9e41c209c1c53a35900a06bf5dca927405	770
15	Nexo 2	3150	**65**	Uniswap V3: USDC-USDT	759
16	CoinList 1	3135	**66**	Uphold.com	747
17	0x3e9afaa4a062a49d64b8ab057b3cb51892e17ecb	3021	**67**	Bitstamp 2	726
18	Crypto.com	2912	**68**	Uniswap V2: USDC-USDT 2	713
19	0xf60c2ea62edbfe808163751dd0d8693dcb30019c	2655	**69**	0xb8001c3ec9aa1985f6c747e25c28324e4a361ec1	683
20	Gate.io	2597	**70**	0x3c02290922a3618a4646e3bbca65853ea45fe7c6	679
21	Huobi 3	2514	**71**	0xc7807e24338b41a34d849492920f2b9d0e4de2cd	662
22	Huobi 2	2512	**72**	Uniswap V2: GAME-USDT 2	661
23	Huobi 36	2505	**73**	0x95e203ba2a7b8aa6acc3e40f5349864c137faf7d	635
24	Huobi 12	2505	**74**	CEX.IO	631
25	Huobi 7	2500	**75**	0x5bccb9eab6059cc6aa779b30ceb4ab9753af9ed1	629
26	0x74de5d4fcbf63e00296fd95d33236b9794016631	2487	**76**	SwissBorg 1	628
27	Huobi 4	2471	**77**	0xe2a07db5a4277112e8caf39c31ad6667d65500d7	618
28	Huobi 9	2446	**78**	0x292f04a44506c2fd49bac032e1ca148c35a478c8	613
29	Huobi 1	2428	**79**	0x5bcbdfb6cc624b959c39a2d16110d1f2d9204f72	613
30	0xd30b438df65f4f788563b2b3611bd6059bff4ad9	2321	**80**	0xb8d7b021772b76df5fd18a20d1a3daebf8615630	611
31	0x48c04ed5691981c42154c6167398f95e8f38a7ff	2062	**81**	0x95b564f3b3bae3f206aa418667ba000afafacc8a	595
32	0x63a395b574d5e23c3dbc6986be5994ef6743afa8	1732	**82**	CoinList 2	594
33	HitBTC 2	1708	**83**	0xc6dbd76f50e358ff56ee2bf11688d9af7091453a	594
34	Bittrex	1590	**84**	Bitmax 3	586
35	0x691f987fd150967d6b6bd7e7b3d04e9dbc1f4efc	1529	**85**	0x2a9613babde7a2b393f253b36338374217da562d	580
36	0x6254b927ecc25ddd233aaecd5296d746b1c006b4	1469	**86**	0x618ffd1cdabee36ce5992a857cc7463f21272bd7	580
37	0x6cc8dcbca746a6e4fdefb98e1d0df903b107fd21	1430	**87**	SumswapV2Pair	538
38	WhiteBIT	1421	**88**	0xa4a0f37031be58a66a8c17185b9a3f1c0634e192	535
39	Uniswap V3: USDT 3	1329	**89**	0x362de5a5b5cd9bb78f53af612d465f551c156799	531
40	0xb89eb49bc337d2bfaee360ec63ad606b6788ceb2	1282	**90**	0x03f19c3317c66024c8327d84117057e8583abd38	529
41	0x3dc7b06badfff6fbe2b0ce72144aef781d16a537	1250	**91**	0xf91a11b31ecd9a93aed7060680b5f7899d7cc98d	528
42	0x107c5b03713d85882b416fd864424d024175cd61	1226	**92**	0xd2eb1db3cdad1045efedb6d384fe6454377494a9	506
43	0xc176761d388caf2f56cf03329d82e1e7c48ae09c	1225	**93**	0x0823beba3f1f0caad19ce9e5724c4f5ce0a2fb97	501
44	0x48210514dc5b1e40faa3f395bc4c246e1b3176d3	1206	**94**	Proxy	497
45	MXC	1181	**95**	0x715c415416fb8f1f19597075366416bdb9d99215	493
46	Tokenlon: PMM	1160	**96**	Remitano 2	492
47	Celsius Network: Wallet 5	1110	**97**	0x6871eacd33fbcfe585009ab64f0795d7152dc5a0	487
48	Paribu 5	1099	**98**	0xfbec25aaecb61348cb9a8735868863330a44c487	482
49	SushiSwap V2: USDT	1095	**99**	Bit2C	480
50	0xa4f327ab8fd2f412ad377614ed1a36ce0b335e49	1057	**100**	Bitzlato	474

Table 10.1: Top 100 addresses by degree centrality

Rank	Name/Address	PageRank	Rank	Name/Address	PageRank
1	Binance 14	0.056194	**51**	HitBTC 3	0.001595
2	0xe59cd29be3be4461d79c0881d238cbe87d64595a	0.01566	**52**	0xa046a8660e66d178ee07ec97c585eeb6aa18c26c	0.001587
3	Binance 18	0.011791	**53**	0x691f987fd150967d6b6bd7e7b3d04e9dbc1f4efc	0.001544
4	Binance 17	0.011034	**54**	0x107c5b03713d85882b416fd864424d024175cd61	0.00147
5	Binance 15	0.009832	**55**	Nexo 2	0.001461
6	Kraken 7	0.008545	**56**	SushiSwap V2: USDT	0.001411
7	Coinbase 5	0.008375	**57**	0xf7b2f3cd946052f8b397f801299b80f053515af9	0.001384
8	FTX Exchange	0.008043	**58**	0xd90f85a53046ceca73f4b16ab66743721bddaff3	0.001372
9	0xfa103c21ea2df71dfb92b0652f8b1d795e51cdef	0.007529	**59**	HitBTC 2	0.001357
10	0xc176761d388caf2f56cf03329d82e1e7c48ae09c	0.007351	**60**	Vypercontract	0.001351
11	Binance 16	0.007336	**61**	0x75fb8a185225231868486ab598acb1fdde6eb1ae	0.001331
12	Coinbase 4	0.007156	**62**	0xa20f10289248717374e9b7776dc368aa526cb6f2	0.00133
13	0x56178a0d5f301baf6cf3e1cd53d9863437345bf9	0.006433	**63**	0xf584f8728b874a6a5c7a8d4d387c9aae9172d621	0.001305
14	0x5041ed759dd4afc3a72b8192c143f72f4724081a	0.005328	**64**	0xf05e2a70346560d3228c7002194bb7c5dc8fe100	0.001297
15	Uniswap V2: USDT 2	0.004709	**65**	Vault	0.001262
16	Huobi 4	0.004316	**66**	0x712d0f306956a6a4b4f9319ad9b9de48c5345996	0.001256
17	Crypto.com 2	0.004159	**67**	0xc7807e24338b41a34d849492920f2b9d0e4de2cd	0.001251
18	Bitfinex 3	0.004138	**68**	0xecbb27eb44033a24e76a56bb23129bdbef4f44e2	0.001251
19	Crypto.com	0.004125	**69**	Wintermute 1	0.00118
20	0xa57bd00134b2850b2a1c55860c9e9ea100fdd6cf	0.003943	**70**	0x0548f59fee79f8832c299e01dca5c76f034f558e	0.001177
21	0x74de5d4fcbf63e00296fd95d33236b9794016631	0.003534	**71**	DMMPool	0.001172
22	Huobi 3	0.003416	**72**	0xb3f923eabaf178fc1bd8e13902fc5c61d3ddef5b	0.001159
23	Huobi 12	0.003406	**73**	0x8f6e98ea87bf57dbde4cfa20c54cf812d5099451	0.001148
24	Huobi 7	0.003399	**74**	0xa75bae74e39dc2f38ee879ec3de9489abe65280e	0.001133
25	BitMart 2	0.003323	**75**	0x88dfbd93f509d0b3dcbfd86ee9d6d036ca199ea5	0.001115
26	0x61f2f664fec20a2fc1d55409cfc85e1baeb943e2	0.003271	**76**	0x0823beba3f1f0caad19ce9e5724c4f5ce0a2fb97	0.001094
27	0x0000000000123685885532dcb685c442dc83126	0.003151	**77**	0x6cc8dcbca746a6e4fdefb98e1d0df903b107fd21	0.001092
28	0xc333e80ef2dec2805f239e3f1e810612d294f771	0.003072	**78**	0xa2af528689e9df84ea482fbb1ab7685a9cf1e9f9	0.00109
29	0x83a127952d266a6ea306c40ac62a4a70668fe3bd	0.002948	**79**	0x4c8cfe078a5b989cea4b330197246ced82764c63	0.001076
30	Huobi 2	0.002893	**80**	Bittrex	0.001073
31	0x3e9afaa4a062a49d64b8ab057b3cb51892e17ecb	0.002832	**81**	0xa294cca691e4c83b1fc0c8d63d9a3eef0a196de1	0.001057
32	Huobi 1	0.002759	**82**	0x63a395b574d5e23c3dbc6986be5994ef6743afa8	0.001019
33	Huobi 36	0.002432	**83**	Compound Tether	0.001
34	0x3a61da6d37493e2f248a6832f49b52af0a6f4fbb	0.00239	**84**	0x522cc3a5a67902f58e9c1e6b2e51651df38b9162	0.000996
35	WhiteBIT	0.00237	**85**	0xf9623fe14733b3ef4f102626a219c9e89a5b5aae	0.000992
36	Huobi 9	0.002354	**86**	0x18877aeddef4fde5be9b818d713ee39a7ca79761	0.00099
37	0xf60c2ea62edbfe808163751dd0d8693dcb30019c	0.002135	**87**	Tokenlon: PMM	0.000981
38	Gate.io	0.002087	**88**	Coinex	0.000956
39	0xa5d07e978398eb1715056d3ca5cb31035c02fdad	0.001964	**89**	0x8ada0767b11e231ea6d90755bc05b6ab9dc1e93a	0.000919
40	Uniswap V2: FEI 3 fei-protocol	0.00192	**90**	0x263568570c148d52712cc879f1667783560d938b	0.000908
41	Celsius Network: Wallet 5	0.0019	**91**	0xd628f7c481c7dd87f674870bec5d7a311fb1d9a2	0.000905
42	0x6254b927ecc25ddd233aaecd5296d746b1c006b4	0.001849	**92**	MXC	0.000902
43	0xb8001c3ec9aa1985f6c747e25c28324e4a361ec1	0.001845	**93**	0xe32fd54add3c631eaf290179539f4bd7cf10d9b8	0.000897
44	0x77f7b398a23ef4cab31dd5503fd8446c4480c70b	0.001798	**94**	Polygon (Matic): ERC20 Bridge polygon-matic bridge	0.000869
45	0x48c04ed5691981c42154c6167398f95e8f38a7ff	0.001723	**95**	0xd9785bf22d7037a5ffdfb3ddb510a090b48145ac	0.000865
46	Bitfinex 2	0.00168	**96**	0x5077bad139cde8e7865a9f6d4c147c21bb13828e	0.00086
47	0x30395d1c24317eb5878dced2ff8e8334350aa056	0.001648	**97**	0x8c1b4060f885d89f4d0c063b341b6a596beee0e7	0.000855
48	0x3507e4978e0eb83315d20df86ca0b976c0e40ccb	0.001639	**98**	0x6753818a769f2024a454fd471d663fac4ddaeefb	0.000854
49	0xf3f702eee065d486d75db946b7c455fc083a06ea	0.001613	**99**	Bit2C	0.000851
50	Poloniex 4	0.001601	**100**	0xbea016fe612dbccfc065c3009f19cbdcd0a495e5	0.000842

Table 10.2: Top 100 addresses by PageRank centrality

Rank	Name/Address	Betweenness	Rank	Name/Address	Betweenness
1	Binance 14	1.67E+10	**51**	Uniswap V3: USDT 3	237853288.1
2	0xe59cd29be3be4461d79c0881d238cbe87d64595a	3.77E+09	**52**	0x107c5b03713d85882b416fd864424d024175cd61	233595219.8
3	Coinbase 5	3.02E+09	**53**	0x4e68ccd3e89f51c3074ca5072bbac773960dfa36	229311163.4
4	Coinbase 4	2.67E+09	**54**	0xb7807b02e17266146389083ff429ac488d34e704	226981178
5	Binance 17	1.91E+09	**55**	0x606de087c52a1e85e8c8922909f2efb6f714ab04	223786386.9
6	Binance 18	1.87E+09	**56**	0x6b71dcaa3fb9a4901491b748074a314dad9e980b	217879757.5
7	Crypto.com 2	1.69E+09	**57**	0x936db7e7c94611c691b20ee95ff7995e2563a6c8	212697962
8	FTX Exchange	1.45E+09	**58**	Polygon (Matic): ERC20 Bridge	208482671.8
9	Binance 15	1.37E+09	**59**	Poloniex 4	204546698.3
10	Kraken 7	1.34E+09	**60**	Uniswap V3: USDC-USDT	193511177.2
11	Binance 16	1.28E+09	**61**	0x2a9613babde7a2b393f253b36338374217da562d	192828538
12	0x5041ed759dd4afc3a72b8192c143f72f4724081a	1.25E+09	**62**	Paribu 5	191666180.3
13	Huobi 7	1.05E+09	**63**	0xb89eb49bc337d2bfaee360ec63ad606b6788ceb2	189969242.9
14	BitMart 2	7.44E+08	**64**	0xa4f327ab8fd2f412ad377614ed1a36ce0b335e49	181770086.9
15	Crypto.com	7.39E+08	**65**	Hotbit 3	177806098.5
16	Huobi 36	6.97E+08	**66**	Vault	169003799.8
17	Huobi 3	6.96E+08	**67**	0x03f19c3317c66024c8327d84117057e8583abd38	167779464
18	Huobi 9	6.68E+08	**68**	0x477b8d5ef7c2c42db84deb555419cd817c336b6f	164231935.9
19	Huobi 12	6.67E+08	**69**	0x74dec05e5b894b0efec69cdf6316971802a2f9a1	155974426.8
20	Huobi 1	6.56E+08	**70**	0xc176761d388caf2f56cf03329d82e1e7c48ae09c	155551035.6
21	Huobi 4	6.39E+08	**71**	0x0823beba3f1f0caad19ce9e5724c4f5ce0a2fb97	154495389.2
22	Huobi 2	6.37E+08	**72**	0xa4a0f37031be58a66a8c17185b9a3f1c0634e192	150762713.2
23	Gate.io	6.22E+08	**73**	0xb02f1329d6a6acef07a763258f8509c2847a0a3e	148363703.4
24	0x48c04ed5691981c42154c6167398f95e8f38a7ff	5.68E+08	**74**	0xc7807e24338b41a34d849492920f2b9d0e4de2cd	147976008.2
25	Nexo 2	5.57E+08	**75**	0xeec0ed9e41c209c1c53a35900a06bf5dca927405	144870959.6
26	0x3e9afaa4a062a49d64b8ab057b3cb51892e17ecb	5.55E+08	**76**	0xbea016fe612dbccfc065c3009f19cbdcd0a495e5	144586706.9
27	HitBTC 2	5.13E+08	**77**	0x5a79f666e951bea013c8dab9774c7d1d50ee8d20	138025088.1
28	HitBTC 3	5.13E+08	**78**	0x3c02290922a3618a4646e3bbca65853ea45fe7c6	137949072
29	0xf60c2ea62edbfe808163751dd0d8693dcb30019c	4.98E+08	**79**	0x8c8d7c46219d9205f056f28fee5950ad564d7465	137801073.1
30	0x0b54420ee63aa04da4cc87064142c6e64b70bb94	4.68E+08	**80**	0x5bcbdfb6cc624b959c39a2d16110d1f2d9204f72	137565465.4
31	0xa5d07e978398eb1715056d3ca5cb31035c02fdad	4.61E+08	**81**	SwissBorg 2	135917859.3
32	0x74de5d4fcbf63e00296fd95d33236b9794016631	4.44E+08	**82**	0xb8001c3ec9aa1985f6c747e25c28324e4a361ec1	135313976.2
33	0xd30b438df65f4f788563b2b3611bd6059bff4ad9	3.84E+08	**83**	SwissBorg 1	135104795.4
34	WhiteBIT	3.69E+08	**84**	0xa25979e9150a8ab187b7db0fccd3d0ea0e7aa03a	130954376.1
35	Celsius Network: Wallet 5	3.63E+08	**85**	Bitstamp 2	128178688.7
36	0xd90f85a53046ceca73f4b16ab66743721bddaff3	3.48E+08	**86**	0xc88f7666330b4b511358b7742dc2a3234710e7b1	126556012
37	0xf3f702eee065d486d75db946b7c455fc083a06ea	3.48E+08	**87**	0x2866608ffb89fde2bacb443fbee74d19439ac8a2	126296052
38	0x613142f57807166e7275ccbd8231c6f02b3ee079	3.48E+08	**88**	0x292f04a44506c2fd49bac032e1ca148c35a478c8	124881745.3
39	MXC	3.47E+08	**89**	0x3a9e6cf4e3157670a3b991c25d6f4fcbd9419c03	122359968.2
40	0x63a395b574d5e23c3dbc6986be5994ef6743afa8	3.01E+08	**90**	MultiSigWalletWithDailyLimit	122327057.2
41	Bittrex	2.96E+08	**91**	0xc97a4ed29f03fd549c4ae79086673523122d2bc5	122294146.2
42	0x6254b927ecc25ddd233aaecd5296d746b1c006b4	2.87E+08	**92**	Bit2C	122264516.4
43	0x691f987fd150967d6b6bd7e7b3d04e9dbc1f4efc	2.84E+08	**93**	Uphold.com	119759475
44	0x3dc7b06badfff6fbe2b0ce72144aef781d16a537	2.8E+08	**94**	0x9239df3e9996c776d539eb9f01a8ae8e7957b3c3	116121892.9
45	0x6753818a769f2024a454fd471d663fac4ddaeefb	2.69E+08	**95**	CEX.IO	114931413.4
46	0x6cc8dcbca746a6e4fdefb98e1d0df903b107fd21	2.58E+08	**96**	0x19af5ca9caa89fe7bceeec2396c4a47c2b0a3fc1	110488123.9
47	0xa75bae74e39dc2f38ee879ec3de9489abe65280e	2.54E+08	**97**	0xe3e43d939c1ebc6485fbc7a7ac900ce168841071	107486460.7
48	0x39bbce3be8c84b1ef7636bca567862790a666f38	2.5E+08	**98**	0x95b564f3b3bae3f206aa418667ba000afafacc8a	106621770.2
49	0xa20f10289248717374e9b7776dc368aa526cb6f2	2.48E+08	**99**	0xf91a11b31ecd9a93aed7060680b5f7899d7cc98d	105868953
50	Coinex	2.44E+08	**100**	0xbb5463f375c7ad90ef3bdd06069f778d1e8e52c6	105560735.4

Table 10.3: Top 100 addresses by betweenness centrality

Rank	Name/Address	Closeness	Rank	Name/Address	Closeness
1	Binance 14	0.370783	**51**	0x038adfb88435bad6012e230c1542cad9dbb74689	0.303982
2	Binance 17	0.344749	**52**	0x3125e57b033b989e306177a8fdb2b7501b56e6bf	0.303968
3	Binance 18	0.343566	**53**	0xdf7e6082ad33ffcc21ab76809f1d2c94e43c24f0	0.303923
4	Binance 15	0.340063	**54**	0x75e67e7d64af256c4ff73934adc91c8aedc1463d	0.303918
5	Binance 16	0.335967	**55**	0xf057d3652e9e3f57237c93b1d4d4a422cf37b1c2	0.303914
6	Crypto.com 2	0.335231	**56**	0x22cdb4ea7d12e6c6fabf8c7096275177bd9ccf26	0.303909
7	0xb8b53751f76492cc32c73553764e07ccd6a98517	0.332728	**57**	Uniswap V2: USDT 2	0.303854
8	0x18877aeddef4fde5be9b818d713ee39a7ca79761	0.33167	**58**	0x7fcffb2fec6a10c01646b99859d0bd696132f0e5	0.303842
9	0x95b564f3b3bae3f206aa418667ba000afafacc8a	0.330233	**59**	0x6ff5b9095f6e2eeafade0178a1af0711510e51cc	0.30384
10	0xfa103c21ea2df71dfb92b0652f8b1d795e51cdef	0.324924	**60**	0xea6433eb6beb3ae074299a6fb488b8842d9a6ab4	0.303832
11	0x48c04ed5691981c42154c6167398f95e8f38a7ff	0.320901	**61**	0x06dc621c7ca1d5c913fc2ae795e3226fbdeac950	0.303814
12	0xa294cca691e4c83b1fc0c8d63d9a3eef0a196de1	0.318579	**62**	0xccb6ddc7fcd02651f5b2f2eed8f886c736fc3b7a	0.303531
13	0xbb98f2a83d78310342da3e63278ce7515d52619d	0.318165	**63**	0x09e6a0f5f24ee4ab4290d3fa557dad0bdb502e22	0.303318
14	0xa2689c1a64a36b7bd047b1cf177a983df869f20a	0.3169	**64**	0x7c177d7d45d7689791800537a5a5a2dffd27495b	0.30305
15	0x0be04ef019cb6a621210f75ffe010f26a88eab7f	0.316749	**65**	0x69404e3baaeee469ee9d89ce3f2420e72276d9cb	0.302969
16	0x52d744f5c4dffcb235888de9d9b68068acf53b3f	0.315851	**66**	0xf1628cfcdfcdbf78b104db69f3827fca85b44828	0.302953
17	0x7e219c169d9f41a9f9f3982ac2e855f60d46433b	0.315728	**67**	0xbe47cefe05265e7878a34b3d903d60ac160f41b3	0.302918
18	0x00cde9024d5b58129ab132b67768b8b44977c98a	0.315598	**68**	0x923678d61cd5c8002822794cb1179b45a1e24de4	0.302905
19	0xe59cd29be3be4461d79c0881d238cbe87d64595a	0.315427	**69**	0x304f4d17435cf45003afea706d1d5d07a87d3d20	0.302852
20	0xff0d4730b6e35cb6bc92d63a3e0139e2638c70d9	0.314971	**70**	0x93b936a8439e2b9f6b746278b3d46f724529122d	0.302846
21	0x39a28a3ed6dacafb6fcc7248e65d51c5058c36d8	0.314386	**71**	0xb8375a171b8b5b446c6d4ee522645fe5d1445f0f	0.302768
22	0xc333e80ef2dec2805f239e3f1e810612d294f771	0.313319	**72**	0xb89eb49bc337d2bfaee360ec63ad606b6788ceb2	0.302727
23	0xadc4eec96c4ebb60baf3a7ff40a5f914d0d12cdf	0.313218	**73**	Paribu 5	0.302597
24	0x618ffd1cdabee36ce5992a857cc7463f21272bd7	0.312415	**74**	0x0bb9743b0bf4bf56973eea7865dbb879b36e25c1	0.302473
25	0xb3c839dbde6b96d37c56ee4f9dad3390d49310aa	0.311883	**75**	Kraken 7	0.30228
26	Coinbase 5	0.311506	**76**	0x96a6357bfc366deb91c4876b10633bf87a0c4e0d	0.302192
27	0x2ddf54edd17e7488a7e2f126405af2092faabb4e	0.311298	**77**	0x6e7dc9b3598af2128316899ee73b23bdd67f86f0	0.302149
28	0x11390cb49e9f6b15746731eb3c32175457f06595	0.309832	**78**	0x7e409ac64a7c145b08b9c85465e9fd11efef59b3	0.302143
29	0xdfb88afc565d884a1ab095a499b1edfa29cda666	0.309753	**79**	0xa4a0f37031be58a66a8c17185b9a3f1c0634e192	0.302133
30	0x16bd48f83cca25e6d3e315433264c028a529e827	0.309184	**80**	0xf83c5fb678ba19d3bf25adaaad90b3e910add4b5	0.302123
31	Coinbase 4	0.309059	**81**	0xdbe73943a6608a4e4f4d910d9319b9cec3ff5f49	0.301979
32	0xfebf53765bbfd79d9eb4ccbaa9ba85491da07350	0.308271	**82**	0x13ffdb156d5e6f8afb6c687c784580db01747faa	0.301901
33	0x5b5ecfc8122ba166b21d6ea26268ef97e09b2e9f	0.308158	**83**	0xd529b8bb671aa459f3fbae0d9d3f9973d7f89820	0.301803
34	0x83a127952d266a6ea306c40ac62a4a70668fe3bd	0.308082	**84**	0x832c431acf68b2d810fee046a22f3653b321bfb6	0.301768
35	0x76854eaed12bdbfbde931f7ea2d5e84667bc064b	0.307875	**85**	0x9239df3e9996c776d539eb9f01a8ae8e7957b3c3	0.301747
36	0xf584f8728b874a6a5c7a8d4d387c9aae9172d621	0.307242	**86**	0x4ff52c8b581482c69c505ab9cd984101cc8faf1a	0.301643
37	0xa378f07d52b586e2ba32c1895542ec72a5bc55c7	0.307012	**87**	0xa85fa9aa54700bef32abd88265e6238004269c94	0.301601
38	0x0aeee23d16084d763e7a65577020c1f3d18804f2	0.305904	**88**	0x63328adb8268a74781d2d9c74671b248a3eca755	0.301573
39	0xc30f4dcc2c1d26d649a2ac5f77b557d741b4a28a	0.305902	**89**	0x6370a8fae7dee4fb63cf16d2e9fb5bbeb7bb47a6	0.301396
40	FTX Exchange	0.305541	**90**	0xc166fb16321f320022d2b98eae85de2f05429e6e	0.301318
41	0xdd61fec137ea26023817c5dab58202fbf17431aa	0.305359	**91**	0x4f124cef2484c7d5aabc619e860ede0763f6acce	0.301222
42	0xde94797e89e18fcacb86bac3ebd4e5924308049f	0.3053	**92**	0x7346f9820c63f0f0d978fb79ae1e3ce6f779ffe1	0.301202
43	0xb7a5dc6d571e9591a439ba33a47e5b09831a246e	0.304985	**93**	0xaf2c7eb4ee0a72c478cc7544a277254a276b6146	0.301168
44	0x2e13bd23bcbee5fa4799a5959c9d3ac760fec947	0.304977	**94**	0xaa6f6dd54542afac6e7b1f995d235aa81c8cb77f	0.301166
45	0xa1f422e732ec2a8b0d6375e3e1eeafc199acbd57	0.304971	**95**	0x08fff51377de51a919021857c31f54caf74f47f1	0.3011
46	0xe78d52399282e9ffe666f6619b31b0ba130cd6ae	0.304902	**96**	0x75e645e50e4a4f8dc9e2c2021e054f29ec9ad3b8	0.3011
47	0xa25979e9150a8ab187b7db0fccd3d0ea0e7aa03a	0.304167	**97**	0x897b425dab19eb886dc6ae2010fe2a0de85308fa	0.300933
48	0x5041ed759dd4afc3a72b8192c143f72f4724081a	0.304142	**98**	0xb04a2232ea59f903a36e3ed4b1d1c90c71dc5814	0.300917
49	0x2b078a13487f160dcb0099df70940a898aacfe81	0.304106	**99**	0xc58bb74606b73c5043b75d7aa25ebe1d5d4e7c72	0.300857
50	0xe8f9d8912f77dcd7b119490f3bfd37187ea19bfe	0.304002	**100**	0x11205f1a38fcbd5491b8a99b416836ae4ca4f001	0.300847

Table 10.4: Top 100 addresses by closeness centrality