The Ethereum network: a graph analysis

Quinten Bruynseraede
R0674455

Louis Van Looy
R0861408

April 19, 2021

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum laoreet dapibus purus. Maecenas eleifend, ipsum eget pretium blandit, nibh diam fermentum turpis, quis semper urna orci eget dolor. Aliquam at vehicula odio. Sed facilisis, odio vitae faucibus consequat, nisi metus condimentum leo, et malesuada sapien metus nec odio. Etiam vel orci non dolor interdum ultricies. Nullam in justo dui. Aenean ut eros non felis molestie lacinia eget ut turpis. Mauris malesuada efficitur libero, at volutpat arcu imperdiet ac. Vivamus gravida sodales felis, varius ultrices magna luctus sit amet. Donec blandit ante neque, at blandit dui vestibulum in. Nam vel lorem.

1 Introduction

The Ethereum network is a blockchain-based network for distributed processing. It contains a thriving ecosystem of distributed applications (dapps), ranging from financial tools to games and gambling. Ethereum's native token is called Ether (ETH), and it has become a acknowledged way of storing and transferring money: a so-called *store-of-value*. In this report, we analyze how ETH is distributed among users, and how it flows between them. In particular, we analyze the role of a group of very wealthy users (called *whales*), and we investigate how important these whales are for the network, and to what degree they contribute to wealth inequality.

In Chapter 2, we introduce some basic concepts related to the ethereum network. Chapter 3 outlines past work about network analysis of cryptocurrency networks, in particular the Ethereum network. In Chapter 4, we describe how we collected the necessary data. Chapter 6 analyzes the network using a set of network analysis tools, and try to show the importance of whales. Chapter 7 focuses on inequality within the Ethereum network. In Chapter 8, we try to analyze clusters in the network, and try to link these clusters to their utility in the network.

2 Ethereum

In the following section, we explain fundamental terms and definitions needed to understand the Ethereum network. Based on this, it is also intended to better understand the network and thus observe and understand the specific characteristics of the particular network.

2.1 Ethereum

First of all a formal definition of Ethereum is needed. Ethereum is an open-source, blockchain-based and decentralized software platform. It makes use of its own cryptocurrency, called Ether. The main novelty of Ethereum is its ability to execute smart contracts and create decentralized, distributed applications without any downtime, fraud, control or involvement from a third party. A smart contract can be seen as a piece of code that describes the rules for a transaction between two parties. Smart contracts on the Ethereum network are Turing-complete (i.e. they can describe arbitrary computations), and are mostly programmed in the Solidity programming language.

2.2 Addresses and transactions

Users store their Ether in an account, uniquely represented by a 20-byte address, such as `0xed9a430d9a11616eb1cb07ebc28c9e20a03bd486`. Transactions contain (among other fields), a sender, a recipient and a transaction value expressed in *wei*. 1 Ether is equal to $10e^{18}$ wei. Smart contracts are also identified by a 20-byte address, and executing these smart contracts can be done by sending them a transaction. All transactions are public by default, but there is no direct link between the identify of a user and his address: the vast majority of transactions happen anonymously. However, many services use addresses that are publicly known. For example, Binance, one of the largest cryptocurrency marketplaces, has at least 20 known addresses [1].

2.3 Whales

An important element within the composition of the Ethereum network are whales. Ma (2021) gives the following definition: "The term *whale* is used to describe an individual or organization that holds a large amount of a particular cryptocurrency." This amount is not clearly defined, but within the cryptocurrency community a threshold of around 1.000 of a particular cryptocurrency is being seen as the standard. It is clear that this value can fluctuate, based on the underlying value of the cryptocurrency. In

[1] https://etherscan.io/accounts/label/binance

the Ethereum network this threshold is set on owners who own 10.000 Ethereum or more, these holders control approximately 70

2.4 The network

The Ethereum network has seen tremendous growth since its inception in 2015, and contains over 350 million transactions at the time of writing. Since June 2020, the daily number of transactions exceeds 1 million [2]. Over 100 million addresses have been registered, and over 600 thousand addresses take part in transactions every day [3]. This means a full analysis is generally not feasible.

3 Related Work

TODO

4 Data collection

Given the size of the Ethereum network, we have to resort to sampling techniques. One type of sampling uses *random walks*, that start at a given point and follow edges while collecting the necessary data. Random walks are particularly useful for networks where it is not possible to delimit a part of the network in advance [1]. Ethereum is such a network: addresses are not ordered or structured in any way. We collect transactions using a Breadth-first search (BFS). To analyze the evolution of the network, we create six separate snapshots: each snapshot spans 1 month and is limited to 1 million transactions.

$\texttt{T} = \emptyset$
Initialize `Queue` with 100 known addresses
while $T.size \leq 1e6$ **do**
  1. Pop an address `A` from `Queue`
  2. Fetch all transactions by `A` during the time period
  3. Add all transactions to `T`
  4. Add all addresses that occur in these transactions to `Queue`
end

Although transactions contain many more features, we only use the columns `from`, `to`, and `value`. Each snapshot can then be considered a graph: nodes are addresses and edges are transactions between addresses. The value of the transaction is the weight of the edge. If multiple transactions occur between the same pair of addresses, we sum their values.

The following table shows the size of each snapshot:

Snapshot	**#txs**	**#nodes**	**#edges**
Jan '17	475 598[4]	81 242	186 942
Nov '17	988 086	548 355	657 697
Sep '18	1 000 258	528 506	635 623
Jul '19	1 006 153	550 393	673 967
May 20	1 003 812	474 030	580 773
Mar '21	1 001 864	566 733	701 482

To account for seasonal effects, the time between two snapshots is always 10 months. We also don't consider the state of the network before January 2017, when it was still relatively new and the number of transactions was low.

5 Analysis of the Ethereum network

First of all, it appears the average degree has remained constant over time (between 1 and 1.25), except for the snapshot Jan'17. We suspect the smaller size and consequentially larger influence of the starting addresses cause this larger average degree in 2017.

Figure 1 shows the degree distribution for each snapshot. The distribution appears to follow a power law at first (estimated $k^{-2.1}$), but as the degree increases, the distribution is more or less uniform. In particular nodes with a degree above 5000 would be very rare in a perfect scale-free network, but we observe them often.

6 Inequality

TODO

7 Clustering

References

[1] Luca Becchetti, Carlos Castillo, Debora Donato, and Adriano Fazzone. A comparison of sampling techniques for web graph characterization, 2006.

[2] https://etherscan.io/chart/tx
[3] https://bitinfocharts.com/comparison/ethereum-activeaddresses.html

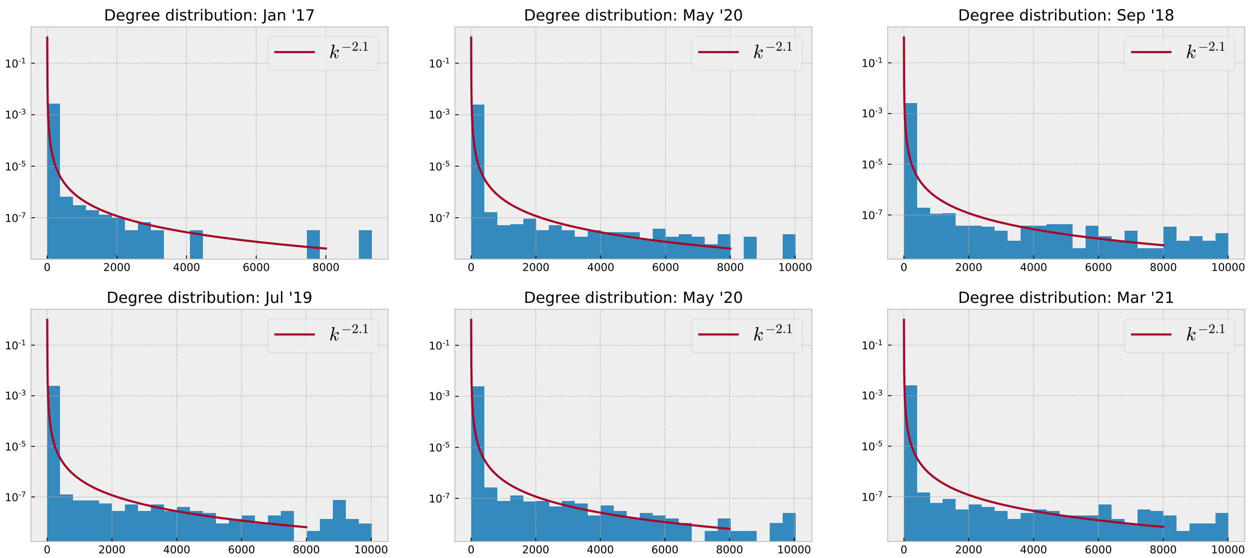

Figure 1: Degree distribution for all snapshots