On the Ethereum blockchain structure: A complex networks theory perspective

On the Ethereum Blockchain Structure: a Complex Networks Theory Perspective

Stefano Ferretti, Gabriele D'Angelo
Department of Computer Science and Engineering (DISI), University of Bologna, Italy
{*s.ferretti, g.dangelo*}@*unibo.it*

***Abstract*—In this paper, we analyze the Ethereum blockchain using the complex networks modeling framework. Accounts acting on the blockchain are represented as nodes, while the interactions among these accounts, recorded on the blockchain, are treated as links in the network. Using this representation, it is possible to derive interesting mathematical characteristics that improve the understanding of the actual interactions happening in the blockchain. Not only, by looking at the history of the blockchain, it is possible to verify if radical changes in the blockchain evolution happened.**

***Index Terms*—Blockchain, Ethereum, Complex Networks**

1. Introduction

The blockchain is arguably one of those technologies that, nowadays, are raising high expectations in terms of possible application domains. It is a global ledger that records transactions efficiently and permanently on a chain of blocks [27]. Each block contains a set of transactions created and dispatched in the system. Furthermore, each block contains a timestamp, a link to the previous block and it is identified by its hash value. All transactions are signed and hashed via cryptographic hash functions. This structure thus provides an unforgeable log containing the history of all the transactions ever made. Nodes participating to the blockchain are connected via a Peer-to-Peer (P2P) network. Each node maintains a replicated version of the entire transaction history.

Several variants of blockchains exist. While Bitcoin still remains the most famous one by the public, Ethereum probably represents one of the most interesting solution. This is due to the fact that Ethereum provides a vast range of use case applications enabled by smart contracts [1], [3], [32], [33], [25], [30], [24], [13]. Ethereum is often described with the term ”world computer”, since this platform enables running distributed applications (i.e. smart contracts) in a distributed manner. It provides a way to create self-executing and self-enforcing contracts. Their execution is triggered via transactions. Once generated, nodes in the P2P system execute the related code. This causes a change of the state. All this is recorded in the blockchain. Thus, through the blockchain all nodes synchronize their replicated state globally, in a manner that is fully verifiable by any system participant. That is why the distributed code run on the blockchain is referred as a smart contract. Once deployed, it cannot be modified. Hence, parties, that agree on the use of this code, are aware that there is no possibility to breach the agreement. (They can, of course, decide to not use that contract anymore, if for some reason the contract becomes obsolete.)

In Ethereum, smart contracts are considered internal accounts, that can interact among themselves and with externally owned accounts, which are in fact users that employ the system. Both these kinds of accounts have their own balance, expressed in a distributed currency referred as Ether. The Ether is the fuel for operating in Ethereum. Every transaction in Ethereum is made possible through a payment made by the clients of the platform to the machines executing the requested operations. This enables several applications, ranging from the exchange of cryptocurrencies, to financial applications, storing and management of tokens and digital assets, notary systems, identity management, voting systems, up to those application that require the traceability of resources and assets [5], [32], [15], [28]. Several works find many application domains in healthcare, supply chain, Internet of Things, etc. [2], [5], [23], [18], [29], [21], [20].

In this paper, we provide an analysis of the Ethereum blockchain. In particular, we employ the modeling techniques of the complex network theory [3], [7], [8]. We represent the flow of transactions happened in the blockchain (or a subset of the blockchain) as a network, where nodes are the Ethereum accounts (i.e. external accounts or smart contracts). In the Ethereum scenario, a transaction can represent some cryptocurrency transfer, the creation of a smart contract, or the invocation of a contract [8]. Each transaction recorded in the blockchain corresponds to the creation of a new link in the network. The rationale behind the analysis is that complex networks provide appropriate modeling to represent a blockchain as a complex system, together with powerful quantitative measures for capturing the essence of its complexity [9], [14], [31], [6], [26], [19].

Varying the number of blocks considered to extract the recorded transactions, we obtain different networks, of different size and complexity. This influences the structure of the network. The investigation, made in this work, leads

[0] The publisher version of this paper is available at https://doi.org/10.1002/cpe.5493. **This is the pre-peer reviewed version of the following article: "Stefano Ferretti, Gabriele D'Angelo. On the Ethereum Blockchain Structure: a Complex Networks Theory Perspective. To appear in Concurrency and Computation: Practice and Experience (Wiley)".**

to observations and insights. For instance, while a majority of nodes has a low degree (i.e. just few amount of links), that demonstrates a poor level of interactions in the blockchain, we notice the presence of several hubs with higher degrees. This information is important to recognize which are the main contributors to the blockchain evolution. While these nodes are important ones, at the same time, they are exposed to a lower level of anonymity, i.e. if we are dealing with an external owned account, it might be easier to discover the identity of the address corresponding to a specific node [26].

The remainder of this paper is organized as follows. Section 2 describes the background and the state of the art about the blockchain technologies, Ethereum and complex network analysis. Section 3 presents the approach used to model the Ethereum blockchain, and the interactions recorded in this distributed ledger, as a complex network. Section 4 presents results obtained from the analysis of the different extracted networks. Finally, Section 5 provides some concluding remarks.

2. Background

2.1. What is a Blockchain

A blockchain is a distributed ledger that records transactions in blocks [4], [11]. Each block contains a set of transactions and it has a link to a previous block, thus creating a chain of chronologically ordered blocks. Transactions within a block are assumed to have happened at the same time. In the typical scenarios, transactions record an exchange of digital currencies, but in fact they can be employed to record any kind of event.

What makes the blockchain technology appealing is that the combination of P2P systems, cryptographic techniques, use of distributed consensus schemes and pseudonymity ensure that the set of confirmed transactions becomes public, traceable and tamper-resistant. The latter property is obtained by linking subsequent blocks together using cryptographic hash functions so that the modification of transaction data in a block B_i would change the hash that is contained in the subsequent block B_{i+1}, thus altering the content of block B_{i+1} and so on. The blockchain is replicated across multiple nodes in a P2P fashion. Therefore, any attempt to alter the blockchain would create an easily detectable inconsistency of all replicas.

The blockchain uses digital pseudonyms (addresses) – usually, a hash of a public key – to provide some level of anonymity. Therefore, everyone can trace the activities of an entity with a given pseudonym, but it is computationally expensive (although not impossible) to associate a pseudonym back to a specific entity or individual.

2.2. Ethereum

Ethereum is a specific blockchain-based software platform that enables the possibility of building and running smart contracts and the so called Distributed Applications (DApps) [27]. Such platform is also the basis for a related virtual currency, called Ether. For the definition of smart contracts, Ethereum provides a Turing complete programming language that allows creating programs and running them on the blockchain [7].

Ethereum operates using accounts and their balances, that change via state transitions. The state denotes the current balances of all accounts, plus other possible extra data. The state is not stored on the blockchain directly, but it is encoded and maintained by accounts in a separate data structure organized as a Merkle Patricia tree. As in all permissionless blockchains, in order to provide anonymity, accounts are pseudonymous and are linked to one or more addresses [17]. There are two types of accounts: externally owned accounts and contracts accounts. Externally owned accounts are controlled by people. Thus, similarly to Bitcoin, each person has his own private key, which is used in order to make transactions in the Ethereum blockchain. Conversely, contract accounts are controlled by some smart contract code. In other words, such accounts are some sort of cyber-entities, having their own balance, that can be triggered through some transactions, coming from an external account (or some other contracts). Once triggered, the code specified in the contract is executed. This code can in turn generate some other transactions. The presence of these smart contracts allows developers to use Ethereum as a general purpose framework to create DApps.

The Ether is the cryptocurrency asset employed in the Ethereum blockchain. In some extent, the Ether is the fuel for operating the distributed applications over Ethereum. Using this cryptocurrency, it is possible to make payments to other accounts or to the machines executing some requested operation. Ether thus enables running DApps, enabling smart contracts, generating tokens during Initial Coin Offering (ICOs), i.e. a type of funding using cryptocurrencies, and also for making standard P2P payments. That's why Ethereum is also referred as "programmable money".

2.3. Complex Networks Analysis

Complex networks theory allows to analyze a given real or synthetic system, and to extract several mathematical properties that describe it. It is quite usual to represent P2P and distributed systems [9], communication networks [10], [14], social networks [12], biological and very other diverse phenomena as complex networks. In order to describe a phenomenon as a network, entities are usually represented as network nodes, while interactions among these entities are links that connect these nodes. Depending on the symmetric or asymmetric nature of the interaction, these links may be undirected or directed, respectively.

In what follows, we briefly introduce the the main metrics, typically employed in complex network theory, that will be used to study the Ethereum blockchain.

2.3.1. Number of nodes. This measure is the total amount of nodes in the network. In our case, that is the total amount of different accounts which were involved in some transactions, in the considered snapshot of the Ethereum blockchain.

2.3.2. Degree distributions. The degree of a node x is the amount of links that connect x with other nodes in the network (included x itself, when a loop is performed). The degree counts the number of addresses a given address had interactions with (i.e. it was involved in one or more transactions).

Weights can be associated to links and exploited to measure the so called weighted degrees. In this case, a weight is assigned to each link, measuring the amount of transactions between two addresses in the considered time range. Thus, the weighted degree is the summation of the weights of links of a given node.

2.3.3. Distance. The average distance is the average shortest path length in a network, i.e. the average number of steps along the shortest paths for all possible pairs of nodes. Distances among nodes are calculated using the standard breadth-first search algorithm, which finds the shortest distance from a single source node to every other node in the network.

This metrics should be considered together with the clustering coefficient. In fact, these two metrics allow determining if the network is a small world or not (as it will be described in the following of this section).

2.3.4. Clustering coefficient. The clustering coefficient is a measure assessing how much nodes in a graph tend to cluster together. It measures to what extent friends of a node are friends of one another too. When two connected nodes have a common neighbor, this triplet of nodes forms a triangle. The clustering coefficient is defined as

$$C = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triplets of nodes}}$$

where a "connected triple" consists of a single node with links reaching a pair of other nodes; or, in other words, a connected triple is a set of three nodes connected by (at least) two links [14]. A triangle of nodes forms three connected triplets, thus explaining the factor of three in the formula. In this context, a triangle of nodes means that an address x had some transactions with other two, say y, z, and at the same time y and z had some transactions as well.

2.3.5. Small worlds. Small world networks are networks that are "highly clustered, like regular lattices; yet, they have small characteristic path lengths, like random graphs". In a small world, most nodes are not linked with each other, but most nodes can be reached from every other by a small number of hops. Indeed, in a small-world network the typical distance between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes.

Given a network, it is possible to verify if it is a small world, by comparing it with a random graph of the same size. A random graph is a network with links randomly generated, based on a simple probabilistic model [14]. Different models can be employed to generate a random graph. According to one of the simplest methods, a random graph can be constructed by creating a set of n isolated nodes; then, we consider every possible pair of nodes x, y, and we add a link (x, y) with probability p, independently of other links. Random graphs exhibit a small average distance among nodes (varying typically as the logarithm of the number of nodes, $\sim \ln(n)$) along with a small clustering coefficient $\sim \frac{\text{mun links}}{n^2}$.

In practice, one can assess whether a network has a small average distance as for a random graph, but a significantly higher clustering coefficient. In this case, the network is a small world. In particular, if one looks at the clustering coefficient (cc) together with the average distance (L) of the considered network, and the clustering coefficient (cc_{RG}) together with the average distance (L_{RG}) of the corresponding random graph, it is possible to measure the small-coefficient as

$$\sigma = \frac{cc/cc_{RG}}{L/L_{RG}}, \tag{1}$$

concluding that the network can be classified as a small world when σ is significantly higher than 1.

3. The Ethereum Blockchain as a Network

It is possible apply the complex network machinery for the analysis of a blockchain. In this case, accounts that interact in the blockchain can be represented as network nodes, while their interactions can be seen as links. More specifically, the interactions represent transactions among different accounts. It is also possible to associate a weight to each link, that may further characterize the interaction. For instance, a counter may be associated to track the number of transactions made in the time interval of consideration. Alternatively, it might represent some other value, such as the currency being transferred between the two accounts.

Figure 1 shows a screenshot of the 3D visualization of approximately 1-hour log of interactions (i.e. 240 blocks) in the Ethereum blockchain network. This figure shows some interesting indications about the network of transactions. In fact, there are several important nodes that are involved in many transactions, while the majority of nodes seem to have a single link entering/exiting from them. This is quite reasonable, since at the current status of this blockchain and its related cryptocurrency, it seems to be unlikely that a typical account participates in more that one transaction per hour. The analyzed network is composed of 28867 nodes, corresponding to the same amount of accounts that have been active in the single hour being considered. The number of links is 32800.

3.1. EtherNet Galaxy: a Software for Blockchain Analysis

The complex networks analysis described above has been performed using a new software called EtherNet

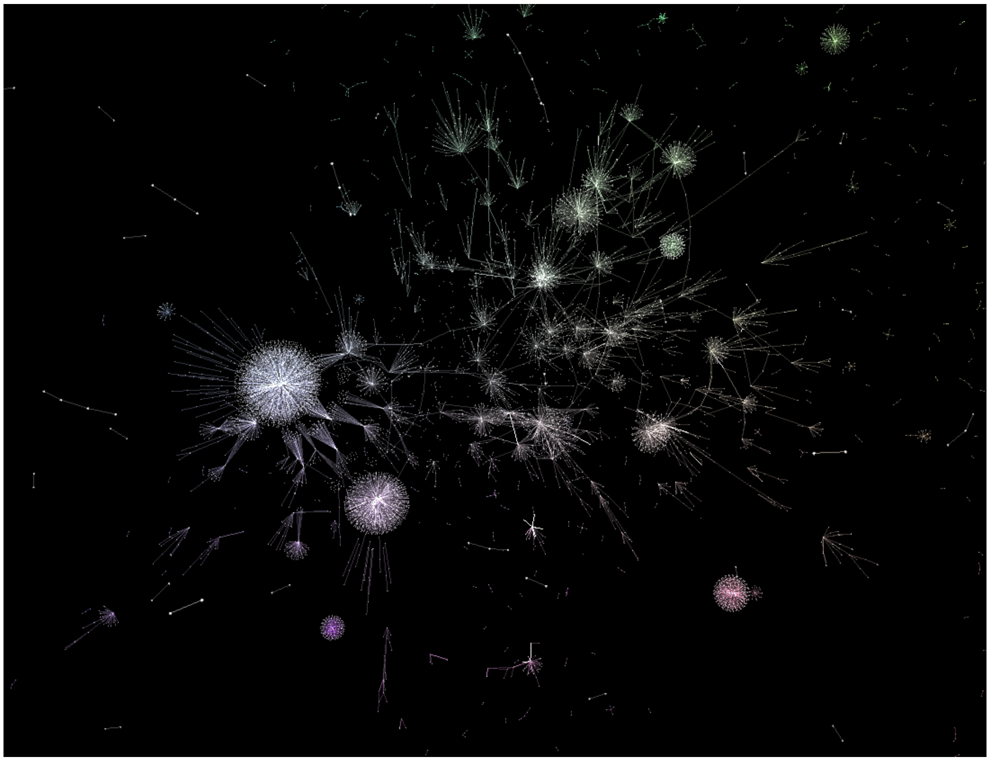

Figure 1: A screenshot of the appearing Ethereum interactions, as a network – 1 hour log (approximately).

Galaxy. A complete description of EtherNet Galaxy is beyond the scope of this paper but in the following a brief description of the main EtherNet Galaxy features is reported. Firstly, EtherNet Galaxy retrieves the Ethereum blockchain data using the APIs provided by Infura (https://infura.io/), a service that delivers RPC access to the Ethereum network. Thanks to this service, EtherNet Galaxy is able to retrieve the information about the blocks (e.g. block number, size, list of transactions, etc.). Secondly, the retrieved blocks are analyzed using the web3-eth package (https://github.com/ethereum/web3.js/). In this stage, the goal is to extract all the transaction encoded in each block and to represent them as a network using the Pajek data format. Finally, the network analysis on the previously generated graphs is performed by EtherNet Galaxy relying on the Python NetworkX software library [16]. The EtherNet Galaxy network analysis software is currently under development, a prototype version has been used for the analysis reported in the following of this paper. At a later stage, EtherNet Galaxy will be made freely available on the research group homepage (https://site.unibo.it/anansi/).

4. Results

Table 1 shows some main metrics related to six networks obtained by considering the set of transactions contained in different numbers of blocks, i.e. 1, 10, 100, 1000, 10000, 100000. A the time of writing, the Ethereum blockchain explorers, e.g. https://bitinfocharts.com/ethereum/, report that the average time between blocks is 14.1 sec, while the average number of blocks per hour is 254. Thus, we can roughly state that the considered networks are related to numbers of transactions ranging from few seconds up to 16 days.

As we expected, if we consider transactions that are contained in a single block, we obtain a very simple network, with few nodes and few edges. The number of nodes is higher than the number of edges; we might thus expect that there are transactions with multiple recipients. Due to the (essentially) random nature of the choice of transactions inserted in a block, we can imagine that the transactions involve different nodes. Thus, the resulting network is very sparse. Indeed, there are no triangles in the network (the clustering coefficient is zero). The amount of network components[1] is relatively high, with respect to the number of nodes in the network.

Let's consider the main component of the 1-block network. It is composed of 19 nodes and 18 links (as reported in the table). We already stated that this network corresponds to a bunch of transactions included in a single block i.e. probably generated in a small time interval. We might have two options here. The first one is that the component has a star structure, meaning that a specific address had an interaction with a set of nodes as shown in Figure 2. Indeed, star structures appear frequently also in the wider

[1] A component is a sub-graph of the main net, in which any two nodes are connected to each other by paths, i.e. the component is composed of nodes connected through links.

TABLE 1: General Metrics of Considered Networks.

# Blocks	# nodes	# edges	avg clus coeff	# components	# nodes largest comp	# edges largest comp
1	55	40	0	15	19	18
10	846	648	0	199	139	139
10^2	7507	7184	0.001	729	4262	4641
10^3	47469	51357	0.006	2848	37201	43682
10^4	284630	347679	0.014	10770	239114	303248
10^5	1467960	2144095	0.036	40276	1321468	1994428

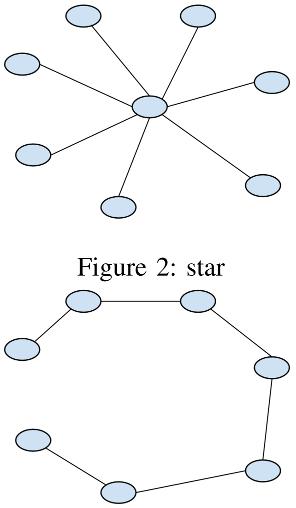

Figure 2: star

Figure 3: chain

Figure 4: Different interaction patterns among Ethereum addresses, depicted as a network.

network reported in Figure 1. The second option is that the interactions correspond to a chain of different transactions (Figure 3). We think that this alternative is more unlikely, since it would mean that during the block generation, the miner selected a bunch of transactions, generated in a given time interval, involving a chain of accounts.

As concerns isolated pairs of connected nodes, or small sized components, these might represent few (test) transactions among different accounts. Alternatively, they might represent the deployment of some (prototype) smart contracts, that is indeed realized through the triggering of a specific transaction in the blockchain, plus some possible (test) interactions with the smart contract. Indeed, at the time of writing, this is a quite common use of the Ethereum blockchain.

When we consider wider networks, built based on higher amounts of blocks, values of the considered metrics increase considerably. Still, all of them show a low average clustering coefficient. All networks show a high number of components, with respect to the amount of nodes. However, when we increase the number of blocks the main component embodies a high majority of network nodes (e.g. $\sim 90\%$ for the biggest net).

4.1. Degree Distribution

Figures 5–10 show the degree distributions of different networks obtained by analyzing an increasing number of 10^n blocks, with $n = 0, \dots, 5$; i.e. from 1 up to 10^5 blocks, respectively. In each figure, we report the degree distribution using a linear scale (left chart) as well as in a log-log scale (right chart). The log-log chart is interesting, since it easily allows to understand if, for instance, the degree distribution follows a power law function (in this case the plot of the degree distribution should appear as a straight line), rather than an heavy tailed distribution, etc.

With just one block, the obtained network is quite simple. There is a limited number of nodes that created transactions (included in the block). Moreover, it is quite unlikely that an account is involved in more than one transaction per block. The degree distribution in Figure 5 confirms this.

When we increase the order of magnitude of considered blocks, then things start changing. Still, the high majority of nodes performed a single transaction (i.e. they have a degree equal to 1). However, the percentage of nodes with higher degrees (i.e. higher amounts of transactions with different nodes) increases. If we look at the chart in log-log scale, we can see an almost linear decrease on the distribution of the degrees, with a long tail, suggesting that those degrees follow a power law function.

It is important to mention that a common practice in cryptocurrencies, especially in Bitcoin, is to create a fresh address for each payment a user receives. This in order to decouple the recipient of different transactions and increase the level of anonymity. Indeed, the wallet of a cryptocurrency is enabled to manage different user addresses/accounts, and a new address makes it more difficult to trace the cryptocurrency trail. Actually, most online wallets automatically create a new address each time a user is involved as the output of a transaction. Moreover, creating a new address for each transaction is a fool-proof way of ensuring that someone has paid the user, because the user gave that address to only that person and no one else.

4.2. Small World Phenomenon

To assess the small world property of the considered networks, we compare the clustering coefficient and the average path length of their main component with that of the equivalent random graph (generated by taking the same amount of nodes, with the same amount of edges randomly distributed among these nodes). Thank to these values, we

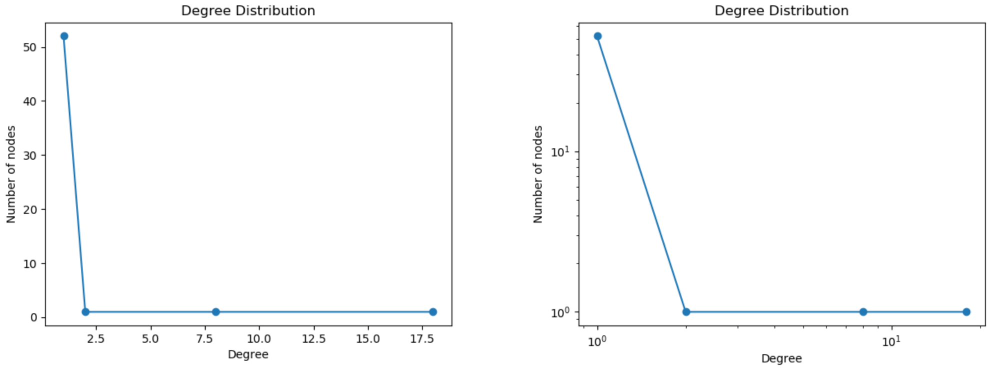

Figure 5: Degree distribution in linear and log scales – 1 block.

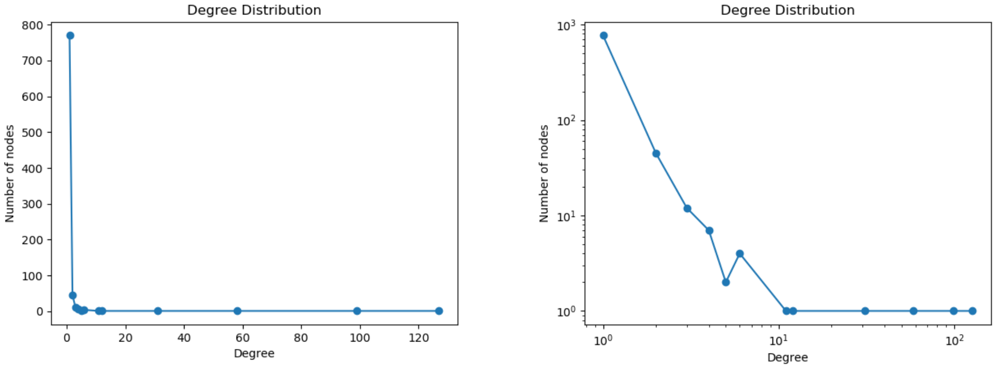

Figure 6: Degree distribution in linear and log scales – 10 blocks.

are able to compute the σ value (Equation 1). All these measures are reported in Table 2.

As already mentioned, with a low amount of blocks, a limited set of transactions is considered. Thus, we have very simple networks with small main components. These components have a null clustering coefficient ("cc" column in Table 2). It is clear that these networks are not small worlds. Also when we increase the amount of blocks, the clustering coefficients of the obtained networks is almost zero. This allows to conclude that even bigger networks are not small worlds. As concerns the 10^3-blocks network, the clustering coefficient of the random graph, obtained with the same amount of edges of the original network, is so small the the final σ value is really high. However, we claim that this is just a numerical outcome and it does not justify stating that this network is a small world.

The fact that the considered networks cannot be associated to small worlds is confirmed by looking at their related diameters, which are reported in Table 3. It is worth noticing that the diameter is calculated on the main component only. When we increase the network size, the diameter increases as well. The increment of the diameter is particularly evident when passing from the 10-blocks network to 10^2-blocks one. In this last case, we notice a network diameter of 23 that is quite far from the "six degrees of separation", that is usually associated to small worlds [14].

4.3. Metrics at Different Snapshots

As a further analysis, we tried to understand if the Ethereum network, obtained using a set of blocks, changes over time, i.e. we tried to understand if the evolution of the Ethereum blockchain has some effects on the related networks. To answer this question, we took different slices of the blockchain, by considering different snapshots starting at block numbers: 1000000, 2000000, 3000000, 4000000, 5000000, 6000000 and 7000000. The size of each of these blockchain slices are 1000 blocks (i.e. approximately 4 hours).

Figure 11 shows different network metrics, when we consider networks obtained from different points of the blockchain. These metrics are the amount of nodes (we report both the total amount of nodes in the net and the number of nodes in the main component), the amount of links (both in the whole network and in the main component), the number of components of the network and the average shortest path length (in the main component). In general, all these charts demonstrate how the network has grown in time, in terms of nodes and interactions. We can notice a spike in the snapshot took at 5000000. It seems that in that

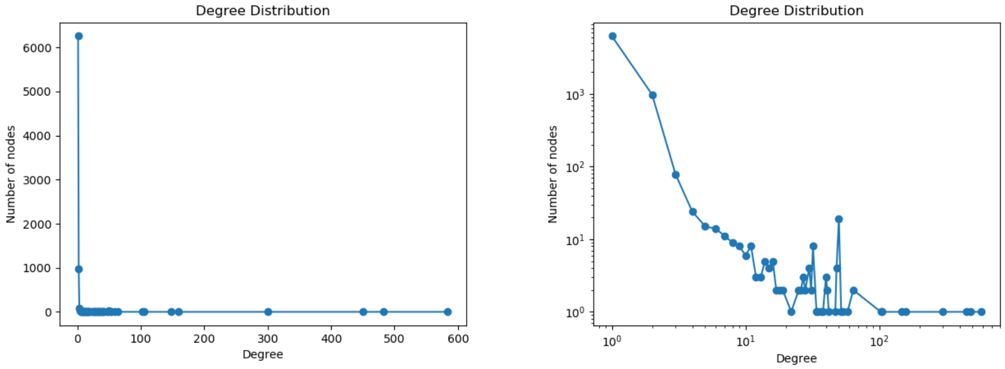

Figure 7: Degree distribution in linear and log scales – 100 blocks.

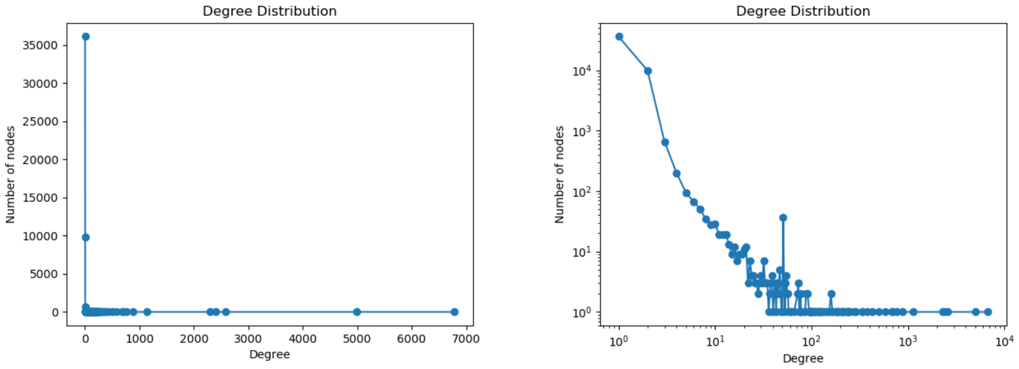

Figure 8: Degree distribution in linear and log scales – 1000 blocks.

TABLE 2: Small worlds: Ethereum networks vs. related random graphs. (Results on main components.)

# Blocks	**# nodes**	**# edges**	**cc**	**L**	**cc RG**	**L RG**	σ
1	19	18	0	1.89	0.05	2.94	0
10	139	139	0	2.26	0.01	4.94	0
10^2	4262	4641	0	7.45	0	8.36	2.07
10^3	37201	43682	0.003	5.66	3×10^{-5}	10.52	225.85
2×10^3	67911	81580	0.005	5.24	1×10^{-5}	11.13	602.16
3×10^3	88184	106842	0.01	5.28	1.37×10^{-5}	11.39	1520.28
4×10^3	108871	108871	0.01	5.29	1.12×10^{-5}	11.60	1770.93
5×10^3	133982	164406	0.01	5.35	9.15×10^{-6}	11.80	2272.96
10^4	239114	303248	0.01	5.28	5.30×10^{-6}	12.38	5891.43

TABLE 3: Networks diameter.

# Blocks	**# nodes (main comp)**	**diameter**
1	19	2
10	139	4
10^2	4262	23
10^3	37201	27
2×10^3	67911	22
3×10^3	88184	24
4×10^3	108871	34
5×10^3	133982	58
10^4	239114	90

slice, a notable amount of interactions occurred. This might be explained by the fact that the considered blocks, in the snapshot identified as 5000000, have been generated in date Jan-30-2018. Indeed, if we look at the exchange rate Ether - US Dollar, shown in Figure 12, in that time period we will notice a spike in the Eth value. It is reasonable to assume that the higher the value of the crypto-currency linked to the blockchain, the higher the activity in the blockchain.

Figure 13 shows results similar to those of Figure 11, but when the blocks intervals' was of 10000 blocks, i.e. approximately 40 hours. Values are different, but the trend is analogous to results related to 1000 blocks' slices.

It is worth mentioning that we measured other metrics as well, such as the average clustering coefficient. In this case, we did not noticed a specific trend. All networks have a very

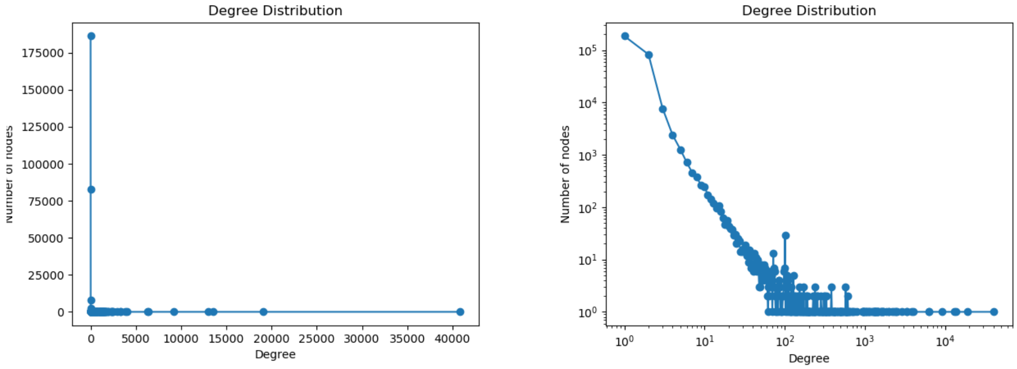

Figure 9: Degree distribution in linear and log scales – 10000 blocks.

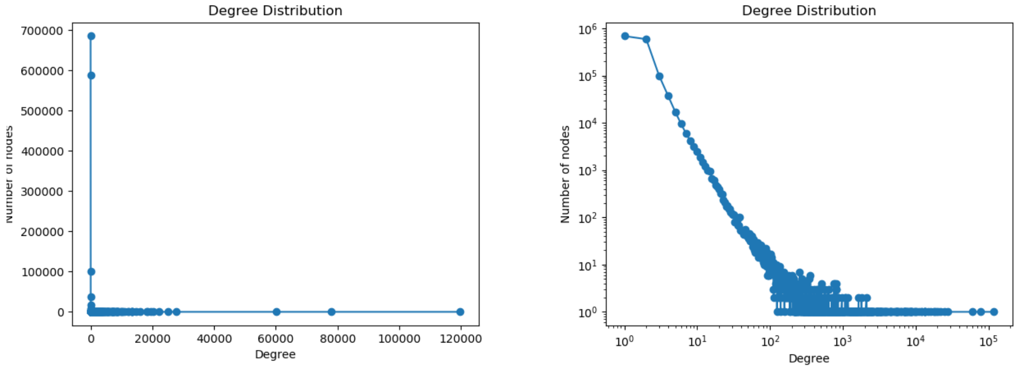

Figure 10: Degree distribution in linear and log scales – 100000 blocks.

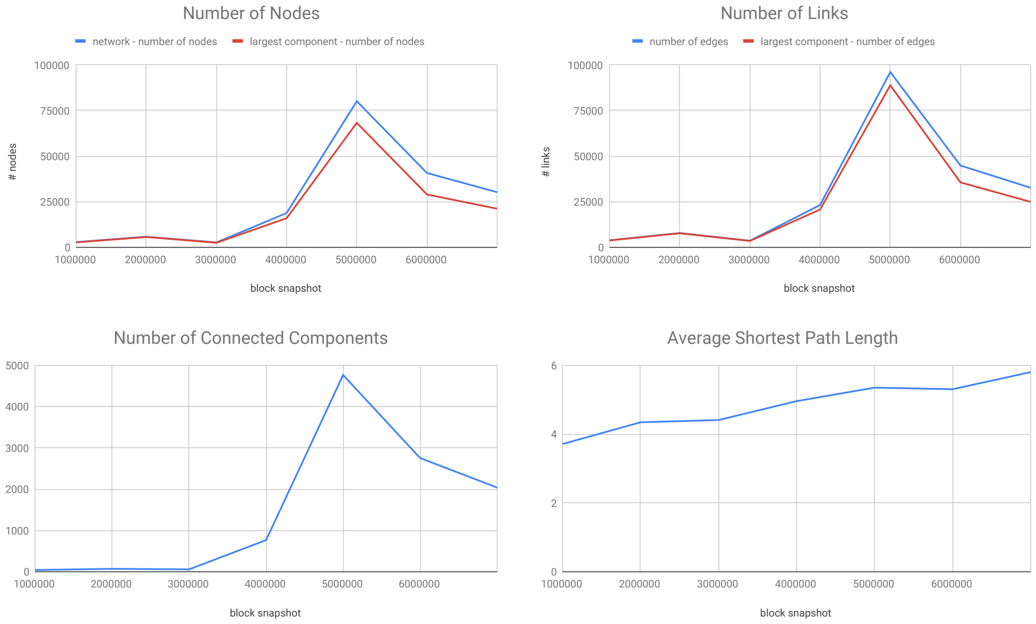

Figure 11: Network statistics at different block snapshots – snapshot size: 1000 blocks.

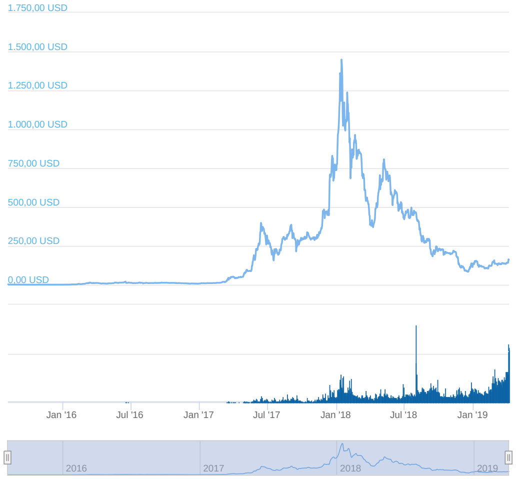

Figure 12: Eth-USD exchange value – source: https://www.coingecko.com.

low average clustering coefficient, similar to the values show in previous subsections.

4.4. Miners Distribution

The analysis of the blockchain allows retrieving diverse typologies of information, related to the generation of blocks. Besides the number of accounts that interact in the blockchain, another interesting metrics is concerned to miners that generate blocks. In particular, it might be interesting to understand if the distribution of miners is truly spread out, or rather if a small niche of miners have the control of the blockchain.

Figure 14 shows the distribution of nodes that mined a certain number of blocks. On the x-axis there is the amount of blocks that have been mined by the same miner. On the y-axis, we have the amount of nodes that mined that number of blocks. In this case, we considered a subset of ~ 180000 blocks, starting backwards from the last available block in the blockchain at the time of the conducted experiments, i.e. the most recent block was the same of results reported at on the first part of this section.

It is possible to observe that, while as expected the majority of miners were able to mine just one block in the considered time interval, there are however certain nodes that have mined a huge number of blocks. In particular, it seems that six nodes mined over 10000 blocks; one node mined 47193 blocks. These nodes are probably mining pools, i.e., a set of nodes who share their processing power, with the aim to split the reward equally, according to the amount of work they contributed to the probability of finding a block. This obtained result is in line with the statistics offered by blockchain explorer web sites, e.g., https://www.etherchain.org/.

5. Conclusions

In this paper, we provided an analysis of the complex network representing the Ethereum blockchain transactions. We varied the amount of blocks considered to build the network. This corresponds to a different temporal range, thus to a different amount of transactions and a different network size. As expected, the wider the network the more likely the presence of hubs in the network, meaning that there are some nodes that are more active in the blockchain. We also considered different temporal intervals, by taking subsets of subsequent blocks back in the chain. This allows to understand how the use of the blockchain changes in time.

Even if Ethereum accounts are anonymous, and hence it is not possible to directly map an external account to a given user, we might assume that hub nodes correspond to well known accounts, that might represent popular smart contracts or external accounts that allow exchanging Ether. Examples are those services that require a real identity to transact, such as online wallet services, currency exchange services, merchants. In public blockchains, such as Ethereum, account anonymity is obtained by employing pseudonymity to represent an account, together with the unlinkability among different interactions of the same user with the system. If an account is a hub node in the network, and it is not a popular smart contract, this means that the related real world entity often employs the same account to transact. In this case, it might be easier to de-anonymize such an account [22].

References

[1] A next-generation smart contract and decentralized application platform. White Paper, 2018. https://github.com/ethereum/wiki/wiki/White-Paper, Accessed on 2018-03-02.

[2] Z. Alhadhrami, S. Alghfeli, M. Alghfeli, J. A. Abedlla, and K. Shuaib. Introducing blockchains for healthcare. In *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pages 1–4, Nov 2017.

[3] A. Anoaica and H. Levard. Quantitative description of internal activity on the ethereum public blockchain. In *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5, Feb 2018.

[4] A. M. Antonopoulos. *Mastering Bitcoin: Unlocking Digital Crypto-Currencies*. O'Reilly Media, Inc., 1st edition, 2014.

[5] Y. N. Aung and T. Tantidham. Review of ethereum: Smart home case study. In *2017 2nd International Conference on Information Technology (INCIT)*, pages 1–4, Nov 2017.

[6] A. Baumann, B. Fabian, and M. Lischke. Exploring the bitcoin network. In *WEBIST (1)*, pages 369–374. SciTePress, 2014.

[7] W. Chan and A. Olmsted. Ethereum transaction graph analysis. In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 498–500, Dec 2017.

[8] T. Chen, Y. Zhu, Z. Li, J. Chen, X. Li, X. Luo, X. Lin, and X. Zhange. Understanding ethereum via graph analysis. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 1484–1492, April 2018.

[9] G. D'Angelo and S. Ferretti. Highly intensive data dissemination in complex networks. *Journal of Parallel and Distributed Computing*, 99:28 – 50, 2017.

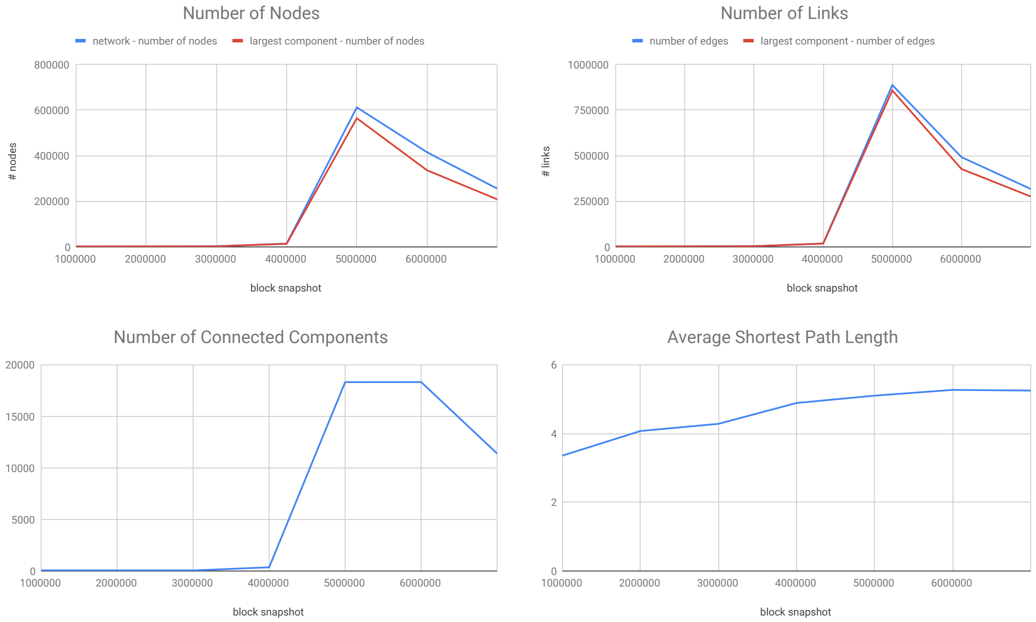

Figure 13: Network statistics at different block snapshots – snapshot size: 10000 blocks.

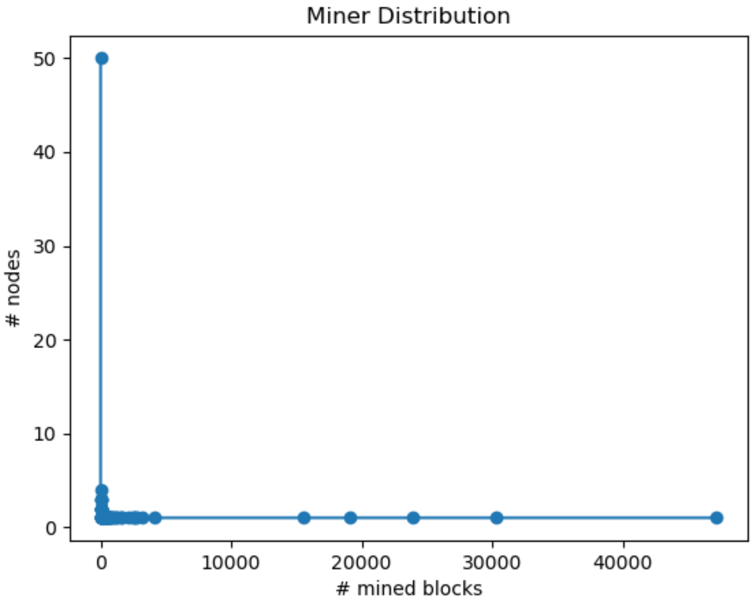

Figure 14: Distribution of of nodes that mined a certain number of blocks.

[10] G. D'Angelo, S. Ferretti, and V. Ghini. Simulation of the internet of things. In *2016 International Conference on High Performance Computing Simulation (HPCS)*, pages 1–8, July 2016.

[11] G. D'Angelo, S. Ferretti, and M. Marzolla. A blockchain-based flight data recorder for cloud accountability. In *Proc. of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, Cry-Block'18, pages 93–98, New York, NY, USA, 2018. ACM.

[12] R. De Michele, S. Ferretti, and M. Furini. On helping broadcasters to promote tv-shows through hashtags. *Multimedia Tools and Applications*, Aug 2018.

[13] A. Durand, P. Gremaud, and J. Pasquier. Resilient, crowd-sourced LPWAN infrastructure using blockchain. In *CRYBLOCK@MobiSys*, pages 25–29. ACM, 2018.

[14] S. Ferretti. Gossiping for resource discovering: An analysis based on complex network theory. *Future Generation Computer Systems*, 29(6):1631 – 1644, 2013.

[15] Z. Gao, L. Xu, G. Turner, B. Patel, N. Diallo, L. Chen, and W. Shi. Blockchain-based identity management with mobile device. In *CRYBLOCK@MobiSys*, pages 66–70. ACM, 2018.

[16] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using networkx. In G. Varoquaux, T. Vaught, and J. Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[17] J. Herrera-Joancomarti. Research and challenges on bitcoin anonymity. In *Proc. of the 9th International Workshop on Data Privacy Management*, volume 8872, 09 2014.

[18] U. Javaid, A. K. Siang, M. N. Aman, and B. Sikdar. Mitigating lot device based ddos attacks using blockchain. In *CRYBLOCK@MobiSys*, pages 71–76. ACM, 2018.

[19] M. A. Javarone and C. S. Wright. From bitcoin to bitcoin cash: a network analysis. In *CRYBLOCK@MobiSys*, pages 77–81. ACM, 2018.

[20] H.-E. Kim, T.-T. Kuo, and L. Ohno-Machado. Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association*, 24(6):1211–1220, 09 2017.

[21] O. López-Pintado, M. Dumas, L. García-Bañuelos, and I. Weber. Dynamic role binding in blockchain-based collaborative business processes. *CoRR*, abs/1812.02909, 2018.

[22] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage. A fistful of bitcoins: Characterizing payments among men with no names. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, pages 127–140, New York, NY, USA, 2013. ACM.

[23] M. Mettler. Blockchain technology in healthcare: The revolution starts here. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–3, Sept 2016.

[24] R. Neisse, G. Steri, and I. Nai-Fovino. A blockchain-based approach for data accountability and provenance tracking. In *Proc. 12th Int. Conf. on Availability, Reliability and Security*, ARES '17, pages 14:1–14:10. ACM, 2017.

[25] E. D. Pascale, J. McMenamy, I. Macaluso, and L. Doyle. Smart contract slas for dense small-cell-as-a-service. *CoRR*, abs/1703.04502, 2017.

[26] F. Reid and M. Harrigan. An analysis of anonymity in the bitcoin system. In *PASSAT/SocialCom 2011, Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), Boston, MA, USA, 9-11 Oct., 2011*, pages 1318–1326, 2011.

[27] S. Rouhani and R. Deters. Performance analysis of ethereum transactions in private blockchain. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 70–74, Nov 2017.

[28] A. Schönhals, T. Hepp, and B. Gipp. Design thinking using the blockchain: Enable traceability of intellectual property in problem-solving processes for open innovation. In *CRYBLOCK@MobiSys*, pages 105–110. ACM, 2018.

[29] M. Selimi, A. R. Kabbinale, A. Ali, L. Navarro, and A. Sathiaseelan. Towards blockchain-enabled wireless mesh networks. In *CRYBLOCK@MobiSys*, pages 13–18. ACM, 2018.

[30] H. Shafagh, L. Burkhalter, A. Hithnawi, and S. Duquennoy. Towards blockchain-based auditable storage and sharing of iot data. In *Proc. 2017 Cloud Computing Security Workshop*, CCSW '17, pages 45–50. ACM, 2017.

[31] S. Somin, G. Gordon, and Y. Altshuler. Network analysis of erc20 tokens trading on ethereum blockchain. In A. J. Morales, C. Gershenson, D. Braha, A. A. Minai, and Y. Bar-Yam, editors, *Unifying Themes in Complex Systems IX*, pages 439–450, Cham, 2018. Springer International Publishing.

[32] E. Yavuz, A. K. Koc, U. C. Cabuk, and G. Dalkilic. Towards secure e-voting using ethereum blockchain. In *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, pages 1–7, March 2018.

[33] M. Zichichi, M. Contu, S. Ferretti, and G. D'Angelo. Likestarter: a smart-contract based social dao for crowdfunding. In *Proc. of the 2nd Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, CryBlock'19. IEEE, 2019.