Measuring Ethereum-Based ERC20 Token Networks

Friedhelm Victor(✉) and Bianca Katharina Lüders

Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
{friedhelm.victor,bianca.lueders}@tu-berlin.de

Abstract. The blockchain and cryptocurrency space has experienced tremendous growth in the past few years. Covered by popular media, the phenomenon of startups launching Initial Coin Offerings (ICOs) to raise funds led to hundreds of virtual tokens being distributed and traded on blockchains and exchanges. The trade of tokens among participants of the network yields *token networks*, whose structure provides valuable insights into the current state and usage of blockchain-based decentralized trading systems. In this paper, we present a descriptive measurement study to quantitatively characterize those networks. Based on the first 6.3 million blocks of the Ethereum blockchain, we provide an overview on more than 64,000 ERC20 token networks and analyze the top 1,000 from a graph perspective. Our results show that even though the entire network of token transfers has been claimed to follow a power-law in its degree distribution, many individual token networks do not: they are frequently dominated by a single hub and spoke pattern. Furthermore, we generally observe very small clustering coefficients and mostly disassortative networks. When considering initial token recipients and path distances to exchanges, we see that a large part of the activity is directed towards these central instances, but many owners never transfer their tokens at all. In conclusion, we believe that our findings about the structure of token distributions on the Ethereum platform may benefit the design of future decentralized asset trade systems and can support and influence regulatory measures.

Keywords: Blockchain · Ethereum · Tokens · Network analysis

1 Introduction

In the past years, blockchains and in particular ICOs have seen increased attention, with startups frequently selling *tokens* to obtain seed funding. Such tokens may represent both digital and physical assets or utilities as entries on the distributed ledger, similar to native digital currencies such as Bitcoin or Ether. They are commonly enabled by ERC20-compliant smart contracts implemented on the Ethereum blockchain. To date, their sale and trade are unregulated in most countries. A lot of research was already dedicated to the analysis of content and communication graphs on different blockchains. In contrast to these,

I. Goldberg and T. Moore (Eds.): FC 2019, LNCS 11598, pp. 113–129, 2019.
https://doi.org/10.1007/978-3-030-32101-7_8

which focused on the trade of native currencies, we investigate the trade of tokens. We define the network between addresses that reflects the distribution and trade of each token as its *token network*, in which each edge represents the transfer of a specified amount of the respective token between two addresses. To the best of our knowledge, no large-scale study of individual token networks on the Ethereum blockchain has been provided to date. We advance approaches developed in the area of network analysis to this new domain and analyze token networks quantitatively from a graph perspective to capture their structure and topology. This allows us to obtain a sound overview of the token landscape. An in-depth understanding of graph structures and usage patterns in the decentralized and unsupervised domain of cryptocurrencies and tokens is necessary to evaluate current token trading systems and serves as a basis for further research.

The remainder of this paper is structured as follows: In Sects. 2 and 3, we provide an overview of the theoretical background, current research results and related work on cryptocurrencies, blockchains and smart contracts. In Sect. 4, we describe our data collection methodology and provide a set of high-level statistics of our data set, followed by an analysis of the token networks based on graph theoretic measures in Sect. 5. Finally, we summarize our paper and provide approaches for future work in Sect. 6.

2 Background

In recent years, the popularity of blockchain-based cryptocurrencies has grown significantly. As of 2018, hundreds of different coins are in circulation, with a large portion of them developed on top of the Ethereum blockchain in the form of tokens, that have recently been the basis for many crowdfunded ventures. A new type of token can be created by implementing a smart contract. While their implementation often follows a standard, their behavior can be implemented arbitrarily. With regulation currently still under development, questions have been raised whether a certain token constitutes a security or a utility, and how they should be treated.

2.1 Ethereum, the EVM and Smart Contracts

Similar to Bitcoin, Ethereum is an open-source, public, distributed, blockchain-based platform with a Proof of Work-based consensus algorithm coupled with rewards, which absolves the need for trusted intermediaries [6]. If popularity were measured by market capitalization, it would be the second most popular blockchain as of September 2018. Ethereum's most significant feature is the Ethereum Virtual Machine (EVM) - a stack-based runtime environment that can execute programs known as *smart contracts*. They can be developed in high-level languages such as Solidity and deployed on the blockchain as bytecode by any participant of the network. The immutable code is reachable through the address of the smart contract account and stored on the ledger, along with all historic state changes. By sending transactions from externally owned accounts (EOA), users can interact with smart contracts and call their functions [27].

2.2 Tokens

The resilience of smart contracts to tampering makes them appealing for many application scenarios - *financial*, *notary*, *game*, *wallet*, and *library* contracts were identified by Bartoletti and Pompianu [4]. The authors further analyzed smart contract design patterns and showed that many of the contracts in the financial category use the token pattern for the representation of fungible assets. In contrast to the native *coins* that typically represent a digital currency, tokens may represent a variety of transferable and countable goods such as digital and physical assets, shares, votes, memberships, or loyalty points. Any third party can create smart contracts and develop, define and distribute their own named asset. A frequent approach to distribute tokens and raise funds is an *initial coin offering* (ICO). The term leans on *Initial Public Offering*, the stock market launch in the traditional economy. Another distribution mode, the so-called *Airdrop*, is designed to distribute tokens without requiring prior investment. Once they have value, the founders can sell additional tokens.

2.3 The ERC20 Token Standard

To establish a common interface for fungible tokens, the ERC20[1] standard was proposed in late 2015. To be compatible, a smart contract needs to implement a set of functions, of which only the signatures, but not the implementations are specified. Within a smart contract's bytecode, these signatures can be identified by their entrypoints, marked by the first 4-bytes of the Keccak hashes of the high level function signature (Table 1). Thus, ERC20-compatible contracts can be identified by means of the corresponding entrypoint hashes in the deployed contract bytecodes.

Table 1. ERC20 signatures and hashes

Classification			Signature	First 4-byte Keccak hash
ERC20	Required	Method	totalSupply()	`18160ddd`
			balanceOf(address)	`70a08231`
			transfer(address,uint256)	`a9059cbb`
			transferFrom(address,address,uint256)	`23b872dd`
			approve(address,uint256)	`095ea7b3`
			allowance(address,address)	`dd62ed3e`
		Event	Transfer(address,address,uint256)	`ddf252ad`
			Approval(address,address,uint256)	`8c5be1e5`
	Optional	Method	name()	`06fdde03`
			symbol()	`95d89b41`
			decimals()	`313ce567`

[1] https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20-token-standard.md.

Pos.	Sender	Receiver	Function call
1	A	X	transfer(B, 3)
2	A	X	transfer(C, 7)
3	B	X	transfer(C, 2)
4	C	X	transfer(D, 4)

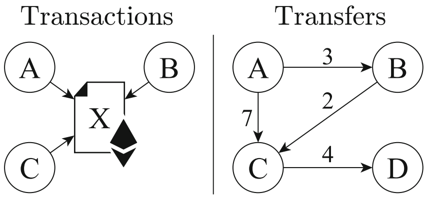

Fig. 1. Transactions to a token contract and corresponding graph perspectives

To send tokens from address A to address B, the owner of address A sends a transaction to a token contract X, calling its `transfer` function. If successful, the balance of both addresses will be updated within the contract, constituting a state change. As balances may also be affected by other functions included by the smart contract developer, the ERC20 standard recommends to emit a `Transfer` event whenever a token transfer has occurred. Figure 1 illustrates the relationship between transactions that call functions and the emitted transfers.

These token transfers yield a graph, in which the nodes are addresses connected by transfers. This graph may also contain addresses that never interacted with the token contract: during deployment or with a specific function, a common way to associate tokens with particular addresses are *initial balances*, which the contract creator allocates to certain addresses upon creating the contract. Some developers have chosen to emit these allocations as transfer events, where the source address is for example set to `0x0`, but mostly, these balance allocations are not emitted as transfer events. Later standard proposals such as ERC621 additionally introduced `Mint` and `Burn` events to increase or decrease balances without requiring a transfer at all. These events change the total supply of the respective token. Although not used widely yet, shortcomings of the ERC20 design that are beyond the scope of this paper are tackled with proposals ERC223, ERC667 and ERC777.

3 Related Work

Several seminal works and studies in the area of graph theory and network analysis, also including the analysis of social networks, as well as in digital currencies and Ethereum smart contracts form the basis of our research.

3.1 Cryptocurrencies and Smart Contracts

The literature on Bitcoin and other cryptocurrencies covers not only the underlying distributed ledger and consensus technologies and protocols, but also the publicly available transactional data, which provides a unique opportunity to analyze real, large-scale financial networks. Various aspects of Bitcoin have been discussed, such as by Barber et al. [3], who investigated the design, success factors, history, strengths and weaknesses. Tschorsch and Scheuermann [24] provided a comprehensive survey on the technical aspects of decentralized digital currencies. Furthermore, Akcora et al. [1] presented a holistic view on distributed ledgers with a focus on graph theoretical aspects.

Going beyond Bitcoin, Bonneau et al. [6] provided the first systematic exposition of the second generation of cryptocurrencies and analyzed *altcoins* that have been implemented as alternate protocols. Similarly, Anderson et al. [2] explored three representative blockchains - Ethereum, Namecoin and Peercoin - which have extended Bitcoin's original mechanism and focused on the features that distinguish them from the pure currency use case. Research on smart contracts frequently focuses on design patterns, applications and security issues [23]. In the first methodic survey and quantitative investigation on their usage and programming, Bartoletti and Pompianu [4] proposed a taxonomy of smart contract application domain categories and identified common programming and design patterns. With a focus on security, Nikolic et al. [20] presented a novel characterization of trace vulnerabilities, which allow to identify contracts as *greedy*, *prodigal* or *suicidal*.

3.2 Blockchain Graph Analysis

Each blockchain can be analyzed from a graph-centric perspective on two layers: as *communication graphs*, which reflect the underlying peer-to-peer communication on the network layer, and as *content graphs*, which reflect transfers of assets on the application layer [1]. For example, Miller et al. [17] investigated the public topology of the Bitcoin peer-to-peer network in a quantitative measurement and analyzed how nodes participate and collaborate in mining pools.

In the analysis of content graphs, techniques from the area of social network analysis are commonly used. For a general overview of social network analysis, the reader is referred to fundamental works such as Newman [19]. Mislove et al. [18] proposed a detailed comparison of the characteristics of multiple online social network graphs at large scale and confirmed the power-law, small-world and scale-free properties of these social networks. A commonly used methodology to detect and validate power laws was presented by Clauset et al. [8].

Content graphs can be modelled on different levels. First, in *transaction graphs*, the nodes represent transactions that happen on the distributed ledger and the edges represent the flow of transferred assets. These graphs start from the genesis block, each transaction can have incoming edges only and a DAG (*directed acyclic graph*) emerges [1,12,21]. In *address graphs*, the nodes denote addresses, and each edge represents a particular transaction between two of them. Address graphs provide a useful abstraction for exploring and tracing flows through the system and identifying recurrent patterns in transactions [12]. A recent approach to investigate the whole address graph spanned up by the trade of all ERC20-compliant tokens on the Ethereum blockchain was presented by Somin, Gordon and Altshuler [23]. The authors consider all trading wallets as the nodes of the network, construct the edges based on buy-sell trades and demonstrate that the degree distribution of the resulting network displays strong power-law properties. Finally, *user* or *entity graphs* reflect the flow of value between real-world entities. In these graphs, each node represents a user or an entity, and each edge represents a transaction between source and target entity [1]. Building these graphs requires to identify and associate public addresses that possibly belong to the same real-world entity. Many approaches to cluster addresses on the Bitcoin blockchain have been presented to date, along with discussions of connected anonymity issues [10,16,21,22]. To the best of our knowledge, no such heuristics exist for Ethereum's account model yet.

Many of the approaches to analyze content graphs arising through the usage of cryptocurrencies rely on methods and assumptions known from the area of social network analysis. Yet, the network generation mechanisms are different. Since the Ethereum network combines aspects of social and financial transaction networks, we also consider analysis approaches that focus on the latter. In this area, Inaoka et al. [13] investigated the network structure of financial transactions on the basis of the logged data of the BOJ-Net. Similarly, Kyriakopoulos et al. [15] analyzed the network of financial transactions of major financial players within Austria and reported the characteristic network parameters. Some of their many empirical findings include the dependency of the network topology on the time scales of observation and the existence of power laws in the cumulative degree distributions.

3.3 Contribution

In summary, these different graph-theoretical approaches provide an intuition for the flow and spread of assets on different blockchains. While previous analyses took either the entire blockchain or the whole network of token trades into consideration, we center our attention on a new type of address graphs: *token networks*, which we define as the network of addresses (nodes) that have owned a specific type of token at any point in time, connected by the transfers of the respective token. Since the tokens are not comparable, neither in their value, which heavily fluctuates over time, nor in their respective total supply, which may further be influenced by `Mint` and `Burn` events, we omit the weight of the transfers, such that we obtain a directed, unweighted graph. Further, due to a lack of approaches for address clustering on Ethereum, we define nodes as addresses in these token networks and assume that they represent different entities, which may be either a user, an exchange, a miner, or another smart contract. A new token network emerges for each newly published ERC20-compliant token contract. Each address may be part of several token networks, and each analyzed token network is essentially an overlay graph of the entire network of Ethereum addresses. To the best of our knowledge, these individual token networks have not been studied yet, and we hope that our measurement and evaluation inspires further research in this area.

4 Data

In this chapter, we describe how we identified ERC20-compatible smart contracts, how we extracted and filtered the transfer events, and provide an overview of the token network landscape in the form of summary statistics.

4.1 Data Collection

The basis for generating token networks are token transfers emitted by ERC20-compatible smart contracts. We used the Parity client[2] for the set-up of a fully synchronized Ethereum node and extracted all transactions, contract addresses and the corresponding smart contract bytecodes from the first 6,300,000 blocks, covering the period from July 30th, 2015 until September 9th, 2018. We identified 7,323,377 smart contract creations, including those that were created by other contracts, of which 75,514 fulfill the criteria introduced in Sect. 2.3 and are thus labeled as ERC20-compatible.

Next, we retrieved the token transfer events emitted by those ERC20-compatible smart contracts. These events can be identified by the corresponding event type and contain information about the source, the target and the amount of tokens that were passed in the respective transfer. In total, we extracted 97,671,089 transfer events. It is noteworthy that the transfer events are only related to 46,970 of the ERC20-compatible smart contracts (62.2%), such that 28,544 of the token contracts have never emitted any transfer events. This does not necessarily imply that the tokens have never been traded on the Ethereum blockchain, but there are no events that document their transfers.

Since it is up to the developer of the smart contract to decide when a transfer event is emitted, not all actual token transfers are logged as such. To account for initial balance allocations, which are only rarely emitted as transfer events (Sect. 2), we added the initial balances as synthetic transfers to our dataset, where the source address is the artificial address `0x0` and the target address corresponds to the address mentioned in the contract bytecode. We could identify the allocation of initial balances in 52,554 ERC20-compatible smart contracts, where each smart contract that uses this method distributes the assets to 2.96 entities on average (median 1). These numbers are comprehensible, since the smart contract developer has an interest so assign a certain amount of tokens to himself and/or his team, which is usually a rather small set of users. These initial balances add 142,673 new token transfers to our dataset, such that we capture a total of 97,813,762 token transfers related to 64,393 ERC20 token contracts.

Figure 2 compares the amount of transactions that were initiated by externally owned accounts (EOAs) to ERC20-compatible smart contracts (white) with the corresponding amount of transactions to all other, non-ERC20-compatible contracts (gray) and the resulting token transfers (black dots) emitted by the ERC20-compatible contracts. All three numbers exhibit a significant increase starting in the beginning of 2017, and the growth indicates an increasing popularity of ERC20-compatible token contracts in terms of contract interactions initiated by EOAs. Since a single interaction with a token contract may yield multiple transfer events, we observe in total more transfer events than ERC20 contract transactions.

[2] https://www.parity.io/.

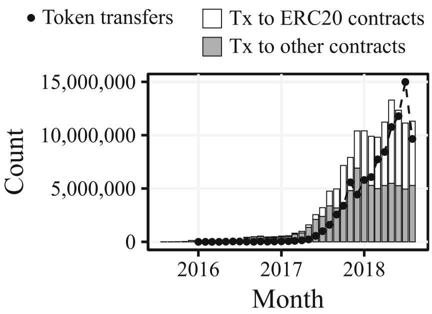

Fig. 2. From March 2018, more than half of transactions are to ERC20 token contracts. One transaction can lead to multiple transfer events – common in Airdops.

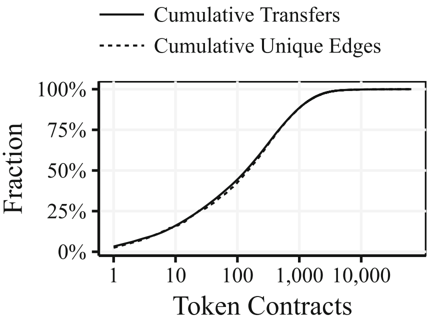

Fig. 3. CDF of transfers and unique edges related to token contracts. Almost 90% of all transfers/edges belong to the top 1,000 token contracts.

4.2 Summary Statistics

The entire set of 64,393 ERC20 token networks captures 19,45 million unique addresses, which corresponds to nearly $45,9\%$ of all addresses on the Ethereum blockchain as of September 9, 2018.

The smallest 599 networks consist of only one node, which may have up to 5 self-edges associated to itself. In general, the size distribution is skewed towards smaller values - while the median is 3 (mean 890), some of the networks capture up to 1.52 million nodes (*Tronix* and *An Etheal Promo*). Other popular token networks which stand out due to their size are *VIU*, *Bitcoin EOS*, and the *Basic Attention Token*. On the other hand, 80.38% of the token networks consist of only 10 nodes or less.

In terms of *edges* per token network, we differentiate between simple, multiple and self-edges, such that $n_{simple} + n_{multi} + n_{self} = n_{edges}$. Besides the networks that consist of only one node, there are five others that have only self-edges – the largest of them (*Explore Coin*) has 46 nodes which are only connected to themselves. Similar to the size in terms of the amount of nodes, 80.33% of the networks have 10 unique edges or less. On average, the networks contain 1519 edges (median 2), the largest network has 3,17 million edges in total (*EOS Token Contract*). The five networks with the largest amount of edges further include *Tronix*, *OMG Token*, *An Etheal Promo*, *BeautyChain*, and these are also the networks with most unique edges, i.e. those with the most connections between different addresses. Still, even in these networks, each node has on average two adjacent edges, which might correspond to obtaining tokens from the contract and then transferring them to an exchange.

In total, the ratio of total edges per node varies from 0.5 to 2631, where the highest ratio is in the *Ether Token*. Removing self-edges and multiple edges, this value drops to at most 8.69 (*Consumer Activity Token*), with a mean of 0.7315, such that we have relatively sparse networks.

5 Analysis

In this section, we present our data selection steps and analyze the structure of the token networks with respect to distributions of degree, density, components, clustering coefficients and assortativity. We then focus on how tokens are received and transferred, with an emphasis on the role of exchanges – providing insight into the activity within token trading networks.

5.1 Data Selection

As discussed in Sect. 4, the majority of token networks consists of only a few transfers and nodes. To remove those from the analysis, since they may not exhibit a concise graph structure comparable to the larger networks and might bias the results, we assess the amount of transfers per token network. Figure 3 shows the cumulative distribution function (CDF) of all transfers, respectively connections between nodes, that the networks add to the total amount. We observe that the top 1000 token networks capture more than 85% of both measures, such that we limit our analysis to these, which account for 86,54 million transfers in total (88.48% of the original amount).

5.2 Degree Distributions

A fundamental property of nodes in a directed graph are their *in-* and *out-*degrees. The frequency distribution of degrees, where p_k is the fraction of nodes with degree k, can provide an insight into the network's structure. Many real-world networks exhibit highly right-skewed degree distributions with a *heavy tail*, which indicates that a significant portion of observations is in the tail and demonstrates the existence of high-degree hubs. Several real-world networks have been confirmed to follow *power laws* in their degree distribution [7,14,18,19]. Power laws are distributions of the form $p_k = Ck^{-\alpha}$, in which the dependent variable, the probability that a node has degree k, varies inversely as a power of the independent variable, the degree k. p_k decreases monotonically [18,19] and decays significantly slower than exponential decays in normal distributions. While the non-negative constant C is fixed by normalization, the parameter α is called the *coefficient* of the power law [9,19] and typically is in the range $2 \leq \alpha \leq 3$.

Using the poweRlaw package in R [11], we estimate parameters for each token network, using maximum likelihood estimation and the *Kolmogorov-Smirnov* statistic to quantify the distance between the observed degree distribution and the estimated power law. We perform goodness-of-fit tests via a bootstrapping to obtain a p value, following the approach of Clauset et al. [8].

Whereas Somin et al. [23] have shown that the full transfer graph consisting of all token networks combined appears to follow a power-law in both in- and outdegree, Figs. 4 and 5 illustrate a different result for the individual token networks. While we can fit a power law model to all of the networks, most of the

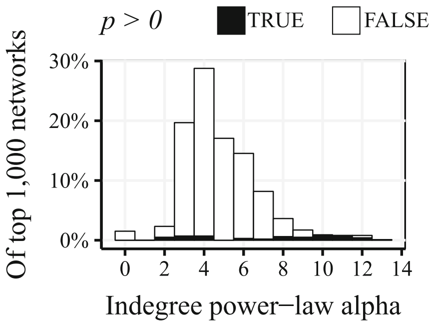

Fig. 4. Estimated indegree power-law coefficients. In most cases, the hypothesis can be rejected.

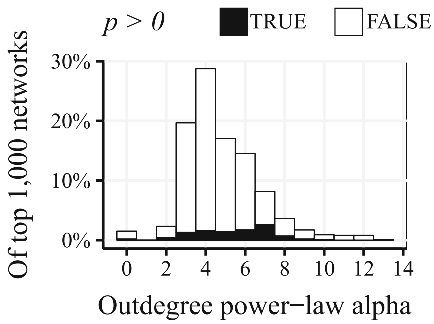

Fig. 5. Estimated outdegree power-law coefficients. The hypothesis can be rejected in fewer cases than for the indegree.

p values obtained via bootstrapping have returned a value of 0, indicating that they likely do not follow a power law.

For those where we cannot reject the power law hypothesis, we suppose that if a network contains multiple exchanges, multiple high indegree addresses are likely to be present. These same addresses frequently also have a high outdegree. Considering that many networks additionally contain a large initial distribution in the form of a star-shaped subgraph, one further address with high outdegree is likely to exist. This may explain why about 10% of token networks appear to follow a power law in their outdegree distribution.

Generally, the fitted power law exponents are very high, indicating quickly decaying degree distributions. This is contraindicative to the power law hypothesis, but adequate for the use case: in token networks, as opposed to social networks, the amount of hubs, i.e. exchanges, is limited institutionally. While social networks allow for an organic growth of "popular" nodes, only a limited number of exchanges are known for securely handling token trades. This reflects an issue of trust - while any user can open an exchange on the basis of pre-defined protocols, most users only trust and trade their tokens on a few well-known ones.

Another aspect that might differentiate the full transfer graph from the individual token networks in terms of power laws in their degree distributions, is that initial token distributions, especially airdrops, frequently choose existing, active addresses. This process, which follows the logic of preferential attachment, leads to these nodes becoming connectors between individual token networks, which adds smaller hubs to the full transfer graph.

5.3 Density and Components

The density d represents the fraction of existent to theoretically possible edges in a network. In general, we observe that a larger number of nodes in a token network leads to a lower graph density (Fig. 6). The network with the highest density, also related to the number of nodes, is the *NPXS Smart Token Relay.*

Further, we investigate the number of weakly connected *components*, disconnected portions of the token networks. Given common token distribution modes (Sect. 2.2), we expect the tokens to be distributed starting from the addresses with the initial balances, and assume that each network consists of a single, large, weakly connected component. We observe that this holds for 75% of the observed token networks when we take the initial balances into account (Sect. 4). Several components indicate that the tokens were not distributed in an ICO, but based on another logic, such as minting, which credits an arbitrary amount of tokens to a specified address and typically emits a `Mint` event (Sect. 2.3). We also find that 29 token networks have more than 100 components, and three consist of more than 3,000 components (*blockwell.ai KYC Casper Token*, *SanDianZhong* and *VGAMES*). This might indicate that many of the nodes in the network received their tokens in a non-standard process, yielding an anomalous graph structure.

5.4 Clustering Coefficients

To measure the clustering coefficient, which indicates the strength of local community structure, two measures are common: (1) the *global* clustering coefficient C_g, which measures the fraction of paths of length two in the network that are closed, and (2) the average of the *local* clustering coefficients, C_{avg_l}, which define for each individual node the share of possible connections among the node's neighbors that actually exist in the network. In either case, the clustering coefficient indicates how much more likely it is to connect to a neighbor's neighbor than to a randomly chosen node [19], and a large clustering coefficient is regarded as an indicator for small-world networks. Values of $C_g = 0.20$ (film actor collaborations), $C_g = 0.09$ (biologist collaborations) and $C_g = 0.16$ (university email communication) are high compared to estimates based on random connections, but typical values for social networks [19]. Similarly, Baumann et al. [5] found that the average local clustering coefficient (C_{avg_l}) in the Bitcoin address graph is fluctuating around 0.1 and thus rather high over time, also indicating a small world network.

For the token networks, we need to take into account that there is, as discussed in Sect. 4.2, a large fraction of nodes with degree one, for which the local clustering coefficient should be set to $C_i = 0$ [19]. If there is a significant number of such nodes, C_{avg_l} would be dominated by these minimum-degree nodes, yielding a poor picture of the overall network properties. Additionally, vertices with a low degree of which 2 or 3 neighbors are connected raises C_{avg_l} disproportionately high. Thus, we rely on C_g, which measures the global cliquishness of the network and provides evidence for a small-world network [26]. For the entire network of token transfers, we observe $C_g = 0.00001062$ and $C_{avg_l} = 0.3042$, which is higher than the known measure for the entire network of Bitcoin addresses [5].

This might indicate that the network of token trades has a higher tendency to form communities, maybe based on users who recommend or send tokens to each other. Similarly, airdrops tend to focus on existing active users, which could further lead to the forming of communities.

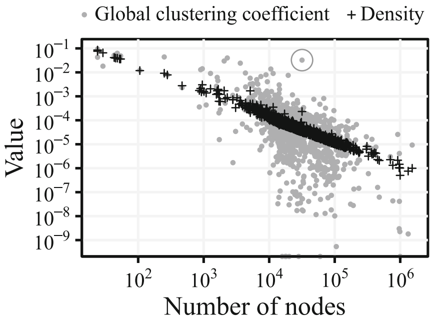

Fig. 6. Distribution of density values and global clustering coefficients vs. network size.

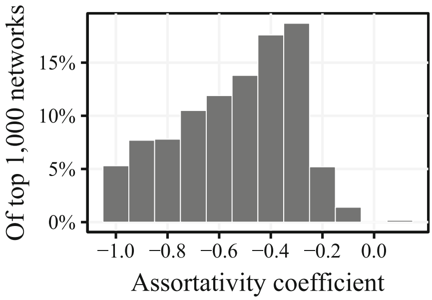

Fig. 7. Distribution of degree assortativity coefficients. All but one token network are disassortative.

For the individual token networks, we observe a mean global clustering coefficient of 0.0008831 and a maximum value of 0.0941 for the *NPXS Smart Token Relay*. This network is rather small, with only 24 nodes. Figure 6 illustrates C_g related to the size of the network, showing a general decrease, and exhibits another outstanding token: *TEST POGO 1* (circled), which has the highest C_g (0.0324) relative to its size. Further, we observe that 707 of the networks in the sample have a higher C_g than the network that connects them. Thus, related to their size, it is more likely that two neighboring nodes are connected to the same third node. On the opposite side, we identify 7 networks with $C_g = 0$, among them the *Funkey Coin* and the *NucleusVisionCore*. Their ratio of simple edges to nodes indicates that they are either very similar or correspond exactly to star schemas - for example, the *FunkeyCoin* has 18106 nodes and 18105 simple edges.

5.5 Degree Assortativity

The assortativity indicates how nodes are connected with respect to a given property, such as the degree. If the degree correlation r_{deg} [25] of a network is positive, nodes tend to connect to other nodes with a similar degree - a network is said to be disassortative if this relationship is inverted, such that high degree nodes tend to be connected to low degree nodes. We calculate the degree assortativity for the simplified, undirected token networks (Fig. 7) and find that almost all of them are disassortative. Those networks that exhibit a degree assortativity of close to $r_{deg} = -1$ resemble star shapes, where most nodes have a connection to only one or a few high degree nodes. The only network with $r_{deg} > 0$ is the *blockwell.ai KYC Casper Token*, potentially due to its high number of small components.

5.6 Network Activity

To further quantify the activity inside a token network, we examine the initial token recipients and determine whether they send their tokens onward. As many tokens are listed on exchanges, and speculating with tokens is a common use case, we also examine whether a token network contains an address that is known to belong to an exchange. For this purpose we manually collected 113 exchange addresses from discussion forums and blockchain explorers such as Etherscan[3].

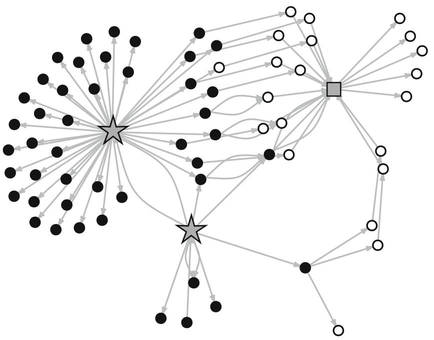

Fig. 8. SoftChainCoin Token Network with distributors (stars), initial recipients (black) and exchange (square).

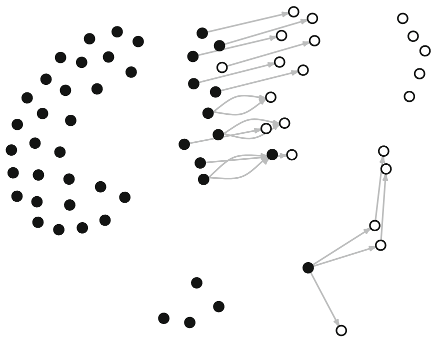

Fig. 9. SoftChainCoin Token Network with distribution and exchange addresses removed.

To illustrate our approach, Fig. 8 shows the small network of a token named *SoftChainCoin*[4]. The star-shaped nodes on the left distributed tokens to the black nodes - the initial recipients (R_i). Some of these, the active initial recipients (R_{ai}), have transferred tokens to other addresses. The active recipients (R_a), including both initial and secondary recipients, then transferred them to an exchange (square-shaped) or to other nodes. We define for each token network:

(a) The fraction of R_i (colored black in Fig. 8) relative to all addresses (R)
(b) The fraction of R_{ai} (that have sent tokens) relative to R_i
(c) The fraction of R_{ai} where there exists a path to an exchange
(d) The fraction of edges remaining, if distribution and exchange addresses are removed (Fig. 9), relative to the number of edges in the original network
(e) The mean minimum path length of those in c to an exchange.

We obtain the set of initial distributing nodes by determining the two nodes with the highest outdegree within the first 10% of transfers seen. We choose two, because manual inspection shows that sometimes tokens are not distributed from the first address itself, as can be seen in Fig. 8. We find that in about 25% of the token networks, the R_i account for 90–100% of all addresses (Fig. 10). These are likely airdrops that did not attract further users. On the other end, also in about 25% of the networks, the R_i account for less than 10% of all addresses, indicating that there are many addresses that joined the network after

[3] https://etherscan.io/.
[4] Token address: 0x86696431d6aca9bae5ce6536ecf5d437f2e6dba2.

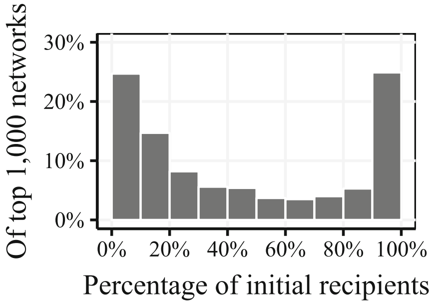

Fig. 10. In ≈25% of the networks (right bar), almost 100% of all nodes are initial recipients. Similarly, in ≈25% of the networks (left bar), nearly all addresses received tokens by other means.

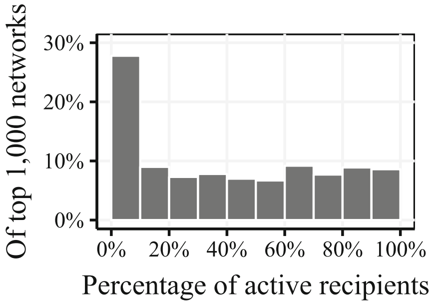

Fig. 11. In more than 25% of the networks (left bar), ≤10% of token recipients transferred their tokens. Conversely, in ≈8% of the networks (right bar), almost every address has issued a transfer.

the initial distribution, or that there has never been a large initial distribution at all. Figure 11 illustrates how many of these R_i have ever sent tokens onward – showed signs of activity (R_{ai}). Here we observe that in more than 25% of the networks, less than 10% of the initial recipients ever transferred their tokens. While it could be argued that these are users simply holding their tokens, another possibility is that these tokens are not wanted, and can be seen as a type of spam. In ≈8% of the networks, this activity percentage is near 100%, indicating that there exist strong incentives to transfer the corresponding token, such as the opportunity to sell the tokens at an exchange.

Figure 12 displays a scatter plot, where each cross represents one token network, positioned by the fraction of R_{ai} (b)), and the fraction of how many of these have an outgoing path to an identified exchange (c)). This fraction is not constant: as more initial recipients are active, more of them tend to send their tokens to an exchange. However, it is worth noting that the mere fact that an exchange offers to trade a certain token may also lead to increased activity directed towards exchanges. Nevertheless, very few token networks show active initial recipients without paths to exchanges, indicating that the main utility of most tokens is their trade on exchanges. If we remove distributing and exchange addresses from the graph, the median fraction of edges remaining (d)) is 42%, indicating that large parts of the networks only exist for that purpose.

For those R_{ai} with paths to exchanges, we determine for each network the shortest path to any identified exchange and compute the average shortest path length between active initial recipients and exchanges. Figure 13 shows that about half of the networks have a mean distance of two transfers. Given that exchanges often create artificial addresses for each customer, this implies that tokens are often sent directly to an exchange, indicating that trading is a main use case.

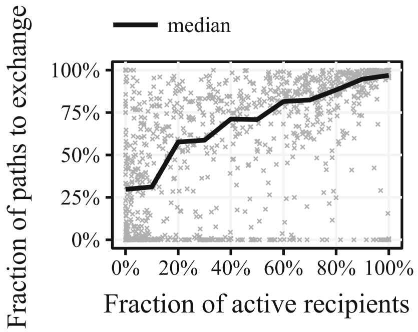

Fig. 12. When more initial recipients are active, more of them have a path to an exchange.

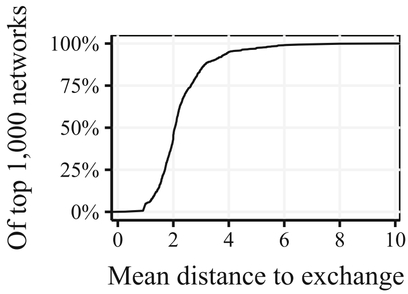

Fig. 13. In more than 50% of networks, those addresses that have a path to an exchange, their mean shortest path is ≤ 2.

6 Conclusion and Future Work

In this paper, we present a measurement study to analyze token networks, enabled through smart contracts on the Ethereum blockchain, from a graph perspective. We find that many follow either a star or a hub-and-spoke pattern. The heavy tails in the degree distributions are not as pronounced as in social networks - the networks tend to contain less and smaller hubs, such that they are mostly dominated by emitting addresses with a large out-degree and exchanges with a large in-degree. The number of exchanges is limited, only a few succeeded to gain the trust of the users. Small values for density and clustering coefficient embody the anonymity prevalent in these networks, as users mostly don't know each other and it's not very common yet to transfer tokens to acquaintances. A main use case of many tokens appears to be their sale, rather than their circulation, and some token networks barely show any activity after the initial distribution. The presented approach is part of our ongoing work on blockchain graph analysis, and our results help to understand current usage patterns and to design future systems.

Further research may refine the common understanding of token networks. While the presented approach observes the token network at the moment of data collection, observing the development of the networks over time might be even more insightful. Similarly, the presented approach is based on the assumption that each Ethereum address represents a single entity - we have not yet taken into consideration that an entity might be represented by several addresses. Furthermore, the forming of communities in the graph could be investigated, and not yet quantitatively available features such as the completeness and trustworthiness of ICO whitepapers could be included. Bitcoin and other cryptocurrencies offer a large field for criminals, while users show a large trust in ICOs even though faced with a total lack of a central contact address in case of a loss - knowledge about typical structures might lead to a differentiation between normal structures and anomalies, which may help to identify potentially fraudulent systems.

References

1. Akcora, C.G., Gel, Y.R., Kantarcioglu, M.: Blockchain: A Graph Primer. arXiv e-prints arXiv:1708.08749, August 2017
2. Anderson, L., Holz, R., Ponomarev, A., Rimba, P., Weber, I.: New kids on the block: an analysis of modern blockchains. arXiv e-prints arXiv:1606.06530, June 2016
3. Barber, S., Boyen, X., Shi, E., Uzun, E.: Bitter to better — how to make bitcoin a better currency. In: Keromytis, A.D. (ed.) FC 2012. LNCS, vol. 7397, pp. 399–414. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32946-3_29
4. Bartoletti, M., Pompianu, L.: An empirical analysis of smart contracts: platforms, applications, and design patterns. In: Brenner, M., et al. (eds.) FC 2017. LNCS, vol. 10323, pp. 494–509. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70278-0_31
5. Baumann, A., Fabian, B., Lischke, M.: Exploring the bitcoin network. In: WebDB 04: Proceedings of the 10th International Conference on Web Information Systems and Technologies (2014)
6. Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J.A., Felten, E.W.: SoK: research perspectives and challenges for bitcoin and cryptocurrencies. In: 2015 IEEE Symposium on Security and Privacy, pp. 104–121, May 2015. https://doi.org/10.1109/SP.2015.14
7. Broder, A., et al.: Graph structure in the web. Comput. Netw. **33**(1), 309–320 (2000). https://doi.org/10.1016/S1389-1286(00)00083-9
8. Clauset, A., Shalizi, C., Newman, M.: Power-law distributions in empirical data. SIAM Rev. **51**(4), 661–703 (2009). https://doi.org/10.1137/070710111
9. Donato, D., Leonardi, S., Millozzi, S., Tsaparas, P.: Mining the inner structure of the web graph. J. Phys. A: Math. Theoret. **41**(22), 224017 (2008)
10. Filtz, E., Polleres, A., Karl, R., Haslhofer, B.: Evolution of the bitcoin address graph. Data Science – Analytics and Applications, pp. 77–82. Springer, Wiesbaden (2017). https://doi.org/10.1007/978-3-658-19287-7_11
11. Gillespie, C.S.: Fitting heavy tailed distributions: the poweRlaw package. J. Stat. Softw. **64**(2), 1–16 (2015). http://www.jstatsoft.org/v64/i02/
12. Haslhofer, B., Karl, R., Filtz, E.: O bitcoin where art thou? insight into large-scale transaction graphs. In: SEMANTICS (Posters, Demos), vol. 1695 (2016)
13. Inaoka, H., Ninomiya, T., Tanigushi, K., Shimizu, T., Takayasu, H.: Fractal network derived from banking transaction. An analysis of network structures formed by financial institutions. Bank of Japan Working Paper Series (04) (2004)
14. Kleinberg, J.M., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A.S.: The web as a graph: measurements, models, and methods. In: Asano, T., Imai, H., Lee, D.T., Nakano, S., Tokuyama, T. (eds.) COCOON 1999. LNCS, vol. 1627, pp. 1–17. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48686-0_1
15. Kyriakopoulos, F., Thurner, S., Puhr, C., Schmitz, S.W.: Network and eigenvalue analysis of financial transaction networks. Eur. Phys. J. B **71**(4), 523 (2009). https://doi.org/10.1140/epjb/e2009-00255-7
16. Meiklejohn, S., et al.: A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 Conference on Internet Measurement Conference IMC 2013, pp. 127–140. ACM, New York (2013). https://doi.org/10.1145/2504730.2504747
17. Miller, A., et al.: Discovering Bitcoin's Public Topology and Influential Nodes (2015)

18. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement IMC 2007, pp. 29–42. ACM, New York (2007). https://doi.org/10.1145/1298306.1298311, http://doi.acm.org/10.1145/1298306.1298311
19. Newman, M.: Networks: An Introduction. Oxford University Press, Oxford (2010)
20. Nikolic, I., Kolluri, A., Sergey, I., Saxena, P., Hobor, A.: Finding The Greedy, Prodigal, and Suicidal Contracts at Scale. arXiv e-prints arXiv:1802.06038, February 2018
21. Reid, F., Harrigan, M.: An analysis of anonymity in the bitcoin system. In: Altshuler, Y., Elovici, Y., Cremers, A., Aharony, N., Pentland, A. (eds.) Security and Privacy in Social Networks, pp. 197–223. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-4139-7_10
22. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In: Sadeghi, A.-R. (ed.) FC 2013. LNCS, vol. 7859, pp. 6–24. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39884-1_2
23. Somin, S., Gordon, G., Altshuler, Y.: Network analysis of ERC20 tokens trading on ethereum blockchain. In: Morales, A.J., Gershenson, C., Braha, D., Minai, A.A., Bar-Yam, Y. (eds.) ICCS 2018. SPC, pp. 439–450. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-96661-8_45
24. Tschorsch, F., Scheuermann, B.: Bitcoin and beyond: a technical survey on decentralized digital currencies. IEEE Commun. Surv. Tutorials **18**(3), 2084–2123 (2016). https://doi.org/10.1109/COMST.2016.2535718
25. Van Steen, M.: Graph theory and complex networks. An Introduction, vol. 144 (2010)
26. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. Nature **393**(6684), 440 (1998)
27. Wood, G.: Ethereum: A secure decentralised generalised transaction ledger. https://github.com/ethereum/yellowpaper