Understanding Polkadot Through Graph Analysis: Transaction Model, Network Properties, and Insights

Hanaa Abbas[1], Maurantonio Caprolu[1], and Roberto Di Pietro[1]

Hamad Bin Khalifa University (HBKU) - College of Science and Engineering (CSE)
Division of Information and Computing Technology (ICT)—Doha, Qatar
{haab09879,macaprolu,rdpietro}@hbku.edu.qa

Abstract. In recent years, considerable efforts have been directed toward investigating the large amount of public transaction data in prominent cryptocurrencies. Nevertheless, aside from Bitcoin and Ethereum, little efforts have been made to investigate other cryptocurrencies, even though the market now comprises thousands, with more than 50 exceeding one billion dollars of capitalization, with some of them sporting innovative technical solutions and governance. This is the case for Polkadot, a relatively new blockchain that promises to solve the shortcomings in scalability and interoperability that encumber many existing blockchain-based systems. In particular, Polkadot relies on a novel multi-chain construction that promises to enable interoperability among heterogeneous blockchains.

This paper presents the first study to formally model and investigate user transactions in the Polkadot network. Our contributions are multifolds: After defining proper and pseudo-spam transactions, we built the transaction graph based on data collected from the launch of the network, in May 2020, until July 2022. The dataset consists of roughly 11 million blocks, including 2 million user accounts and 7.6 million transactions. We applied a selected set of graph metrics, such as degree distribution, strongly/weakly connected components, density, and several centrality measures, to the collected data. In addition, we also investigated a few interesting idiosyncratic indicators, such as the accounts' balance over time and improper transactions. Our results shed light on the topology of the network, which resembles a heavy-tailed power-law distribution, demonstrate that Polkadot is affected by the rich get richer conundrum, and provide other insights into the financial ecosystem of the network. The approach, methodology, and metrics proposed in this work, while being applied to Polkadot, can also be applied to other cryptocurrencies, hence having a high potential impact and the possibility to further research in the cryptocurrency field.

Keywords: Polkadot · Multi-chain Blockchain · Cryptocurrency · Graph Analysis · Metrics · Network Science · Decentralization · DeFi.

1 Introduction

Over the years, blockchain-based cryptocurrencies have witnessed rapid acceleration in terms of protocols evolution, market capitalization growth, and widespread public and business acceptance. Consequently, considerable efforts have been directed toward investigating the large amount of transactions data in cryptocurrency blockchains. Many complex systems are modeled using network science (or complex network theory) in various applications, such as computer networks, social networks, linguistics and even biology. By applying graph analysis to cryptocurrency networks, researchers were able to discover groundbreaking insights, uncover interesting properties, and characterize major activities on these systems. When these techniques have been applied to cryptocurrencies, some works revealed security concerns manifested in the form of unusual economical patterns. For instance, in Bitcoin, Ron and Shamir (2013) [20] discovered abnormally long and "fork-merge" chains in the transaction graph, which led to the identification of some malicious entities possibly abusing Bitcoin for money laundering, fraud, or other illegal activities. Graph analysis also allows identifying the topological properties of the network; where most cryptocurrency networks are usually found to exhibit small-world structures and power-law distributions [7]. Other studies used clustering algorithms to find hidden relations between different accounts to deanonymize users [14] and investigate unknown transaction patterns [4]. So far, all these techniques have been applied only to the two most diffused cryptocurrencies, Bitcoin and Ethereum. However, the current cryptocurrency landscape includes several other projects that, for capitalization and architectural advantages, certainly deserve the same level of attention. Moreover, these recent proposals also introduce elements of novelties, since they try to address the technical limitations the first proponents have discovered with time, as well as novel governance mechanisms. These latter features, in particular, require to be investigated with scientific method.

In this study, we investigate Polkadot, a recent cryptocurrency launched in May 2020. Despite its recent mint, it has successfully secured a spot amongst the top 10 cryptocurrencies by market capitalization[1]. Polkadot is known for being a fully "*sharded*" blockchain, whose design principles are based on sharding [18,23]—a database splitting technique—that enables multiple chains to process their transactions in parallel. Each blockchain shard is called a "*parachain*" which is connected to the *Relay Chain*. Parachains are heterogeneous blockchains that can be customized per project needs; for example, to host smart contracts or bridges [24]. The Relay Chain acts as the main hub of the system, orchestrating the network's Nominated Proof-of-Stake (NPoS) consensus [2] which requires the cooperation of DOT holders, validators (block authors), and nominators. Furthermore, Polkadot serves as an interoperability platform; i.e., it allows cross-communication between heterogeneous blockchains including external ones, such as Bitcoin and Ethereum.

[1] Data sourced from https://coinmarketcap.com/

Due to the novelty of multi-chains, there is a definite need to investigate their network operations, especially within an active ecosystem such as Polkadot. Despite achieving a good standing in the market, Polkadot has not yet received the same level of attention from academia commanded by other proposals, such as Bitcoin and Ethereum. In fact, to the best of our knowledge, this paper presents the first study on Polkadot that leverages graph analysis to characterize its transactions network. In detail, we investigate the Polkadot network using graph analysis to identify major network characteristics, including, but not limited to, statistical and topological properties. We examine how DOT, Polkadot's native currency, are transferred between user accounts. We collect all transactions that were committed on Polkadot's Relay Chain from Genesis to #11,320,000. Although it is to be noted that the transfer function was enabled on Polkadot on August 18, 2020 (block height #1,205,128). From the data, we construct the transactions graph and measure common graph metrics, such as: degree distribution, strongly/weakly connected components (SCC/WCC), and degree centrality. We believe that our analysis, enriched with data driven considerations, can help forecast the prospect growth and uses of both Polkadot and similar multi-chain blockchains, as well as opening up a few novel investigation avenues.

Contributions. Our main contributions are as follows:

1. We model the transactions among regular users in the Polkadot network. To this end, we first provide a formal definition of a Polkadot transaction, further divided into proper and improper transactions. Then, we model the transactions corresponding to money flow as a weighted directed multigraph.
2. We parse the Polkadot ledger, from the genesis block (May 2020) to block 11,320,000 (July 2022), to build the transaction graph representing the money flow among users.
3. We analyze the transaction graph by measuring global and local metrics. We obtain many new observations and insights on the structure of the network, useful to better understand the Polkadot ecosystem.
4. We identify and quantitatively analyze two different types of abnormal transactions, that we call self-loop and zero-transfer transactions, highlighting their patterns in terms of daily frequency and transaction values.
5. We empirically verify that Polkadot is affected by the *rich get richer* problem by studying user balances over time.
6. To the best of our knowledge, this is the first study that, leveraging graph theory and network science, analyzes transaction data and measures statistical properties of the Polkadot network.
7. The code used to collect the data analysed in this study is released as open source[2].

Paper Organization. The remainder of the paper is organized as follows. Section 2 explores related work in the literature. We model the transactions among

[2] A link pointing to the source code for building and analyzing the graph will be provided in the camera-ready version

Polkadot users in Section 3, then we describe the process of building the transaction graph in Section 4. In sections 5, 6, 7, and 8, we present and discuss the results of our analysis. Lastly, we report some concluding remarks in Section 9.

2 Related Work

Several works utilized graph-based analysis to investigate prominent blockchain-based networks, mainly Bitcoin and Ethereum [15,21]. Graph-based modeling allows to reveal insights into cryptocurrency transactions and user interactions, including other important tasks such as: cryptocurrency price prediction [13], address clustering [9,22,16], user deanonymization [10], attack forensics, detection of malicious activities such as phishing scams, counterfeit tokens, or money laundering [6], and detection of anomalies (e.g., in smart contracts execution) [5,11]. The graphs are built from the blockchains' publicly available transactions data. However, the architectural differences existing between transaction-based blockchains (e.g., Bitcoin) and account-based blockchains (e.g., Ethereum and Polkadot) require different graph analysis approaches. In this paper, we focus on account-based methods.

In account-based networks, native currency or tokens are represented as a balance that can be deposited to or withdrawn from the user's account. Each transaction can have only one input and one output. A node in the transaction graph represents a unique address and an edge represents a transaction. Since there are no works thus far pertaining to Polkadot, we summarize works from the literature about Ethereum. [8] found that Ethereum transactions volume, components size, incoming or outgoing transaction relations can be approximated by a power-law distribution, which exhibits a heavy-tailed structure. Additionally, [15] found that the growth rate (size of nodes and edges) and graph density are correlated with the price of ETH. Also, the degree distribution of the network follows a power law, and the transaction network is non-assortative. Non-assortativity means that nodes do not tend to communicate with only low-degree or only high-degree nodes [17].

There are a few works in the scientific literature that investigate Polkadot. The work in [1] presented a data-driven study that details the architecture of Polkadot and identifies several of its limitations and design contradictions. Their investigation shows that due to the restriction on the number of allowed validators in the network, a high minimum stake requirement was enforced which varied with the size of the validators set. In addition, a majority of the validators were found to charge 100% commission, thus excluding nominators from monetary incentivization and violating the basic principles of the NPoS economic security. Our work investigates Polkadot from a different perspective through graph-based modeling. Graph analysis allows us to extract refined insights into not only the structure of the network but also the transaction patterns, allowing us to highlight a few abnormal features in the transactions graph.

3 Modeling The Polkadot Transaction Graph

In this section, we model the economic interactions among users in the Polkadot environment. To this end, we first formally define a transaction, either in its proper or abnormal form. This distinction allows us to formally separate transactions that have effectively moved money between two accounts from those that, even if successful, have not had any real effect on the involved balances.

3.1 Polkadot Transactions

Polkadot uses the term "extrinsics" to refer to state changes emerging from the outside world, which include balance transfers. However, for the sake of simplicity, we refer to `balances.transfer` extrinsics as "transactions" in the rest of the paper. We define a transaction in Polkadot as follows:

Definition 1 (Transaction). *A transaction is a signed extrinsic submitted to the blockchain by a user account via a* ***`balances.transfer`*** *call or part of a* ***`utility.batch`*** *call, where its general attributes are:*

- *signed* $= True$*;*
- *module_id* $=$ *"Balances";*
- *and, call_id in*$($"$transfer$", "$transfer\_keep\_alive$", "$transfer\_all$"$)$*;*

Furthermore, transactions can be formally divided as proper and improper, according the the definitions provided in the following. Let A be the set of all addresses present in the Polkadot ledger, and *ExtrinsicSuccess* is the system event triggered if the transaction is successful. We model a transaction t as a tuple $(In, Out, \lambda, \tau, \phi)$, where $In, Out \in A$ and $\lambda, \phi \in \mathbb{R}^+$, meaning that the account In is paying, at the time τ, λ DOTs to the account Out. In addition, ϕ represents the fee payed for issuing the transaction.

Definition 2 (Proper Transaction). *We say that t is a proper transaction if it satisfies the following properties:*

- $ExtrinsicSuccess = 1$*;*
- $Value > 0$*; and,*
- $In \neq Out$*.*

Definition 3 (Pseudospam Transaction). *We say that t is a pseudo-spam, also called improper, transaction if it satisfies the following properties:*

- $ExtrinsicSuccess = 1$*; and,*
- $Value = 0$ *or* $In = Out$*.*

In other words, a transaction is considered proper when the given transaction amount, greater than zero, is withdrawn from the sender's account and deposited to the receiver's account successfully. Conversely, an improper transaction is a successful transaction with no impact on the account's balance other than the deduction of the transaction fee, because the sender and receiver addresses are the same and/or the value of the transaction is 0.

Following our analysis of the entire Polkadot ledger, we have identified two forms of pseudospam transactions: (1) *zero transfers* where the transaction value is zero DOT; and, (2) *self-loops* where the destination address is the same as the sender's address. It is important to emphasize that an improper transaction is not a failed transaction; in fact, in the Polkadot network, all transactions are stored on the blockchain, even if they have failed. However, only a successful transaction returns an $ExtrinsicSuccess$. Consequently, we do not consider failed transactions as abnormal. Examples of failed transactions include: transactions whose destination address was not found, or those attempting a balance transfer while having insufficient funds to cover the transaction fee or the transfer value. In addition, we disregard Balances extrinsics that called methods intended for use by Root origin only (Note: Sudo user was removed only after the NPoS scheme was enabled in June 2021). Even though the use of sudo-level functions might have had a malicious intent, such transactions were scarcely found in the dataset, and more importantly, they have failed.

3.2 Polkadot Transaction Graph

Since Polkadot is an account-based blockchain, similar to Ethereum, the money flow among users can be formally modeled as a weighted directed multigraph $M := (A, T)$, where A is the set of all addresses, i.e., user accounts, and T is the set of successful transactions, as defined in definitions 2 and 3.

Fig. 1 shows an example of the Polkadot transaction graph. A multi-graph allows an arbitrary number of edges to exist between a pair of nodes in any direction (e.g., Nodes A-B) and also supports self-loops (e.g., Node D). In this example, Node C is the most central node—all other nodes are connected through it. The graph is weighted, where weights are attributes that describe the graph's edges. The attributes include the transaction value in DOT and transaction timestamp. Incorporating timestamps in the graph analysis is essential for investigating temporal properties and evolution of the network, e.g., monthly progress.

4 Building the Transaction Graph

To build the transaction graph, we followed a methodology that includes multiple steps. First, we parsed the Polkadot ledger and we imported the transaction data into a relational database. Then, we queried from the database all the transactions that met the conditions listed above in Definition 1. Finally, using the NetworkX [3] python library, we built a *MultiDiGraph* and we analyzed it under different perspectives.

For the experimental part of this work, we set up a development environment on a DELL workstation, running a Windows 10 PRO OS, that includes a python

[3] https://networkx.org/

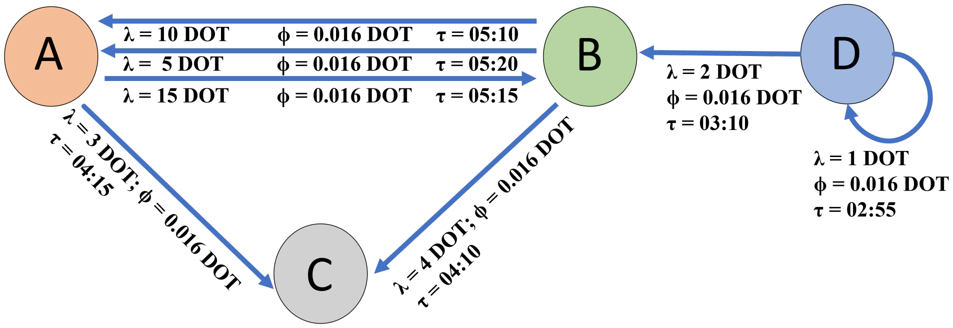


Fig. 1: Graph Representation of a Weighted MultiDiGraph in Polkadot.

(v3.10.2) IDE, a MySQL database (v8.0.28), and a Polkadot full node. The hardware specifications are as follows: Intel(R) Core(TM) i9-9900KS CPU@4 GHz, 64 GB RAM, and 2 TB SSD. We configured the full Polkadot node on an Ubuntu 20.04 LTS running on the Windows Subsystem for Linux (WSL), since Substrate—the framework on which Polkadot is built—is not natively compatible with Windows. We run the node in archive mode to access past states of the chain at any point in time. After fully syncing the blockchain storage on the node, we query and parse the blocks data into the MySQL database, starting from genesis and ending with block #11,320,000. Nonetheless, the actual analysis of transactions data starts at block #1,205,128, because Polkadot balance transfers were enabled only after the specified block height. Even though an archive node takes up large disk space (the collected data corresponding to 11.32 million blocks occupies up to 457 GB on disk), we opted for running our own node instead of querying the data from publicly available RPCs to guarantee data integrity and validity.

We implemented a software that comprises two main components: (1) a blockchain data parser; and (2) a graph analyzer. Our code base in Python follows a modular approach: The data parser queries blocks stored on Polkadot's Relay Chain, along with their extrinsics and events data, and stores the collected data on a MySQL database, whereas the graph analyzer generates a directed multi-graph abstraction of the transactions network. From the generated graph, the analyzer computes relevant metrics that define the network structure and characteristics. We also perform statistical analysis of the transactions data through direct SQL queries. To interface with the Polkadot node, we use two open-source Python libraries implemented by Parity: `substrate-interface` (API for Substrate nodes, which provides different methods for querying data storage and interacting with the chain) and `scalecodec`—needed for decoding/encoding SCALE Codec format that is used by the Substrate runtime. Moreover, we choose NetworkX library to perform the network analysis in Python, since it offers a vast choice of algorithms and tools to produce various metrics, including but not limited to: clustering, connectivity, assortativity, connected components and graph flows. It also supports graph visualization and serialization into different formats such as GraphML, JSON, GIS Shapefile, or a Python Pickle object. Our graph builder module relies on graph pickling functionality to store the graph object and deserialize it for faster processing.

5 Transaction Graph Analysis

In this section, we perform an in-depth study of the transaction network based on various graph properties which can be classified as global properties, i.e., related to the whole graph, and local properties, i.e., related to single nodes. In the following section, we elaborate on what the metrics suggest in terms of the network's structure.

Global Properties. The graph has a total of 2,149,679 nodes (corresponding to unique addresses) and 7,613,325 edges (corresponding to transactions), which include pseudospam transactions to be explored in more detail in the next section (Section 6). In the following analysis, we omitted pseudospam transactions then computed the graph metrics accordingly. Excluding pseudospam transactions reduced the count of edges and nodes by 53,121 (transactions) and 1,722 (accounts), respectively. This is an interesting finding as it suggests that 1,722 accounts have been involved with only pseudospam transactions throughout the history of the network. Among the global properties, we studied the graph's connected components, in addition to assortativity, reciprocity, density, clustering, and transitivity, displayed in Table 2.

Connected Components. Graph connectivity is an important measure of the network's resilience. A graph is said to be connected if there exists a path between every pair of nodes. A connected component is a subgraph in which every node is reachable from every other node. For Strongly Connected Components (SCCs), edge direction is taken into account, whereas for Weakly Connected Components (WCCs), direction is ignored. In Fig. 2, we plot the distributions of SCCs in blue and WCCs in green. For both, the result demonstrates that the network is composed of a single giant component—the largest connected subgraph—and many, much smaller components. The components size distribution resembles power-law distribution and is heavy tailed as shown in Fig. 2. This indicates that the network has a few central nodes (hubs) involved in a very large number of transactions with other nodes, forming a giant connected subgraph; while, the majority of the other nodes transact with just a small number of nodes. The hubs in this network carry out a significant role; that is, connecting a significant number of users together. Table 1 lists the node composition of the giant SCC and WCC components, each consisting of 62% and 99.97% of the nodes, respectively. Almost all nodes in the network can be reached from another node by some path, ignoring edge direction.

Degree Assortativity Coefficient. Assortativity measures the correlation between nodes in the graph with respect to their degree. For Polkadot's transactions network, the assortativity coefficient is reported as -0.255, indicating weak disassortativity. A negative assortativity value indicates that the graph's degrees are

Table 1: Summary of Graph Connected Components: SCC and WCC

#SCC	Giant SCC		#WCC	Giant WCC	
	#Nodes (% of nodes)	#Edges (% of edges)		#Nodes (% of nodes)	#Edges (% of edges)
806747	1,332,655 (62%)	5,870,608 (77.7%)	257	2,147,265 (99.97%)	7,559,472 (99.99%)

Table 2: Polkadot Transactions Graph's Global Properties

#Nodes	#Edges	Assortativity	Reciprocity	Density	Clustering	Transitivity
2,147,957	7,560,204	-0.255	0.017	1.64e-6	0.256	9.07e-6

negatively correlated. Notably, it indicates that high-degree nodes (aka 'hubs', such as crypto market exchanges) tend to form connections with nodes of lower degrees and that the network's topology does not behave like the so-called "rich club" phenomenon [25]. High degree nodes connect with smaller ones rather than with similarly high degree nodes.

Reciprocity. Reciprocity measures the likelihood of nodes in a directed network to be mutually linked (i.e., having bidirectional edges) [12]. Reciprocity is computed as 0.017, which is a value approaching 0. This indicates that just a small number of nodes transact in both directions. Even though the majority of the nodes are somehow connected in Polkadot, they tend to transact mostly in a uni-directional manner.

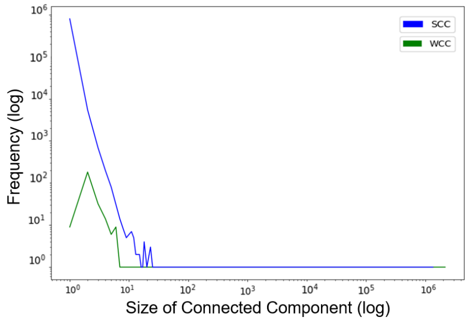


Fig. 2: Connected Components Size Distribution

Density, Clustering, and Transitivity. These three metrics are computed over the undirected graph. *Density* is the ratio of existing edges divided by the maximum possible edges in a graph [12]. The small density value (1.64e-6) indicates a less-dense graph that has more nodes than edges. Meaning, it is likely that users tend to create new accounts while executing new transactions to increase their anonymity [15]. *Global clustering coefficient* evaluates the extent to which nodes in a graph tend to cluster together [5]. The coefficient approximates to 0.256 ($\approx\frac{1}{4}$), indicating that user accounts are likely to form clusters; i.e., if two accounts transact with a third account, it is likely that the former will also transact with the other two. *Transitivity* can be used to find the community structure in blockchain graphs [12]. The transitivity of the graph, a small value in the order of 10^{-6}, suggests the lack of community structure possibly due to the presence of high-degree nodes that are "loner-stars" connected mainly to low-degree nodes.

Local Properties. Next, we investigate common local properties which are node degree distribution (including in- and out- degree) and degree centrality.

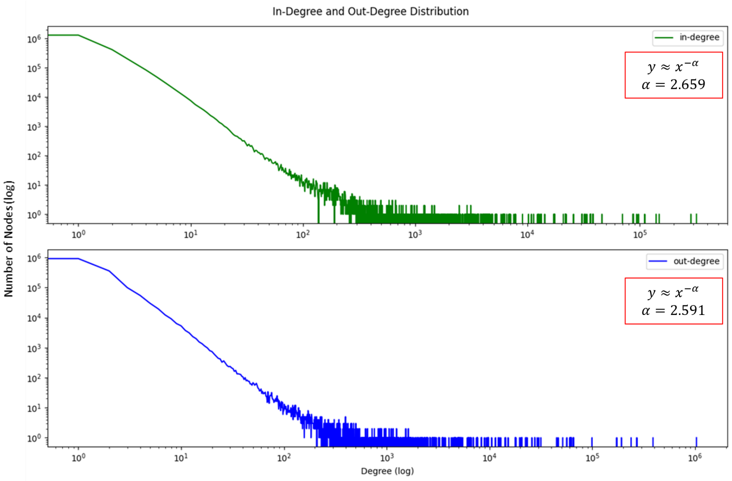


Fig. 3: Transactions Graph In-Degree and Out-Degree Distribution

Degree Distribution. For cryptocurrency networks, the degree distribution provides a high-level outlook about the transaction relations and how nodes are connected in the network. The in-degree and out-degree values of a node correspond to incoming and outgoing transactions, respectively. Fig. 3 shows the in- and out- degree distributions, in log-log scale, of the transaction network. For both distributions, the power-law model ($y \sim x^{-\alpha}$) provides a reasonable fit. The tail/end segment is heavier than pure power law distributions, indicating that the number of high-degree nodes (influential nodes, e.g., market exchanges) is relatively much smaller than low-degree nodes (e.g., regular users). The larger the value of α, shown in Fig. 3, the less variable the node degrees are.

Degree Centrality. Centrality measures help to identify the most important nodes in a network. Table 3 lists the top 10 accounts based on normalized degree centrality, which is the fraction of addresses each node is connected to. We also list the in-degree and out-degree coefficients whose sum adds up to the degree value. The max degree centrality belongs to address $1exaAg...T6EGdE$. Upon further search, we found that the address has been identified by the online community as belonging to Binance [19], a prominent cryptocurrency exchange marketplace. This Binance node has been inactive since January 2022; however, before it went inactive, it transferred large sums of DOT to a new address $1qnJN7FViy3H...8GT7$ (listed in row 3), which we believe is the new Binance node—it has achieved high degree centrality in a relatively short time span.

Table 3: Top-10 Most Important Nodes Evaluated By Degree Centrality

#	**Account**	**Known Identity/Role**	**Degree**	**In-degree**	**Out-degree**
1	1exaAg2VJRQ...EGdE	Binance	0.614	0.133	0.481
2	12xtAYsRUrm...XkLW	Nominator	0.275	0.149	0.126
3	1qnJN7FViy3H...8GT7	Binance	0.227	0.045	0.182
4	15kUt2i86LH...XAkX	N/A	0.163	0.052	0.111
5	15SbxvcrYSQz...jy82	N/A	0.150	0.069	0.081
6	16hp43x8DUZt...4oEd	N/A	0.090	0.044	0.046
7	14Kazg6SFiUC...dQhv	N/A	0.090	≈ 0	0.090
8	12wVuvpApgp...Lchb	N/A	0.065	0.065	≈ 0
9	16HNPJqej7E...L8cj	N/A	0.049	0.018	0.031
10	157PD8GV7pJ...B2KR	N/A	0.049	0.019	0.030

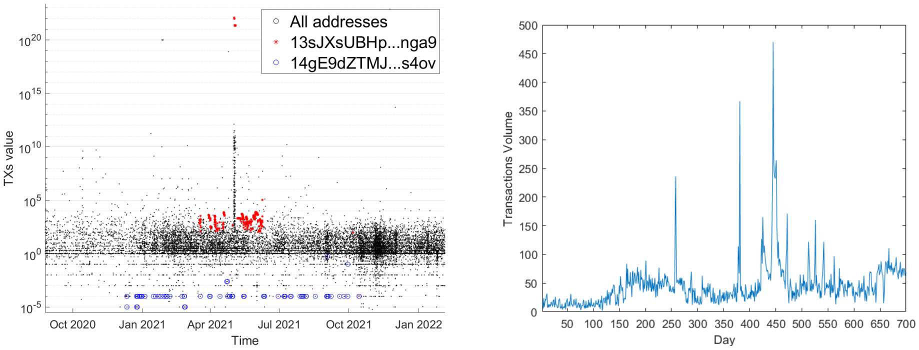
6 Statistical Analysis of Self-loop Transactions

The collected data contains 31,961 (0.41%) self-loop and 4,677 (0.06%) zero-transfer transactions out of 7,613,325 transactions. Both of these transaction types account for much less than 1% of all transactions; however, it is important to investigate them since they do not comply with typical economical interactions. In this paper, we focus mostly on self-transfers since they occur more often.

First, we investigated self-loop transactions in the literature. We found one mention in [12], where the authors interpreted the presence of self-loops in Ethereum according to two trivial scenarios: users verifying if it is possible to send Ether to themselves, or due to a mistake while specifying the receiver address. However, in the case of Polkadot, further investigation is needed to understand the cause of this trend, as the frequency of those transactions suggests different scenarios.

Fig. 4 shows the value and volume of self-loop transactions over time. Self-loop transactions were found to exist on a daily basis with arbitrary values (sometimes constant and sometimes following a pattern) and occur throughout the day. Fig. 4a reveals interesting patterns in self-loop values and peculiar user behaviors. For example, 1 is the most frequent transaction value, constantly used over time, together with other multiples and sub-multiples of 10. In addition, the figure also highlights the values adopted by the two accounts with the highest number of self-loops. The first one, represented with red asterisks, issued 635 transactions over three months, with different values, sometimes decreasing according to a specific pattern. The second one, represented with blue circles, issued 145 transactions over a year, almost all with the same value of 0.0001.

From Fig. 4b, instead, it can be observed that self-loop transactions appear on a daily basis in the Polkadot ledger. In particular, every day we can observe around 50 self-loops, with a few huge spikes, and almost the double during the last observed months.



(a) Self-loop Transaction: Values over Time

(b) Self-loop Transactions: Volume over Time

Fig. 4: Polkadot Self-loop Transactions: (a) Values over Time — all users (black circles) and the two most active accounts (red asterisks and blue circles); (b) Volume over Time

7 Analysis of Polkadot Accounts' Balance

We investigated the distribution of the total balance in DOT for all Polkadot accounts. Overall, there is a total of 1,042,149 active accounts in the network as of July 25, 2022. As shown in Fig. 5a, the distribution of DOTs over all accounts in Polkadot indeed resembles power-law distribution with a heavy-tailed structure. The majority of the accounts (over 1 million accounts) hold small balances, in the range 0-499K DOTs, and only a few own balances float in the range from 500K to over 50 million DOTs—50M DOTs have a market value of 300+ millions USD as of the 19th of October 2022.

We also examined the percentage of DOT held per account type (See Table 4): nominators, validators, council members, and others which may include regular users and proxy accounts. Proxy accounts are addresses created to perform a limited number of actions on behalf of the main account. Nominators own the largest fraction of DOT ($\approx$ 57%), whereas validators on the other hand hold only $\approx$ 0.1% of all available DOTs.

The previous observation is interesting given that, as shown in [1], over 60% of the validators in April 2022 charged 100% commission and retained block/era

Table 4: Balance Share per Account Type (in July 2022)

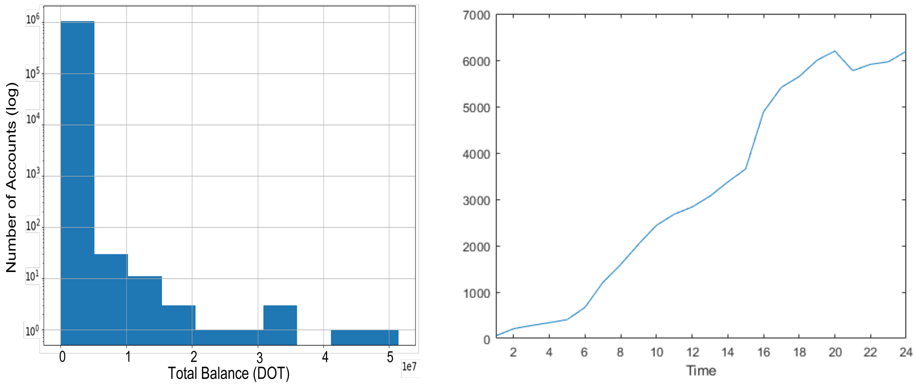
Type	**Count**	**Share**
Nominators	21,404	57.275%
Validators	297	0.152%
Council	13	0.006%
Others	1,020,435	42,567%

rewards to themselves. Hence, it was expected that validators should comparatively have higher balances, but in reality validators contribute little (around 0.2%) to total staking. We find that the typical interaction of nominators and validators in Polkadot is as depicted in Fig 6. Nominators declare their intent to vote for their validator(s) by staking their DOTs. The validators collect a large-enough stake from nominators that allows them to join the active set. After every era (24 hours), the era rewards are relatively equally distributed to all validators. 100%-commissioned validators retain rewards to themselves, whereas other validators can trigger a payout action to nominators according to their share in the total stake. In the case of Binance—the world's largest crypto exchange [3], the rewards amassed by its validators are forwarded to an intermediary address (called rewards address) which then forwards all its balance to the exchange address (top central node as listed in Table 3). We would like to point out that this behavior does not violate the protocols set out by Polkadot, nor does it pose major security risks because block production is not affected by validator stake [24]. These observations only identify limitations towards a 'true' decentralization of the network, due to the presence of highly capitalized, centralized, crypto exchanges [1].

To investigate the "rich get richer" phenomenon in Polkadot, we measured the users' balance evolution over time. A user is considered rich if his/her balance is higher than the average user balance. Formally, the hypothesis is that *the k richest users at time t are richer that the k richest users at time $t' < t$* [7]. To verify this hypothesis, we first define the Wealth Ratio (wr) as the average balance of the k richest users over the average balance of all the other ($|A| - k$) active users in the Polkadot network. Then, we check if the k richest accounts in M_t are richer than the k richest accounts in $M_{t'}$ by computing wr over time, as follows:

$$wr_t = \frac{\sum_{a \in K_t} \frac{b_t(a)}{|K_t|}}{\sum_{a \in \{A_t \setminus K_t\}} \frac{b_t(a)}{|A_t \setminus K_t|}} \tag{1}$$

where $b_t(a)$ is the balance of account a at time t, and A_t and K_t are the set of all active accounts and the set of the k richest accounts, respectively, at time t. With M_t we refer to the graph induced by transactions having timestamp less than t. For our investigation, we set $k = 100$, while t varies appropriately to



(a) Distribution of DOT

(b) Wealth ratio over time

Fig. 5: Analysis of Polkadot Accounts' Balance. (a) Account balance distribution; (b) Ratio between the Top-100 richest accounts average balance with respect to all (Active) accounts' average balances.

consider monthly snapshots of the Polkadot ledger over the observation period. Fig. 5b shows that *wr* clearly increases over time. This means that the disparity between richest nodes and all the other accounts grows over time, empirically confirming the *rich get richer* hypothesis.

8 Discussion

It is a common phenomenon for real-world networks to contain hubs that are highly connected to many nodes. The presence of hubs gives the degree and component size distribution a long (heavy) tail, indicating that: there are a few nodes, with a much higher degree than most other nodes, also at the center of the network's giant components. These characteristics, specifically the power law approximation, are associated with what is known as a scale-free network [8].

Based on what discussed in Section 5, we can conclude that Polkadot's topology resembles a scale-free network, where at its center is Binance, a crypto market exchange, that dominates the network in terms of centrality and influence. As well-known in the literature, networks with power law degree distributions may introduce potential vulnerabilities. Indeed, if the central hubs, or the nodes with high degrees, are controlled or compromised, the entire network's functionality will get affected [8]. Having exchange centers and mining/staking pools with stronger connectivity than other nodes eventually leads to concentration/centralization of power, which is a phenomenon that is not desirable in decentralized blockchains. In the specific case of Polkadot, the most central entity in the network, Binance, also actively participates in the consensus protocol with nominator/validators accounts, potentially exacerbating the vulnerabilities above mentioned.

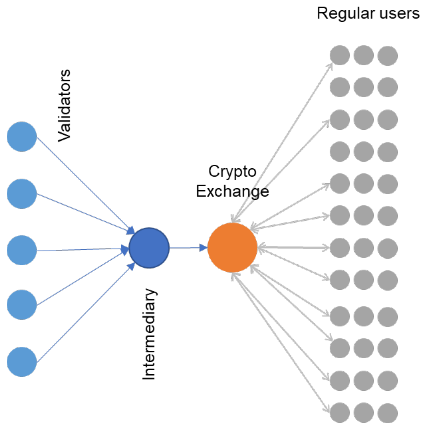


Fig. 6: Typical Interaction among Binance-supported validators, Binance-owned accounts, and regular users.

Other interesting insights on the Polkadot environment come from Section 7, where we showed that about 57% of DOT's total supply is owned by nominators, which accounts for only 2% of active users. In addition, we found that the disparity between rich accounts and regular users is increasing over time, demonstrating that Polkadot suffers from the *rich get richer* phenomenon.

9 Conclusion and Future Work

To the best of our knowledge, this study is the first to formally model Polkadot's transactions data, probing statistical and structural properties of the network, and investigating its properties. By means of graph analysis, we have identified that Polkadot resembles a scale-free network and discovered the presence of a hub, attributable to Binance, dominating the network in terms of centrality and influence. We have also identified abnormal transaction patterns, which we term "pseudo-spam", that include two categories: self-loops (sender address is the same as the receiver address) and zero-transfers (transfer value equals to zero DOT). Both categories effectively have no economic value or impact on the owner's account balance. However, they still frequently appear in the ledger and, sometimes, exhibit fuzzy patterns that deserve further investigation in future work. In addition, we investigated the users' balance over time, finding that the distribution of DOT over all accounts resembles a heavy-tailed power-law distribution, and that the Polkadot network, as many other cryptocurrencies, is affected by the *rich get richer* problem.

The contributions provided in this paper, other than shedding light on a novel proposal in the cryptocurrency ecosystem (multichains), also highlight a few existing critical structural issues and point out transactions' suspicious patterns, possibly stimulating further research in the field.

Acknowledgements This publication was partially supported by the Qatar National Research Fund (QNRF), a member of The Qatar Foundation, through the awards [NPRP-S-11-0109-180242] and [NPRP11C-1229-170007]. The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the QNRF.

References

1. Abbas, H., Caprolu, M., Di Pietro, R.: Analysis of polkadot: Architecture, internals, and contradictions. In: 2022 IEEE International Conference on Blockchain (Blockchain). pp. 61–70 (2022). https://doi.org/10.1109/Blockchain55522.2022.00042
2. Ali, I.M., Caprolu, M., Di Pietro, R.: Foundations, properties, and security applications of puzzles: A survey. ACM Comput. Surv. **53**(4) (Aug 2020). https://doi.org/10.1145/3396374, https://doi.org/10.1145/3396374
3. Aysan, A.F., Khan, A.U.I., Topuz, H., Tunali, A.S.: Survival of the fittest: A natural experiment from crypto exchanges. The Singapore Economic Review pp. 1–20 (2021)
4. Caprolu, M., Pontecorvi, M., Signorini, M., Segarra, C., Di Pietro, R.: Analysis and patterns of unknown transactions in bitcoin. In: 2021 IEEE International Conference on Blockchain (Blockchain). pp. 170–179 (2021). https://doi.org/10.1109/Blockchain53845.2021.00031
5. Chen, T., Li, Z., Zhu, Y., Chen, J., Luo, X., Lui, J.C.S., Lin, X., Zhang, X.: Understanding ethereum via graph analysis. ACM Transactions on Internet Technology (TOIT) **20**(2), 1–32 (2020)
6. Di Battista, G., Di Donato, V., Patrignani, M., Pizzonia, M., Roselli, V., Tamassia, R.: Bitconeview: visualization of flows in the bitcoin transaction graph. In: 2015 IEEE Symposium on Visualization for Cyber Security (VizSec). pp. 1–8. IEEE (2015)
7. Di Francesco Maesa, D., Marino, A., Ricci, L.: Data-driven analysis of bitcoin properties: exploiting the users graph. International Journal of Data Science and Analytics **6**(1), 63–80 (2018)
8. Guo, D., Dong, J., Wang, K.: Graph structure and statistical properties of ethereum transaction relationships. Information Sciences **492**, 58–71 (2019)
9. Harrigan, M., Fretter, C.: The unreasonable effectiveness of address clustering. In: 2016 Intl IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld). pp. 368–373. IEEE, Toulouse, France (July 2016)
10. Jawaheri, H.A., Sabah, M.A., Boshmaf, Y., Erbad, A.: Deanonymizing tor hidden service users through bitcoin transactions analysis. Computers & Security **89**, 101684 (2020). https://doi.org/https://doi.org/10.1016/j.cose.2019.101684, https://www.sciencedirect.com/science/article/pii/S0167404818309908
11. Khan, A.: Graph analysis of the ethereum blockchain data: A survey of datasets, methods, and future work. In: 2022 IEEE International Conference on Blockchain (Blockchain). pp. 250–257 (2022). https://doi.org/10.1109/Blockchain55522.2022.00019

12. Lee, X.T., Khan, A., Sen Gupta, S., Ong, Y.H., Liu, X.: Measurements, analyses, and insights on the entire ethereum blockchain network. In: Proceedings of The Web Conference 2020. pp. 155–166 (2020)
13. Lin, D., Wu, J., Yuan, Q., Zheng, Z.: Modeling and understanding ethereum transaction records via a complex network approach. IEEE Transactions on Circuits and Systems II: Express Briefs **67**(11), 2737–2741 (2020). https://doi.org/10.1109/TCSII.2020.2968376
14. Maesa, D.D.F., Marino, A., Ricci, L.: An analysis of the bitcoin users graph: inferring unusual behaviours. In: International Workshop on Complex Networks and their Applications. pp. 749–760. Springer (2016)
15. Motamed, A.P., Bahrak, B.: Quantitative analysis of cryptocurrencies transaction graph. Applied Network Science **4**(1), 1–21 (2019)
16. Neudecker, T., Hartenstein, H.: Could network information facilitate address clustering in bitcoin? In: Brenner, M., Rohloff, K., Bonneau, J., Miller, A., Ryan, P.Y., Teague, V., Bracciali, A., Sala, M., Pintore, F., Jakobsson, M. (eds.) Financial Cryptography and Data Security. pp. 155–169. Springer International Publishing, Cham (2017)
17. Piraveenan, M.R.: Topological analysis of complex networks using assortativity. University of Sydney (2010)
18. Polkadot: Polkadot v1.0: Sharding and economic security. https://polkadot.network/blog/polkadot-v1-0-sharding-and-economic-security/, accessed: 2022-10-10
19. Polkadot.js: Polkadot.js phishing known addresses. https://github.com/polkadot-js/phishing/blob/master/known.json, accessed: 2022-10-10
20. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In: International Conference on Financial Cryptography and Data Security. pp. 6–24. Springer (2013)
21. Serena, L., Ferretti, S., D'Angelo, G.: Cryptocurrencies activity as a complex network: Analysis of transactions graphs. Peer-to-Peer Networking and Applications pp. 1–15 (2021)
22. Victor, F.: Address clustering heuristics for ethereum. In: Bonneau, J., Heninger, N. (eds.) Financial Cryptography and Data Security. pp. 617–633. Springer International Publishing, Cham (2020)
23. Wang, G., Shi, Z.J., Nixon, M., Han, S.: Sok: Sharding on blockchain. In: Proceedings of the 1st ACM Conference on Advances in Financial Technologies. pp. 41–61 (2019)
24. Wood, G.: Polkadot: Vision for a heterogeneous multi-chain framework. White Paper **21**, 2327–4662 (2016)
25. Zhou, S., Mondragón, R.J.: The rich-club phenomenon in the internet topology. IEEE communications letters **8**(3), 180–182 (2004)