Bow-tie structure of the Polkadot transfer network

Piskorec, Matija ; James Murphy, Ben Domenic ; Rüegsegger, Florian ; Niya, Sina Rafati ; Tessone, Claudio J

DOI: https://doi.org/10.1109/icbc56567.2023.10174939

Originally published at:
Piskorec, Matija; James Murphy, Ben Domenic; Rüegsegger, Florian; Niya, Sina Rafati; Tessone, Claudio J (2023). Bow-tie structure of the Polkadot transfer network. In: 2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Dubai, United Arab Emirates, 1 May 2023 - 5 May 2023. Institute of Electrical and Electronics Engineers, 10174939.
DOI: https://doi.org/10.1109/icbc56567.2023.10174939

Bow-tie structure of the Polkadot transfer network

Matija Piškorec*†, Ben Domenic James Murphy‡, Florian Rüegsegger‡, Sina Rafati Niya*, Claudio J. Tessone*§
* Blockchain and Distributed Ledger Technologies Group, University of Zurich, Zurich, Switzerland
† Rudjer Boskovic Institute, Zagreb, Croatia
‡ University of Zurich, Zurich, Switzerland
§ UZH Blockchain Center, University of Zurich, Zurich, Switzerland

***Abstract*—While there are many data collection and analysis tools for Ethereum - the largest smart contract blockchain by market capitalization, development of similar tools for other smart contract blockchains is lacking. Reasons for this are non-existent standards, changing specifications due to rapid development, common usage of the off-chain storage, and lack of developers. One of such blockchains is Polkadot - a layer-zero blockchain featuring a single relay chain whose role is to secure smart contract transactions on multiple other parachains. In this paper we describe a data collection pipeline for Polkadot blockchain that we then use to perform an analysis of the bow-tie structure of its transfer network over time, with special emphasis on the role of nominators and validators in this structure. We find evidence that the Polkadot ecosystem is slowly maturing from a system dominated by nominators and validators, both of which require some technical skill as well as willingness to bond sufficient amount of funds, into a system increasingly populated by regular users using the financial services of Polkadot.**

***Index Terms*—Polkadot, Proof-of-Stake, blockchain data collection, transfer network, bow-tie network structure**

I. Introduction

In this paper we describe a data collection pipeline and provide analysis of the staking dynamics with respect to the bow-tie structure of the transfer network of the Polkadot blockchain - a multi-chain smart contract blockchain released in 2020 [1]–[3]. Our data collection pipeline is designed to continually store recently validated blocks in a streaming fashion, and to provide persistence and easy retrieval of the preprocessed data. A custom data collection pipeline for Polkadot is crucial if we are interested in the analysis of the whole Polkadot ecosystem, as blockchain nodes in general are only designed to facilitate data retrieval needed for functioning of the blockchain network.

Polkadot's consensus mechanism is a Nominated Proof-of-Stake (NPoS) [4], [5] where nominators stake their DOT's - a native currency of the Polkadot blockchain, with the validators which participate in consensus on their behalf. If a validator is elected into the active set they actively validate blocks and consequently earn rewards for themselves and possibly for the nominators backing them. The validator election for the active set occurs every era, which equals to one day [6]. The main motivating idea behind this design is to enhance security, efficiency and fair representation. Much of the staking dynamics data for Polkadot is stored off-chain and in a non-documented fashion, making the retrieval of the staking data more challenging. The nominators and validators are connected to each other by the nature of their staking relationship, but also on the level of Peer-to-Peer (P2P) network. While analysis of P2P network in similar blockchain systems such as Bitcoin is noteworthy [7], [8], in this paper we concentrate on the analysis of the transfer network between all users in the Polkadot, as defined by the transfer of DOT's between different accounts. Analyzing the network of transactions is useful to understand exogenous information, such as price formation [9].

The bow-tie model is a commonly used for describing a directed network structure in Web [10] and financial networks [11], and was recently applied to the Bitcoin transaction network as well [12]. It decomposes the network into several components which have different economic interpretations: 1) Strongly Connected Component (SCC) where all nodes are reachable between each other, 2) In-component where nodes can reach the SCC but not the other way around and, 3) Out-component where nodes are reachable from SCC but not the other way around. The interpretation of these components in the context of decentralized financial network is that the SCC represents tightly connected nodes which provide transactional services to the users, the In-component represents nodes which provide liquidity to these services and the Out-component represents end users which are only holding to their assets. In this paper we are particularly interested in the composition of these components with respect to the nominators and validators in the Polkadot staking ecosystem. The change in the size and nominator and validator composition of these components can suggest that economic properties of Polkadot ecosystem are changing. In general the change in the structure of transaction network, for example an observed increase of transaction network centralization as evidenced in many other blockchain systems [13]–[15], can provide more insight into the fundamental forces governing the evolution of large decentralized financial systems.

II. Methodology

A. Data collection pipeline

The elements that make up our data pipeline are divided into three main layers: 1) data collection, 2) data persistence and 3) query layer. The architecture is composed of the standard Apache Kafka and Spark pipeline shown in Fig. 1. The pipeline is designed to allow for trivial horizontal scaling

978-8-3503-1019-1/23/$31.00 ©2023 IEEE

if one requires better performance, although our current implementation works on a single server. We setup an archive node which stores the full history of the blockchain, which currently is approximately 12.5M blocks and takes up 620 GB of memory.

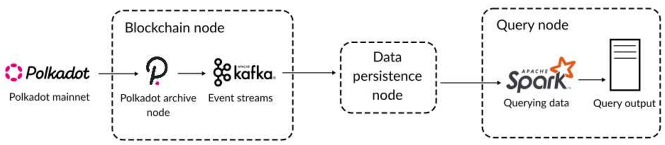

Fig. 1. Architecture of our Polkadot data collection pipeline

We use Apache Kafka - an open-source framework for streaming data processing, to get block data from the Polkadot node in a JSON compatible format using the Python Substrate Interface library[1]. The block data is then preprocessed using a custom data model and stored into a Postgres database. Data model includes entities which we want to easily retrieve for analysis, for example `Block`, `Extrinsic`, `Event`, `Account`, `Controller`, `Balance`, `Transfer`, `ValidatorPool`, `Validator` and `Nominator`. We also define a special concept `Aggregator` to speed up some common queries and which includes a block number and the total sum of extrinsics, events, accounts, transfers and stakes up that particular block. The preprocessed data can then be queried with tools such as PySpark[2] which enable horizontal scaling of computational resources.

B. Transfer network analysis

We use graph-tool [3] - a highly efficient Python library for network analysis, to calculate Strongly Connected Component (SCC) and In- and Out-components of the Polkadot transfer network (Fig. 2). For constructing the transfer network we use a `Transfer` extrinsic from Polkadot blockchain which records transfers of DOT's between two accounts on the relay chain. This is a directed network where two accounts are connected if one initiated a transfer to another.

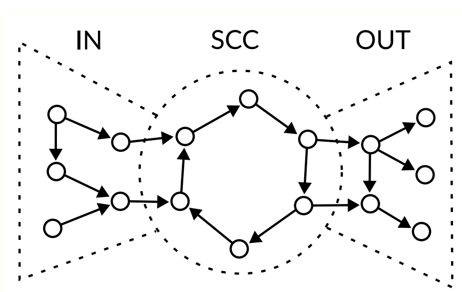

Fig. 2. Bow-tie model for a directed network. Nodes in the SCC component are all reachable between each other. Nodes in the In-component are only able to reach SCC but not the other way around, while those in the Out-component are reachable from SCC but not the other way around.

[1]https://github.com/polkascan/py-substrate-interface
[2]https://spark.apache.org/docs/latest/api/python/
[3]https://graph-tool.skewed.de/

Tab. I shows an example query output for the staking dynamics in Polkadot, with one entry connecting nominators with their respective validators and the amounts of their stakes and the staking rewards in each era. Fig. 3 shows an example of nominator/validator network visualization for era 80. Note that this is not a transfer network but rather a staking relationship network between nominators and validators! In the transfer network a nominator will be connected to a validator only if there is a `Transfer` even between them. We use a similar query to construct transfer networks between all accounts on a monthly basis since the beginning of Polkadot operation in 2020, both in cumulative fashion and in disjoint one month windows.

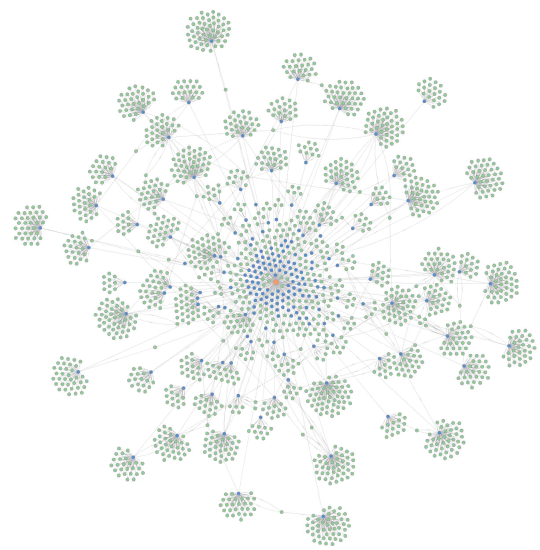

Fig. 3. Validator pool network of era 80. The central node represents the validator pool node - a representation of a pool of several hundred randomly selected validators for a given era. The central nodes of smaller clusters represent the actual validators and the nodes surrounding them represent the nominators that bond their DOT's with them.

For all transfer networks obtained in this way we calculate the absolute and relative sizes (as a percentage of the full transfer network size) of SCC and In- and Out-components in time, and the relative number of nominators and validators in each of the components in time. This allows us to observe the changes in the network structure over time with respect to the bow-tie model, similarly as observing changes in other structural properties of the transaction networks in other blockchains such as degree distribution, preferential attachment and centralization [13]–[15].

TABLE I
AN EXAMPLE QUERY OUTPUT ON THE STAKING DYNAMICS IN POLKADOT

id	nominator	validator	reward	destination	era	stake
1	13zPXt...	11uMPb...	280745...	Stash	1	190210...
2	14P6qJ...	11uMPb...	200732...	null	1	136000...
3	1563fH...	11uMPb...	368992...	null	1	250000...
4	15834Y...	11uMPb...	737985...	null	1	500000...

III. Results

In this section we outline the results of the analysis of Polkadot transfer network. Fig. 4 shows the cumulative sizes of the SCC and the In- and Out-components. We observe that while the absolute sizes of all components grow, the relative sizes (as compared to the full transfer network) of In- and Out-component are actually decreasing. On the other hand, the relative size of the SCC component keeps increasing, indicating that more and more accounts are transitioning to the SCC component. This suggests that the interconnectedness of the Polkadot transfer network is increasing as well, as more and more accounts become a member of the SCC component.

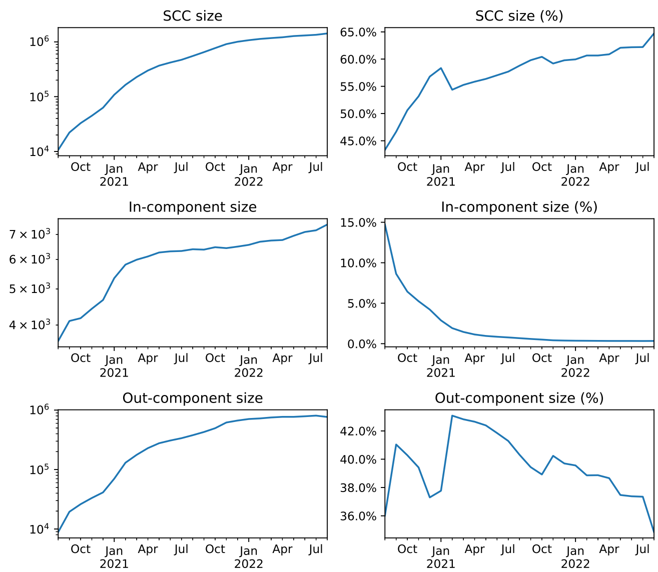

Fig. 4. Cumulative sizes of the SCC and the In- and Out-components. The absolute sizes of all components grow in time, as is the relative size (as compared to the full transfer network) of the SCC component. However, the relative sizes of In- and Out-component are decreasing, indicating that more and more accounts are transitioning to the SCC component.

Fig. 5 shows the cumulative number of nominators and validators in the Polkadot ecosystem. Again, we observe that while the absolute number of both nominators and validators increase, their relative numbers are decreasing. This indicates that the nominators and validators dominated early Polkadot ecosystem, while later the ecosystem becomes more dominated with common users not necessarily interested in staking but rather in using the services of the Polkadot ecosystem. A better confirmation of this would come from investigating the transfer networks of the Polkadot's parachains which host the decentralized applications (DApps) in the Polkadot ecosystem, unlike the Relay Chain which manages only DOT transfer between accounts.

Fig. 6 shows the cumulative relative number of nominators and validators in the SCC and In- and Out-components over time. We can observe that although the number of both nominators and validators increase in time (Fig. 5) their relative number in all of these components is decreasing. There are only occasional time-restricted exceptions such as relative

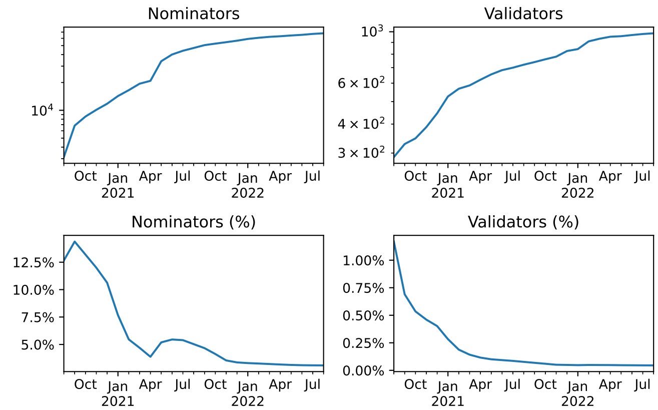

Fig. 5. Cumulative number of nominators and validators in the Polkadot ecosystem. While the absolute number of both nominators and validators increases their relative numbers are decreasing over time, meaning that more and more common users not primarily interested in staking are joining the Polkadot network.

sudden increase in the number of nominators in the period from April to October 2021. This corresponds to the period of a general crypto bull market where many new users entered the Polkadot ecosystem and joined as nominators, which takes less technical skill and also significantly less DOT's as compared to joining as a validator.

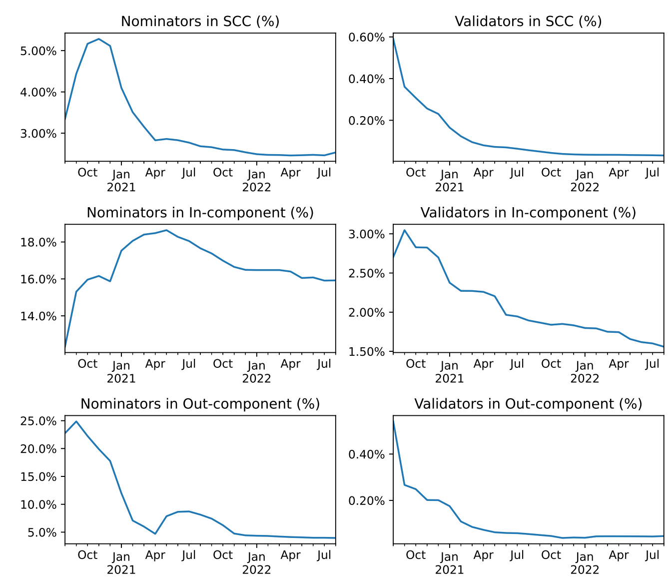

Fig. 6. Cumulative relative number of nominators and validators in the SCC and In- and Out-components over time. We observe that their fraction decreases over time, with occasional sudden increases - for example in time period from April and December 2021. This is the period when many new users joined the Polkadot ecosystem and bonded their newly acquired DOT's as nominators - a process which is much simpler than becoming a validator, which explains the lack of similar peak in the validator graph.

Fig. 7 shows the cumulative and monthly size of the whole Polkadot transfer network. Consistent with previous results we observe that there is a significant peak in transfer network size near the end of 2021 when many new users joined the Polkadot ecosystem and started bonding their DOT's as nominators. Fig. 7 also shows the relative sizes of the SCC and In- and Out-components in disjoint one-month windows. We observe that transfer network and individual components are decreasing in size throughout the whole of 2022. Only for the the In-component could be argued that it keeps relatively consistent size on the monthly basis, probably due to the consistent creation of new accounts. The relative drop on a monthly scale of SCC component is probably due to the increased activity in the Polkadot ecosystem as a whole.

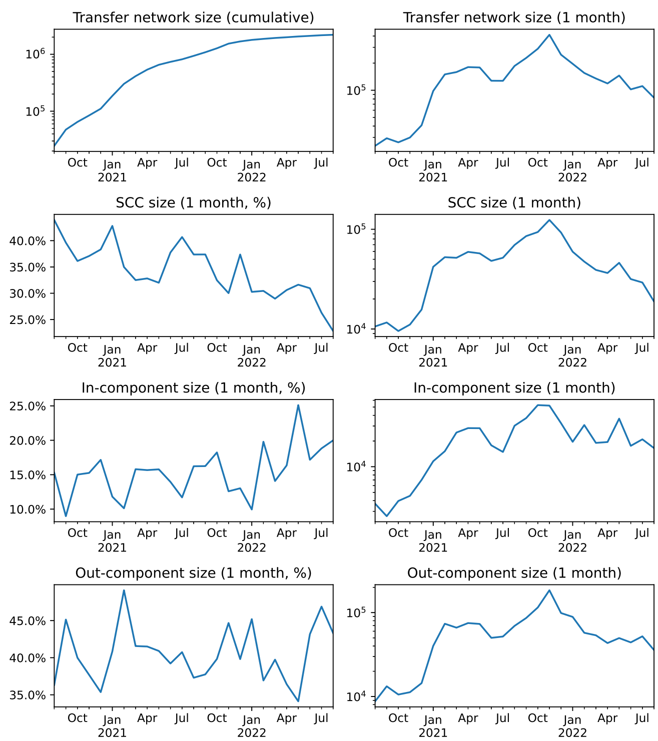

Fig. 7. Cumulative and monthly size of the whole Polkadot transfer network. In addition to this, it also shows the relative sizes of the SCC and In- and Out-components in disjoint one month windows. We can observe that the relative sizes of the transfer network and individual components are generally decreasing on a monthly basis throughout the 2022, with the sole exception of In-component which reflects the continuous influx of new users to the Polkadot ecosystem.

IV. Conclusion

Our data collection pipeline for Polkadot blockchain is designed to support easy storage and retrieval of staking and reward data, changes in balance of individual accounts, and a transfer network or validator and nominator network. In the future we plan to extend the data collection process to Polkadot parachains which power the decentralized applications in the Polkadot ecosystem, and to expand the data model so that it includes a richer set of concepts from the Polkadot ecosystem, for example Crowdloan and Democracy systems.

We use the collected data to perform bow-tie transfer network analysis over time by dividing it into three distinct components - the SCC, In- and Out-components. Each of these has an intuitive interpretation in economic terms. While SCC reflects the pool of highly active users and service providers, the In-component represents the liquidity providers and the Out-component the end users of the services which are mostly holding on to their DOT's. We also analyze the relative number of nominators and validators which participate in the Polkadot's staking ecosystem within each of these three components. Our general conclusion is that the Polkadot ecosystem is slowly maturing from a system dominated by nominators and validators, both of which require some technical skill as well as willingness to bond sufficient amount of funds, into a system increasingly populated by regular users using the financial services of Polkadot.

Blockchain systems are often criticized due to their lack of utility and real world use cases. Our bow-tie network structure analysis of the Polkadot transfer network suggests that this is an inevitable consequence of the technological system where early adoption is slow and dominated by experts and early investors, unlike regular users which are driven by utility.

References

[1] Gavin Wood, "Polkadot: Vision for a heterogeneous multi-chain framework", [Online], Available: https://polkadot.network/PolkaDotPaper.pdf, [Accessed: 8th December 2022]

[2] H. Abbas, M. Caprolu and R. Di Pietro, "Analysis of Polkadot: Architecture, Internals, and Contradictions," 2022 IEEE International Conference on Blockchain (Blockchain), 2022, pp. 61-70, doi: 10.1109/Blockchain55522.2022.00019.

[3] J. Burdges et. al., "Overview of Polkadot and its Design Considerations". arXiv preprint arXiv:2005.13456. (2020)

[4] C. Badertscher, P. Gaži, A. Kiayias, A. Russell, and V. Zikas, "Ouroboros genesis: Composable proof-of-stake blockchains with dynamic availability", Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, pp. 913-930

[5] F. Spychiger, P. Tasca, and C.J. Tessone, "Unveiling the Importance and Evolution of Design Components Through the "Tree of Blockchain"", Frontiers in Blockchains, vol. 3, pp. 613476, doi: 10.3389/fbloc.2020.613476

[6] "Nominated Proof-of-Stake", [Online], Available: https://research.web3.foundation/en/latest/polkadot/NPoS/index.html, [Accessed: 26th September 2022]

[7] M. Grundmann, M. Baumstark and H. Hartenstein, "On the Peer Degree Distribution of the Bitcoin P2P Network," 2022 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 2022, pp. 1-5, doi: 10.1109/ICBC54727.2022.9805511.

[8] M. Essaid, S. Park and H. Ju, "Visualising Bitcoin's Dynamic P2P Network Topoogy and Performance," 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 2019, pp. 141-145, doi: 10.1109/BLOC.2019.8751305.

[9] A. Bovet, C. Campajola, F. Mottes, V. Restocchi, N. Vallarano, T. Squartini, C.J. Tessone, "The evolving liaisons between the transaction networks of Bitcoin and its price dynamics", arXiv preprint arXiv:1907.03577 (2019)

[10] Y. Fujita, Y. Kichikawa, Y. Fujiwara, W. Souma and H. Iyetomi, "Local bow-tie structure of the web", Applied Network Science, 4:15 (2019)

[11] S. Vitali, J.B. Glattfelder and S. Battiston, "The network of global corporate control", PloS One, vol. 6, 10 (2011)

[12] D.D.F. Maesa, A. Marino and L. Ricci, "The bow tie structure of the Bitcoin users graph", Applied Network Science, vol. 4, 1 (2019)

[13] N. Vallarano, C.J. Tessone and T. Squartini. "Bitcoin Transaction Networks: an overview of recent results.", Frontiers in Physics 8 (2020): 286.

[14] F.M. De Collibus, M. Piškorec, A. Partida and C.J. Tessone, "The structural role of smart contracts and exchanges in the centralisation of ethereum-based cryptoassets". Entropy, 24(8), 1048. (2022)

[15] C. Campajola, R. Cristodaro, F.M. De Collibus, T. Yan, N. Vallarano and C.J. Tessone, "The evolution of centralisation on cryptocurrency platforms". arXiv preprint arXiv:2206.05081. (2022)