Reputation and Pagerank on the Ethereum Blockchain: Technical Report

Scott K. Imig
Octan Network, Hanoi, Vietnam.

Email: scott@scottonchain.xyz (S.K.I.)
Manuscript received November 30, 2023; accepted December 6, 2023; published December 13, 2023.
DOI: 10.18178/IJBTA.2023.1.2.89-94

Abstract: Motivated by the need for a credibly neutral reputation mechanism on the blockchain, we study Pagerank and other properties of two Ethereum addresses, considering one address as trustworthy and the other as problematic. We show that an unmodified Pagerank algorithm does not distinguish these cases, while asserting that Pagerank is nonetheless an important starting point for such a mechanism. We give several directions for extending or modifying the mechanism to cover such cases.

Keywords: credible neutrality, blockchain reputation, algorithmic reputation, ranking

1. Introduction

Reputation on the blockchain is an active area of research and development, promising significant advantages for both consumers and industry. At the current time, end users bear responsibility for evaluating the trustworthiness of entities and interactions. However, users have limited tools at their disposal to aid in this assessment. Furthermore, lack of a transparent reputation mechanism is an artificial constraint on innovation for decentralized applications such as lending. A standardized mechanism for evaluating reputation is therefore conducive to utility and innovation across users and industry.

As proposed by Armstrong [1], internet reputation algorithms, such as Pagerank, apply naturally to the digraph structure of blockchain transactions. Ongoing research into algorithmic reputation on the blockchain, such as [2], along with industry advancements in this direction [3] demonstrate the potential of this approach as a direct benefit to end users, developers, and industry.

2. Reputation, Uniswap, and Forsage

The definition of reputation on the blockchain is nuanced and sensitive to social, political, and philosophical factors. For example, a smart contract privacy layer might be considered trustworthy by one party and problematic by another.

Leaving aside the difficult general question, we focus on the Pagerank of two addresses with high traffic, considering the first as trustworthy, and the other as problematic.

We consider the address of the original Uniswap v3 SwapRouter,

0xe592427a0aece92de3edee1f18e0157c05861564,

as highly trustworthy. By contrast, we consider the address of the Forsage smart contract,

0x5acc84a3e955bdd76467d3348077d003f00ffb97,

as problematic. While this is uncontroversial according to many users and established standards, we provide a brief rationale.

2.1. Uniswap

Since 2018, Uniswap has been the leading decentralized trading protocol on Ethereum. It is cited as an example of a trustworthy decentralized exchange (dex) in academic literature (e.g. [4, 5]) and is frequently included in lists of trustworthy dex platforms. Moreover, Uniswap Labs maintains a physical presence in New York City, with its founder and representatives actively engaging in public collaboration and discourse within the cryptocurrency community and beyond.

2.2. Forsage

By contrast, the Forsage smart contract is frequently studied in the literature as example of a fraudulent onchain pyramid scheme [6–8]. Research has shown that 88% of participants have incurred losses in the scheme [6], and that Forsage's marketing and social media is often false or misleading. Forsage is flagged with an Etherscan warning, and individuals promoting Forsage have faced legal indictments for their involvement.

For the remainder of this note, we abbreviate on-chainthe addresses in question as "0xe59 Uniswap" and "0x5ac Forsage" or simply "Uniswap" and "Forsage" when the context is clear.

3. Pagerank

The Pagerank graph algorithm was famously proposed by Brin and Page in 1995 [9]. It is the foundation of the original Google static ranking algorithm. The algorithm applies generally to any directed graph, and has found applications to real-world scenarios well beyond internet ranking, for example, detecting vulnerabilities in electric power grids [10].

Blockchain addresses and transactions are naturally modeled with a digraph as in Fig. 1, with nodes representing addresses and edges representing transactions. For the purpose of this report, we use an unweighted digraph, treating multiple transactions between two addresses as a single edge, and ignoring any transactions from an address to itself.

The most common formulation of Pagerank is as an iterative algorithm:

1) For each node n, initialize the Pagerank value $PR_0(n)$ to

$$PR_0(n) = 1/N,$$

where N is the number of nodes in the graph.

2) At each iteration i, update the Pagerank $PR_i(n)$ associated with node n to:

$$PR_i(n) = \frac{1-d}{N} + d\sum_p \frac{PR_{i-1}(p)}{D(p)}$$

with the sum running over all predecessors p of n. Here, $D(p)$ represents the out-degree of p, and d is a damping factor, usually set to 0.85 for internet applications.

3) The final Pagerank $PR(n)$ is the limit of $PR_i(n)$.

In the context of internet ranking, the damping factor d is motivated as the probability that a random surfer continues clicking links from a given page. The analogous behavior in blockchain is less clear, so we simplify the Pagerank calculation by setting $d = 1$:

$$PR_i(n) = \sum_p \frac{PR_{i-1}(p)}{D(p)}.$$

In this formulation, the vector of Pageranks $PR(n)$ is the eigenvector of the graph's transition matrix corresponding to eigenvalue 1, which is the largest eigenvalue.

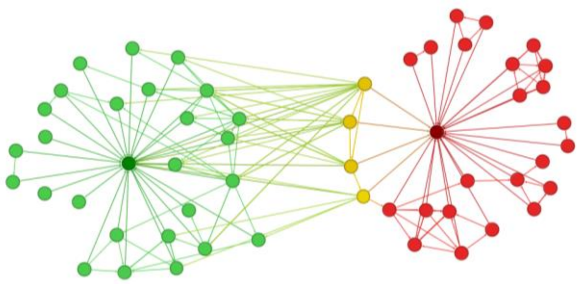

Fig. 1. Subgraph of predecessors of 0xe59 Uniswap (green) and 0x5ac Forsage (red), including transactions between predecessors.

3.1. 0xe59 Uniswap and 0x5ac Forsage

Table 1 shows highly ranked Ethereum addresses and their Pageranks[1]. Notably, the Pagerank of 0x5ac Forsage is much higher than that of 0xe59 Uniswap, placing it at rank 19 as compared with rank 60. This shows that Pagerank alone is not sufficient to distinguish a trustworthy address from a problematic address. Note that Uniswap Labs has many addresses, some ranked higher than 0xe59 or 0x5ac.

Table 1. Pagerank for Highly Ranked Ethereum Addresses as of September 2023

Rank	Address	Pagerank	Etherscan Description
1	0xa090e606e30bd747d4e6245a1517ebe430f0057e	0.051304	Coinbase: Miscellaneous
2	0xa0b86991c6218b36c1d19d4a2e9eb0ce3606eb48	0.019605	Circle: USD Coin
3	0xdac17f958d2ee523a2206206994597c13d831ec7	0.014651	Tether: USDT Stablecoin
...	...	...	...
18	0x74381d4533cc43121abfef7566010dd9fb7c9f7a	0.001643	*none*
19	0x5acc84a3e955bdd76467d3348077d003f00ffb97	0.001582	**Forsage.io**
20	0x4e5b2e1dc63f6b91cb6cd759936495434c7e972f	0.001547	FixedFloat
...	...	...	...
59	0xb04c0eb29c72cebc467b9d4944d29116fa02c44a	0.000684	*none*
60	0xe592427a0aece92de3edee1f18e0157c05861564	0.000682	**Uniswap V3:Router**
61	0x4945ce2d1b5bd904cac839b7fdabafd19cab982b	0.000680	Bitrefill: Payment Gateway
...	...	...	...

The in-degree of a vertex is closely related to Pagerank [11, 12]. We compare these rankings in Table 2. Uniswap is, by this metric, higher than Forsage, but they are very close, appearing at ranks 21 and 23. Again, ranking by in-degree does not differentiate reputation.

Table 2. Indegrees for popular Ethereum addresses, from Google BigQuery public Ethereum data, as of September 1, 2023. Multiple transactions between addresses are treated as a single edge in the digraph.

Rank	Address	Indegree	Etherscan Description
1	0xdac17f958d2ee523a2206206994597c13d831ec7	26,558,521	Tether: USDT Stablecoin
2	0xa090e606e30bd747d4e6245a1517ebe430f0057e	17,134,324	Coinbase: Miscellaneou
3	0xa0b86991c6218b36c1d19d4a2e9eb0ce3606eb48	8,677,496	Circle: USD Coin
...	...	...	...
20	0x7be8076f4ea4a4ad08075c2508e481d6c946d12b	1,252,400	Opensea: Wyvern Exchange v1
21	0xe592427a0aece92de3edee1f18e0157c05861564	1,220,302	**Uniswap V3: Router**
22	0xabea9132b05a70803a4e85094fd0e1800777fbef	1,105,086	Zksync
23	0x5acc84a3e955bdd76467d3348077d003f00ffb97	1,065,472	**Forsage.io**
24	0xcad621da75a66c7a8f4ff86d30a2bf981bfc8fdd	1,061,988	Kucoin 10
...	...	...	...

This result shows that an unmodified Pagerank algorithm does not differentiate trustworthy addresses, and that the algorithm would need adjustment for the new domain. Internet search engines make similar adjustments to modify and extend Pagerank, mitigating manipulation and improving relevance. One approach is to use Pagerank as an input to an ensemble algorithm, deriving a score from a variety of features and models. Another approach is to modify the Pagerank score itself, for example, by zeroing out the score for contracts that meet certain criteria, modifying the transaction graph, or evolving the algorithm. In the next sections, we give some preliminary directions for such future research.

3.2. Time-Based Features

The addresses in question exhibit important differences, as measured by traffic after deployment. Referring to Fig. 2, the 0x5ac Forsage address took several months to acquire traffic, then spiked very quickly before traffic fell off. In Fig. 3, by contrast, 0xe59 Uniswap had immediate substantial traffic after deployment

[1] The transaction graph uses all transactions on Ethereum mainnet, for all time, as of September 17, 2023. Sum of squared error convergence < 4×10^{-5}

in April of 2021. Address 0xe59 Uniswap also has a larger right-hand tail than 0x5ac Forsage, but early traffic patterns are more useful in practice than later traffic patterns.

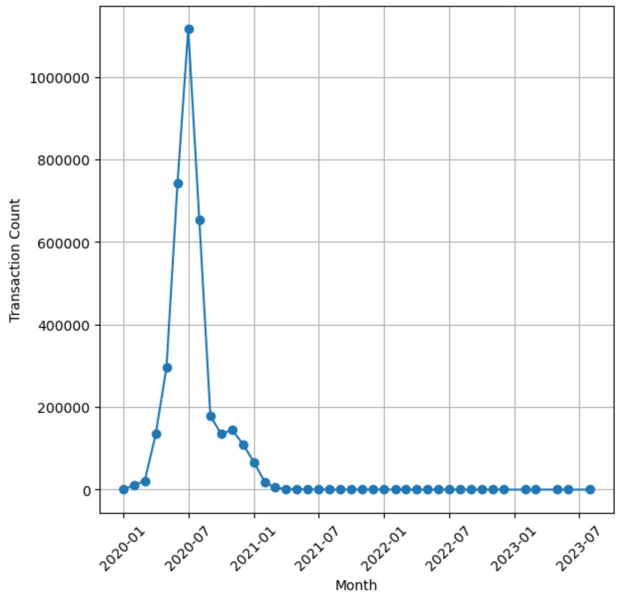

Fig. 2. 0x5ac Forsage.

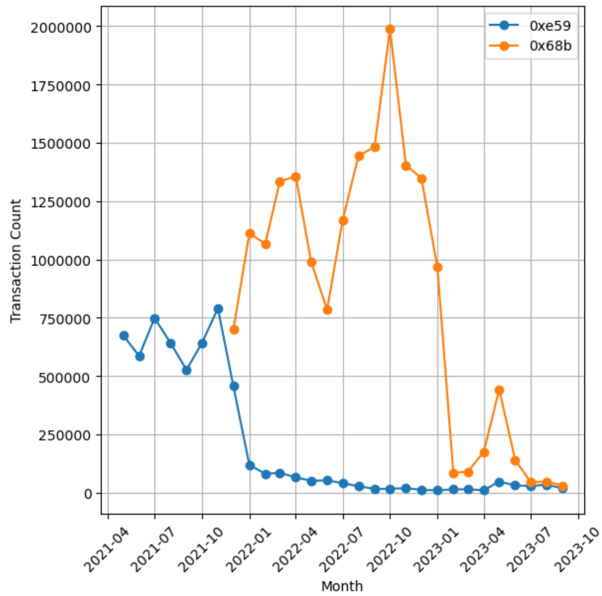

Fig. 3. 0xe59 Uniswap and 0x68b Uniswap.

Note that both addresses show a quick drop in traffic. For the Uniswap address, this is because traffic transferred to an updated Uniswap SwapRouter[2], which was deployed in December 2021. This is visible in the graph as the line starting in December 2021.

A useful time-based model, which has gained currency in popular discussion, is the Lindy Effect [13]. In situations where the Lindy Effect applies, expected future lifetime is proportional to current age. Mathematically, the model for the Lindy Effect is

$$E[T - t_0 | T >= t_0] \propto t_0,$$

where T is a random variable representing total lifetime and t_0 is current age. This formulation makes the Lindy Effect measurable and trainable. A mathematical model stemming from this idea, with a threshold for

[2] 0x68b3465833fb72a70ecdf485e0e4c7bd8665fc45

transaction volume over time, can differentiate time-based differences such as those shown in Figs 2 and 3.

Such a time-based score can augment Pagerank in different ways, for example, by modifying the initial value in step 1 of the Pagerank algorithm, or as an additional input to an ensemble model.

3.3. Contract-Based Features

The blockchain digraph includes rich information beyond transactions. Specifically, the reputation of a contract's deployer is related to the reputation of the contract itself. The Uniswap contract is deployed by the same address as other Uniswap contracts, and similarly with Forsage.

A naive approach for applying the deployer-contract relationship to Pagerank is to allow the deployer's score to accrue to the deployed contract by introducing an artificial edge. However, in this example, the deployers' scores are extremely low: 1.1×10^{-8} for the Uniswap deployer and 1.5×10^{-9} for the Forsage deployer, putting them at ranks 6,883,353 and 26,352,887 respectively. These deployers have very low in-degree, and the most naive approach does not improve the ranking of the contracts.

The deployer-contract relationship also lends itself to less naive approaches. For example, a graph diffusion algorithm on the undirected deployment graph would distribute reputation between deployed contracts with a common deployer.

Beyond the deployer, the contract has many features available for an algorithmic reputation computation. One rich source of information is the code of the contract itself, which can be audited by human or artificial intelligence. Another is the contract's output, including all internal transactions.

3.4. Social Mechanisms and Credible Neutrality

Ethereum has an active social layer, including individual end users, developers, and institutions. The social layer already provides certain reputation mechanisms, including blocklists and warnings. The biggest challenge to widely invoking a social layer for reputation scoring is that of maintaining credible neutrality. Quoting from [14]:

Essentially, a mechanism is credibly neutral if, just by looking at the mechanism's design, it is easy to see that the mechanism does not discriminate for or against any specific people.

As mentioned, the definition of reputation is nuanced and controversial. Any centralized entity has to navigate this nuance, making credible neutrality impossible.

That said, less centralized social mechanisms do lend themselves to credible neutrality. One example is collaborative filtering by regularized singular value decomposition [15, 16]. This "Netflix algorithm" for assigning star scores to unseen content is a social mechanism which meets the criterion for credible neutrality. It shows that social mechanisms can be credibly neutral, while providing a possible starting point for the blockchain.

4. Conclusion

Modern web ranking algorithms modify Pagerank substantially to mitigate manipulation, personalize results to the user, and otherwise improve on relevance. In the same manner, reputation scoring on the blockchain requires an evolution from raw Pagerank. Some of the desirable properties of this evolved mechanism will be:

1) The mechanism is credibly neutral.
2) The output of the mechanism agrees with the least controversial cases, such as the addresses considered in this report.
3) The mechanism is resistant to manipulation.

The evolution of algorithmic reputation on blockchain may require a marketplace of alternatives, similar to the way Ethereum Layer 2 solutions have evolved. Any final mechanism is likely to involve algorithmic or credibly neutral social mechanisms.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgments

The author would like to extend sincere gratitude to Dr. Thuat Do, for discussions on the topic of blockchain

reputation which have significantly impacted this work.

All data is from the Ethereum public dataset maintained by Google.

References

[1] Armstrong, "Request for builders: Startups I would build today," *Coinbase Blog*, 2023.
[2] D. Do, "PageRank and HodgeRank on ethereum transactions: A measure for social credit," *International Journal of Software Innovation*, vol. 11, no. 1, 2023.
[3] O. Network, "What is reputation ranking system and how does it work?" *Octan Network Blog*, 2023
[4] H. Z. Han, "Trust in DeFi: An empirical study of the decentralized exchange," *Social Science Research Network*, 2022.
[5] S. Muruganantham, "Blockchain-based application: Decentralized financial technologies for exchanging crypto currency," *International Conference on Advances in Computing, Communication and Applied Informatics*, 2022.
[6] Kell, Yosaf, Allen, Meiklejohn, and Juels, "Forsage: Anatomy of a smart contract pyramid scheme," *Financial Cryptography and Data Security*, 2023.
[7] Bandara, Blockrath, Hobeck, Klinkmüller, Pufahl, Rebesky, van der Aalst, Weber "Event logs of ethereum-based applications," 2021.
[8] Fan, Fu, Xu, Cheng, "Al-SPSD: Anti-leakage smart Ponzi schemes detection in blockchain," *Information Processing and Management*, 2021.
[9] Brin, Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 1995.
[10] Ma, Shen, Liu, Mei, "Fast screening of vulnerable transmission lines in power grids: A pagerank-based approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, 2017.
[11] Litvak, Scheinhardt, Volkovich, "In-degree and PageRank: Why do they follow similar power laws?, *Internet Mathematics*, vol. 4, no. 3-4, 2007.
[12] Fortunato, Boguñá, Flammini, Menczer, "Approximating PageRank from in-degree," *International Workshop on Algorithms and Models for the Web-Graph*, 2006.
[13] Eliazar, "Lindy's law," *Physica A: Statistical Mechanics and its Applications*, vol. 485, 2017.
[14] Buterin, "Credible neutrality as a guiding principle," Nakamoto.com, 2020.
[15] Funk, "Netflix Update: Try This at Home", *Blog Post*, 2006.
[16] Koren, Bell, Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Computer*, 2009.