Trustworthiness and Pagerank on the Ethereum Blockchain: Technical Report

Scott Imig
scott@scottonchain.xyz
September 21, 2023

Abstract

Motivated by the need for a credibly neutral trust mechanism on the blockchain, we study Pagerank and other properties of two Ethereum addresses, considering one address is 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.

1   Introduction

Trust and 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 trustworthiness of entities and interactions. However, users have limited tools at their disposal to aid in this assessment. Furthermore, lack of a transparent trust mechanism is an artificial constraint on innovation for decentralized applications such as lending. A standardized mechanism for evaluating trustworthiness is therefore conducive to utility and innovation across users and industry.

As proposed by Armstrong ([AR], Summer 2023), 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 [DD], along with industry advancements in this direction ([OC]) demonstrate the potential of this approach as a direct benefit to end users, developers, and industry.

2   Trustworthiness, Uniswap, and Forsage

The definition of trustworthiness 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. [HH, SM]) 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.

2.1.1   Forsage

By contrast, the Forsage smart contract is frequently studied in the literature as example of a fraudulent onchain pyramid scheme ([KY], [BB], [FF]). Research

has shown that 88% of participants have incurred losses in the scheme [KY], 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 the 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 [BP]. 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 ([MS]).

Blockchain addresses and transactions are naturally modeled with a digraph, 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. Pagerank applies to this digraph in the obvious way.

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

i. 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.

ii. 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.

iii. 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.

3.1 0xe59 Uniswap and 0x5ac Forsage

Table 1 shows highly ranked Ethereum addresses and their Pageranks. 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, such as the address at Rank 11 in Table 1.

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

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.

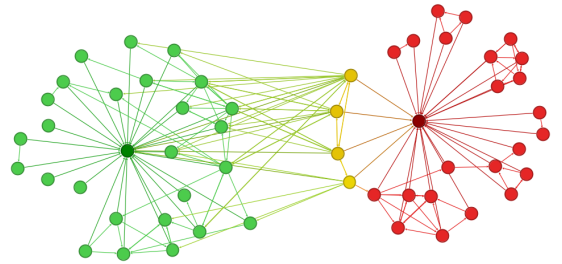In the next sections, we give some preliminary directions for such future research.

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

---

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

Rank	Address	Pagerank	Description
1	0xa090e606e30bd747d4e6245a1517ebe430f0057e	0.051304	Coinbase: Miscellaneous
2	0xa0b86991c6218b36c1d19d4a2e9eb0ce3606eb48	0.019605	Circle: USD Coin
3	0xdac17f958d2ee523a2206206994597c13d831ec7	0.014651	Tether: USDT Stablecoin
. . .	. . .	. . .	. . .
11	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
. . .	. . .	. . .	. . .

Table 1: Pagerank for highly ranked Ethereum addresses as of September 2023[1]

3.2 Time-based Features

The addresses in question exhibit important differences in as measured by traffic after deployment. Referring to Figure 2, the 0x5ac Forsage address took several months to acquire traffic, then spiked very quickly before traffic fell off. In Figure 3, by contrast, 0xe59 Uniswap had immediate substantial traffic after deployment in April of 2021. Address 0xe59 Uniswap also has a more significant right-hand tail than 0x5ac Forsage, but for obvious reasons early traffic patterns are more useful in practice than later traffic patterns.

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* [EL]. 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 transaction volume over time, can differentiate time-based differences such as those shown in Figures 2 and 3.

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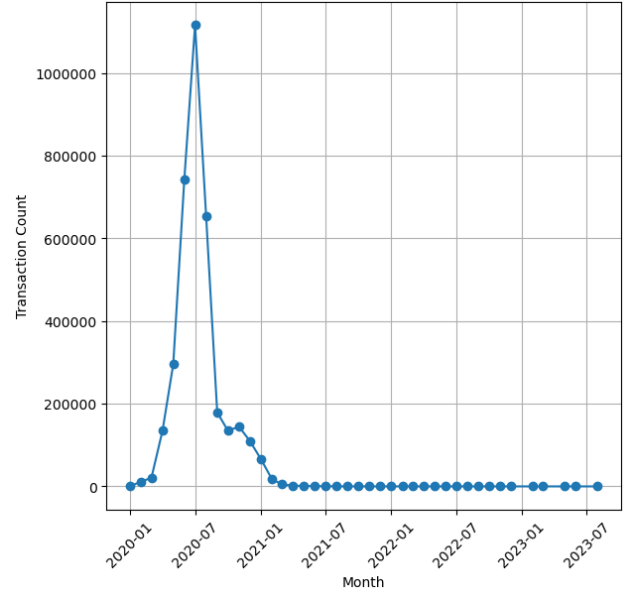

Figure 2: 0x5ac Forsage

3.3 Contract-based Features

The blockchain digraph includes rich information beyond transactions. Specifically, the trustworthiness of a contract's *deployer* is related to the trustworthiness of the contract itself. The Uniswap contract is deployed by the same address[3] as other Uniswap contracts, and similarly with Forsage.

A naive approach for applying the deployer-contract relationship to Pagerank is 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 For-

---

[2] 0x68b3465833fb72a70ecdf485e0e4c7bd8665fc45

[3] 0x6c9fc64a53c1b71fb3f9af64d1ae3a4931a5f4e9

Rank	Address	Indegree	Description
1	0xdac17f958d2ee523a2206206994597c13d831ec7	26,558,521	Tether: USDT Stablecoin
2	0xa090e606e30bd747d4e6245a1517ebe430f0057e	17,134,324	Coinbase: Miscellaneous
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
...	...	...	...

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.
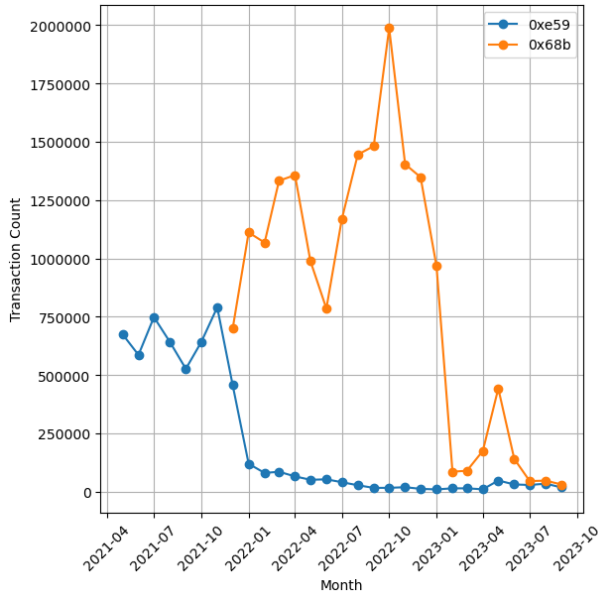

Figure 3: 0xe59 Uniswap and 0x68b Uniswap

sage 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 doesn't improve the ranking of the contracts.

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

Beyond the deployer-contract relationship, the contract has many features available for an algorithmic trust 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 trust mechanisms, including blocklists and warnings. The biggest challenge to widely invoking a social layer for trust scoring is that of maintaining credible neutrality. Quoting from [BU]:

> 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 trustworthiness 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 (regularized SVD, [FU, KB]). This "Netflix algorithm" for assigning star scores to unseen content is a social mechanism and algorithm which meets the criterion for credible neutrality with broad applications in recommender systems and beyond. This example shows that social mechanisms can be credibly neutral, while providing a starting point for a credibly neutral social mechanism for trustworthiness for 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, trustworthiness scoring on the blockchain requires an evolution from raw Pagerank.

Some of the desirable properties of this evolved mechanism will be:

i. The mechanism is credibly neutral.

ii. The output of the mechanism agrees with the least controversial cases, such as the addresses considered in this report.

iii. The mechanism is resistant to manipulation.

The evolution of algorithmic trustworthiness 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.

4.0.1 Acknowledgements

The author would like to extend sincere gratitude to Thuat Do, author of [DD], 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

[AR]   Armstrong, "Request for Builders: Startups I Would Build Today", Coinbase Blog, 2023

[BB]   Bandara, Blockrath, Hobeck, Klinkmüller, Pufahl, Rebesky, van der Aalst, Weber "Event Logs of Ethereum-Based Applications", 2021

[BP]   Brin, Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", 1995

[BU]   Buterin, "Credible Neutrality As A Guiding Principle", Nakamoto.com, 2020

[DN]   Do, Nguyen, Pham, "Delegated Proof of Reputation: a Novel Blockchain Consensus", Proceedings of the 1st International Electronics Communication Conference, 2019

[DD]   Do, Do, "PageRank and HodgeRank on Ethereum Transactions: A Measure for Social Credit", International Journal of Software Innovation, v.11 n.1, 2023

[EL]   Eliazar , "Lindy's Law", Physica A: Statistical Mechanics and its Applications, v.485, 2017

[FB]   "Approximating PageRank from In-Degree", International Workshop on Algorithms and Models for the Web-Graph, 2006

[FF]   Fan, Fu, Xu, Cheng, "Al-SPSD: Anti-leakage smart Ponzi schemes detection in blockchain", Information Processing and Management, 2021

[FU]   Funk, "Netflix Update: Try this at home", Blog post, 2006

[HH]   Han, Huong, Zhong, "Trust in defi: an empirical study of the decentralized exchange", Social Science Research Network, 2022

[KB]   Koren, Bell, Volinsky, "Matrix Factorization Techniques for Recommender Systems", IEEE Computer, 2009

[KY]   Kell, Yosaf, Allen, Meiklejohn, Juels, "Forsage: Anatomy of a Smart-Contract Pyramid Scheme", Financial Cryptography and Data Security, 2023

[LS]   Litvak, Scheinhardt, Volkovich, "In-Degree and PageRank: Why Do They Follow Similar Power Laws?", Internet Mathematics v.4 n.3-4, 2007

[MI]   Mitra, "How can we enhance reputation in Blockchain consensus for Industry 4.0–A proposed approach by extending the PageRank algorithm", International Journal of Information Management Data Insights, v.2 n.2, 2022

[MS]   Ma, Shen, Liu, Mei, "Fast Screening of Vulnerable Transmission Lines in Power Grids: A PageRank-Based Approach", IEEE Transactions on Smart Grid, v.10 n.1, 2017

[OC]   Octan Network, "What is Reputation Ranking System and How Does It Work?", Octan Network Blog, 2023

[SM]   Shamili, Muruganantham, "Blockchain based Application: Decentralized Financial Technologies for Exchanging Crypto Currency", International Conference on Advances in Computing, Communication and Applied Informatics, 2022