

Ethereum Data Analysis Literature Review

Sanjif Rajaratnam

I. INTRODUCTION

The purpose of this research project is to analyze the data from the Ethereum blockchain to find interesting connections. Ethereum, like Bitcoin, has a blockchain-based distributing platform at its core. However, Ethereum builds on the Bitcoin technology by including “smart” contracts. Ethereum uses a cryptocurrency known as “ether” that can be used to pay for contracts that are ran on the network based on their computation cost (“gas”), or as virtual currency. Miners can run the computations and verify transactions to create and add new blocks to the network in exchange for gas.

There are currently three potential research questions proposed based on research: (1) Analyzing the computation cost (“gas”) of contracts on the network, and the value of gas to the miner, (2) Analyzing the relationship between the network structure of Ethereum and market exchange rates, and (3) Determining types of users: e.g. miners, vendors, consumers, etc.

II. PAPERS SEARCH METHODOLOGY

The first set of papers were found by searching the IEEE Xplore Database, AMINER, and the ACM Digital Library for a set of keywords related to Ethereum and Bitcoin data analysis. Ethereum is fairly new (established in the summer of 2015) so it had very few papers associated with it. There were much more papers related to Bitcoins. The search was then broadened to include non-cryptocurrency related search terms centered around financial transaction analysis, time-series analysis, graph

search, and ledger analysis to get more papers.

Then the related works and references of the provided papers and papers found from the databases above were parsed to find more relevant papers.

III. RELATIONSHIP TO PROJECT

The papers were then analyzed to see if they could be used with the provided dataset. Non-compatible papers were still acceptable if they provided a means of getting further data.

The papers were then organized into their related research question set.

IV. BIBLIOGRAPHY

A. *Analyzing Computation Cost of Contracts*

The research question presented is to analyze the value of gas to miners given the computation cost of a transaction. This is addressed in O’Dwyer and Malone [1]. O’Dwyer and Malone [1] set out to determine if it is worth mining Bitcoins by analyzing the energy consumption associated with Bitcoin mining. They start by describing the bitcoin mining process, and the different levels of difficulties and rewards. They compare the energy usage to the exchange rate and find the energy consumption of the Bitcoin mining network. This analysis can be transferred to Ethereum quite easily because of Ethereum’s concept of gas which represents computational cost for running scripts. This type of analysis may require the historical exchange rate of Ethereum to be fully completed.

B. Analyzing the Relationship Between the Network Structure and Exchange Rate

The research question presented here is to analyze the relationship between Ethereum's network structure and its exchange rate. Kondor et al. [2] do this same analysis with Bitcoins. Kondor et al. [3] set out to understand the time series of prices of goods and assets by modeling the underlying system of interacting agents. They do this by analyzing Bitcoin's complete list of transactions. They use Principal Component Analysis at different time instances of the block chain to show that structural changes in the network is usually accompanied by changes in the Bitcoin exchange rate. They show how they get and clean their data and how they detected structural changes. Then they showed the result and their analysis. This type of analysis will require the historical exchange rate of Ethereum to be done.

C. Determining Types of Users

The research question presented here is finding types of Ethereum users from their transactional histories. This would involve finding itemsets of users through from and to account ids in transactions. This would also involve clustering users into different sets depending on the itemsets found. This is addressed in quite a few papers in the papers but these papers are presented as being the most relevant.

I) Clustering

Yang et al. [3] set out to attack the problem of categorical data clustering with transactional data with their CLOPE algorithm. Their goal was to efficiently group together similar transactions. They provide pseudo-code and show implementation of their clustering algorithm. They also then test their algorithm on various datasets to prove its effectiveness.

II) Finding Itemsets

Burdick et al. [4] set out to mine maximal frequent itemsets from a transactional database with their MAFIA algorithm. They assume that the entire database will fit into main memory. They discuss related algorithms and how there is better. Their algorithm uses a depth-first tree method to traverse the database. They provide pseudo-code and

different method for pruning the tree. They also show experimental results with real database and compared their results with other existing algorithms.

Han et al. [5] present a frequent pattern tree (FP-tree) structure and FP-tree based mining method. They specify how their tree is designed and constructed. They provide the algorithm and the observations required to construct their FP-tree. Then they specify how one can mine frequent patterns with their tree. Finally, they analyze the performance of their method.

Finally, Yen and Chen [6] presents a strategy for mining association rules to discover large sets that appear together in a sufficient number of transactions.

V. REFERENCES

- [1] O'dwyer, K.j., and D. Malone. "Bitcoin Mining and Its Energy Footprint." 25th IET Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communities Technologies (ISSC 2014/CICT 2014), 2014, doi:10.1049/cp.2014.0699.
- [2] Kondor, Daniel, et al. "Inferring the Interplay between Network Structure and Market Effects in Bitcoin." *New Journal of Physics*, vol. 16, no. 12, Feb. 2014, p. 125003., doi:10.1088/1367-2630/16/12/125003
- [3] Yang, Yiling, Xudong Guan, and Jinyuan You. "Clope: A Fast and Effective Clustering Algorithm for Transactional Data." *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '02* (2002): n. pag. Web.
- [4] Burdick, D., M. Calimlim, and J. Gehrke. "MAFIA: A Maximal Frequent Itemset Algorithm for Transactional Databases." *Proceedings 17th International Conference on Data Engineering* (n.d.): n. pag. Web.
- [5] Han, Jiawei, Jian Pei, and Yiwen Yin. "Mining Frequent Patterns without Candidate Generation."

Proceedings of the 2000 ACM SIGMOD
International Conference on Management of Data -
SIGMOD '00 (2000): n. pag. Web.

[6] Yen, Show-Jane, and A.l.p. Chen. "A Graph-Based Approach for Discovering Various Types of Association Rules." IEEE Transactions on Knowledge and Data Engineering, vol. 13, no. 5, 2001, pp. 839–845., doi:10.1109/69.956106.