Project Presentation

Alessandro Lombardi

Artificial Intelligence - University of Bologna

08-05-2020

Outline

Introduction

The idea

Analyze Ethereum Blockchain as a network of transactions between addresses.

The Ethereum Blockchain

Some stats to get an idea of the involved numbers

- latest **block** height value higher than 10 million
- more than 96 million unique **addresses**
- more than 698 million **transactions**
- Ethereum blocks are mined every 20 seconds

Source: `https://etherscan.io/`


Ethereum Daily Transactions Chart
Source: Etherscan.io
Click and drag in the plot area to zoom in
1 500 k
1 250 k
1 000 k
750 k
500 k
250 k
0
Transactions per Day
2016
2017
2018
2019
2020

Addresses

In Ethereum and Solidity, an address is a 20 byte (160 bits or 40 hex characters) alphanumeric string. It corresponds to the last 20 bytes of the Keccak-256 hash of the public key.

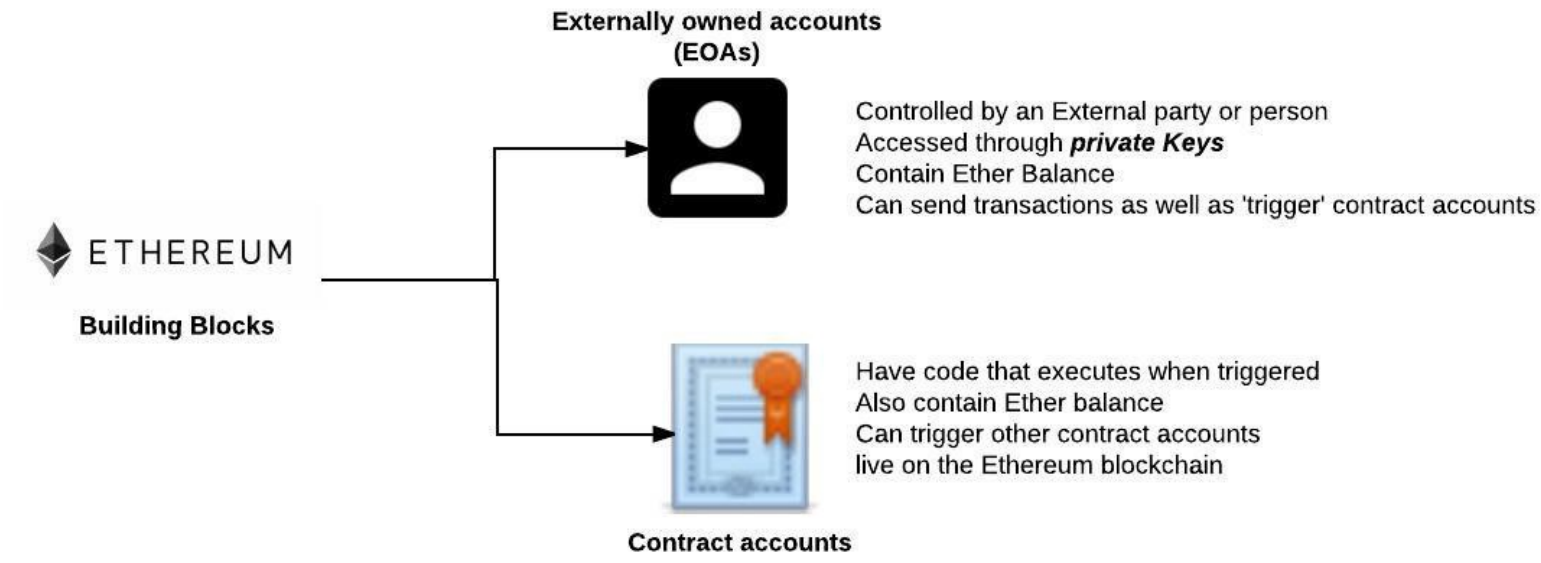

Figure 2: Source: Abhishek Chakravarty, Hackernoon.com

Transactions

While transactions are used for different purposes, the transaction structure is the same. The amount of fund is expressed in wei (1 ether $= 10^{18}$ weis). Contracts can receive transfers just like externally controlled accounts, but they can also receive more complicated transactions that actually run parts of their code and update their state.

Types

- **Fund Transfer Between EOA** used when an EOA is transferring fund to another EOA.
- **Deploy a Contract on Ethereum Network** used to deploy a compiled contract
- **Execute a Function on a Deployed Contract**

Clustering coefficients

Undirected graph $G = (V, E)$ without self loops and multiple edges.

Transitivity

The fraction of all possible triangles (closed triplets) present in G. Possible triangles are identified by the number of "triads" (two edges with a shared vertex, open and closed triplets): $3\frac{|\delta(G)|}{|\tau(G)|}$

Global clustering coefficient

The sum of fractions of number of triangles over number of open or closed triples passing on each vertex with degree greater than 2: $\frac{1}{|V_2|}\sum_{i \in V_2} \frac{|\delta(v_i)|}{|\tau(v_i)|}$

Local clustering coefficient

Quantifies how close the neighbours of a vertex in a graph are to being a clique (complete graph): $\frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{|N_i|(|N_i| - 1)}$

Where E is the set of edges and V the set of vertices. N_i is the set of neighbors of the node $v_i \in V$. V_2 is the set $\{v_i \in V : |N_i| \geq 2\}$. The set of triangles is expressed with the function $\delta(x)$ and the set of open and closed triplets with $\tau(x)$.

Modularity

Measure the strength of division of a network into **modules (also called groups, clusters or communities)**. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.

$\sum_{i=1}^{c}(e_{ii} - a_i^2)$, where e_{ij} is the fraction of edges in the community i to community j and a_i is $\sum_j e_{ij}$

Fraction of edges that fall within communities, minus the expected value of the same quantity if edges fall at random without regard for the community structure, which is the total degree of the cluster.

Community Detection

Two different concepts

- Clustering
- Partitioning

Clustering methods

- **Label Propagation**
- Louvain Method
- Minimum Cut Three
- Dynamic Clustering

Results

Clustering Coefficients

The measured global coefficient is between 0.02 and 0.03 and almost the 95% of vertices have a local cluster coefficient equal to 0. That may depends also on the way the data is collected and distributed temporally and spatially in the Blockchain the network seems to be very sparse.

Modularity

The graph could be divided in many connected components, in a network of more than 10 thousands vertices there are almost 1 thousands groups. The biggest connected component includes almost half of the vertices and the modularity of its clustering in almost 700 communities is near 0 (from -0.5 to 1).

Community detection

Figure 3: The greatest connected component rendered on Gephi

Further works

Deanonymize the network

Scraping from the web more features to associate to the addresses, (forums, datasets etc..)

Cluster users

Apply Machine Learning/Data Mining techniques to cluster addresses based on activity (Miners, ICO, etc..) in addition to the study of the topology structure

References

- Transactions in Ethereum, KC Tam, Medium
- Transitivity, Massimo Franceschet, University of Udine
- Graph Partitioning and Graph Clustering in Theory and Practice, Christian Schulz, Institute for Theoretical Informatics Karlsruhe Institute of Technology