Visualising Bitcoin's Dynamic P2P Network Topology and Performance

Meryam Essaid
Department of Computer Engineering
Keimyung University
Daegu, Republic of korea
meryamesd@stu.kmu.ac.kr

Sejin Park
Department of Computer Engineering
Keimyung University
Daegu, Republic of korea
baksejin@kmu.ac.kr

Hongteak Ju
Department of Computer Engineering
Keimyung University
Daegu, Republic of korea
juht@kmu.ac.kr

***Abstract*— In massive, dynamic and distributed P2P networks like Bitcoin, where thousands of updates occur per second, it is hard to obtain an accurate topology representing the structure of the network as a graph with nodes and links by using the traditional local measurement approaches based on batches, offline data, or on the discovery of the topology around a small set of nodes and then combine them to discover an approximate network topology. All of which present some limitation when applying them on blockchain-based networks. In this paper, we propose a topology discovery system, which performs a real-time data collection and analysis for Bitcoin P2P links with the use of a customized version of the Page-Rank algorithm that assembles incoming nodes information for deeper graph analysis processing. The topology discovery system allows us to gain knowledge on the Bitcoin network size, the network stability in term of well-connected Bitcoin nodes, as well as some data regarding the Bitcoin nodes geolocation.**

***Index Terms*—Bitcoin, network Topology, PageRank Algorithm, Nodes geolocation**

I. Introduction

Bitcoin has recently attracted significant attention owing to its fundamental nature of combining cryptographic technologies [1, 2], digital monetary systems [3], and blockchain technology [4, 5]. Bitcoin operates to distribute the ledger among all the participants in a flooding P2P network. Bitcoin neither has nor requires a central server or authority. Users running the Bitcoin communicate directly with each other in one of the largest P2P networks over the TCP. Through this distributed network, the Bitcoin protocol is specified by the behavior of its clients. When a node broadcasts a transaction or block, it selects 8 nodes randomly from its local addrMan list, a list that contains data of active peers in the network. After the peers' selection, the transaction/block is multicast to them. These 8 nodes, in turn, will validate the transaction/block and each node will forward it to another 8 nodes and so on till every node in the network has a copy of the transaction/block. In similar unstructured network, resource discovery is not very efficient and may cause severe performance and security issues [6].

Therefore, the knowledge of the network topology is useful and crucial to examine the network robustness and can be employed in several areas such as nodes location [7], management [8, 9], routing [10], and diagnosis [11]. Diverse strategies have been suggested to automatically picture network topology, containing nodes and communication links description [12, 13]. Those strategies are usually used in centralized networks, with the description being created by a single node, traditional strategies either for centralized or distributed platforms assume that the topology stays at least for a finite period unchanged.

During the last decade, few approaches have examined the Bitcoin network. Decker et al. [14] measured the rate of information propagation throughout the Bitcoin network. Babaioff et al. [15] pointed out that peers have encouragement not to participate in the network by not broadcasting events, but this could be solved by paying fees to relay peers. Biryukov et al. [16] disclose a technique that also uses Bitcoin address propagation messages to detect peer links. Their technique, however, is somewhat invasive since it involves polluting each node's table of potential peers with fake entries. Andrew Miller et al. [17] studied the underlying network topology by gathering and analyzing information that nodes readily provide. They have introduced AddressProbe, a technique that discovers peer-to-peer links in Bitcoin and applies this to the main network. To support AddressProbe and other tools, they have developed CoinScope, an infrastructure to manage short, but large-scale experiments in Bitcoin. They also analyzed the measured topology to discover both high degree nodes and a well-connected giant component.

However, in Bitcoin network, the topology is dynamic, with nodes and communication links being added/removed continuously. In such a decentralized and dynamic network, instead of a task which eventually reaches an end, a permanent topology discovery process is needed. Unlike previous approaches, In this research, we have developed a real-time Bitcoin-based topology discovery system consists of three main components; a bridge nodes enabling the communication with the Bitcoin network, a discovery client, that collected the Bitcoin P2P network features, and topology analysis server which decided as to whether process and rank a node. The system can determine in real-time which nodes require deeper graph analysis due to the use of a customized Page-rank algorithm [18]; the system reaches the real-time scaling by the split of the analysis mechanism to several layers, each of the used layers has a well-defined ranking features help to determine in which a node should be processed. The data in this paper provides a clear insight into the Bitcoin network size, topology visualization, Bitcoin nodes geolocation.

The rest of the paper is organized as follows: Section 2 introduces the Bitcoin network and the technology behind it. The used Bitcoin topology discovery methodology is described in Section 3. Section 4 reports the analysis of the Bitcoin network. Finally, Section 5 discusses the conclusion and future work.

II. The Bitcoin Background

A. Bitcoin Blockchain

Digital currency based on cryptography [26] is not a new idea, but until recently it did not attract much attention. It changed rapidly with the introduction of Bitcoin in 2009. Nowadays Bitcoin [25] is the dominant cryptocurrency system. Bitcoin is a decentralized digital currency which does

978-1-7281-1328-9/19/$31.00 ©2019 IEEE

not rely on a trusted issuing entity but rather on a peer-to-peer network with peer minting Bitcoins by the brute-forcing dual SHA-256 hash function. The system is primarily composed of an agreed protocol for broadcasting exchanges of value between tokenized participants of a peer-to-peer network. To make the generation process of bitcoins (BTC) computationally hard, the Bitcoin protocol [19] requires the miners to generate the hash value of the mined block starting with a certain number of zeros (as an example of the proof of work [20] concept). The Bitcoin network provides peer discovery and reputation mechanisms to achieve stability among its homogeneous nodes. By design, the ledger and its updates are public to allow a real-time majority consensus [38] to form the current state of the system.

B. Bitcoin's P2P Network

In the Bitcoin network, there is no authentication functionality. Thus, each node keeps a list of IP addresses associated with its connections. Bitcoin peers try to maintain by default a well-specific number of connections, nodes behind NAT, Tor or firewall try to maintain by default a minimum of 8 outgoing connections and can accept up to 117 incoming connections (thus having up to 125 connections in total), while directly-connected nodes (nodes not using firewall, NAT or Tor) can accept up to 12 outbound connection and have average of 145 inbound connections. In our case, we call the local used Bitcoin nodes as a Bridge nodes and the peers to which it establishes connections with as an Entry Node.

A newly joining node broadcast PING message advertising its presence. Whenever an existing node receives the PING message, it forwards it to its entries and initiates a back-propagated PONG message. The PONG message contains information about the recipient node such as its IP address and other data. To become a network member, a node must open one or many connections with nodes that are already in the network. In the dynamic Bitcoin environment, nodes often join and leave the network and peers' connections are unreliable. To cope with these unstopped changes, after joining the network, a node periodically PINGs its neighbors to discover other participating nodes. Nodes decide where to connect to the network based only on local information; Utilizing the local IP addresses list "peers.dat", a disconnected node can always reconnect to the network. thus, the entire network forms a dynamic, self-organizing network of independent entities. This virtual, application-level network has Bitcoin servants as its nodes and open TCP connections as its links.

III. Bitcoin Topology Discovery Methodology

A. Architecture

Our ambition was to make the Bitcoin topology discovery process as uncomplicated and straightforward as possible, allowing for a transparent and flexible implementation. The suggested approach is based on the following main three components:

- The Bridge nodes, a Bitcoin full nodes serves as a bridge between our local discovery network and the Bitcoin network.
- The Discovery client, a customized Bitcoin node that establishes connections between the bridge nodes and all the connected Bitcoin node in order to get all the IP addresses in their "peers.dat" list.
- The Topology analysis server uses a modified Page-Rank algorithm to analyze the collected data in order to discover the connection nature among the Bitcoin node.

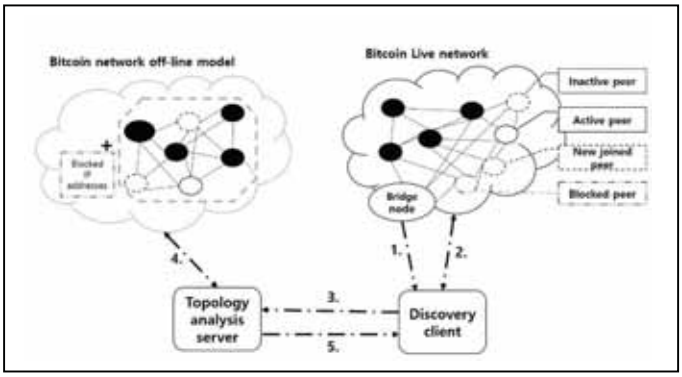

Fig. 1. Bitcoin topology discovery and analysis architecture

A high-level description of the proposed architecture is given in Figure 1. The discovery process is executed as follows, (1.) The Bridge node sends a request to the Discovery client to run the discovery scripts. (2.) The client scans the network and all the active nodes. A file is created for each discovered node containing its data. (3.) The files are transferred to the server. (4.) The data is processed by comparing the gotten data from the live Bitcoin network with the off-line model stored in the local database. (5.) The digest lists, the topology visualization model. A detailed description of the used components, scripts are presented in the upcoming section.

B. Data Collection

To measure and analyze the Bitcoin topology, we developed a system that collects data from the main Bitcoin network by utilizing both the Bitcoin Core historical (bridge node) and the Real-time data (discovery client) retrieval methods.

The bridge to the Bitcoin network is established by installing the Bitcoin core client and configured as a full node called Bridge node, which collects all needed data for building the network topology graphs as well as the performance evaluation. The Discovery Client has been implemented on top of the Bridge node. The discovery system consists of three phases, i) the processing phase based on scanning the Bitcoin DNS seeds, ii)the network scanning phase and iii) peers scanning phase. These individual scans run in a sequential fashion, where the processing scanner spawns instances of the network scanner, which in turn spawns instances of the peers' scanner. The discovery client initiates the discovery process by digging initial IP addresses list of all the pre-coded DNS seeds. Then, it perform a TCP scan to each node in the initial list by sending a generic join-in message (ping/pong) to check the reachability of the nodes.

In the case of reachable nodes, the client process the Bitcoin handshake scan via exchanging the protocol version among the nodes to examine the node state through clustering the reachable nodes to active nodes if the client receives a *verack* message or inactive in case of no answer is sent back. Lastly, the client scans each active peer to obtain its IP address list by performing the peer scan based on the *getaddr/addr* messages. The discovered IP addresses are added to the local Bitcoin IPs list, for each entry in the list the client performs the three-level scans to discover each of its known peers.

In Bitcoin, nodes connect over IPv4, IPv6 and Tor. The used methodology give us access to large datasets. However, the nodes excluded from the clustering tend to be either part of isolated clusters (Tor), not connected or not reachable nodes.

C. Data Analysis

As described in the above sections, the dynamic nature of the Bitcoin P2P network, where nodes join/leave the network frequently, drive us to develop a topology analysis server that recuperates the data gathered by the discovery client and process it by comparing the gotten data from the main Bitcoin network with the stored off-line topology model. The analysis server uses a modified version of PageRank (Algorithm 1) to build an offline topology model based on the initially discovered points, and by digesting the discovered data from each iteration the model is updated to cover all the Bitcoin nodes and visualize real-time Bitcoin network topology. The used discovery method helps in reducing the discovering-time and discover the topology graph for almost all reachable nodes. The topology analysis script is designed as a runner and scheduled to run after each discovery client scan iteration is completed.

Algorithm 1: Topology model

```
While receiving node is true:
append node CN to the OffLineTopologyModel
compute undirected PageRank update
compute rank(CN) based on PageRank and node wright
for node CN in OffLineTopologyModeline:
        if size(CN) > 10th percentage of size(OffLineTopologyModel):
CN = Super-Node.
```

PageRank is an algorithm that measures the transitive influence or connectivity of nodes [22]. In this paper we propose a real-time Bitcoin analysis system, the system achieves scale by separating the analysis process into several layers, by using a customized PageRank algorithm that takes into consideration nodes weight as the sampling function, while the use of nodes' connections enables the system to decide in real time which nodes require a deeper graph analysis. Each layer has a well-defined analysis and ranking features help in decide in which layer a node should be processed. As the correlations of Bitcoin nodes in real-time are stochastic, we start to adjacency matrices such that rank doesn't split up among the nodes' neighbors.

IV. Bitcoin Network Analysis

In this research, our goal was to enable the discovery client to establish a connection to each reachable active node in the network. To provide this capability, we configure the client node as a Python-based customized node that establishes connection with a well-defined set of nodes, and collects the needed data, e.g., the node's IP, the node's IP database, the node's ledger state.

In this section, we describe the Bitcoin network size, topology visualization, Bitcoin hybrid architecture, as well as the well-connected node performance evaluation.

TABLE I. Collected data statistics

	IPs	***Percentage***
Total IP addresses	162,319	X.X% (*)
Unique IP addresses	136,023	83.80%
Double IP addresses	26,295	16.20%
IPv4	146,087	90%
IPv6	8,116	5%
Tor	8,116	5%
Reachable peers	87,652	54%
Unreachable peers	74,667	46%
Misbehaving peers	1,623	1%

A. Network Size

By running the nodes discovery system for 45 days from 2018/08/06 till 2018/09/19, we had detected over one hundred sixty thousand IP addresses [23]. Of this, we found approximately one hundred thirty thousand unique IP address with eighty seven thousand reachable bitcoin nodes [25]. Table 1 provides more information of the collected data and number of the discovered nodes.

(*) Even though the collected data is enormous, as Bitcoin nodes often join/leave the network and the protocol characteristics, having a global observation of the entire Bitcoin network size is quite challenging due to the following reasons:

- The nature of the Bitcoin client standard implementation limits the amount of the obtained data when basing the discovery process on *getaddr/addr* with a set of 2500 IP addresses in average (between 23% of the entire local data set).
- In the current system implementation, we can process the peer's scan for only clear connected peers and the peers behind a firewall, however, TOR nodes are still out of our scope, due to the complicated structure of the onion network.

B. Bitcoin network Topology Visualization

Concerning the topology visualization of the Bitcoin network, during the 45 days of the experiment, we performed a daily scan for 8 hours, with a total of 45 snapshots. The initial scan allows us to discover over forty thousand nodes, which was used to build the first offline model, while during the last scan the observed network had almost eighteen thousand eight hundred reachable node. with an average degree equal to 4.7%. Table 1 shows the successful connections to unique peers. The average of successful connection attempts is around 54%. This means, that about 46% of the IP addresses broadcasted in the Bitcoin network belong to a non-active, non-routable or non-reachable nodes.

Regarding community structure, we used the Louvain algorithm [27] to detect the communities among Bitcoin discovered nodes, the modularity in Bitcoin main-net graph sample is higher compared to any random graph samples, regardless of the used model [28, 29]. Thus, the Bitcoin network shows more community structures compare to what should be expected from a random graph network, Figure 5 illustrates the visualization of the communities found in the last scan of Bitcoin main net, with the color of the node indicating the community it belongs to. As shown in figure 5, within the Bitcoin network there are ten large communities, whit the largest ones (Black, Blue, Yellow, Green) containing almost 40% of the nodes, we also have noticed that some of the small communities (Red and Dark-Green) group a total of 30% of the well-connected nodes (16% of the discovered

nodes), while each of the other 8 communities includes a maximum of 5 super-nodes. As we based our detecting community structure on modularity calculation. Note that we have a certain error, it has been obvious that modularity suffers a resolution limit, thus, it is unable to detect small communities. Therefore, nodes in grey (grey dotes in figure 5) were part of numerous communities under 3% of the total discovered nodes.

During the experiment, we noticed a massive overlapping between the created snapshots, where more than 16% of the discovered nodes exists in all the gotten snapshots. Since in Bitcoin P2P network, users establish a connection by initially performing outbound connections to one or more publicly reachable nodes, and because Bitcoin referred to as distributed ledger, a database that's distributed throughout a network. Information is continually shared and reconciled throughout multiple nodes and each one has an identical copy of the database. Transactions within this database are audited and agreed upon by consensus of all its users. The permanently connected peers mentioned above could be seen as Bitcoin's backbone. therefore, those nodes give more stability to the entire Bitcoin ecosystem and can be considered as the Bitcoin trust nodes.

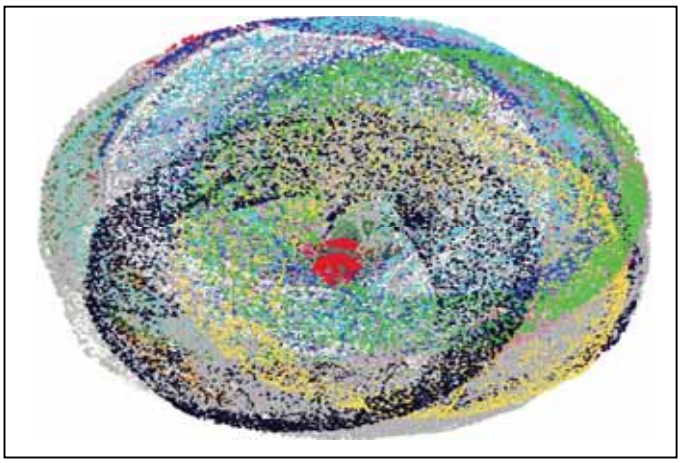

Fig. 2. Bitcoin Network graph visualization. final snapshot . Node size represents degree connectivity, where nodes with a higher degree have larger circles. Node color present the the community it belongs to.

C. Bitcoin Nodes Geolocation and AS :

The nodes can be easily grouped by mean of geographic regions (Geolocation) or Autonomous Systems (AS). We used an IP Geolocation API [30] to obtain both information regarding the geographical location of the node in Bitcoin main-net and their distribution among autonomous systems. Table 2 provides data concerning the Bitcoin nodes distribution over the top 10 countries. During the last scan of the network, most of the discovered node IP addresses were located in the United States with more than 28%, China 24.7%. The Tor IP addresses and some other nodes IPs around 2.7% of the discovered IP addresses during the last scan couldn't be geolocated due to their used internet infrastructure. Figure 3 provides the distribution of Bitcoin active reachable nodes among the 1,639 discovered autonomous systems. As shown in the graph more than 60% of all the discovered reachable nodes in Bitcoin reside in only 50 ASs, we also noticed that 7% of the AS contained a SINGLE Bitcoin node and almost 32% of Bitcoin active routable peers are run in the data centers.

TABLE II. DISTRIBUTION OF BITCOIN NODE IP ADDRESSES BY COUNTRY

Country	*Total Nodes*	*Percentage*
UNITED STATES	5353	28.30%
PEOPLE'S REPUBLIC OF CHINA	4663	24.70%
GERMANY	2185	11.60%
FRANCE	814	4.30%
NETHERLANDS	765	4.10%
UNITED KINGDOM	531	2.80%
RUSSIAN FEDERATION	512	2.70%
CANADA	492	2.60%
ITALY	220	1.20%
SWITZERLAND	211	1.10%

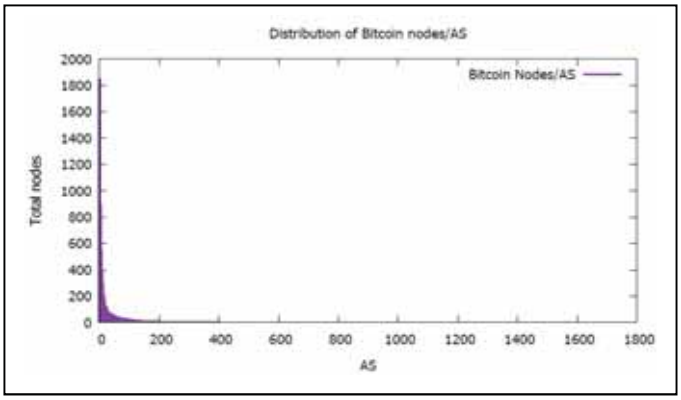

Fig. 3. Distribution of Bitcoin nodes by AS

V. CONCLUSION AND FUTURE WORK

In this paper, we presented a Bitcoin P2P topology discovery framework which collects data from the main Bitcoin network and analyzed the performance metrics of the discovery nodes. The use discovery strategy was based on a client/server architecture. The server is responsible for managing the list of nodes to be contacted, assigning work to clients, and assembling the network topology graph based on a modified ranking algorithm. While the client collects data by discovering the network topology around the list of initial points assigned from the server by using both the Bitcoin core client and a custom Python-based client for real-time data collection.

Further, in our future work, we will focus more on understanding the graph structure of the Bitcoin network by using NeuroTopology-based discovery algorithms that are more efficient than the ranking algorithm. Extract more useful information from the topology such as diameter, influential peers, identifying the vulnerabilities arising from clustering, etc. We also aim to Include Onion peers, and consider the extended Bitcoin network where peers implement other protocols such as Stratum.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01059786), and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.2018-0-00539, Development of Blockchain Transaction Monitoring and Analysis Technology).

REFERENCES

[1] Blockchain basics: Cryptography, [Online]. Available: https://lisk.io/academy/blockchain-basics/how-does-blockchain-work/blockchain-cryptography-explained

[2] Security Evaluation of Cryptographic Technology, [Online]. Available: http://www.nict.go.jp/publication/shuppan/kihou-journal/journal-vol58no3_4/journal-vol58no3-4_0407.pdf

[3] The great illusion of digital currencies, [Online]. Available: https://helda.helsinki.fi/bof/bitstream/handle/123456789/15564/BoFER_1_2018.pdf

[4] Blockchain, [Online]. Available: https://www.investopedia.com/terms/b/blockchain.asp

[5] Blockchain: The new technology of trust, [Online]. Available: https://www.goldmansachs.com/insights/pages/blockchain/

[6] Fletcher G.H.L., Sheth H.A., Börner K. Unstructured Peer-to-Peer Networks: Topological Properties and Search Performance. In: Moro G., Bergamaschi S., Aberer K. (eds) Agents and Peer-to-Peer Computing. AP2PC 2004. Lecture Notes in Computer Science, vol 3601. Springer, Berlin, Heidelberg, (2005).

[7] O. Babaoglu et al. Anthill, "A framework for the development of agent-based P2P Systems". In Int. Conference on Distributed Computing Systems. IEEE Computer Society. Austria, (2002).

[8] Cisco Systems, "chapter 7: Network Management Basics. Cisco Press", Inc. Internetworking Technologies Handbook, 4th edition, (2003).

[9] A. Auvinen et al. "New topology management algorithms for unstuctured P2P Networks". The second international Conference on Internet and Web Applications and Services, (May 2007).

[10] K. A. Amin, Armin R Mikler. "Resource Efficient and Scalable Routing Using Intelligent Mobile Agent". Master's thesis, USA: University of North Texas, May 2003. [Online]. Available: https://www.researchgate.net/publication/2491030_Towards_Resource_Efficient_and_Scalable_Routing_An_Agent-based_Approach

[11] E. P. Duarte et al. "A Distributed Network Connectivity Algorithm". In Proceedings of the Sixth IEEE International Symposium on Autonomous Decentralized Systems (ISADS'2003), pages 285–292, Italy, (2003).

[12] Y. Breitbart et al. "Topology Discovery in Heterogeneous IP Networks: The NetInventory System". IEEE/ACM Transactions on Networking, 12(3):401–414, (2004).

[13] Y. Bejerano et al. "Physical Topology Discovery for Large Multi-Subnet Networks". In Annual Joint Conf. of the IEEE Computer and Communications Societies (INFOCOM'2003), USA, (2003).

[14] Decker et al. "Information propagation in the Bitcoin network". 2013 IEEE Thirteenth International Conference, IEEE, pp. 1–10, (2013),

[15] Babaioff et al. "On Bitcoin and red balloons". ACM conference on electronic commerce, pp. 56–73. (2012)

[16] Biryukov et al. "Deanonymisation of clients in Bitcoin P2P network". CoRR abs/1405.7418 (2014).

[17] Andrew Miller et al. "Discovering Bitcoin's Public Topology and Influential Nodes"

[18] Karthikeyan Sankaralinga et al. "Distributed Pagerank for P2P System", the 12th International Symposium on High Performance Distributed Computing, (2003).

[19] Bitcoin Protocol , [Online], Available: https://en.bitcoin.it/wiki/Protocol_documentation

[20] Proof-of-Work, Explained, [Online]. Available: https://cointelegraph.com/explained/proof-of-work-explained

[21] Review of blockchain consensus mechanisms, [Online]. Available: https://blog.wavesplatform.com/review-of-blockchain-consensus-mechanisms-f575afae38f2

[22] The PageRank algorithm, [Online]. Available: https://neo4j.com/docs/graph-algorithms/current/algorithms/page-rank/

[23] BTCD, [Online], Availible: https://github.com/btcsuite/btcd

[24] Bitcoin Core P2P Network, [Online], Available: https://en.bitcoin.it/wiki/Bitcoin_Core_0.11_(ch_4):_P2P_Network

[25] "Bitcoin: A Peer-to-Peer Electronic Cash System", [Online]. Available: https://bitcoin.org/bitcoin.pdf

[26] D. Chaum et al. "Untraceable electronic cash. In Proceedings on Advances in Cryptology" (CRYPTO '88). Springer, (1988).

[27] Louvain Algorithm, [Online], Available: https://neo4j.com/docs/graph-algorithms/current/algorithms/louvain/

[28] Community Detection in Python I, [Online], Available: https://yoyoinwanderland.github.io/2017/08/08/Community-Detection-in-Python/

[29] Community Detection in Python II, [Online], Available: https://github.com/topics/community-detection-algorithm?l=python

[30] IP Geolocation API, [Online], Available: https://ipstack.com/