From Bitcoin to Bitcoin Cash: a network analysis

Marco Alberto Javarone
nChain
London W1W 8AP, UK
University of Kent
School of Computing, University of Kent, UK
marcojavarone@gmail.com

Craig Steven Wright
nChain
London W1W 8AP, UK
craig@ncrypt.com

ABSTRACT

In the last years, Bitcoins and Blockchain technologies are gathering a wide attention from different scientific communities. Notably, thanks to widespread industrial applications and to the continuous introduction of cryptocurrencies, even the public opinion is increasing its attention towards this field. The underlying structure of these technologies constitutes one of their core concepts. In particular, they are based on peer-to-peer networks. Accordingly, all nodes lay at the same level, so that there is no place for privileged actors as, for instance, banking institutions in classical financial networks. In this work, we perform a preliminary investigation on two networks, i.e. the Bitcoin network and the Bitcoin Cash network. Notably, we aim to analyze their global structure and to evaluate if they are provided with a small-world behavior. Results suggest that the principle known as 'fittest-gets-richer', combined with a continuous increasing of connections, might constitute the mechanism leading these networks to reach their current structure. In addition, further observations open the way to new investigations into this direction.

Keywords

Bitcoin; Bitcoin Cash; Complex Networks

1. INTRODUCTION

Nowadays, a number of services and platforms are based on distributed networks. In particular, one of the major benefits of the latters is given by the partition of a computational workload among multiple nodes, so that each one can perform an autonomous processing. As result, on a global level, a distributed network allows to implement the so-called 'parallel computing' [1]. When this kind of network is not controlled by a (or a few) central unit (e.g. a node that coordinates the whole system, or a part of the network), it can be also defined as 'decentralized'. Blockchain is a modern technology, described for the first time in [2], that can be briefly defined as a decentralized and distributed ledger. The latter contains records of transactions and implements different functionalities mostly based on the modern cryptography. In the last years, this technology is finding application in several industrial sectors [3], spanning from finance to healthcare, having then an impact at different societal levels. In this context, bitcoins are exchanged among users and the transactions are recorded in the Blockchain. In addition, as observed during last months, bitcoins are exponentially increasing their value, and a number of new cryptocurrencies is continuously generated [4]. The absence of a central control (e.g. a Banking Institution) appears to be one of the major and most notable results of Blockchain but, at the same time, can constitute a motivation of concern and skepticism for those who do not understand the underlying mechanism. Since these technologies are strictly based on a networked system, as older sharing platforms (e.g. those commonly used for file sharing), it is worth to investigate the related topological properties. Notably, a global topological view allows both to obtain a deeper knowledge on the structure, and to evaluate the dynamics of stochastic processes, as spreading and percolation, that can be of interest for new applications, security reasons, and so on. Therefore, in this work, we aim to analyze the structure of this kind of networks, with a focus on Bitcoin and Bitcoin Cash. Notably, the latter results from the former, and constitutes a new cryptocurrency (available in the market from August 2017). Our results indicate that these two networks share some topological similarities, and that some generative models as 'preferential attachment' [5] or 'fittest-gets-richer' [6, 7] might be adopted for representing their evolution. In particular, even if both networks are 'peer-to-peer', parameters like the 'fitness' can be useful for discriminating among nodes, e.g. from those provided with high computational resources to those with low power (e.g. Raspberry pi). The remainder the paper is organized as follows: Section 2 briefly summarizes the dynamics of the Bitcoin Network. Section 3 introduces some concepts of complex network analysis and illustrates two generative models that might be useful for the considered networks. Section 4 shows results of numerical investigations on the real datasets. Finally, 5 ends the paper.

2. THE BITCOIN NETWORK

In this section, we provide a very brief introduction to the Bitcoin Network [8]. In particular, we aim to link the basic mechanisms of the network, without to consider all the local processes that occur during a single transaction, with the emergence of a connectivity pattern in the real datasets (below described) we consider in this analysis. In few words, the Bitcoin Network represents the set of nodes running the bitcoin P2P protocol. In this network, each node can have a specific role depending on its functionality, e.g. routing, mining, Blockchain database and so on. Usually, a node that performs all these functions is defined as 'full node'. All nodes play as router in order to simplify spreading processes over the network. Full nodes can verify any transaction without asking for external references. Mining nodes take part to a kind of 'competition' for solving the proof-of-work algorithm. Here, in order to increase the success probability, nodes can cluster together forming mining pools. Now, when a new node joins the network, it needs to get connected with some of the pre-existing nodes. In particular, the new node needs to discover at least one node in the network. This process is completely random, i.e. the new node connects with one randomly chosen among pre-existing nodes. It is worth to highlight that the Client used to join the network contains the list of some nodes (i.e. seed nodes). However the first nodes to consider for new connections can be also those connected to these seed nodes. Eventually, in order to be connected to the whole network in a reliable way, the new node connects to a few nodes, then forming different paths. While one single connection can be a bit too little for ensuring reliability, usually nodes need only few neighbors, saving network resources.

3. NETWORK ANALYSIS

As stated before, in this context we are dealing with a networked system. As result, the modern mathematical framework that allows to analyze this kind of systems is the Theory of Complex Networks [9]. Due to its relevance and ubiquity in a number of fields, we provide a quick introduction. Notably, we focus on global properties for characterizing the structure of a network and on few generative mechanisms that can be of interest for studying the Bitcoin Network. Let us begin introducing simple concepts. A complex network is a graph characterized by non-trivial topological features. In more general terms, a graph is a mathematical entity that allows to represent specific relations among a collection of objects. More formally, a graph G is described as $G = (N, E)$, with N set of nodes and E set of edges (or links). Nodes are the main elements of a system and can be described by a label. On the other hand, the edges represent connections among nodes and can map specific relations, e.g. friendship in a social network, correlations in EEG networks, direct links among web-sites in the WEB, and so on and so forth. A graph can be 'directed' or 'undirected', i.e., we can have symmetrical relations among nodes or not. A simple undirected edge is drawn between two nodes when there is a symmetric relation (e.g. friendship) whereas, in the second case, the edge takes the form of an arrow (e.g. a website linking to another website). In addition, a graph can be 'weighted' or 'unweighted'. Notably, if a numerical value is associated to edges, for example because the related relations can be somehow weighted (e.g. representing an intensity), the graph is weighted. For instance, considering an airline network where the airports are represented as nodes and the routes as edges, weights can be computed according to the geographical distance between airports. The whole set of edges, of a N nodes graph, is collected in a $N \times N$ matrix, defined 'adjacency matrix'. Undirected graphs have a symmetric adjacency matrix, and unweighted graphs are represented by a binary adjacency matrix. In particular, the adjacency matrix A of an unweighted graph is composed by the elements:

$$a_{ij} = \begin{cases} 1 & \text{if } e_{ij} \text{ is defined} \\ 0 & \text{if } e_{ij} \text{ is not defined} \end{cases} \tag{1}$$

Instead, a weighted graph is represented by a real matrix. Now, we present with more details some global network properties that can be derived by analyzing the adjacency matrix, i.e. the degree distribution, the clustering coefficient, and the path length.

3.1 Degree distribution

Nodes of a graph can have many connections (i.e. many edges). Usually, the number of connections of a node is called degree, and it is denoted as k. An important property widely used to asses the structure of a network is the degree distribution $P(k)$ [5]. The latter represents the probability that a randomly selected node has the degree equal to k, i.e., it is connected with k nodes. Remarkably, the degree distribution uncovers a number of information related to a network. In addition, there are classes of random networks that are related to their degree distribution. For instance, Erdös-Renyi graphs and Scale-free networks, that we briefly describe below. Notably, the first model of random networks has been proposed by Paul Erdös and Alfred Renyi [10]. They defined a famous model known as Erdös-Renyi graph, or simply E-R graph. Their model considers a graph with N nodes and a probability p to generate each edge, so that there are around $p \cdot \frac{N(N-1)}{2}$ edges, resulting in a binomial degree distribution:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{n-1-k} \tag{2}$$

now, if $N \to \inf$ and $Np = const$, this degree distribution converges to a Poissonian distribution:

$$P(k) \sim e^{-pN} \cdot \frac{(pN)^k}{k!} \tag{3}$$

While this model represents an early attempt to describe real systems, a more advanced one is the Barabasi-Albert model (BA model) [5] that focuses on scale-free networks, i.e. structures characterized by a $P(k)$ that follows a power-law function as:

$$P(k) \sim c \cdot k^{-\gamma} \tag{4}$$

with c normalizing constant and γ parameter of the distribution known as scaling parameter. In this kind of networks, few nodes (called hubs) have many connections (i.e, a high degree), while the majority of nodes only few. The BA model considers N nodes and a further parameter, e.g. m, representing the minimum number of edges per node. In the thermodynamic limit, the BA model leads the network to be fitted by a power-law function with scaling parameter $\gamma = 3$, and to an average degree equal to $2m$. Two main generative mechanisms can be related to the dynamics of the BA model: the first-mover-wins and the fittest gets richer.

First-mover-wins

The first-mover-wins constitutes the basic mechanism of the BA model, summarized as follows:

1. Define N number of nodes and m minimum number of edges for each node

2. Add a new node and link it with other m preexisisting nodes. Pre-existing nodes are selected according to the following equation:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{5}$$

   with $\Pi(k_i)$ probability that the new node generates a link with the i-th node having a k_i degree.

This mechanism is not able to capture any difference among nodes.

Fittest gets richer

Bianconi and Barabasi [6] described this mechanism introducing a fitness parameter η. Here, the fitness parameter represents the ability of a node to compete for new links. In particular, the degree of the i-th node is proportional to:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \tag{6}$$

with k_i degree of the ith node. In particular, new nodes tend to link with pre-existing nodes having high values of (η, k). In doing so, even new nodes can reach a high degree when provided with a good fitness.

3.2 Clustering Coefficient

The clustering coefficient [5] allows to assess if nodes tend to cluster together. This phenomenon is common in many real networks as social networks, where it is possible to observe the emergence of circles of friends where all people know each other. The clustering coefficient can be computed as:

$$C = \frac{3 \times Tn}{Tp} \tag{7}$$

Tn is the number of triangles in the network and Tp is the number of connected triples of nodes. A connected triple is a single node with edges running to an unordered pair of others. The value of C lies in the range $0 \leq C \leq 1$. A further way to compute this parameter has been proposed by Watts and Strogatz [11], computing this quantity as a local value:

$$C_i = \frac{Tn_i}{Tp_i} \tag{8}$$

with Tn_i number of triangles connected to node i and Tp_i number of triples centered on node i. In this case, the local C of nodes with a degree equal to 0 or 1 is set to 0. Therefore, the global clustering coefficient of a network becomes:

$$C = \frac{1}{n} \sum_i C_i \tag{9}$$

This parameter allows to measure the density of triangles in a network, and can be computed for directed and undirected networks.

Path Length

The distance between two nodes, belonging to the same network, can be computed considering the edges (and their weights in weighted networks) between them. In particular, a geodesic corresponds to the minimum path between two nodes. At the same time, the distance between two nodes is infinite if there are no paths in the network for connecting them. Finding the shortest path can be computationally expensive, so different heuristics and meta-heuristics can be adopted. For instance, the Dijkstra's algorithm [12] and the Floyd-Warshall algorithm [13]. Computing the average shortest path length provides a first clue of small-worldness of a network [11]. Notably, a network can be defined as 'small-world' when the average distance between pairs of nodes L scales with the size of the network, i.e. $L \propto \ln N$.

4. RESULTS

In this section, we describe results of our preliminary investigations on two kinds of networks, i.e. Bitcoin network and Bitcoin Cash network (see a pictorial representation in fig. 1). The dataset related to Bitcoin network contains 7025 nodes and refers to the month

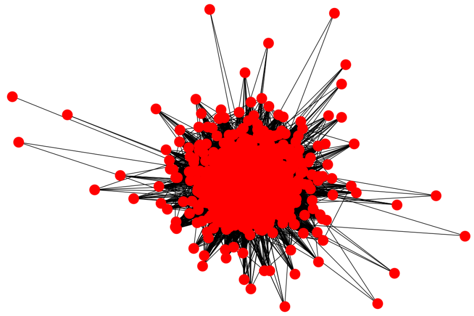

Figure 1: Pictorial representation of the Bitcoin Cash Network related to the month of August 2017. A giant core strongly connected clearly emerges.

of April 2016, while those related to Bitcoin Cash refer to August and December 2017, and contain 963 and 1454 nodes, respectively. According to their dynamics, these networks are directed, i.e. nodes are connected by arrows. As preliminary analysis, we focus on the degree distribution. In particular, the out-degree distribution representing that of arrows leaving a node, and the in-degree distribution representing the distribution of arrows reaching a node. Results related to the out-degree distribution (i.e. $P(k_{out})$) of the two Bitcoin Cash networks are shown in fig. 2, while the in-degree (i.e. $P(k_{in})$) of the same networks are shown in fig. 3.

The scaling parameters (i.e. γ) used for fitting these networks are all smaller than 1 (see Table 4). Notably, they have been computed after a binning process, considering 50 samples (see [14] for further details). These results indicate that both kinds of networks (i.e. Bitcoin and Bitcoin Cash) have an out-degree distribution that can be fitted by a power-law, and a in-degree distribution that seems to be exponential (i.e. all nodes have a similar in-degree). Regarding the out-degree distributions, since the scaling parameter is smaller than 1 both in Bitcoin and Bitcoin Cash Network, they cannot be properly defined as scale-free. Then, we analyze the average clustering coefficient (i.e. C) and the average shortest path length (i.e. SPL) of the three networks. Table 4 reports the related numerical values.

Network Parameters

Network	N	γ	C	SPL
BC Apr 2016	7025	0.4	0.486	1.816
BCH Aug 2017	963	0.58	0.576	1.7267
BCH Dec 2017	1454	0.57	0.544	1.7605

It is worth to highlight that all networks seem to be provided with a small-world behavior, since their SPL is smaller than the logarithm of their size (i.e. N).

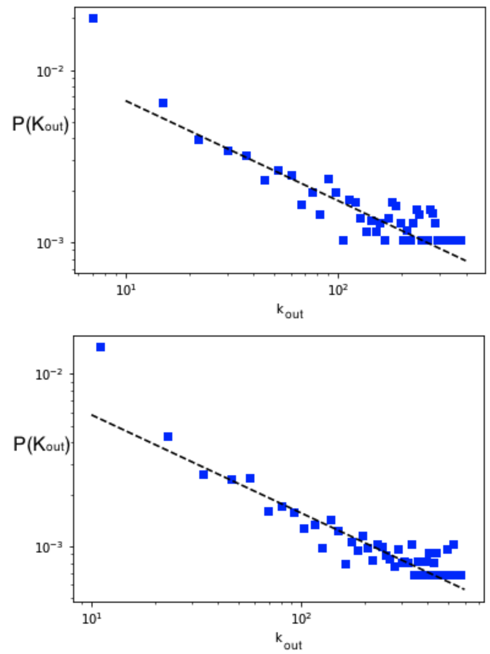

Figure 2: Out-degree distribution $P(k_{out})$. On the top: Bitcoin Cash network on August 2017. On the bottom: Bitcoin Cash Network on December 2017. The black dotted lines represent the power-law fitting both curves.

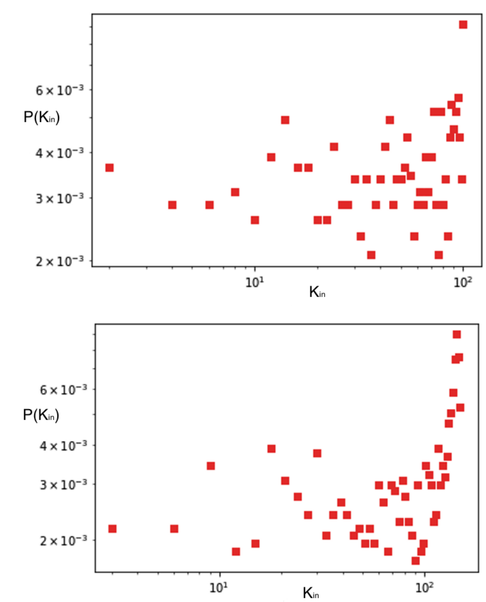

Figure 3: In-degree distribution $P(k_{in})$. On the top: Bitcoin Cash network on August 2017. On the bottom: Bitcoin Cash Network on December 2017.

5. DISCUSSION AND CONCLUSION

In this work, we perform a preliminary analysis of a real Bitcoin Network and two Bitcoin Cash networks. Our aim is both analyzing their connectivity pattern and trying to understand 'how' evolved their structure. In order to reach this goal, we need to remind how a Bitcoin network works, as described in Section 2. In particular, as we know, since in principle all nodes have to save the whole 'ledger' of transactions, the in-degree distribution, i.e. the information received, must be more or less uniform (remarkably an exponential distribution characterizes homogeneous structures). On the other hand, since storing connections requires network resources (e.g. memory), in general most nodes remain connected to few neighbors. At the same time, powerful nodes can store more node connections than others. Thus, considering the generative mechanisms before described, i.e. 'fittest-gets-richer' and 'first-mover-wins', we think that the former might be more suitable for modeling the evolution of the out-degree distribution. In particular, the fitness can be related to the computational power of nodes, i.e. the higher the fitness the higher the available resources (and then the out-degree). In addition, even in the light of the increasing amount of Raspberry Pi, appearing in the network, we suggest that a 'fitness based' model can be effective in describing the dynamics of the Bitcoin network and the Bitcoin Cash network. Moreover, since the structure of the networks (considering their out-degree distribution), is very similar to scale-free networks, we have a first clue that they can be 'small-world' [11]. Here, a small-worldness behavior would be essential for the reliability of these networks. Therefore, we analyzed to important parameter, i.e. the average clustering coefficient and the average shortest path length. While the latter confirms that all these three networks are 'small-world', the former can be used also for future investigations. Notably, we aim to compare the average clustering coefficient with that achieved in a E-R networks generated with the same number of nodes and a similar number or edges. In particular, this analysis might constitute a further proof, combined with the average shortest path length, of the small-world behavior of a network. Finally, it is important to observe that beyond the difference in the size (i.e. in the amount of nodes), the Bitcoin network and the two Bitcoin Cash networks appear very similar. In general, that difference can be expected, being the Bitcoin Cash network recently proposed. To conclude, beyond the further analyses above mentioned, as for future work we deem relevant to investigate also the evolution of stochastic processes on these structures. In particular, analyzing spreading phenomena, and or percolation, can be useful for a better understanding on how these modern technologies behave on sharing data and information.

6. REFERENCES

[1] Keckler, S.W., et al.: GPUs and the future of parallel computing. *IEEE Micro* **31-5** 7 –17 (2011)

[2] Satoshi, N.: Bitcoin: A Peer-to-Peer Electronic Cash System. *https://bitcoin.org/bitcoin.pdf* (2009)

[3] Crosby, M., et al.: Blockchain technology: Beyond bitcoin. *Applied Innovation Review* **2** (2016)

[4] ElBahrawy, A., et al.: Evolutionary dynamics of the cryptocurrency market. *R. Soc. open sci.* **4**, 170623 (2017)

[5] Albert,R. and Barabasi, A.L.: Statistical Mechanics of Complex Networks. *Rev. Mod. Phys* **74**, 47–97 (2002)

[6] Bianconi, G. and Barabasi, A.L.: Bose-Einstein condensation in complex networks. *Physical Review Letters* **86**, 5632 (2001)

[7] Javarone, M.A., Armano, G.: Quantum Classical Transition in Complex Networks. *Journal of Statistical Mechanics: Theory and Experiment* **4**, P04019 (2013)

[8] Antonopoulos, A.M.: Mastering Bitcoin: Unlocking Digital Cryptocurrencies. *OReilly* (2014)

[9] Estrada, E.: The Structure of Complex Networks: Theory and Applications. *Oxford University Press* (2011)

[10] P. Erdos, P. and Renyi, A., On the Evolution of Random Graphs. *pubblication of the mathematical institute of the hungarian academy of sciences* 17–61 (1960)

[11] Watts, D. J. and Strogatz, S. H., Collective dynamics of "small-world" networks. *Nature* 440-442 (1998)

[12] Mohring, R.H., Schilling, H., Schutz, B., Wagner, D., Willhalm, T., Partitioning Graphs to Speed Up DijkstraâĂŹs Algorithm. Experimental and Efficient Algorithms LNCS, 3503, 189–202 (2005)

[13] Han, S.C., Franchetti, F., Pushel, M., Program Generation for all-pairs shortes path problem. Proc. of the 15th Int.Conf. on Parallel architectures and compilation techniques, 22–232 (2006)

[14] Clauset, A., Shalizi, C.R., Newman, M.: Power-Law Distributions in Empirical Data. *SIAM* **51(4)** 661âĂŞ703 (2009)