Blockchain centralization and trust coefficients.

noise@sumus.team

December 1, 2019.

Abstract

The paper proposes a quantitative assessment of the blockchain network centralization in the form of a coefficient calculated on the basis of the network parameters. According to the coefficient values, it is estimated whether the network operates randomly or centrally, or it is decentralized. An estimate is obtained for the optimal number of network nodes using consensus-based algorithms for the best combination of the network's operation speed and the level of trust in the results of its operation.

1 Introduction

One of the main questions that should be answered when assessing the blockchain network operation is to ensure equality of nodes in the network and uniform loading of their blocks at closing. This state of the network can be called decentralized. In an ideal case, all nodes that work for a long enough time cover the same number of blocks in a decentralized network.

The decentralized state of the network is opposed by two other equally undesirable states. The first is the random operation of the network, when many nodes close a small number of blocks each, but there are fewer nodes where each node closes many blocks, so that these nodes close about as many blocks as the nodes from the first set close. The second is the centralized operation of the network, where most nodes close one or two blocks, and several nodes close almost all blocks.

The task is to build a quantitative assessment of the blockchain network, which allows us to get a quick answer to the question of the level of centralization/decentralization/randomness of the network over a certain time interval, based on the simplest parameters of the network's operation.

2 Assessment of the centralized operation of the blockchain network

The set $\aleph$ of cardinality α is the set of blocks closed at the time t^*, $B_n \subset A_N$ is the set of nodes with cardinality n that participated in closing the blocks from the set $\aleph$ [1], A_N is the set of all nodes of the network.

We consider the lattice function $y(i)$, where $i = 0, 1, \ldots, \alpha$, which is constructed according to the following rule: $y(1) = n_1$ which is the number of nodes that closed only one block; $y(2) = n_2$ is the number of nodes that have closed only 2 blocks, $\ldots$, $y(\left[\frac{\alpha}{n}\right]) = n\frac{\alpha}{n}$ is the number of nodes that closed $\left[\frac{\alpha}{n}\right]$ blocks, and $y(\alpha) = n_\alpha$ is the number of nodes that closed α blocks each. Obviously, n_α is 0 or 1; $y(0) = n_0$ is the number of nodes that did not close a single block (did not work); $i = 1, \ldots, \alpha$, as a rule, $\alpha \gg n$.

We also consider some important special cases of the function $y(i)$ values.

a) If $n_\alpha = 1$, then the network is maximally centralized, and all blocks were closed by one node.

b) If $n_\alpha = 0$, then the study can be continued.

c) If $n_1 = n$, then $n = \alpha$, and each node closed one block at a time, and the condition $\alpha \gg n$ is not fulfilled.

d) If $n\frac{\alpha}{n} = n$, then each node closed $\left[\frac{\alpha}{n}\right]$ nodes, which corresponds to maximum decentralization degree without randomness. Case c) also refers to the decentralized state of the network, but cannot be maintained for a long time.

e) If n_0 is close to n ($n_0 < n$), then the blockchain has a lot of idle nodes, and the problem statement needs to be reviewed.

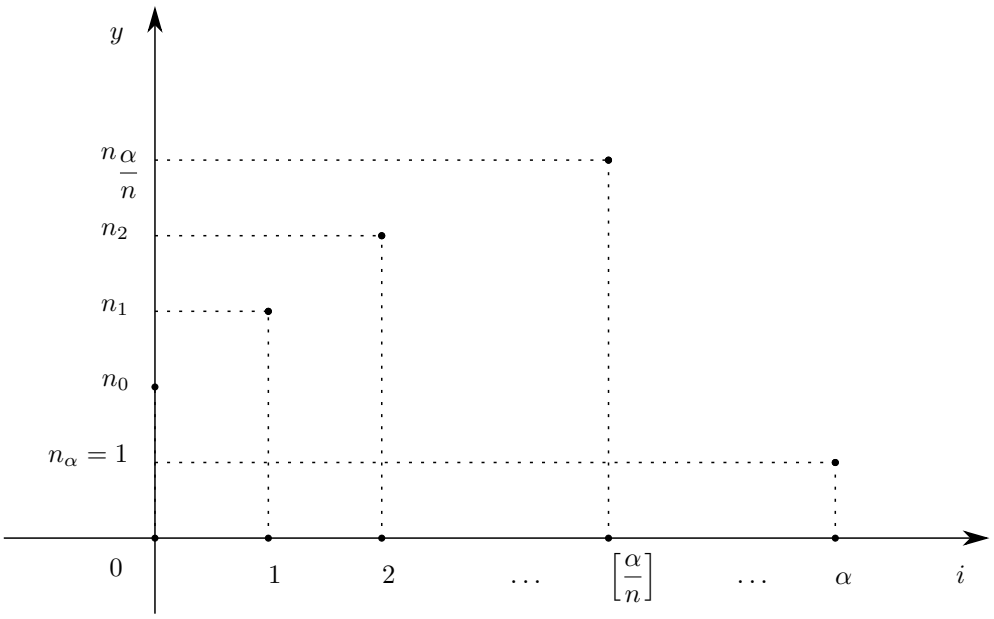

Figure 1: Dependence of the number of nodes closing the same number of blocks on the number of blocks closed by one node.

We consider a simplified version of the centralization coefficient. Let $i^* = \underset{i=0,\ldots,\alpha}{\operatorname{argmax}}(y(i))$. We assume that i^* is unique.

We introduce the centralization coefficient

$$K_y = 1 - \frac{\alpha}{n \cdot i^*},\ K_y \in [1 - \frac{\alpha}{n}, 1] \tag{1}$$

If $K_y < 0$, then this indicates a tendency to random system operation, for example, when $i^* = 1;\ \alpha \gg n$

$$K_{y\,\min} = 1 - \frac{\alpha}{n} < 0$$

When the $K_y = 0$, the most complete decentralization without randomness is achieved, $i^* = \frac{\alpha}{n}$. If $y(i^*) = n$, this corresponds to complete decentralization.

If $K_y > 0$, then this indicates a tendency toward centralization, with $K_{y\,\max} = 1 - \frac{1}{n} > 0$, $i^* = \alpha$, and $n_\alpha = 1$.

If $K^* = \max\{|K_{y\,\min}|, K_{y\,\max}\}$, then the normalized centralization coefficient is

$$K_{y\,\text{norm}} = \frac{K_y}{K^*} \tag{2}$$

The coefficient K_y can be calculated using instead of i^* the average number of blocks closed by one node $\langle i \rangle$, which is determined by the following expression:

$$\langle i \rangle = \underset{\tilde{i}=0,\ldots,\alpha}{\operatorname{argmin}} \left| \sum_{i=0}^{\tilde{i}} y(i) \cdot i - \sum_{i=\tilde{i}+1}^{\alpha} y(i) \cdot i \right| \tag{3}$$

$$K_y = 1 - \frac{\alpha}{n \cdot \langle i \rangle} \tag{4}$$

Using (4) requires more calculations, but the result will be a more accurate estimate of centralization. The most preferred case is the coincidence of the values of K_y obtained by formulas (1) and (4). If these values are essentially different, the network cannot be considered as decentralized.

3 Assessment of trust in the blockchain network.

As was pointed out in [1], all blockchain nodes constitute the set A_N divided into K disjoint subsets B_n^i

$$A_N = \bigcup_{i=1}^{K} B_n^i \tag{5}$$

Nodes included in each B_n^i in accordance and with the hypothesis of a uniform distribution of their numbers. Each set B_n^i has the same cardinality n and, in its meaning, is equivalent to the set from section **2**.

How many nodes should we take in B_n^i so that it causes sufficient trust in the results of the network operation? The greater n is the better for trusting in the system, but the less n is relative to N, the lower are the computational costs for consensus building. Then the larger is each of the quantities n, $\frac{N}{n}$, the smaller is $z(n) = \frac{1}{n} + \frac{n}{N}$. Having found the minimum of this function by n, we obtain the optimal (4).

$$n_{\text{opt}} = \sqrt{N} \tag{6}$$

We can consider $z(n)$ as the coefficient of "mistrust". Then $K_{\text{dr}} = \frac{1}{z(n)}$ would be the trust coefficient without loss of efficiency, reaching a maximum at $n = n_{\text{opt}}$.

$$K_{\text{dr}} = \frac{n \cdot N}{n^2 + N};\ K_{\text{dr max}} = \frac{\sqrt{N}}{2} \tag{7}$$

Normalized coefficient K_{dr} is

$$K_{\text{drn}} = \frac{K_{\text{dr}}}{K_{\text{dr max}}} = \frac{2n\sqrt{N}}{n^2 + N} \in [0, 1] \tag{8}$$

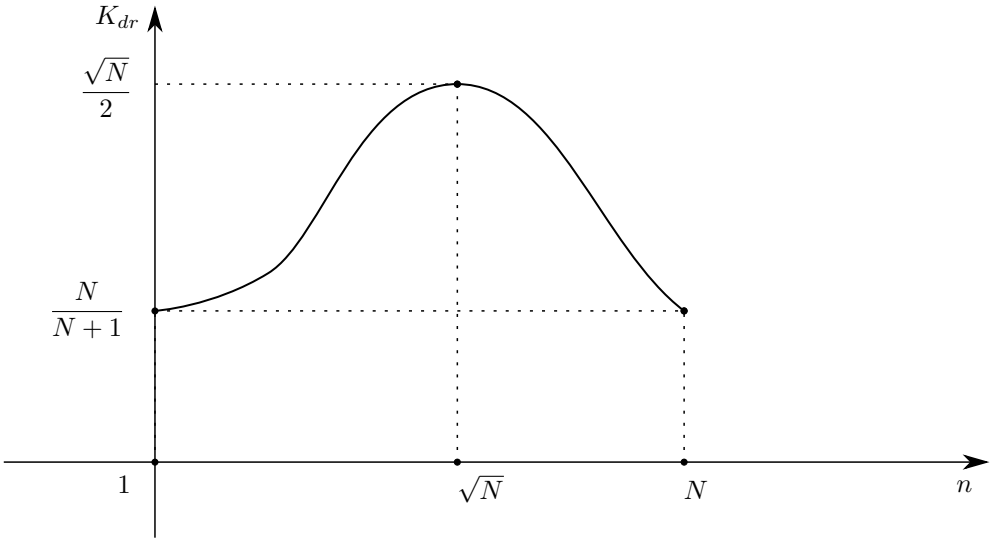

Figure 2: Dependence of the trust coefficient K_{dr} on the number of nodes in B_n.

4 On the optimal number of nodes in the set of "escort + master" in blockchain networks using the *SDBFT* algorithm

If it is necessary to determine the optimal number of nodes in the set $B_{n'}$ consisting of a master node and escort nodes for a blockchain with the set of nodes B_n, then we can reason the same way as in the previous section, replacing N with n and n with n' in formula (4). Then we get

$$n'_{\mathrm{opt}} = \sqrt{n} \tag{9}$$

If a trust coefficient for $B_{n'}$ is required, then with the indicated changes of variables we obtain analogues of formulas (7), (8) taking into account the limitations of the *SDBFT* algorithm.

$$n' \geq 5 \tag{10}$$

5 An example of applying centralization and trust ratios

The operation data of the BITCOIN blockchain [2] for three consecutive rather long time intervals was taken. For each of them, the function $y(i)$ was constructed and the centralization coefficient (4) was calculated.

The first interval: $\alpha = 210,000$; $n = 186,660$; $\left[\frac{\alpha}{n}\right] = 1$; $i^* = 1$; $K_y = -0.125$. The blockchain operates, being almost decentralized close to case c). Such an operating mode cannot be maintained for an arbitrarily long time, since it requires almost complete fulfillment of the condition $n = \alpha$.

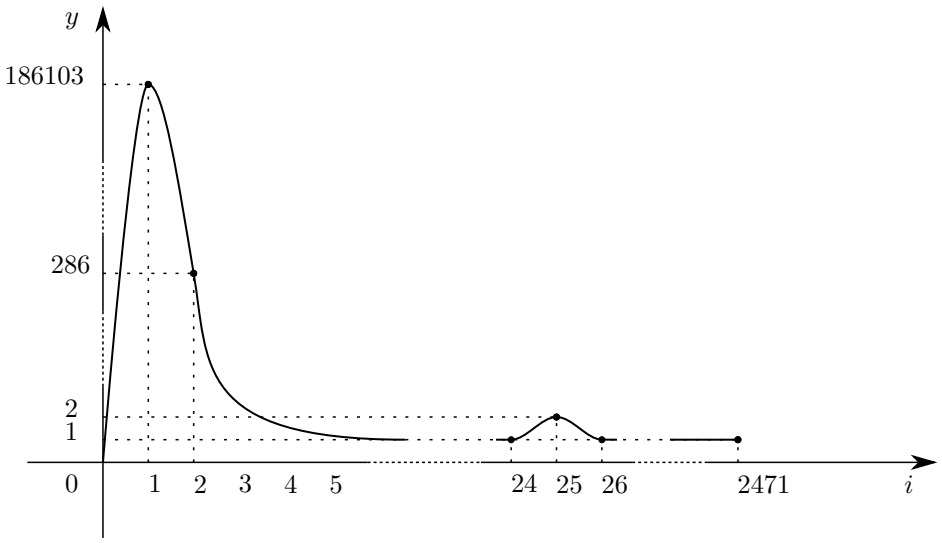

Figure 3: Dependence of the number of nodes on the number of blocks they closed for the first interval.

The second interval: $\alpha = 210,000$; $n = 9,774$; $\left[\frac{\alpha}{n}\right] = 21$; $K_y = 0.996$, which corresponds to the centralized operation of the blockchain.

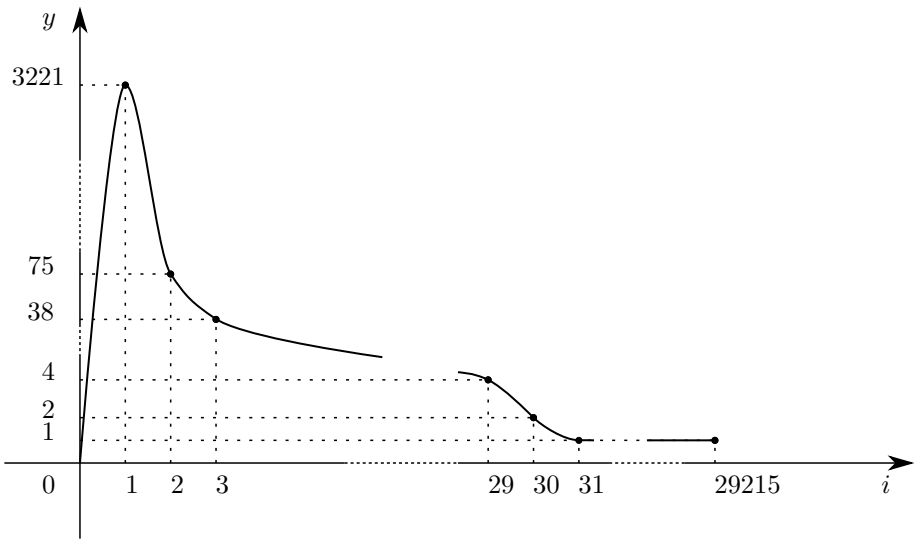

Figure 4: Dependence of the number of nodes on the number of blocks they closed for the second interval.

Third interval: $\alpha = 188930$; $n = 512$; $\left[\frac{\alpha}{n}\right] = 369$; $K_y = 0.954$, which corresponds to the centralized operation of the blockchain.

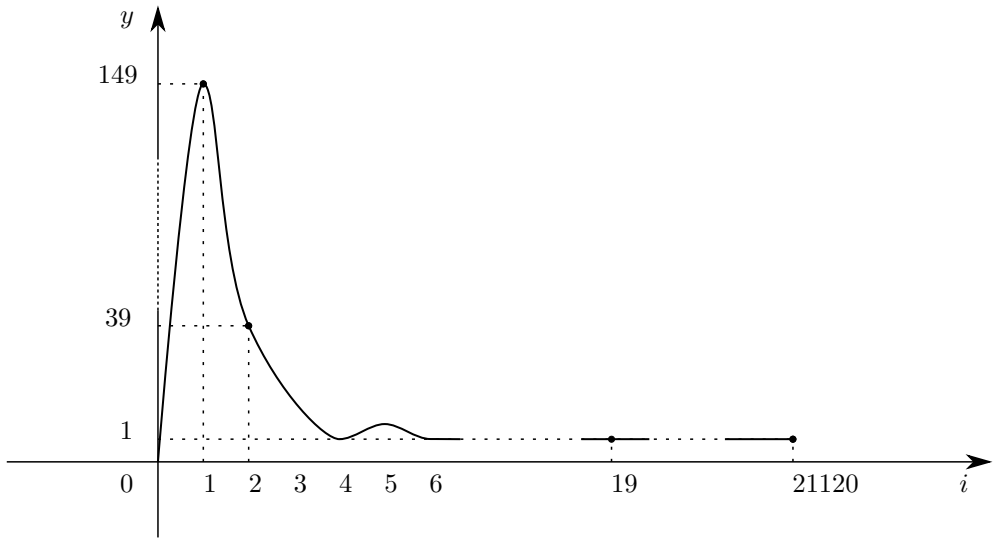

Figure 5: Dependence of the number of nodes on the number of blocks they closed for the third interval.

6 Conclusion

The paper offers a convenient assessment procedure for the uniformity of loading blockchain network nodes in the process of closing blocks. This estimate is quite simple and is calculated as a coefficient taking into account the number of closed blocks, the number of nodes that closed a certain number of blocks, and is called the centralization coefficient.

In the normalized form, this coefficient takes on values from -1 to 1, which corresponds to the network state from full randomness (-1) to full centralization (1) which allows to quickly estimate, whether the network blockchain corresponds at a certain time interval the main blockchain concept, or equality, and node independence (decentralization) which corresponds to a zero coefficient value.

As shown by studies of the BITCOIN network at three time intervals, this network can remain in a decentralized state for a long time, provided that the number of nodes is linearly related to the number of blocks they close.

However, at other time intervals, the network comes to centralization if this dependence is violated. The paper also offers a trust assessment for the system operation, provided it is decentralized. The assessment is also based on the cardinality of the set of nodes ensuring the formation of an escort and the possibility of choosing a master node to generate consensus, and on the cardinality of the set of all nodes included in the network.

References

[1] a@sumus.team, k@sumus.team, rr@sumus.team. “Consensus Algorithm for Bigger Blockchain Networks” (2018).

[2] Satoshi Nakamoto (2009). “Bitcoin: A Peer-to-Peer Electronic Cash System”. www.bitcoin.org .