How to measure decentralization of nodes

b.unlimited@protonmail.ch

May 10, 2018

Abstract

Let p_{total} and μ_{total} be the total population and GDP pr. capita of the world. Let C be a list of all countries in the world, and N be a list of all running nodes in the network, and let μ_c, p_c and n_c be the GDP pr. capita, population and number of nodes in country c.
Then a usefull measure of the decentralization of the nodes in the network is give by

$$D(N) = 1 - \frac{1}{p_{total} \cdot \mu_{total}} \sum_{c \in C} \frac{\mu_c \cdot \Big(\max(0, p_c - n_c)\Big)^2}{p_c}$$

where $D(N) \in [0,1]$ and a higher value of $D(N)$ indicates more decentralization.

Contents

1 Introduction

During the Bitcoin scaling debates the issue(s) of (de)centralization have come up numerous times.
The two biggest centralization threats seem to be hash-power centralization and centralization of nodes (although there are others as well).
The issue of hash-power centralization is (at least at first glance) easy to eastimate. If a few miners together controll 50% or more of the hash-power it should be reason for alarm.
The other issue of centralization of (non-mining) nodes is more dificult to gauge.
In order to qualify the discussion of centralisation of nodes we introduce a measure of the centralisation and de-centralisation of these nodes. Do notice that since you can't really tell from the outside if a node is used for mining we will consider all nodes, mining and non-mining alike, the same.

1.1 Why decentralisation

Why do we want decentrilization of nodes in the first place? Well, if all the nodes in the network are centralized in one legal juristiction or in one physical place, it only takes the actions of one government, or one event of flooding or *lightning*strike, to shut down the entire network. So ideally we would like the nodes in the network to be distributed accross the globe.

2 Mathematical model

2.1 First draft

Let's assume we can measure centralization C(N) as a function of all the nodes in the network. Here $N = \{n_1,n_2,\ldots,n_k\}$ is a list of all the nodes in the network at a given time.
We require $C(N) \in [0,1]$, where $C(N) = 1$ would be a completely centralized network, and $C(N) = 0$ would be a completely decentrilized network.
Also, we will assume that the de-centralization as a function of all the nodes in the network is the given as

$$D(N) = 1 - C(N)$$

What would be a situation where everyone would agree that the centralization is zero (and the decentralization is one)?
Let us imagine a situation where every person on the planet is running their own personal full node at their home. Although not very realistic this would surely give us $C(N) = 0$ and $D(N) = 1$.
The other extreme would be zero nodes running[1]. That would indeed be very centralized, so that would should give us $C(N) = 1$ and $D(N) = 0$.
One simple way to model this would be to simply compare the number of running nodes to the number of people on the planet. If C is a list of all countries and p_c is the population of country c, then the total population is give by

$$p_{total} = \sum_{c \in C} p_c$$

This could then work as a first draft

$$C(N) = \frac{1}{p_{total}} \Big(\max(0, p_{total} - n_{total}) \Big) \tag{1}$$

2.2 Countries, states and provinces

The above method won't differentiate between the two situations where

- every person on the planet is running their own node
- a number of nodes equal to the total population are being run in the same datacenter

One way to get around this would be to do the comparisson of number of nodes and population on a country (or state) level.

$$C(N) = \frac{1}{p_{total}} \sum_{c \in C} \Big(\max(0, p_c - n_c) \Big)$$

Where n_c is the number of nodes in country c.
Here the sum over *countries* can be extended with states / provines for the biggest and most populous countries. So for China, India and the USA you could further split into states or regions.

[1] or two nodes running, else it really stops being a network.

2.3 Distribution of nodes, and economy

One could argue that 150 nodes distributed in 150 different countries is vastly more decentrelized than 150 nodes in one country. The function below takes this into account

$$C(N) = \frac{1}{p_{total}} \sum_{c \in C} \frac{\Big(\max(0, p_c - n_c)\Big)^2}{p_c}$$

Finally it makes sence to weigh each country by GDP pr. capita. Let μ_c be the GDP pr. capita of country c. Then $p_c \cdot \mu_c = GDP_c$ and $p_{total} \cdot \mu_{total} = GDP_{total}$.

Now consider

$$C(N) = \frac{1}{p_{total} \cdot \mu_{total}} \sum_{c \in C} \mu_c \cdot \frac{\Big(\max(0, p_c - n_c)\Big)^2}{p_c} \tag{2}$$

This is a measure of the total centralization of the nodes in the network. Let us define the decentralisation as:

$$D(N) = 1 - C(N) = 1 - \frac{1}{p_{total} \cdot \mu_{total}} \sum_{c \in C} \frac{\mu_c \cdot \Big(\max(0, p_c - n_c)\Big)^2}{p_c} \tag{3}$$

3 Exercises

It is left as an exercise for the reader to:

- Write a small program to calculate the value of $D(N)$ of your favorite network using the formula (3) above.
  - Get a list of countries (and / or states), and their population and GDP.
  - Takes a list of all nodes in the network
  - Figure out the country of each node (or state in a country, at least for the three biggest countries, China, India and US).
  - Calculates the value of $D(N)$ from the formula above.

- Given that both the population and GDP of a given country, and indeed the entire world, are changing, consider if it is reasonable to use $D(N,t)$ to tell us something about the changes in de-centralization of the nodes in the network over time. Here t denotes time.
- Locate all spelling mistakes in the text above, and report back to the author :)

`b.unlimited@protonmail.ch`