FRTN30 Network Dynamics

Assignment 1

(Van Duy Dang - va7200da-s)

May 4, 2023

1 Centrality in Input-Output Network of Goods

We know that it is crucial to identify the sectors to which financial aid should be allocated. In this part, 47 sectors during the year 2000 of Sweden and Indonesia will be analyzed. For instance, 'Agriculture', 'Food products', 'Motor vehicles'. The objective is to find the three most central sectors using the following centrality measures:

- The in-degree and out-degree centrality.
- The eigenvector centrality on the largest connected component.
- The Katz centrality, with $\beta = 0.15$ and with two different values on μ.

a) The in-degree and out-degree centrality

We find the central sectors by calculating the sum of the columns and the rows of the adjacency matrix. The *out_degree* and *in_degree* of a node i are defined as follows:

$$w_i = \sum_{j \in \mathcal{V}} W_{ij} \qquad\qquad w_i^- = \sum_{j \in \mathcal{V}} W_{ji}$$

Country Name	Most central sector (in-degree centrality)	Most central sector (out-degree centrality)
Sweden	19 Radio 21 Motor vehicles 43 Other Business Activities	43 Other Business Activities 39 Real estate activities 31 Wholesale & retail trade; repairs
Indonesia	4 Food products 30 Construction 31 Wholesale & retail trade; repairs	31 Wholesale & retail trade; repairs 1 Agriculture 2 Mining and quarrying (energy)

Table 1: Central sectors table based on the in-degree and out-degree centrality.

Sweden is strong in high technologies and "Real estate activities". While Indonesia is strong in agriculture, so the data seems reasonable. In addition, according to [1] and [2], "Construction", "Mining and quarrying", and "Wholesale & retail trade, repairs" are also major contributors to the Indonesian economy.

b) The eigenvector centrality on the largest connected component

We find the eigenvector centrality z by

$$\lambda z = W' z$$

We will use the *eig* function in MATLAB to calculate the eigenvalues and eigenvectors of the adjacency matrix. The eigenvector centrality z is calculated by taking the absolute values of the components of the eigenvector corresponding to the largest eigenvalue.

Country Name	Most central sector (eigenvector centrality)
Sweden	21 Motor vehicles 19 Radio 43 Other Business Activities
Indonesia	4 Food products 32 Hotels & restaurants 1 Agriculture

Table 2: Central sectors table based on the eigenvector centrality.

The sectors of Sweden are the same, only switching the order of "Radio" and "Motor vehicles". On the other hand, "Hotels & restaurants" is added to the central sectors of Indonesia. It seems logical since the tourism industry is also famous there.

c) The Katz centrality

The eigenvector centrality has a limitation that the addition of a self-loop with a large weight to a node can lead to a significant increase in its eigenvector centrality. We overcome this drawback by using the Katz centrality with $\beta \in (0, 1]$ and μ is a non-negative vector to be thought of some intrinsic centrality.

We calculate the Katz centrality vector as

$$z^{(\beta)} = (\frac{1-\beta}{\lambda_W})W'z^{(\beta)} + \beta\mu.$$

where λ_W is the dominant eigenvalue of W'. The centrality vector can be represented as

$$z^{(\beta)} = (I - \lambda_W^{-1}(1-\beta)W')^{-1}\beta\mu.$$

Country Name	Most central sector with $\mu = \mathbb{1}$	Most central sector with with $\mu = 1$ for the "Wholesale & retail trade; repairs" sector only
Sweden	21 Motor vehicles 19 Radio 43 Other Business Activities	31 Wholesale & retail trade; repairs 21 Motor vehicles 19 Radio
Indonesia	4 Food products 32 Hotels & restaurants 1 Agriculture	4 Food products 31 Wholesale & retail trade; repairs 32 Hotels & restaurants

Table 3: Central sectors table based on the in-degree and out-degree centrality.

With $\mu = \mathbb{1}$, all nodes have identical intrinsic centrality and the result is the same as the previous task.

When we set $\mu = 1$ for only the "Wholesale & retail trade; repairs" sector and zero for the rest, then this sector will have a higher value than its value for $\mu = \mathbb{1}$. We can clearly see that in the above table.

2. Influence on Twitter

a) PageRank

In this task, we compute the PageRank iteratively and find the five most central nodes. Firstly, the data needs to be handled before computing:

- A 6893x6881 sparse matrix W created by *spconvert*() is not a square matrix. Therefore, we need to run the command $W(6893, 6893) = 0$; to make W a square matrix.
- In the PageRank algorithm, nodes with out-degree of 0 will have a PageRank score of 0, which can lead to issues in the computation of the scores for other nodes in the graph. To avoid this problem, we check the out-degrees of all nodes. If any node has an out-degree of 0, a self-loop is added to that node by setting the corresponding diagonal element of W to 1. This ensures that all nodes have at least one outgoing edge and are connected to the rest of the graph, which makes the computation of PageRank more robust and stable.

The Pagerank centrality vector can be calculated as:

$$z^{(\beta)} = \beta \sum_{k \geq 0} (1-\beta)^k (P')^k \mu.$$

where P is the normalized adjacency matrix, $\beta = 0.15$ and $\mu = \mathbb{1}$.

We need to choose k so that the algorithm can converge. However, increasing the number of iterations also increases the computational cost of the algorithm. The data that we analyze is small so k = 100 seems a good value to make sure the algorithm converges.

The five most central nodes can be found from the below table with k = 100

Rank	**Username**	**Node**
1	@gustavnilsson	1
2	@AVPapadopoulos	2
3	@Asienfoset	112
4	@Vikingafoset	9
5	@bianca_grossi94	26

Table 4: Central sectors table based on the PageRank centrality.

b) Discrete-time consensus algorithm

Let $\mathcal{G} = (\mathcal{V}, \varepsilon, W)$ be a graph and let $\mathcal{S} \subseteq \mathcal{V}$ be a nonempty subset of nodes.

The normalized weight matrix P is partitioned as

$$P = \begin{array}{cc} \begin{array}{cc} \text{R} & \text{S} \end{array} & \\ \begin{bmatrix} Q & E \\ F & G \end{bmatrix} & \begin{array}{c} \text{R} \\ \text{S} \end{array} \end{array}$$

where $\mathcal{R} = \mathcal{V} \backslash \mathcal{S}$ is the set of regular nodes.

In addition, the state vectors of the distributed averaging is partitioned as

$$x(t) = \begin{array}{c} \begin{bmatrix} \underline{x}(t) \\ u(t) \end{bmatrix} \end{array} \begin{array}{l} \mathrm{R} \\ \mathrm{S} \end{array}$$

where $u(t) \in \mathbb{R}^{\mathrm{S}}$. The distributed averaging with inputs can be calculated as

$$\underline{x}(t+1) = Q\underline{x} + Eu.$$

We choose the two most central nodes as stubborn nodes, node 1 has value 1 and node 2 has value 0. Then we initiate the rest to the neutral value 0.5. Moreover, we choose 3 nodes to plot the opinions to see how it changes over time.

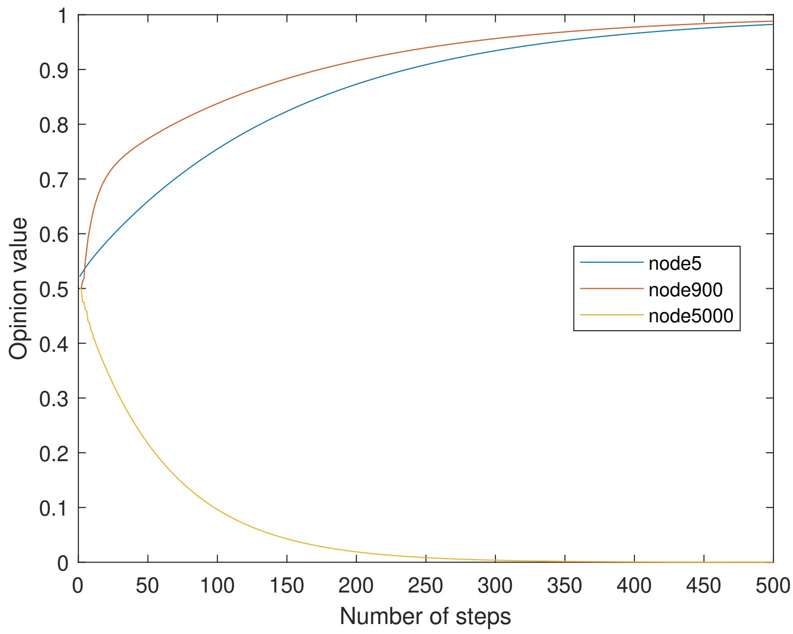

Figure 1: Opinions change over time

As we can see, these nodes start from the neutral value 0.5. After 500 steps, node 5 and node 900 converge to 1. On the contrary, node 5000 converges to 0.

We can draw a histogram to see the opinion distribution

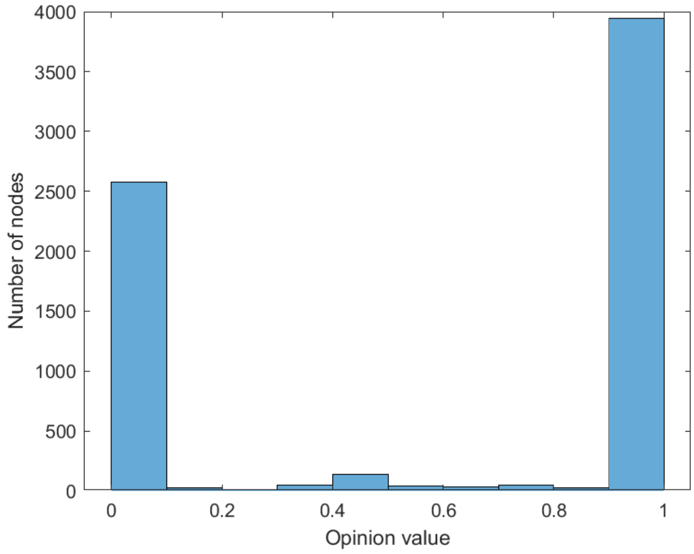

Figure 2: Opinion distribution

It looks logical because node 1 is the most central node so there are many nodes that converge to 1.

2c) Investigate how the choice of nodes with respect to their PageRank change the stationary opinion distribution

We select 3 sets of stubborn nodes to observe how it changes the stationary opinion distribution.

- The third most central node and one random node, 100

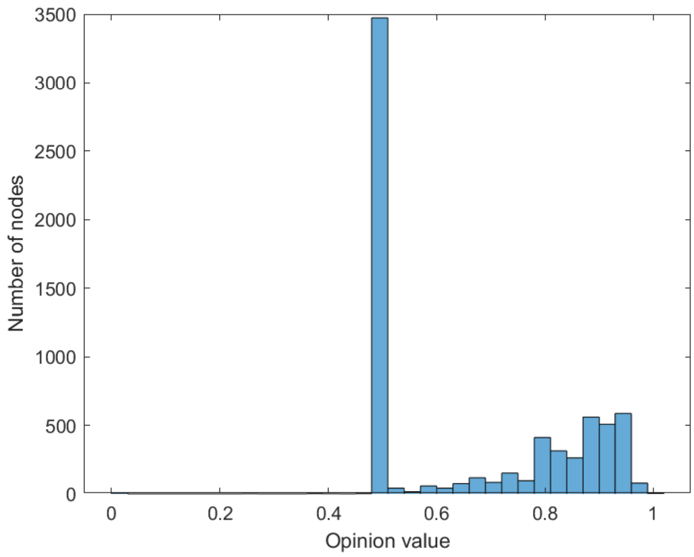

Figure 3: Opinion distribution of set 1

- The least central node and one random node, 1000

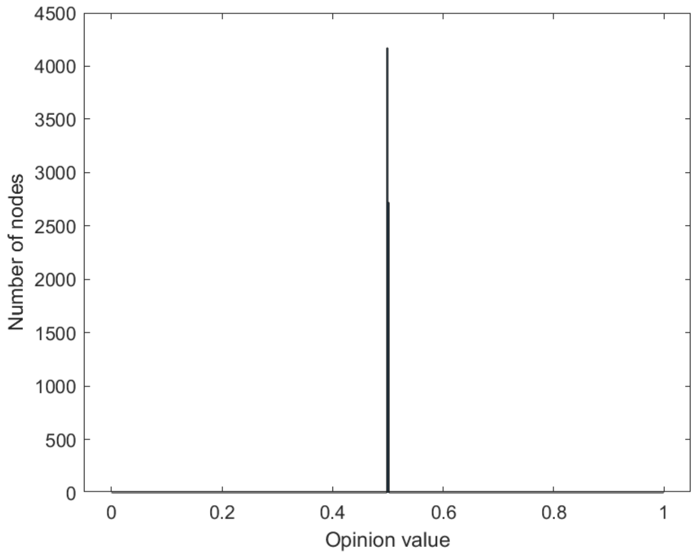

Figure 4: Opinion distribution of set 2

- The 2 random nodes, 2000 and 3000

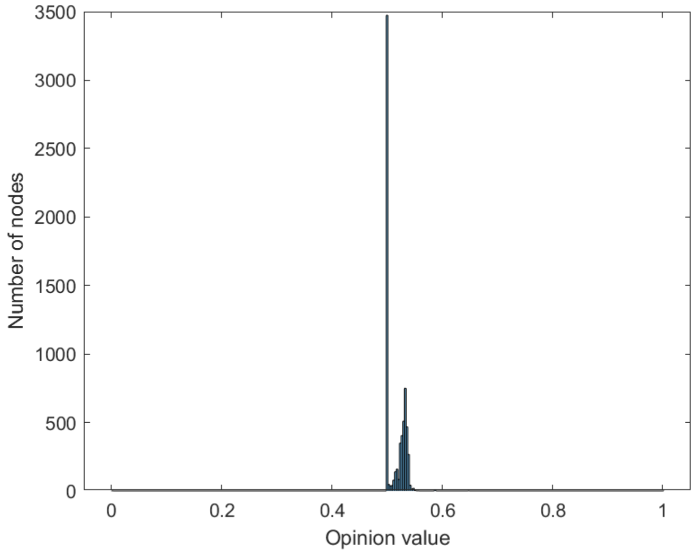

Figure 5: Opinion distribution of set 3

In set 1, we select the third most central node so we still can see part of the opinion values are distributed toward the value 1. On the other hand, in set 2 and 3, the stationary opinion distribution is moved toward the value 0.5 as the stubborn nodes are not too influential.

In my opinion, the opinion values are distributed mostly toward the value 0.5 for all 3 cases, so these distributions converge.

References

[1] https://www.researchinindonesia.com/

[2] https://www.pwc.com/id/en/pwc-publications/industries-publications/energy–utilities—mining-publications/mining-guide-2019.html