Notes of Network Centrality Measures and their Application

Nils Hallerfelt

April 2023

1 Centrality in Input-Output Network of goods

1.1 Introduction

For this problem we will look at the flow of intermediate inputs, i.e. the flows between vertices only contain the sales of goods or services that are directly consumed in the production process. An immediate consequence of this is that the in-flow and out-flow from a given node in the graph is not necessarily equal.

The data is structured as follows: In IOdownload.mat there are three cells. The cell econ contains a list of 39 countries and which year the data is obtained. The struct io contains the 39 input-matrices, one for each country, where each country has a 47×47 matrix such that index (i, j) represents the intermediate flow from sector i to sector j. The data for a single country can thus be modelled as weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V} = \{1, \cdots, 47\}$ are the vertices, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ are the edges, $\mathcal{W} \in \mathbb{R}^{\mathcal{V} \times \mathcal{V}}$ is the weight matrix.

1.2 The graph $\mathcal{G}$ and the Corresponding weight matrix $\mathcal{W}$

Assume we are given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ from the given data. We will first give the graph $\mathcal{G}$ some general properties that helps in the theoretical analysis of it.

1. As the flow from a sector i to a sector j cannot be assumed to be 1 or 0, $\mathcal{G}$ is a weighted.
2. The flow from an edge i to a node j is not necessarily the same as the flow node j to i: $f(i, j) \neq f(j, i)$. So $\mathcal{G}$ is directed.
3. $\mathcal{G}$ can have self-loops (a sector can produce flow that it itself uses).
4. $\mathcal{G}$ is not necessarily balanced, as the intermediate flows between nodes are considered.
5. $\mathcal{G}$ is not strongly connected, as analysis by Tarjans SCC (strongly connected components) algorithm have shown. We will later see how this problem can be handled by examining the condensation graph $\mathcal{H}$ of $\mathcal{G}$.

So we will in the theory assume that $\mathcal{G}$ is weighted, directed, may contain self-loops, and is not strongly connected.

1.3 The weight matrix $\mathcal{W}$ and its associate $\mathcal{P}$

We have from the start a natural connection between our graph $\mathbf{G}$ and its weight matrix $\mathcal{W}$. The properties of $\mathcal{G}$ imposes that $\mathcal{W}$ is in general non-symmetric, $diag(\mathcal{W}) \neq \mathbf{0}$, and $\mathcal{W} \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$. We will now associate two additional matrices to $\mathcal{G}$, the *Normalized weight-matrix* and the *Laplacian matrix*.

1.3.1 The Normalized Weight-matrix $\mathcal{P}$

We will assume without loss of generality that each vertex in $\mathcal{G}$ has positive out-degree, i.e. that $\omega_i = \sum_{j\in\mathcal{V}} \mathcal{W}_{i,j} > 0$ for all i. We can make this assumption, as if some vertex i has zero out-degree we may construct a real-positive self-loop for this vertex. Let $D = diag(\omega)$, where $\omega \in \mathbb{R}^{|\mathcal{V}|}$. The normalized weight matrix is then defined as:

$$P = D^{-1}\mathcal{W} \tag{1}$$

For P to exist, D must be invertible. By the assumption that $\omega_i > 0\,\forall i \in \mathcal{V}$, D is invertible by definition.

Note that P can equivalently be written in indexed form:

$$P_{i,j} = \frac{W_{i,j}}{\omega_i}. \tag{2}$$

This highlights why P is called the normalized weight-matrix: the weight/flow over the edge (i, j) is normalized by the out-weight ω_i (the out-weight of (i, j) tail). Thus the sum over row i of P can be written as:

$$\sum_{j\in\mathcal{V}} P_{i,j} = \sum_{j\in\mathcal{V}} \frac{\mathcal{W}_{i,j}}{\omega_i} = \frac{1}{\omega_i}\sum_{j\in\mathcal{V}} \mathcal{W}_{i,j} = \frac{1}{\sum_{j\in\mathcal{V}} \mathcal{W}_{i,j}} \sum_{j\in\mathcal{V}} \mathcal{W}_{i,j} = 1 \tag{3}$$

Thus D^{-1} can be seen as mapping $D^{-1} : \mathbb{R}^{\mathcal{V}\times\mathcal{V}} \to \Delta^{(|\mathcal{V}|-1)\times|\mathcal{V}|}$, i.e. a mapping to a probability-simplex $\Delta^{|\mathcal{V}|-1}$ for every row $j \in \mathcal{V}$. So P is a stochastic matrix, and an immediate consequence of this is that $P\mathbb{1} = \mathbb{1}$

1.4 Some Relevant Perron-Frobenius Theory

Theorem 1. *Let $M \in \mathbb{R}_+^{n\times n}$. Then there exists a non-negative real eigenvalue λ_M and non-negative $x, y \neq \mathbf{0}$ such that:*

1. $Mx = \lambda_M x$, $M'y = \lambda_M y$.
2. $\max\{\omega_{min}, \omega^-_{min}\} \leq \lambda_M \min\{\omega_{max}, \omega^-_{min}\}$.
3. *Every eigenvalue λ of M is such that $|\lambda| \leq \lambda_M$, i.e. λ_M is the dominant eigenvalue.*

Proof: Giacomo Como and Fabio Fagnani, Lecture Notes on Network Dynamics, page 25.

Corollary 1. *Let $\mathcal{G})(\mathcal{V}, \mathcal{E}, \mathcal{W}$ be a graph, and assume that the out-degree of every vertex i is strictly positive. Then there exists a positive dominant eigenvalue $\lambda_{\mathcal{W}} > 0$ with associated non-negative right eigenvector $\mathbf{v} = \lambda_{\mathcal{W}}^{-1}\mathcal{W}\mathbf{v}$ and left eigenvector $\mathbf{u} = \lambda_{\mathcal{W}}^{-1}\mathcal{W}'\mathbf{u}$.*

Proof: Giacomo Como and Fabio Fagnani, Lecture Notes on Network Dynamics, page 26.

1.5 Some Relevant Network Centrality Measures

In this section some theory about different centrality measures will be given. The centrality measures that will be focused on is the *in/out - centrality*, *Eigenvector centrality*, and *Katz centrality*.

1.5.1 In/Out - Centrality

Maybe the most natural centrality measure for a vertex i is its degree centrality - i.e. the out-degree ω_i and in-degree ω_i^-. By definition of balanced graphs, the in-degree and out-degree coincides if the graph is balanced. However, we will assume that the graph $\mathcal{G}$ is unbalanced, so in general there will exists $i \in \mathcal{V}$ such that $\omega_i \neq \omega_i^-$. Vertices with high in-degree centrality are those that are heavily targeted by other vertices in the graph, and they can be considered to be influential or important in the network. Conversely, vertices with high out-degree centrality are those that send out large quantities, and can as well be seen as influential or important in the network. One of the drawbacks(more will be discussed later) is that the in/out-centrality does not take into account what nodes they are connected to, just how much direct connection it has to other nodes in its direct neighborhood. Each vertex in this neighborhood is in a sense "anonymous", the in/our-centrality does not take into account the centrality/importance of these vertices.

1.5.2 Eigenvector Centrality

One way to overcome the "anonymous" problem is to use the Eigenvector centrality, which extends the in-degree centrality. The idea is rather natural: assume we want a centrality measure for vertex i, and it has connections $(j,i),(k,i)$. If the centrality of vertex j is higher than the centrality of vertex k, then the edge (j,i) should contribute more to the centrality of node i than the link (k,i) does (relatively to the flows over the edges). Lets formalize this idea.

Let u_i be the centrality of vertex i. Furthermore, let $j \in \mathcal{IN}(i)$ (the in-neighborhood of i).

$$u_i = \alpha \sum_{j \in \mathcal{IN}(i)} \mathcal{W}_{j,i} u_j = \alpha \mathcal{W}_{\sim,i} \mathbf{u} \tag{4}$$

where alpha is some proportionality constant. Suppose we do this for each $i \in \mathcal{V}$, and let $\alpha = \dfrac{1}{\lambda} \geq 0$. We then get:

$$\mathbf{u} = \frac{1}{\lambda} \mathcal{W}' \mathbf{u} \Leftrightarrow \mathcal{W}' \mathbf{u} = \lambda \mathbf{u}. \tag{5}$$

So $\mathbf{u}$ is an eigenvector of $\mathcal{W}'$, corresponding to an eigenvalue λ. By choosing $\lambda = \lambda_{\mathcal{W}}$, it follows from Corollary 1 that $\mathbf{u} = \lambda_{\mathcal{W}}^{-1} \mathcal{W}' \mathbf{u}$ is an eigenvector of $\mathcal{W}'$ with corresponding eigenvalue $\lambda_{\mathcal{W}}$

If we assume $\mathcal{G}$ to be strongly connected and impose the normalization $\mathbf{u}\mathbb{1} = 1$, then $\mathbf{u}$ is the unique eigenvector centrality of $\mathcal{G}$.

The general drawback of the eigenvector centrality measure is that vertices, or a collection of vertices, may modify its own centrality by either adding self-loops or loops between each other of arbitrary large weights. Then the centrality of this vertex, or the collection of vertices, grows larger while the nodes outside the "cooperation" stays the same, resulting in an unfair measurement.

1.5.3 Katz centrality

The Katz centrality measurements aims at overcoming the issue of "deceptive" vertices, by allowing some centrality independent of the in-neighborhood (note that a vertex can be in its own in-neighborhood if there are selfloops, $i \in \mathcal{IN}(i)$).

Let μ be a vector of some native centrality, such that μ_i is some non-negative native centrality of node i. Also, let $\beta \in (0, 1]$ be fixed parameter. The Katz centrality $\mathbf{u}^{(\beta)}$ is then defined as the solution to the following equation:

$$\mathbf{u}^{(\beta)} = \frac{1-\beta}{\lambda_{\mathcal{W}}} \mathcal{W}' \mathbf{u}^{(\beta)} + \beta\mu \tag{6}$$

To solve this equation we would like to take the right inverse of $I - \frac{1-\beta}{\lambda_{\mathcal{W}}}\mathcal{W}'$. As $\lambda_{\mathcal{W}}$ is the dominating eigenvalue of $\mathcal{W}'$, and $1 - \beta$ is contracting, the eigenvalues of $\frac{1-\beta}{\lambda_{\mathcal{W}}}\mathcal{W}'$ must be smaller than one and the matrix will be invertible. Thus the Katz centrality will always be unique under these assumptions on $\mathcal{G}$, and can be determined as:

$$\mathbf{u}^{(\beta)} = \left(I - \frac{1-\beta}{\lambda_{\mathcal{W}}}\mathcal{W}'\right)^{-1} \beta\boldsymbol{\mu} \tag{7}$$

From this we can see the interpretation of parameter β. If beta equals its upper limit 1, $\mathcal{W}_s$ disappears from the equation determining $\mathbf{u}_s^{(\beta)}$, so then the Katz centrality does not take the graph structure in account at all. On the other hand, if we let $\beta \to 0$, the Katz centrality reduces to the eigenvector centrality. In this sense, the Katz centrality is a generalization of the eigenvector centrality, where β is a parameter allowing us to choose how much of the graph topology we want to take into account.

1.6 Results

1.6.1 In/Out Centrality

To calculate the the out-centrality and in-centrality, we calculated $w = \mathcal{W}\mathbb{1}$, $w^- = \mathcal{W}'\mathbb{1}$ for both Indonesia and Sweden. Then the three most central nodes were picked for both countries. The results for Indonesia and Sweden can be seen in the first two rows of Table 1 and Table 2, respectively.

High out-centrality would indicate that many other sectors are using this sectors commodities in their production, while high in-centrality would indicate that this sector is important many other sectors in terms consuming. So the sectors with high out-centrality could be a good place for financial aid of logistical/structural importance, so that the production system would not fail. The in-centrality would be of importance if the aid is to avoid lower demand in the system (if a sector with high in-centrality fails many other would take damage in their selling pipelines), maybe causing waves of problem's. Of course further economical analysis of the actual state of each sector would also be needed.

1.6.2 Eigenvector Centrality

The eigenvector centrality for both Indonesia and Sweden was calculated for the largest strongly connected component of the graph for each country. To find the strongly connected components of each graph, *Tarjan's strongly connected components algorithm* was used. Then $V^* = \text{argmax}_{V_i \in \text{scc}(\mathcal{W})}\{|V_i|\}$ was chosen as the largest scc. The eigenvector corresponding to the dominant eigenvalue of $\mathcal{W}_*'$ (the transpose of the weight matrix associated to the V^* subgraph) was calculated:

$$\mathcal{W}_*'\mathbf{u} = \lambda_{\mathcal{W}_*'}\mathbf{u}$$

After imposing the normalization $\mathbf{u}'\mathbb{1} = 1$, and noting that as $\mathcal{W}_*'$ is the weight matrix of a strongly connected component, this measure of centrality will be unique on the subgraph associated to the weight matrix $\mathcal{W}_*'$. Then the three most central nodes for Indonesia and Sweden according to this measure were extracted. The results for Indonesia and Sweden can be seen in the third row of Table 1 and Table 2, respectively.

Eigenvector centrality is a measure of node importance in a network based on the concept that connections to high-scoring nodes contribute more to a node's score than connections to low-scoring nodes.

Interpreting the results of eigenvector centrality involves understanding the scores assigned to each node and the network structure that contributes to those scores. In general, nodes with higher eigenvector centrality scores are considered more important or influential in the network, as they are connected to other important nodes. These nodes may be important hubs or connectors within the network. As such, the analysis is much like the in-centrality, with the extra information about whether the node contributes to the selling pipeline of other important nodes. This is of course very relevant information for where financial aid would go, as it takes the networks centrality structure into account, and more important sectors should be prioritized over less important sectors (the argument is kind of recursive :)).

1.6.3 Katz Centrality

Lastly, the Katz-centrality for Sweden and Indonesia was calculated for

$$\beta = 0.15, \boldsymbol{\mu}_1 = \mathbb{1}, \boldsymbol{\mu_2} = \begin{cases} \mu_i = 0 \text{ for } i \neq 31 \\ \mu_{31} = 1 \end{cases}$$

using equation 7. The results for Indonesia and Sweden can be seen in the fourth and fifth row of Table 1 and 2, respectively.

One advantage of Katz centrality is that it can capture the influence of nodes that may not have many direct connections but are still important because they are part of longer chains of connections. This can be useful in identifying nodes that may be important mediators or bridges between different parts of the network.

Indonesia			
Centrality Measure	First	Second	Third
Out	'31 Wholesale retail'	'1 Agriculture'	'2 Mining'
In	'4 Food products'	'30 Construction'	'31 Wholesale retail'
Eigenvector	'4 Food products'	'32 Hotels, rest.'	'1 Agriculture'
Katz 1	'4 Food products'	'32 Hotels, rest.'	'1 Agriculture'
Katz 2	'4 Food products'	'31 Wholesale retail'	'32 Hotels, rest.'

Table 1: The five most central sectors in Indonesia for the different centrality measures. Katz 1 uses μ_1 and Katz 2 uses μ_2.

Sweden			
Centrality Measure	First	Second	Third
Out	’43 Other Business’	’39 Real estate’	’31 Wholesale retail’
In	’19 Radio’	’21 Motor vehicles’	’43 Other Business’
Eigenvector	’21 Motor vehicles’	’19 Radio’	’43 Other Business’
Katz 1	’21 Motor vehicles’	’19 Radio’	’43 Other Business’
Katz 2	’31 Wholesale retail’	’21 Motor vehicles’	’19 Radio’

Table 2: The five most central sectors in Sweden for the different centrality measures. Katz 1 uses μ_1 and Katz 2 uses μ_2.

2 Influence on Twitter

Here a subgraph of the Twitter network will be analysed. The data was obtained by crawling from a subset of an initial users followers, then those followers followers, etc. The data is in a three-column table with links, where the first two columns represent the tail and the head of the link, and the third column the weight of the link. Here a link (i, j) represents that i follows j. For our purposes, this data was converted to a weight matrix (making it square by appending zeros). Furthermore, as the normalized weight matrix will be used extensively in the analysis, the weight matrix was adjusted so that it contains self-loops for nodes which had zero out-degree, as discussed in section 1.3.1. After these adjustments, the twitter graph is denoted $\mathcal{G} = \big(\mathcal{V}, \mathcal{E}, \mathcal{W}\big)$.

2.1 Results

2.1.1 Iterative PageRank-centrality

The PageRank centrality (or the more general Bonacich) measure has a close resemblance to the Katz centrality measure. The difference is that, when doing the PageRank, one is using the normalized weight matrix P' instead of the weight matrix $\mathcal{W}'$. This gets us:

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

Observe that, as the dominant eigenvalue of P is less than 1, $I - (1-\beta)P'$ is invertible. We thus find the solution according to:

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

We can also express the solution as a geometric series expansion to get an iterative solution:

$$z^{(\beta)} = \sum_{k\geq 0} \beta(1-\beta)^k (P')^{(k)}\mu. \tag{8}$$

For the PageRank-algorithm (8) the parameters was set to $\beta = 0.15$, and the intrinsic centrality $\mu = \mathbb{1}$. The five most central nodes according to the PageRank-measure can be found in Table 3. The iteration was stopped when no nodes PageRank-value was updated by more than 10^{-18}.

Twitter Subgraph			
	Node	User	PageRank-Value
First	1	'@gustavnilsson'	718
Second	2	'@AVPapadopoulos'	122
Third	112	'@Asienfoset'	82
Fourth	9	'@Vikingafoset'	71
Fifth	26	'@bianca_grossi94'	63

Table 3: The five most central nodes in the Twitter subgraph, according to the PageRank-centrality measure. The corresponding username and the PageRank-value is also presented.

2.1.2 Consensus-Algorithm with Two Stubborn Nodes

In this section the convergence of five randomly chosen nodes was analysed using a consensus algorithm (distributed linear averaging) with inputs. The inputs were two stubborn nodes, node 1 and node 2. These were given opinions 0 and 1 respectively, i.e. opposite opinions. The randomly chosen nodes were $1420, 4854, 2497, 6573, 194$. In Figure 1, all regular nodes in the graph was initialized to be indifferent in their opinions, which corresponds to a value of 0.5. Then the same procedure was repeated, but this time the regular nodes were given opinions uniformly random between 0 and 1.

Comparing Figure 1 and Figure 2, we can see that they converge towards the same values. FOLLOWS FROM PROPOSITION; S GLOBALLY REACHABLE. However, the convergence behaves somewhat different in the beginning, where we can see that the opinions oscillate more for the initialized random opinions.

Notice that a disagreement among the agents occurs in both cases, not surprisingly as two highly influential nodes are given opposite opposite stubborn opinions.

2.1.3 Choosing Stubborn Nodes w.r.t. their PageRank

In this section we will investigate how the stationary distribution changes for different inputs of stubborn nodes. The choices of stubborn nodes are done with respect to their PageRank-centrality, where in particular different combinations PageRank-centrality's are of interest.

To do this, the normalized weight matrix was partitioned into blocks, according to:

$$P = \begin{bmatrix} Q & E \\ F & G \end{bmatrix}, \quad x(t) = \begin{bmatrix} \underline{x}(t) \\ u(t) \end{bmatrix}$$

and the distributed linear averaging becomes, by having the stubborn nodes in E and u a constant vector:

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

To find the stationary distribution, one can either make use of a iterative procedure, or solve the equation directly. Here, the equation was solved directly:

$$\underline{x} = (I - Q)^{-1} Eu$$

Firstly we look at the stationary distribution when both stubborn nodes have high PageRank. To be more precise, the two nodes with highest PageRank were chosen, i.e. node 1 and node 2. Node 1 was given opinion 1, and node 2 was given opinion 0. The result can be seen in Figure 3. We can here see a disagreement.

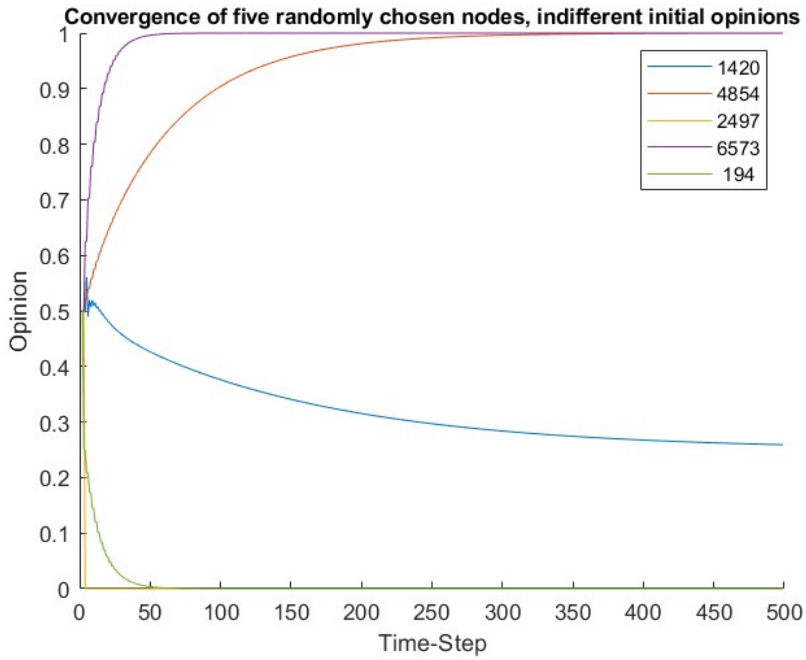

Figure 1: The convergence of five randomly chosen nodes in the graph, with an input of the two stubborn nodes 1 and 2. Each node in the graph had an initial opinion of 0.5 (indifferent opinion). For illustrating purposes only the first 500 time-steps are plotted.

Then we chose the stubborn nodes such that Node 1 had opinion 1 and Node 6893 had opinion 0. Node 1 had the highest PageRank, and Node 6893 had the lowest (tie). The result can be seen in Figure 4. Here consensus is reached among the agents.

The last test that was done using two nodes in the middle of the pack. Although in the middle of the pack, both these nodes have a low PageRank-value, around 0.17. The result is a disappointing NaN, as the the matrix $(I - Q)$ is not invertible, so the stationary distribution does not seem to exist. This indicates that the set of stubborn nodes are not globally reachable, as if they were, a stationary distribution would exist.

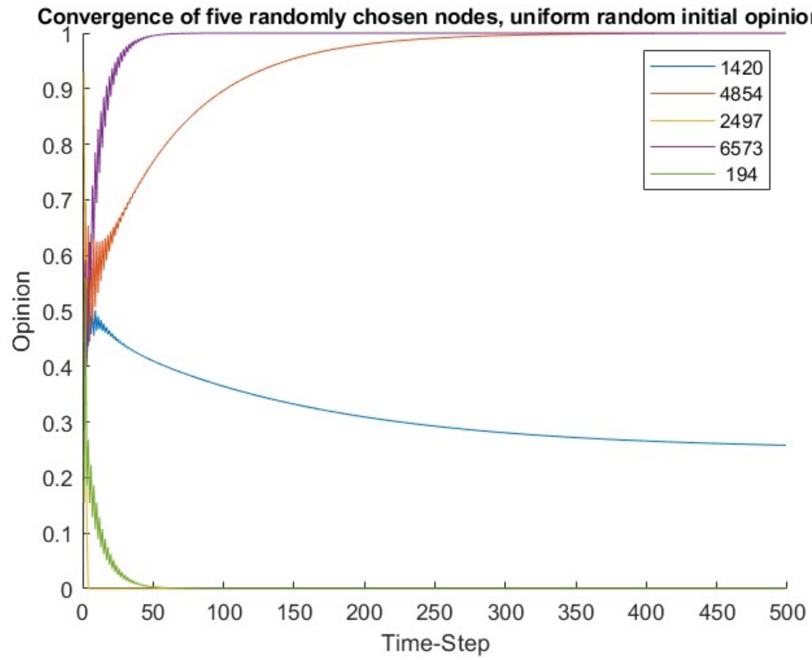

Figure 2: The convergence of five randomly chosen nodes in the graph, with an input of the two stubborn nodes 1 and 2. Each node in the graph was given a uniformly random opinion. For illustrating purposes only the first 500 time-steps are plotted.

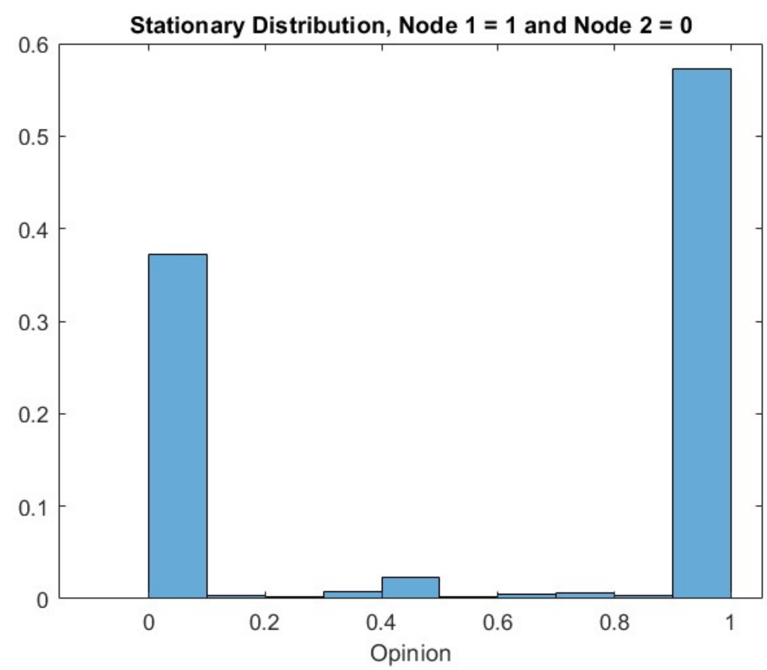

Figure 3: Stationary distribution using two highly influential nodes (according to the PageRank-measure).

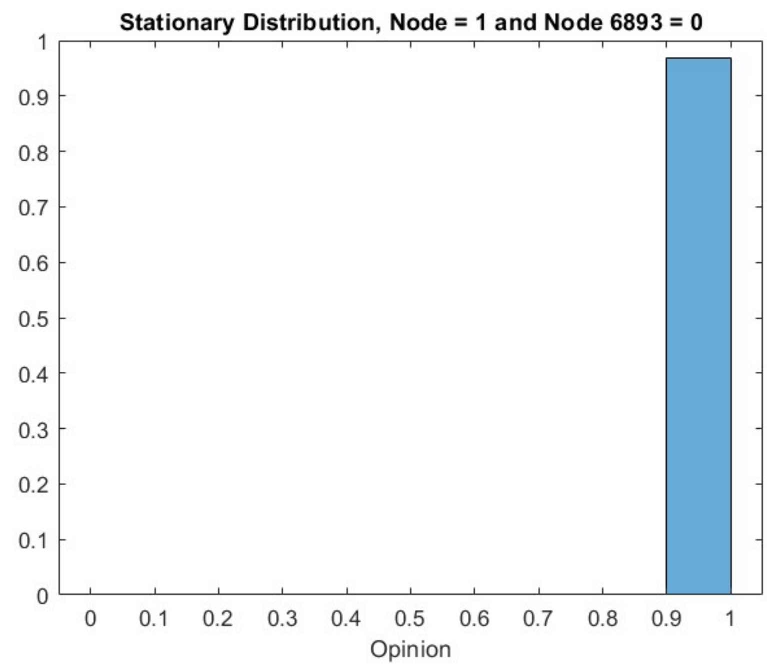

Figure 4: Stationary distribution using one highly influential node and one with low influence. The distribution does not sum to 1 because of numerical precision issues.