FRTN30 Network Dynamics: Hand-In 1

Alfred Bornefalk, al1718bo-s@student.lu.se, 2022-05-06

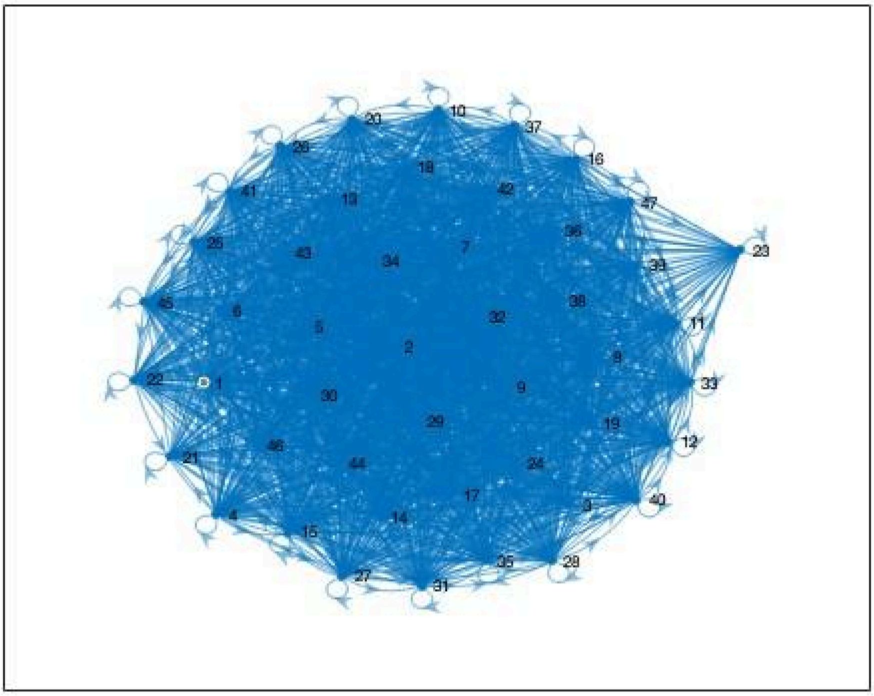

Table of Contents

1. Centrality in Input-Output Network of Goods

In the first section of the assignment, we are asked to determine the three most central sectors in Sweden and Indonesia, respectively, based on different centrality measures. All values concern the year 2000. In all of the centrality measures investigated, we will consider the *weighted* adjacency matrix. This is because the analysis carried out when working with the unweighted adjacency matrix barely yields any satisfying conclusions. For instance, in the case of Sweden, all 41 connected sectors are deemed as important (central) for the Swedish economy for each centrality measure. Additionally, when comparing the relative centrality of the most central nodes for each measure, the analysis becomes more precise when the weighted adjacency matrix is utilized.

1.1. In-Degree and Out-Degree Centrality

Here, we are asked to compute the three most central sectors based on in-degree and out-degree centrality, respectively. These measures are, respectively, defined as

$$w_i^- = \sum_{j \in V} W_{ji}, \text{ and } w_i = \sum_{j \in V} W_{ji}$$

In undirected graphs, there is no distinction between in- and out-degree. In our case, the graphs for both Sweden and Indonesia are directed, enabling us to use both measures without forced duplication of the results.

1.1.1. Sweden

After performing the in-degree centrality calculations for the weighted adjacency matrix for the Swedish economy, the degrees were sorted in descending order. The three most central sectors – i.e., the corresponding index to three largest elements in the sorted array – based on in-degree centrality are presented in the table below.

Table 1: In-degree centrality in Sweden.

Rank	Sector	Centrality
1	Radio	1.42×10^5
2	Motor Vehicles	1.41×10^5
3	Other Business Activities	1.35×10^5

Table 1 showcases that the three most central sectors in Sweden, based on the in-degree centrality given by the weighted adjacency matrix, are, respectively, *Radio*, *Motor Vehicles*, and *Other Business Activities*. Comparing their relative centrality, we can see that their centralities are rather similar, with radio and motor vehicles for instance differing by less than one percent.

The same approach as described with the in-degree centrality was applied to determine the three most central sectors according to out-degree centrality. Again, the table is given below.

Table 2: Out-degree centrality in Sweden.

Rank	Sector	Centrality
1	Other Business Activities	2.58×10^5
2	Real Estate Activities	1.15×10^5
3	Wholesale & Retail Trade: Repairs’	1.06×10^5

Table 2 showcases that the three most central sectors in Sweden, based on the out-degree centrality given by the weighted adjacency matrix, are, respectively, *Other Business Activities*, *Real Estate Activities*, and *Wholesale & Retail Trade: Repairs’*. It can be noted that *Other Business Activities* also appeared in table 1. This time around, said sector also stands out by its relative centrality, which is more than double the amount of *Real Estate Activities*.

1.1.2. Indonesia

The exact same approach for both in-degree and out-degree centrality was computed for Indonesia; the only difference being that the weighted adjacency matrix concerned the Indonesian economy. The three most central sectors based on in-degree centrality are presented in the table below.

Table 3: In-degree centrality in Indonesia.

Rank	Sector	Centrality
1	Food Products	2.15×10^8
2	Construction	1.51×10^8
3	Wholesale & Retail Trade: Repairs’	1.14×10^8

Table 3 showcases that the three most central sectors in Indonesia, based on the in-degree centrality given by the weighted adjacency matrix, are, respectively, *Food Products*, *Construction*, and *Wholesale & Retail Trade: Repairs’*. Looking at the relative centralities, we can see that *Food Products* have a centrality in excess of 40 percent above *Construction*, and almost twice as high as *Wholesale & Retail Trade: Repairs’*.

The same approach as described with the in-degree centrality was applied to determine the three most central sectors according to out-degree centrality. Again, the table is given below.

Table 4: Out-degree centrality in Indonesia.

Rank	Sector	Centrality
1	Wholesale & Retail Trade: Repairs'	1.78×10^8
2	Agriculture	1.75×10^8
3	Mining and Quarrying (Energy)	1.10×10^8

Table 4 showcases that the three most central sectors in Indonesia, based on the in-degree centrality given by the weighted adjacency matrix, are, respectively, *Wholesale & Retail Trade: Repairs*, *Agriculture*, and *Mining and Quarrying (Energy)*. It can be noted that *Wholesale & Retail Trade: Repairs'* also appeared in table 4. Additionally, the sector appeared in the Swedish economy with the out-degree centrality measure (table 2). The difference in percentage terms between *Wholesale & Retail Trade: Repairs* and *Agriculture* is small, whereas both of these sectors have a centrality score more than 50 percent above *Mining and Quarrying (Energy)*.

1.2. Eigenvector Centrality on the Largest Connected Component

Here, we are asked to compute the three most central sectors based on the eigenvector centrality on the largest connected component. Theorem 2.1 in the lecture notes states that, given a non-negative square matrix *M*, there exists a non-negative real eigenvalue $\lambda_M > 0$ and non-negative vectors x and y such that

$$Mx = \lambda_M x, M'y = \lambda_M y$$

Additionally, it is important to recall that the splitting in connected components constitutes a partition of the node set V, i.e. we have that

$$V = V_1 \cup \dots \cup V_k, V_h \cap V_l = \emptyset, h \neq l$$

Thus, the largest connected component is simply the connected component with the largest number of nodes.

1.2.1. Sweden

After performing the eigenvector centrality calculations for the weighted adjacency matrix of the largest connected component for the Swedish economy, the degrees were sorted in descending order. The three most central sectors, together with their centrality score, are presented in the table below.

Table 5: Eigenvector centrality of the largest connected component in Sweden.

Rank	Sector	Centrality
1	Other Business Activities	.7072
2	Pulp	.2925
3	Real Estate Activities	.2169

Table 5 showcases that the three most central sectors in Sweden, based on the eigenvector centrality given by the weighted adjacency matrix of the largest connected component, are, respectively, *Other Business Activities*, *Pulp*, and *Real Estate Activities*. The first mentioned has appeared as one of the most central sectors in the Swedish economy in both previous measures, and the last one mentioned ranked second in out-degree centrality. Looking at the values of their centrality, it is further suggested that *Other Business Activities* is a crucial sector at this point in time, since it is, with margin, more than twice as high as *Pulp*, and also more than three times bigger than the score of *Real Estate Activities*.

1.2.2. Indonesia

The exact same approach for the eigenvector centrality based on the largest connected component (which happened to be the same as the previous graph used; the economy at this time was already connected) was computed for Indonesia; the only difference being that the weighted adjacency matrix concerned the Indonesian economy. The three most central sectors based on said centrality are presented in the table below.

Table 6: Eigenvector centrality of the largest connected component in Indonesia.

Rank	Sector	Centrality
1	Agriculture	.7658
2	Food Products	.4588
3	Wholesale & Retail Trade: Repairs’	.3851

Table 6 showcases that the three most central sectors in Indonesia, based on the eigenvector centrality given by the weighted adjacency matrix of the largest connected component, are, respectively, *Agriculture*, *Food Products*, and *Wholesale & Retail Trade: Repairs’*. All of these have appeared previously in the centrality measures for the Indonesian economy. Scrutinizing their respective centrality values, *Agriculture* stands out, being around ⅔ higher than *Food Products*, and roughly twice the size of *Wholesale & Retail Trade: Repairs’*.

1.3. Two Different Calculations of Katz Centrality

Here, we are asked to compute the three most central sectors based on two different Katz centralities. In both cases, $\beta = 1$; in the first case, $\mu_i = 1 \forall i$, and in the other case, $\mu_i = 1$ for the sector *Wholesale & Retail Trade: Repairs'*, and 0 otherwise. The Katz centrality vector is unique and given by

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

where λ_W is the dominant eigenvalue of W'.

1.3.1. Sweden

After performing the Katz centrality calculations for the weighted adjacency matrix in case 1 for the Swedish economy, the degrees were sorted in descending order. The three most central sectors, together with their centrality score, are presented in the table below.

Table 7: Katz centrality in Sweden (case 1).

Rank	Sector	Centrality
1	Radio	2.34×10^5
2	Motor Vehicles	2.32×10^5
3	Other Business Activities	2.22×10^5

Table 7 showcases that the three most central sectors in Sweden, based on the Katz centrality (case 1) given by the weighted adjacency matrix, are, respectively, *Radio, Motor Vehicles*, and *Other Business Activities*. The difference between *Radio* and *Motor Vehicles* is almost miniscule, with the former being one percent above the latter, and *Other Business Activities* only trailing *Radio* with roughly five percent. An interesting observation is that the ranking (in terms of the top three, at least) concluded from this centrality measure coincides with the one based on in-degree centrality.

In case 2, the weights varied as explained in the introduction of this subsection; sorting the degrees in descending order yielded the table provided below.

Table 8: Katz centrality in Sweden (case 2).

Rank	Sector	Centrality
1	Motor Vehicles	1.52×10^4
2	Radio	1.35×10^4

3	Machinery & Equipment	1.15×10^4

Table 8 showcases that the three most central sectors in Sweden, based on the Katz centrality (case 2) given by the weighted adjacency matrix, are, respectively, *Motor Vehicles, Radio*, and *Machinery & Equipment*. It can be noted that both *Motor Vehicles* and *Radio* also appeared in table 1 and table 7. Rather surprisingly, *Wholesale & Retail Trade: Repairs'* does not appear in the top three, although its intrinsic centrality far exceeds the others (being set to one and the rest to zero in μ). An interpretation of this fact could be that when *Wholesale & Retail Trade* is given a large intrinsic centrality, as compared to case 1 where all sectors are given equal intrinsic centralities, its neighbors become positively impacted. If said neighbors then are central in terms of in- and out-degree for instance, this could then result in *Wholesale & Retail Trade: Repairs'* not appearing amongst the most important central sectors. (Here, it was calculated to be the eighth most central one, which is still in the top 20 percent.)

1.3.2. Indonesia

The exact same approach for the Katz centrality was computed for Indonesia; the only difference being that the weighted adjacency matrix concerned the Indonesian economy. After performing the Katz centrality calculations for the weighted adjacency matrix in case 1 for the Indonesian economy, the degrees were sorted in descending order. The three most central sectors, together with their centrality score, are presented in the table below.

Table 9: Katz centrality in Indonesia (case 1).

Rank	Sector	Centrality
1	Food Products	2.00×10^8
2	Construction	1.41×10^8
3	Wholesale & Retail Trade: Repairs'	1.06×10^8

Table 9 showcases that the three most central sectors in Indonesia, based on the Katz centrality (case 1) given by the weighted adjacency matrix, are, respectively, *Food Products, Construction*, and *Wholesale & Retail Trade: Repairs'*. As was the case when computing the in-degree centrality, *Food Products* stand out between the three, and *Construction* is also equipped with roughly a ⅓ higher centrality score than *Wholesale & Retail Trade: Repairs'*. Once more, the top three ranking concluded from this centrality measure coincides with the one based on in-degree centrality.

In case 2, the weights again varied as explained in the introduction of this subsection; sorting the degrees in descending order yielded the table provided below.

Table 10: Katz centrality in Indonesia (case 2).

Rank	Sector	Centrality
1	Food Products	3.83×10^7
2	Hotels & Restaurants	1.87×10^7
3	Construction	1.75×10^7

Table 10 showcases that the three most central sectors in Indonesia, based on the Katz centrality (case 2) given by the weighted adjacency matrix, are, respectively, *Food Products*, *Hotels & Restaurants*, and *Construction*. The first and last of these have appeared previously in the centrality measures for the Indonesian economy. Remarkably, *Food Products* have a centrality more than twice the size of *Hotels & Restaurants*. We can once again observe the counter-intuitive result that *Wholesale & Retail Trade: Repairs'* does not appear in the table although it was given the intrinsic centrality previously discussed. (This time around it came close however, being ranked as the fourth most central sector.) Assuming that the calculations performed by the novice who is handing in this very assignment, the discussion previously is still relevant, i.e. that the intrinsic importance of *Wholesale & Retail Trade: Repairs'* could benefit some of the other nodes in the network even more.

1.4. Concluding Comment

In Sweden, there are three sectors that appear amongst the top three most important sectors in more than two centrality measurements; these are *Radio*, *Motor Vehicles*, and *Other Business Activities*. Thus, these are deemed to be the most important sectors in the Swedish economy for the time period being investigated, perhaps with *Other Business Activities* being the single most important one due to the fact that its centrality score sometimes is extremely dominant (see table 2 and 5).

In Indonesia, there are two sectors that appear amongst the top three most important sectors in four centrality measurements; these are *Food Products*, and *Wholesale & Retail Trade: Repairs'*. Naturally, these are deemed to belong to the three most important sectors in the Indonesian economy for the time period being investigated. Separating the two in terms of the most important one is rather difficult, since both of them dominate at least one centrality measure. Furthermore, *Construction* is the only sector to appear thrice. However, *Agriculture* appears twice, and is also at one point (table 6) deemed to be the most important sector, and quite significantly so. Since this is never the case for *Construction*, *Agriculture* is included in the overall three most important sectors, behind the aforementioned two.

Again, all the conclusions presented in the first section are under the assumptions that the computations are in fact correct. If one or more of the ten tables above are incorrect, which they very well may be, the analysis is subject to change. With the information at hand, the author has presented the "best guess", as is (almost) always the case in scientific reports.

2. Influence on Twitter

In the second section of the assignment, we are provided with a subgraph of the Twitter network, where a link (i, j) means that i follows j. More specifically, our subset contains a bit over 6,000 users, and we will investigate the importance of the nodes (users) and how the opinions change based on different manipulations of the network.

2.1. The Five Most Central Nodes According to PageRank

Here, we are asked to *iteratively* compute the PageRank, with $\beta = .15, \mu = 1 \forall i$, and ultimately conclude which the five most central nodes are. PageRank is a centrality vector that was first introduced to measure the relative importance of webpages in the WWW, and it its centrality vector can be expressed as

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

With P' in our case being the transformed adjacency matrix obtainable from the provided data file, the most central nodes were iteratively calculated and then sorted. The results from these operations are presented in the table below.

Table 9: PageRank of the subgraph of Twitter users.

Rank	User
1	876641
2	99123165
3	483996827
4	1128920917
5	4417595361

Table 9 showcases that the five most central users in our Twitter subgraph, based on the PageRank centrality are, respectively, *876641*, *99123165*, *483996827*, *1128920917*, and *4417595361*.

2.2. How the Opinions Change Over Time

Now, we are to simulate the discrete-time consensus algorithm with two stubborn nodes (one taking the value of 1, and the other the value of 0). All the other initial opinions for the rest of the nodes (Twitter users) are drawn uniformly between 0 and 1. With the provided Twitter network, we have that there exists a unique invariant probability distribution π such that

$$\pi = P'\pi$$

It is particularly implied that

$$\pi' x(t) = \pi' x(0), t = 0, 1, \ldots$$

Thus, as t grows larger, we get that, if the opinion vector $x(t)$ actually converges, it does so to a consensus vector in which $x_i = \alpha_i \forall i$. It is this (potential) consensus vector that we aim to find in this subsection.

We begin our analysis by arbitrarily choosing two stubborn nodes, one with value 1, and the other with value 0. Then, to initialize the actual stubbornness, we remove the edges from said nodes (since their opinion by construction shall never change), and then create an adjacency matrix W out of our modified graph G. Since the non-stubborn nodes can copy the opinions of the stubborn nodes, our adjacency matrix is still square.

To proceed, we, again, choose three arbitrary nodes to observe. Then, we calculate the normalized adjacency matrix P by first calculating the out-degree for all nodes. Thereafter, in order to address the nodes with an out-degree of 0, we add a self-loop to every node in the system. This is indeed a realistic modeling choice, since there ought to be cases where a user interacts with other users during a day and still keeps his or her opinion. If not, it tells us something about the integrity of people's opinions. Finally, after updating the out-degree vector such that it neither has any elements equal to zero, we obtain P' by multiplying a diagonal matrix with the out-degrees as diagonal elements with the adjacency matrix. Hence, the normalized adjacency matrix is square with the same number of rows and columns as W.

Now, we have reached the actual simulation phase. We randomize all the users initial opinion uniformly between 0 and 1, with the exception of the stubborn nodes. (By design, there will be nothing random regarding these values at all.) As mentioned previously, when t grows larger, we will get a consensus vector *if* $x(t)$ actually converges. For illustrative purposes, 300 iterations were chosen to be computed, where π is continuously updated by multiplying the transposed normalized adjacency matrix with the current probability distribution. The resulting time plot of the changes of opinions is displayed in the figure below.

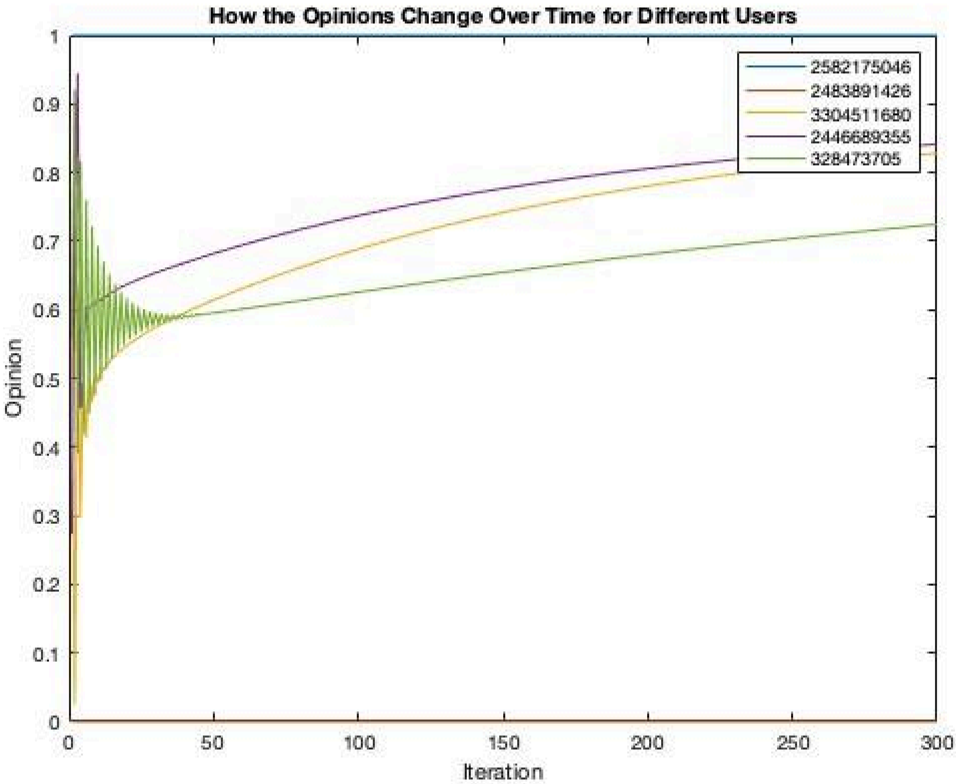

Figure 1: Plot of how the opinions change over time for different users. The stubborn node with value 1 is user 2582175046 (blue), whilst the stubborn node with value 0 is user 2483891426 (orange). The purple, yellow, and green graphs represent, respectively, users 2446689355, 3304511680, and 328473705.

Figure 1 shows that, as per design, the opinions of the stubborn nodes 2582175046 and 243891426 remain constant. For the other Twitter users, the opinions for this set of initial randomized opinions seem to converge to a value approximately equal to .85. An amount of iterations exceeding 300 would show this even clearer, but with the loss of insights in the more dramatic changes in the opinions for the first few iterations. (In particular, user 328473705 converges much slower than the two other non-stubborn users.) Clearly, figure 1 suggests that non-stubborn users on Twitter bear transient opinions, which means that the placement of stubborn nodes in the network may greatly impact the overall opinions.

2.3. Changes to the Stationary Opinion Distribution

In the very last part of this assignment, we are to investigate how the choice of nodes with respect to their PageRank changes the stationary opinion distribution. Three different cases were simulated; in the first, the node with the highest PageRank has the stubborn opinion 1, and the second highest ranked bears the stubborn opinion 0. Similarly, in the second case, the most central node is fixed at 1, whereas the fifth most central is fixed at 0. Finally, the third case is designed such that the third most central node according to the PageRank criteria has the stubborn opinion 1, and the fourth one the opinion 0. The simulated stationary opinion distributions of these three pairs of choices are displayed in the figures 2 to 4.

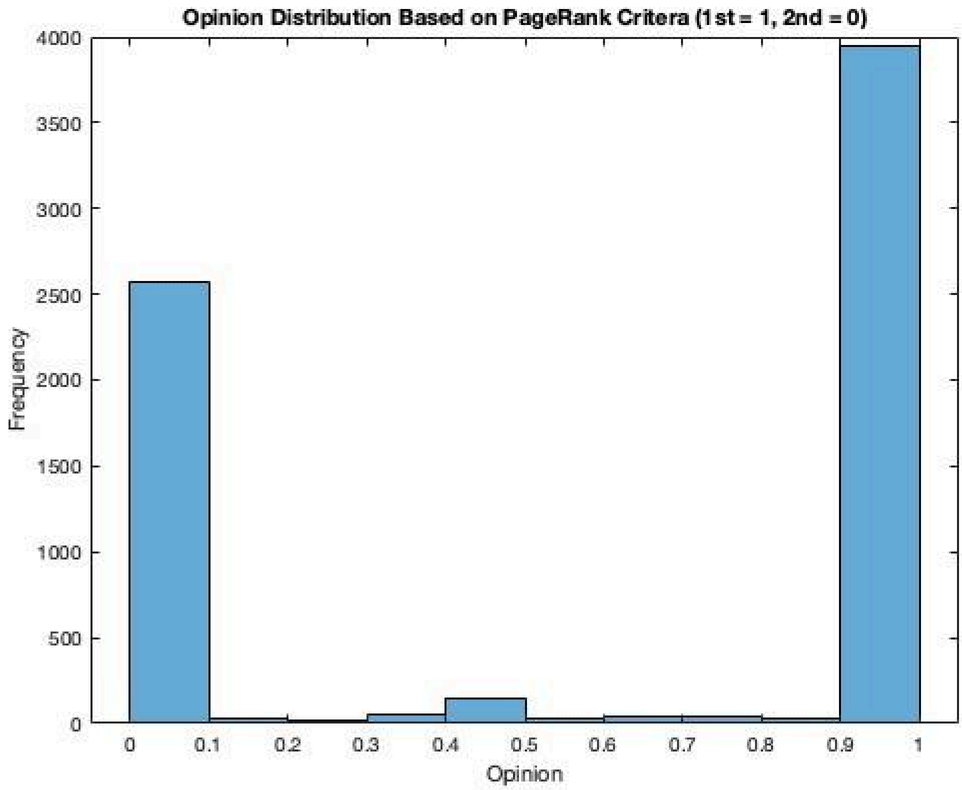

Figure 2: Histogram plot of the opinion distribution based on the PageRank criteria, where the most central user holds the stubborn opinion 1, and the second most central one holds the opinion 0.

From figure 2, we can see that the user ranked first with PageRank indeed poses more influence on the Twitter network at hand than the Twitter user who ranked second, since the opinion distribution is skewed to the right. Comparing the histogram staple furthest to the right (opinions closest to the most central user) with the staples to the left (opinions closest to the second most central user), we can see that the right-hand staple is almost 60% as big.

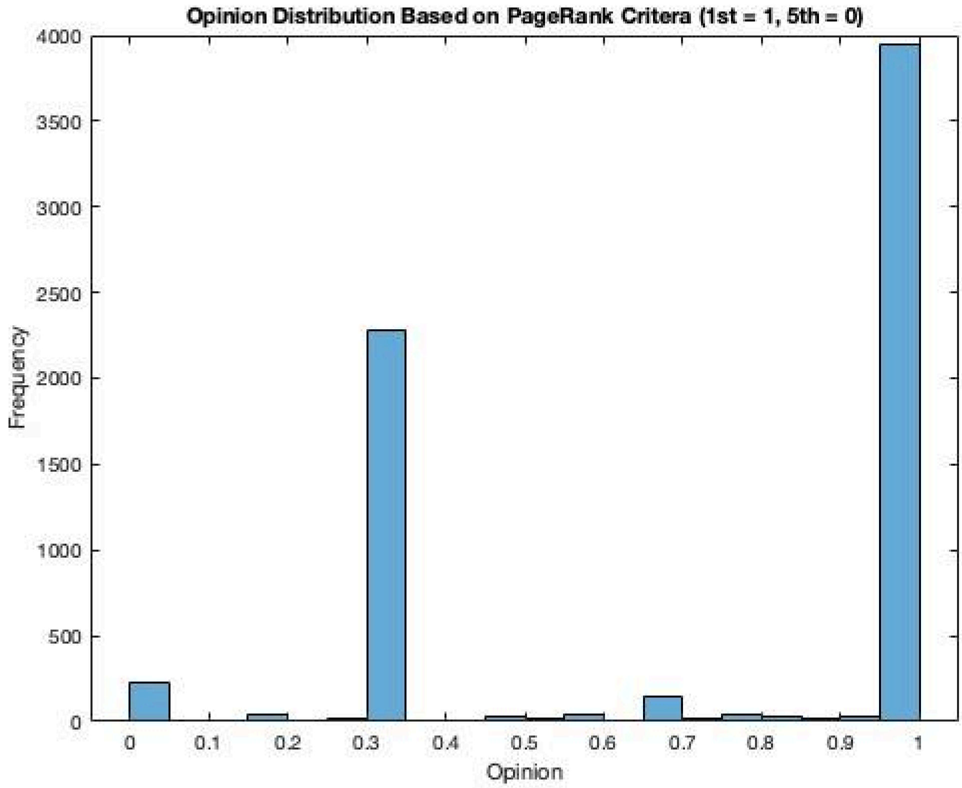

Figure 3: Histogram plot of the opinion distribution based on the PageRank criteria, where the most central user holds the stubborn opinion 1, and the fifth most central one holds the opinion 0.

When we compare the changes to the stationary opinion distribution observable in figure 3 with figure 2, it is sufficient to say that the most central node dominates the fifth most central one in a much clearer fashion than the aforementioned case. There are still as many users located in the staple furthest to the right, but now the frequency of this opinion is around 20 times the size of the one furthest to the left. The biggest change seems to be that when the second most important node held the stubborn opinion 0, the users who did not tend towards the most central user instead mostly ended up furthest to the left. Now, we can see that the majority of these users instead hold an opinion in the range between .30 and .35. This indicates that for this very network, the share of users that will copy the most central user's opinion may be limited, but the lower the importance of the user that holds the opposite when stubborn opinion, the closer the average opinion comes to 1. (Or 0, if the value of the stubborn opinions in question were distributed the other way around.)

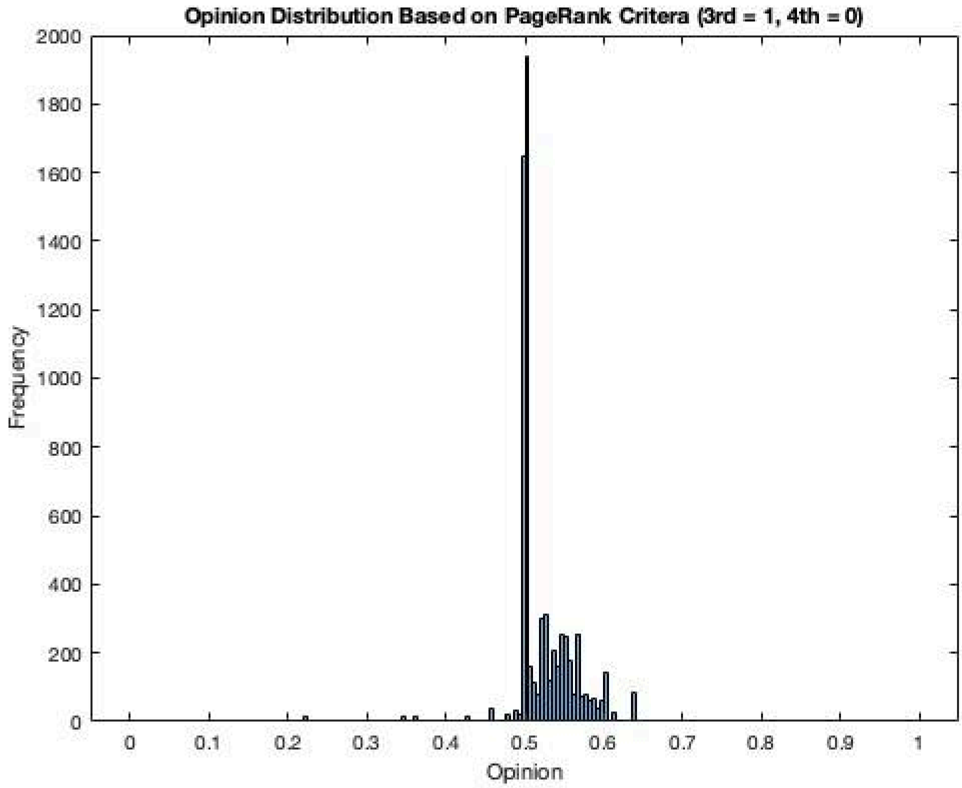

Figure 4: Histogram plot of the opinion distribution based on the PageRank criteria, where the third most central user holds the stubborn opinion 1, and the fourth most central one holds the opinion 0.

Lastly, we can directly see from figure 4 that there is only a small difference in the influence of the third and fourth highest ranked user, since the distribution is only slightly tilted to the right. More interestingly, the amount of users that end up in the end points are negligible; the majority actually meet in the middle, i.e. at .50. Thus, not having the most central user as the holder of a stubborn opinion seems to create a less polarized climate of opinions in the provided network. When most users sway around an opinion in between the two most extreme ones, they are more open towards both sides of the opinion spectrum, whereas a user who has adopted an opinion of 1 could very well have an abysmal reaction when being exposed to an opinion close to 0.

To conclude, it is important to note that a common fallacy is that triangulation (that is, meeting in the middle) is always the best. Let us consider an example where opinion 0 means that one supports the autonomy of neighboring states, whereas opinion 1 is equivalent to supporting several full-scale, bloody invasions. Is it then really in the best interest for the network that the stationary opinion distribution is centered around the middle? And if not, who can say so with full objectivity? The takeaway is that network dynamics provides useful, data-driven insights in several dimensions of society, but what is done with said information will likely more times than not come down to the subjectivity of the decision-maker(s).