Department of Computer Science

CS 429/529 – Dynamic and Social Network Analysis

Assignment 3

Alperen CAN

21601740

Instructor: Miray KAŞ

Teaching Assistant: Gözde YAZICI

Bilkent University

03.11.2022 (dd/mm/yy)

EXERCISE 1:

In order to compute network and node level metrics, I decided to use ORA. Having imported the *karate.xml* file, from *properties* menu, I have treated the data as undirected (symmetric), binary, and without self-loops as below:

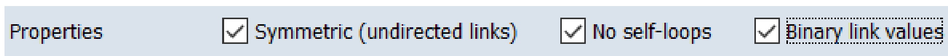

Figure 1: Properties

Consequently, the values of the combobox in *When computing measures* menu has become true, and they are automatically disabled:

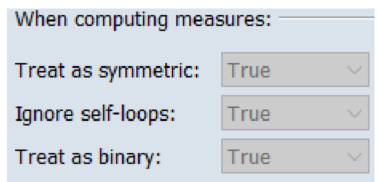

Figure 2: When Computing Measures Menu

a) Degree Distribution

Degree distribution is the probability distribution of the degrees over the whole network. Since I could not obtain the graph by using ORA, I used Gephi for this part only. However, I continued to use ORA for the following parts.

Average degree is found 4,588. The distribution graph is below:

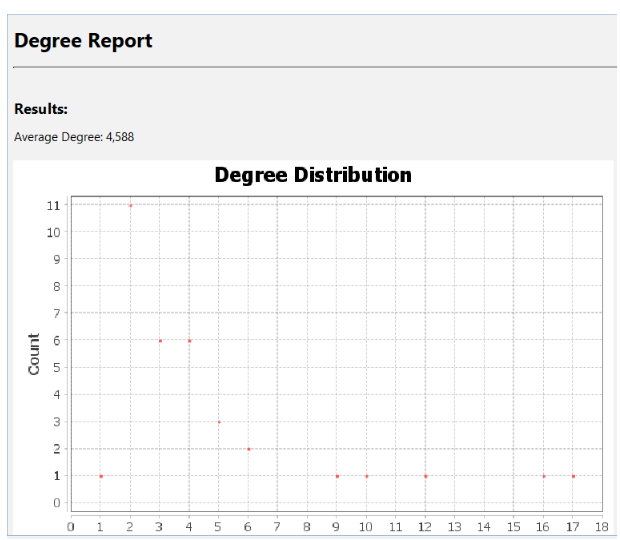

Figure 3: Degree Distribution Graph Created by Gephi

b) Network Density

Density of a network is calculated by as a ratio of existing links over all possible links.

$$D = \frac{\mathrm{L}}{\mathrm{N} * \frac{(\mathrm{N}-1)}{2}}$$

In this undirected network, the number of links in this network is L= 78, whereas number of nodes is N=34. Therefore, density is calculated as:

$$D = \frac{78}{34 * \frac{(33)}{2}} = 0{,}13903743$$

Indeed, the network density calculated by ORA matches with this result:

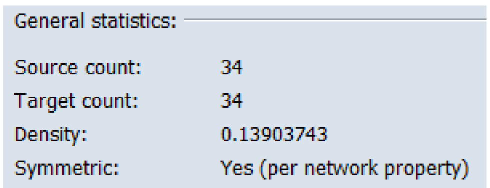

Figure 4: Network Density Calculated by ORA

Later, by clicking *Generate Reports* and choosing *Centrality Measures,* I have obtained an html file:

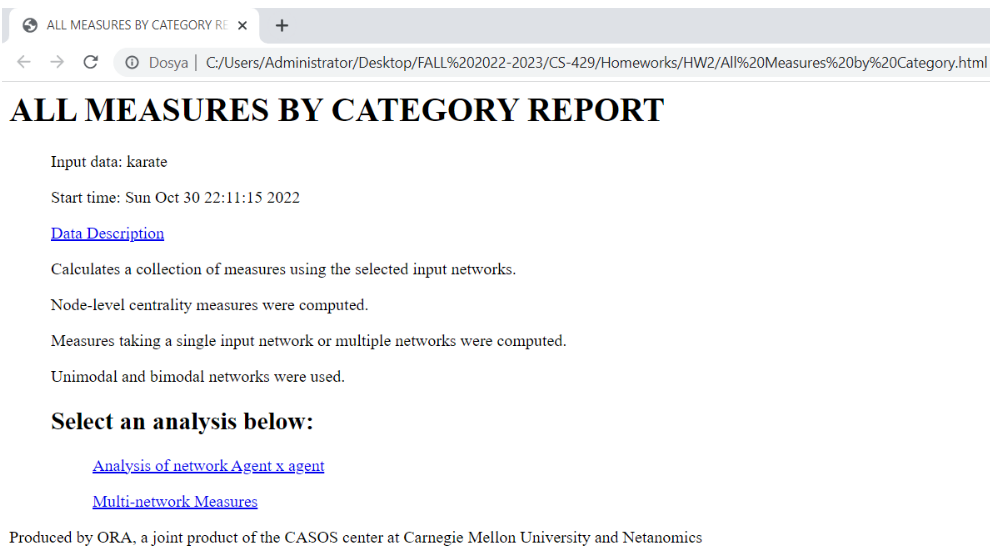

Figure 5: Generated html of karate.xml

Later, by clicking *Analysis of Network Agent x Agent,* I obtained measure list containing several centrality values, including degree centrality, betweenness centrality, closeness centrality etc.

c) Degree Centrality

Degree centrality is commonly thought of as a measure of influence or importance, that is, nodes with high degree centrality have the opportunity to influence and be influenced directly. It is a node level metric, and calculated as

In Degree Centrality + Out Degree Centrality

for a node, where in degree centrality refers to the number of edges incoming towards that node, and out degree centrality is the number of edges outgoing from that node.

The statistics for the unscaled version of total degree centrality for karate network is as follows:

Measure	Min	Mean	Max	Std.Dev
Centrality, In-Degree [Unscaled]	1	4,588	17	3,820
Centrality, Out-Degree [Unscaled]	1	4,588	17	3,820
Centrality, Total-Degree [Unscaled]	1	4,588	17	3,820

Figure 6: General Degree Centrality Information for the Network

I also observed that each in, out and total degree centrality have the same values, which is due to the undirected feature of the network.

And, the total degree centrality values for the top 10 nodes are as below (unscaled):

Centrality, Total-Degree

Individuals or organizations who are 'in the know' are those who are linked to many others and so, by virtue of their position have access to the ideas, thoughts, beliefs of many others. Individuals who are 'in the know' are identified by degree centrality in the relevant social network. Those who are ranked high on this metrics have more connections to others in the same network. The scientific name of this measure is total degree centrality and it is calculated on the agent by agent matrices.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Input network(s): agent x agent

Copy Print Search:

Rank	Agent	Value	Unscaled
1	34	0,515	17
2	1	0,485	16
3	33	0,364	12
4	3	0,303	10
5	2	0,273	9
6	32	0,182	6
7	4	0,182	6
8	14	0,152	5
9	24	0,152	5
10	9	0,152	5

Figure 7: Top 10 Nodes with Highest Degree Centrality Values

d) Betweenness Centrality

Betweenness centrality is related to how often a node lies along the shortest path between two other nodes. Its potential usage covers detecting gate-keeping, brokering, controlling the flow, and bridging separate parts of the network. It is calculated as,

$$b(v) = \sum \left(\frac{\sigma_{st}(v)}{\sigma_{st}} \right)$$

where s ≠ v ≠ t. The numerator represents the number of shortest paths between s and t that passes through v, whereas the denominator represents number of shortest paths between s and t. However, since it is expensive and very hard to compute, getting help from the network tools will be a good choice.

The statistics for the unscaled version of betweenness centrality for karate network is as follows:

Measure	Min	Mean	Max	Std.Dev
Centrality, Betweenness [Unscaled]	0	23,235	231,071	48,863

Figure 8: General Betweenness Centrality Information for the Network

And, betweenness centrality values for the top 10 nodes are as below (unscaled):

Centrality, Betweenness

The Betweenness Centrality of node v in a network is defined as: across all node pairs that have a shortest path containing v, the percentage that pass through v. When the data is weighted, the higher the weight the more value the link has. Individuals or organizations that are potentially influential are positioned to broker connections between groups and to bring to bear the influence of one group on another or serve as a gatekeeper between groups. This agent occurs on many of the shortest paths between other agents. The scientific name of this measure is betweenness centrality and it is calculated on agent by agent matrices.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Input network(s): agent x agent

Copy Print Search:

Rank	Agent	Value	Unscaled
1	1	0,438	231,071
2	34	0,304	160,552
3	33	0,145	76,690
4	3	0,144	75,851
5	32	0,138	73,010
6	9	0,056	29,529
7	2	0,054	28,479
8	14	0,046	24,216
9	20	0,032	17,147
10	6	0,030	15,833

Figure 9: Top 10 Nodes with Highest Betweenness Centrality Values

e) Closeness Centrality

Closeness centrality is about the nodes which is usually in the middle of the network by being close to many friends. These nodes can be considered as central players who hear things first. The usage may cover, the efficiency of a node in reaching out to everyone quickly.

$$C_i = \frac{1}{\sum_{j=1}^{N} d(i,j)}$$

where d(i,j) is the distance between node i and j.

The statistics for the unscaled version of closeness centrality for karate network is as follows:

Measure	Min	Mean	Max	Std.Dev
Centrality, Closeness [Unscaled]	0,009	0,013	0,017	0,002

Figure 10: General Closeness Centrality Information for the Network

And, closeness centrality values for the top 10 nodes are as below (unscaled):

Centrality, Closeness

The closeness of a node to the other nodes in a network (also called out-closeness). Loosely, Closeness is the inverse of the sum of distances in the network from a node to all other nodes.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Input network(s): agent x agent

Copy Print Search:

Rank	Agent	Value	Unscaled
1	1	0,569	0,017
2	3	0,559	0,017
3	34	0,550	0,017
4	32	0,541	0,016
5	14	0,516	0,016
6	33	0,516	0,016
7	9	0,516	0,016
8	20	0,500	0,015
9	2	0,485	0,015
10	4	0,465	0,014

Figure 11: Top 10 Nodes with Highest Closeness Centrality Values

f) Page Rank Centrality

Page rank centrality calculates the importance of a node, based on the importance of its incoming links. It is based on Eigenvector centrality.

The statistics for the unscaled version of closeness centrality for karate network is as follows:

Measure	Min	Mean	Max	Std.Dev
Centrality, PageRank	0,007	0,029	0,108	0,024

Figure 12: General Page Rank Centrality Information for the Network

And, page rank centrality values for the top 10 nodes are as below:

Centrality, PageRank

Calculates the importance of a node based on the importance of its in-coming neighbors. The input network links are normalized and interpreted as the probability of a transition from node i to node j. The PageRank of a node can be interpreted as the fraction of times a node would be visitied when traversing the network according to the network of probabilities.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Input network(s): agent x agent

Copy Print Search:

Rank	Agent	Value
1	34	0,108
2	1	0,102
3	33	0,076
4	3	0,063
5	2	0,057
6	32	0,038
7	4	0,038
8	24	0,032
9	9	0,032
10	14	0,032

Figure 13: Top 10 Nodes with Highest Page Rank Centrality Values

EXERCISE 2:

Below, the karate network has nodes colored by group, and sized by betweenness centrality as requested:

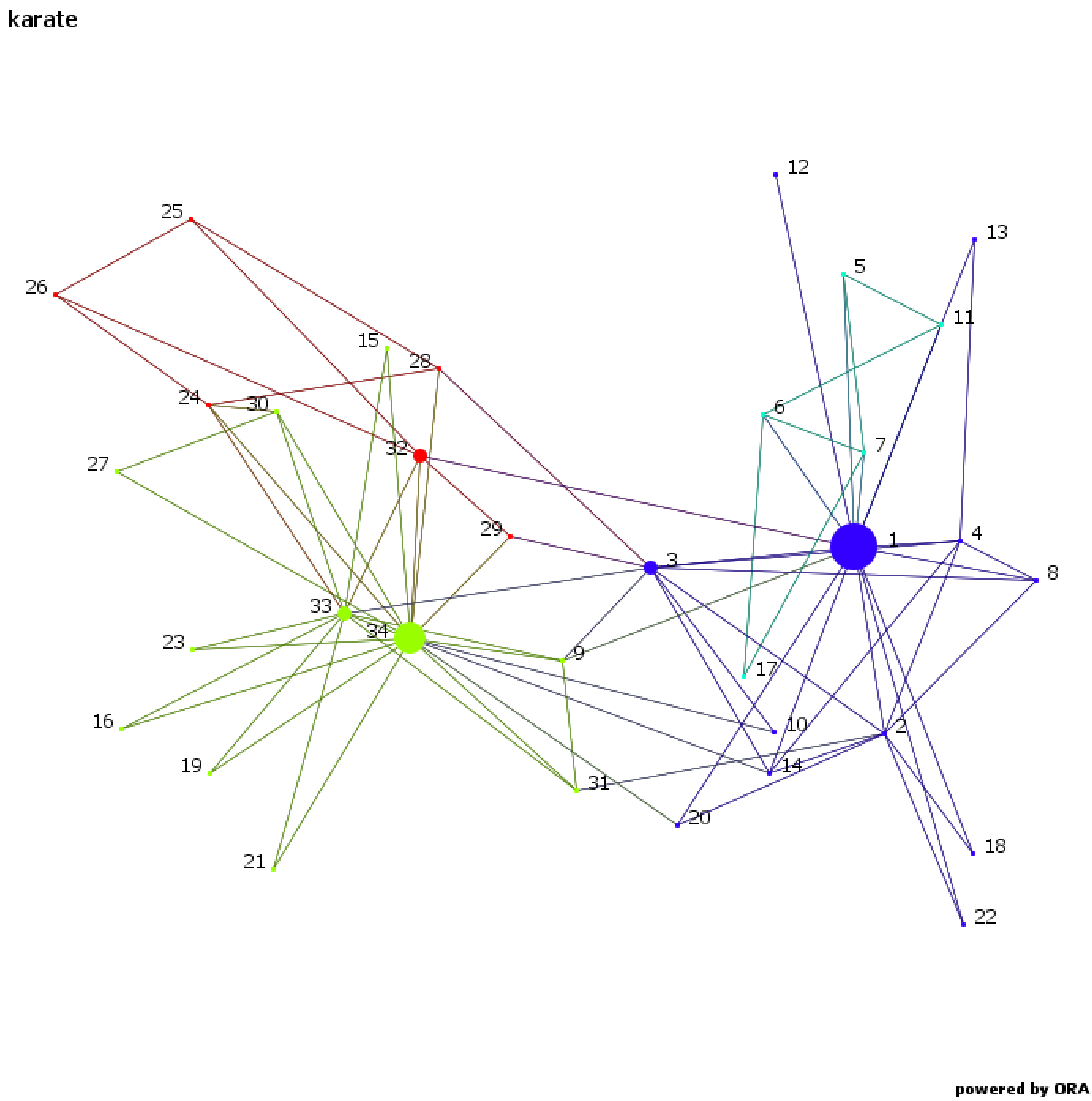

Figure 14: Visualization of the Network with Node Sizes Based on Betweenness Centrality

EXERCISE 3:

Highest ranked 3 nodes for each centrality metrics are as follows:

Degree centrality: 34, 1, 33

Betweenness centrality: 1, 34, 33

Closeness centrality: 1, 3, 34

Page rank centrality: 34, 1, 33

Above, I observed that for degree, betweenness and page rank centralities the top 3 nodes consist of same nodes, which are node 1, node 33, and node 34. Even the order of them is the same in degree and page rank centralities. For the betweenness centrality, the order is different but the nodes are the same as mentioned. Only for the closeness centrality we observed that there is node 3 instead of node 33. So we can assert that different centralities yield in almost the same results.

Before interpreting the result, I would like to give information for each centrality for clarification of paragraph below. Degree centrality represents the measure of influence or importance based on the number of incoming and outgoing edges, betweenness centrality is related to how often a node lies along the shortest path between two other nodes, and page rank centrality calculates the importance of a node, based on the importance of its incoming links.

Here, I want to assert that as Ms. Kaş mentioned in the class during the *network vs individual level analysis (Slide no 17 in Slide Set 4),* individual behaviors are not independent of the network they are in. Also, individuals network position is not independent of the network structure. Therefore, for example, I can argue that due to the structure of network the nodes which have high incoming links (therefore have a high page rank centrality) will also have a high degree centrality, since degree centrality is also dependent to the number of incoming links. Thus, for me, the main idea is that different metrics also have a close connection with each other, which together results in powerful nodes. For that reason, seeing those same or similar powerful nodes for different centrality metrics was not unexpected for me. However, it does not necessarily mean other nodes cannot be the most important based on different metrics, such as node 3 in closeness centrality, due to the fact that networks sometimes may be very complicated, especially in the real life.