Name : Adil Ashish Kumar

Emory MSBA 2019-2020

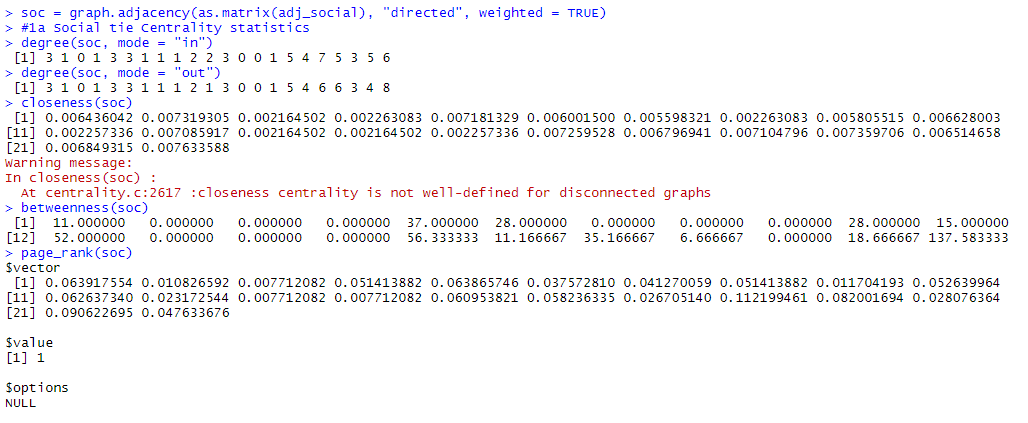
Social Networks ISOM 673 – Empirical Assignment 1

1. First, consider the social and task ties as separate networks.

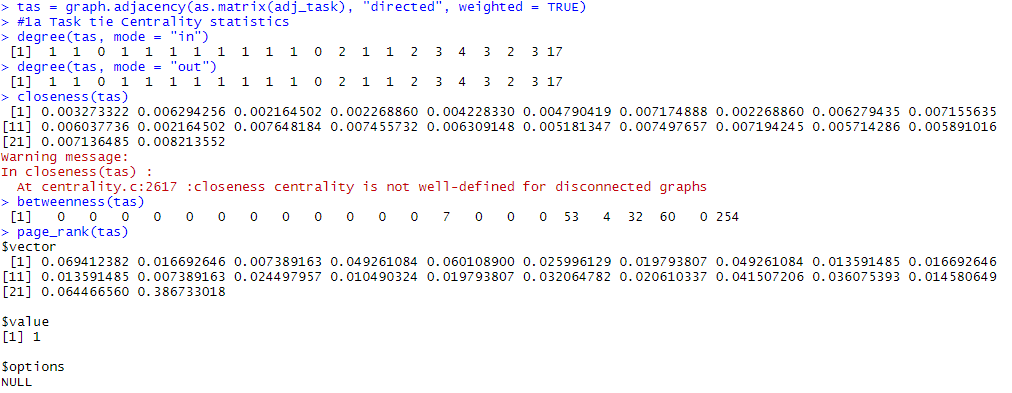
(A) Use igraph to generate indegree, outdegree, closeness, betweenness, and PageRank centrality statistics for each individual the social and task networks.

To begin, I created 2 separate adjacency matrices for each of both social and task ties. I then calculated the centrality statistics for both social and task ties independently using the degree, betweenness, closeness and page\_rank functions from igraph package.

Below are the centrality statistics for the social tie:

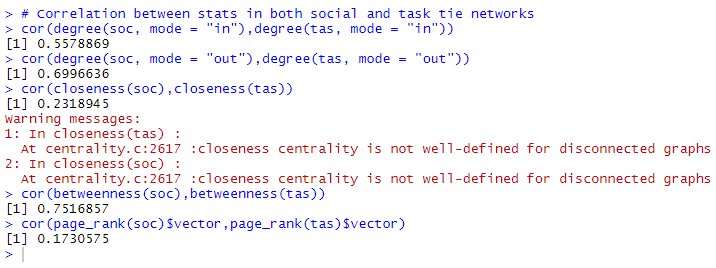


Below are the centrality statistics for the task tie:



(B) Compute the correlations of the five centrality measures you generate for the social network with the five measures generated for the task network. Which measures in the task network are most closely related to those in the socializing network? Name at least one insight can you draw from the relationships between these five measures across the two networks.

Below is the correlation between the 5 centrality statistics of both social and task tie networks



The correlation between in degree of social and task tie is 0.56. This indicates a moderate correlation between the in degrees, based on the magnitude of the correlation. Since the correlation is positive, we can say that an increase in in degree of one network would follow an increase in in degree of the other.

The correlation between the out degrees of social and task tie is 0.69, which indicates a slightly strong correlation. The degree gives us a measure of frequency and direction of communication activity. Looking at the correlation values of both in and out degree we can infer that strength of one tie tends to moderately influence the strength of the other tie and leads to higher frequency and communication activity.

The correlation of closeness between the two ties is 0.23, which indicates a weak correlation. Closeness gives an idea of reachability. Looking at the low correlation we can infer that reachability achieved due to one tie does not really improve reachability with respect to other tie.

The correlation between the betweenness of both ties is highest among all correlations. Betweenness indicates ability of bridging and brokerage. From the high correlation we can infer that the ability to bridge disconnected groups in one tie could improve the ability to do the same in the other tie.

The correlation between page rank of both social and task ties is the least among the correlations between centrality statistics.

2. Next, consider the social and task ties together, as two distinct types of ties comprising one

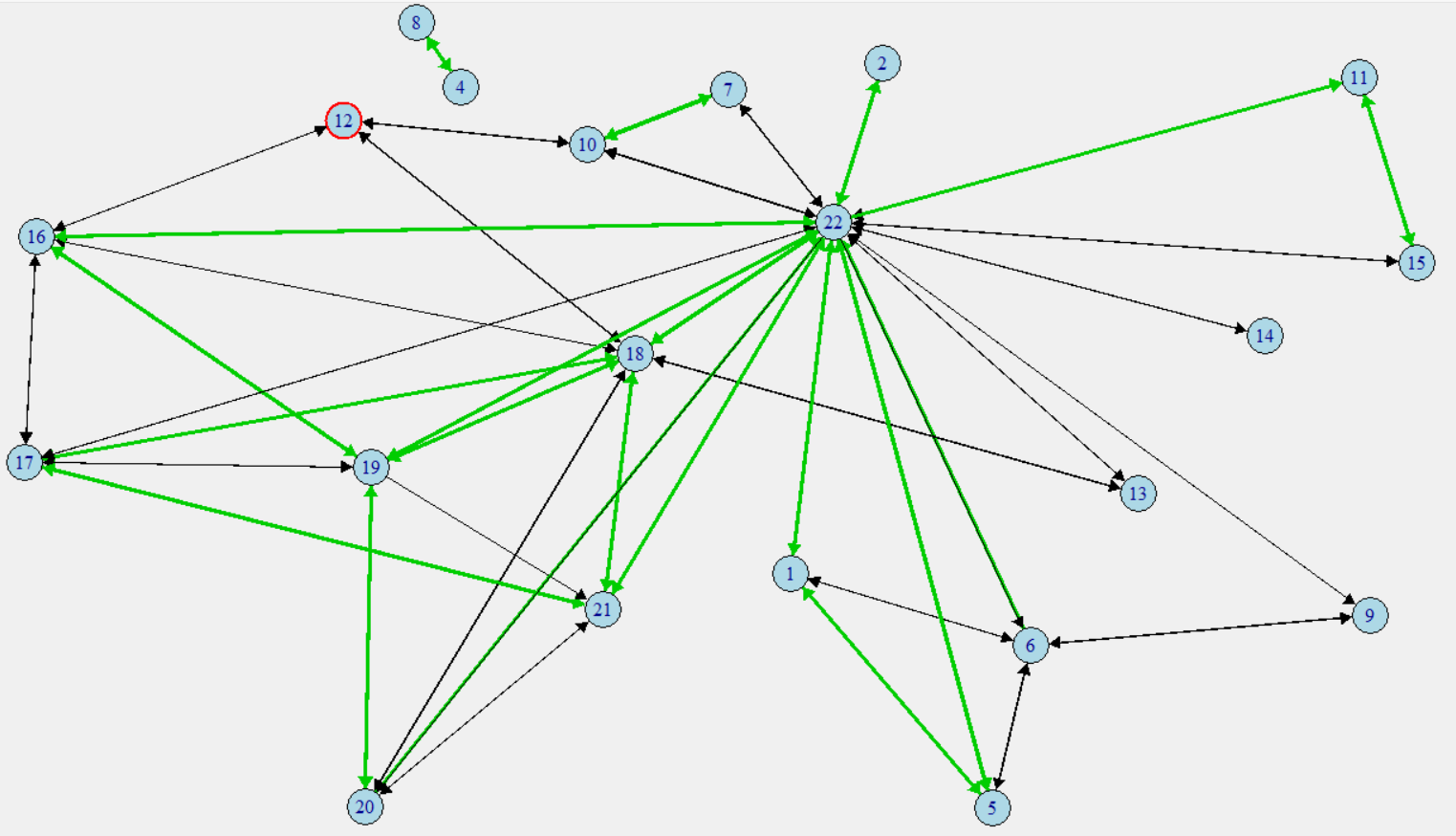
network.

(A) Suppose that a tie is strong if it is above the mean strength for that type, conditional on

the tie existing\_do not include weights of 0 in the calculation of the mean. Under this definition, does the network satisfy Strong Triadic Closure? Come up with a solution that illustrates this (1) visually, in a plot, as well as (2) programmatically, by giving the number or proportion of ties that are violation of Strong Triadic Closure.

In order to visualize it graphically, I first calculated the mean and median strength values after excluding 0s for each tie. I then used the mean/median values per tie to define a tie as strong/weak. All ties that were over the mean /median strength were characterized as strong, and of the remaining, all non 0 ties were identified as weak. This was done for both type of ties. I then combined the 2 ties into 1 tie column. I used the following logic – 1) if a node had at least 1 strong tie (social/task) , I identified it as having a strong tie 2) If a node had no strong tie but at least 1 weak tie, I identified it as a weak tie. 3) If a node had neither strong or weak tie, then it had no tie at all.

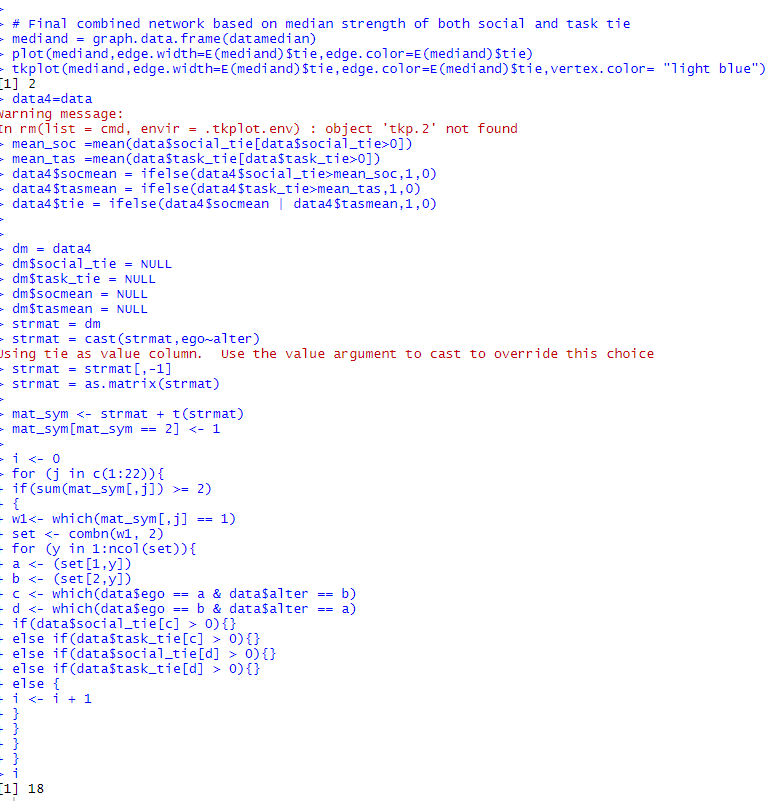
Based on the mean criterion, below is the combined network: The green arrows, which are thicker, indicate strong ties. The black arrows, which are thinner, indicate weak ties. Upon visual inspection we notice that 19 has a strong tie with both 16 and 20, but no tie exists between 16 and 20. Similar case with 22, 16 and 2. This indicates a violation of strong triadic disclosure. Thus the network does not satisfy strong triadic disclosure.



Programmatic:

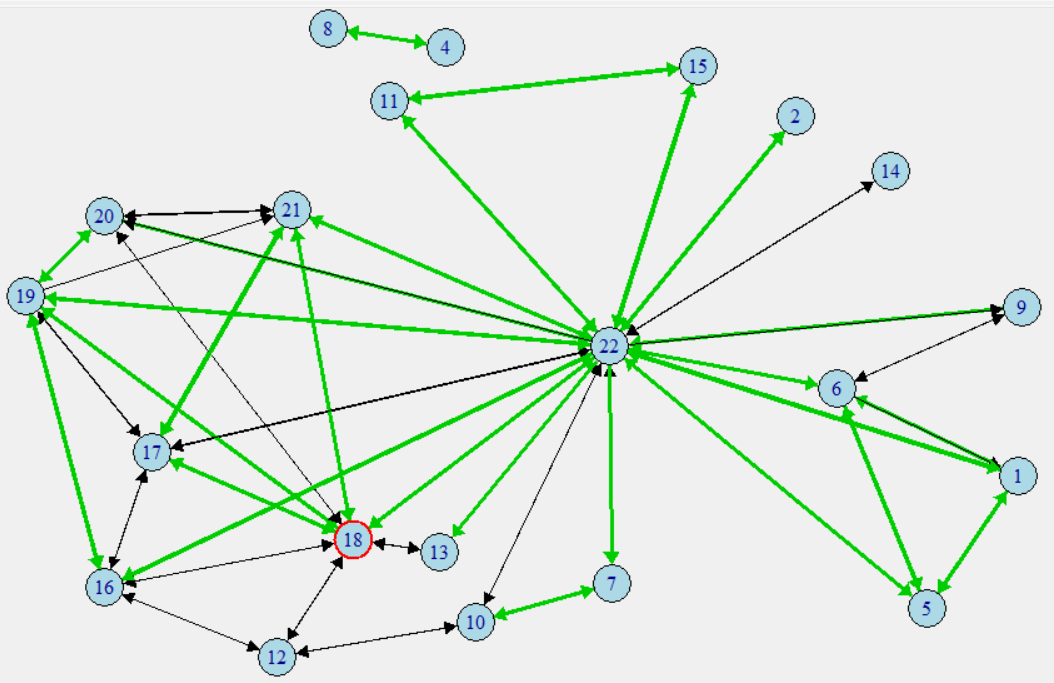
To identify violations of strong triadic disclosure, I used a for loop to iterate and identify violations. I used a matrix indicating just strong ties to find points where there could be possible violations and then used a for loop to identify and count the number of instances that were in violation of strong triadic disclosure.

Below is the code output, we can see that there were 18 instances where strong triadic closure was violated



(B) Now suppose that a tie is strong if it is above the median strength for that type, conditional on the tie existing. Under this definition, does the network satisfy Strong Triadic Closure? What insights does this illustrate about these interactions within the network?

Based on the median criterion, below is the plot for the combined network. Upon visual inspection we notice that 22 has a strong tie with both 2 and 15, but no tie exists between 2 and 15. This indicates a violation of strong triadic disclosure. Similar case with 22, 13 and 7. Thus the network does not satisfy strong triadic disclosure.

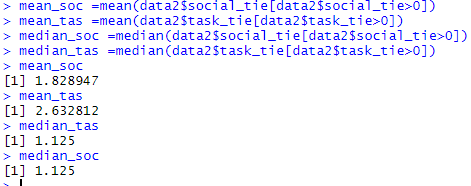


Programmatic:

For the median criterion, I used the same logic as for the mean criterion to find instances of violation of strong triadic closure programmatically. From the code output we can see that there are 78 cases of violation of strong triadic closure.



Based on the code output, we see that the no of strong triadic closure violations are more in case of using median as compared to mean to indicate strong/weak ties.

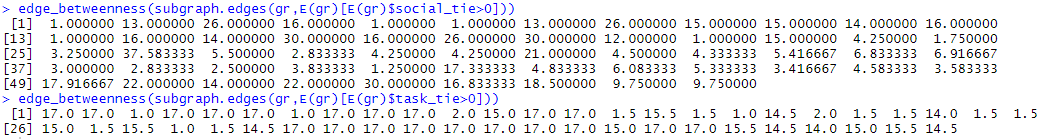


Looking at the mean and median values the above result does make sense. The mean values are higher than median values. This means that the median criterion cutoff to classify a tie as strong/weak is lower than that of the mean. Thus, there would be more strong ties in our network when we use median to classify a tie as strong/weak in comparison to mean. Since there are mores strong ties in the network when using median as opposed to mean, it makes sense that we have more violations of strong triadic closure when using median. While we might have more strong and weak ties when the strength criterion is a bit lower while using median, the existence of a tie still remains the same. This again justifies the fact that we have more violations of strong triadic closure while using the median.

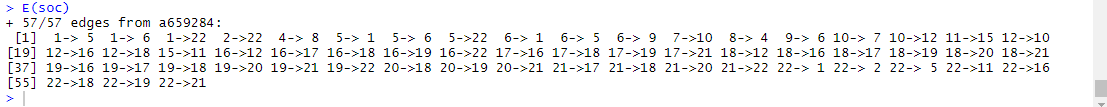
3. Continue to treat the social and task ties as two distinct types ties comprising one network.

(A) It is also possible to compute betweenness on the edges in a network, as well as the vertices. This is a good measure of the flow of information and resources through a network. Calculate the edge-level betweenness for both of the types of tie.

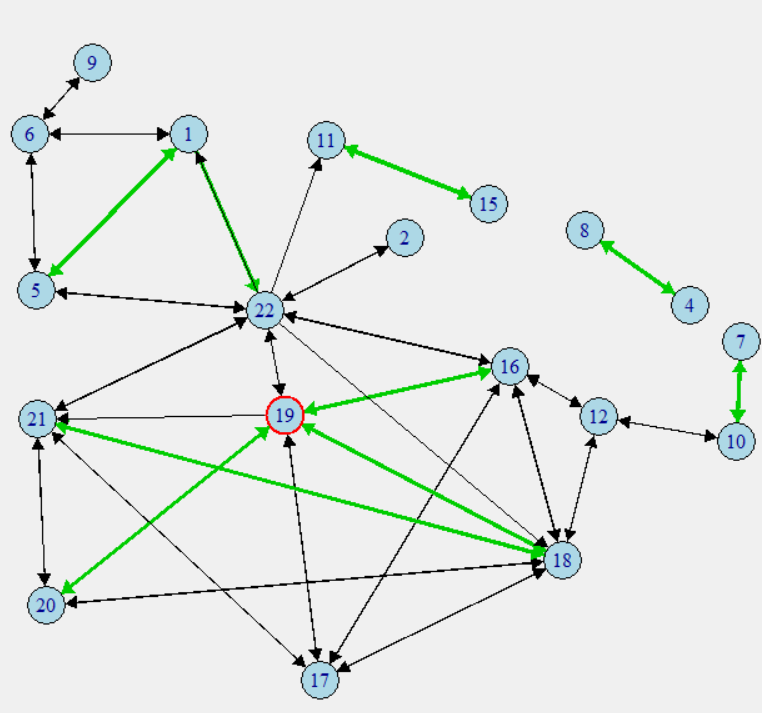
I used the edge betweenness function from igraph to calculate the edge level betweenness for both social and task ties. Below is the code output.



(B) Does it seem like edges with high betweenness tend to be strong or weak ties, according to our two definitions above? Does this result make sense?



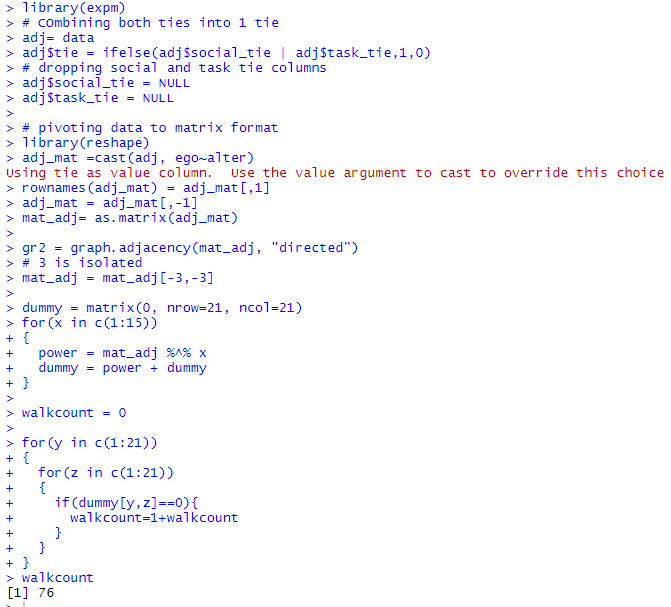
From the above edge betweenness for social tie, we notice highest values for 16-22 (37.583). In the below combined network for strong/weak ties for social tie defined by mean, we can see that 16 and 22 have a weak tie. Looking at another high betweenness value 12-16(30), we see below there exists a weak tie. Nodes 22-11 (30) also has a weak tie from below. 1-22(26) however has a strong tie as seen below. 5-22(26) also has a weak tie. It seems that in this case, most of the edges with high betweenness had weak ties. Since betweenness measures the probability that a share of geodesics from vj to vk may pass through vi, the result is in line with the fact that all edges with high betweenness at least had a tie, even though it was mostly a weak tie. Generally, edges with high betweenness should at have a tie that exists. This does make sense because of the way betweenness works.

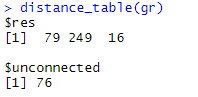


4) Continue to treat the social and task ties as two distinct types of ties comprising one network. How many pairs of nodes do not have walks between one another? Find a solution that performs this calculation directly on the matrix\_it is possible to verify this solution via igraph afterward.

In order to calculate the no of pairs of nodes which do not have walks between them, I used the concept of multiplying the adjacency matrix to a certain power to identify all cells where there would be 0s. These 0s would indicate the absence of a walk between the nodes.

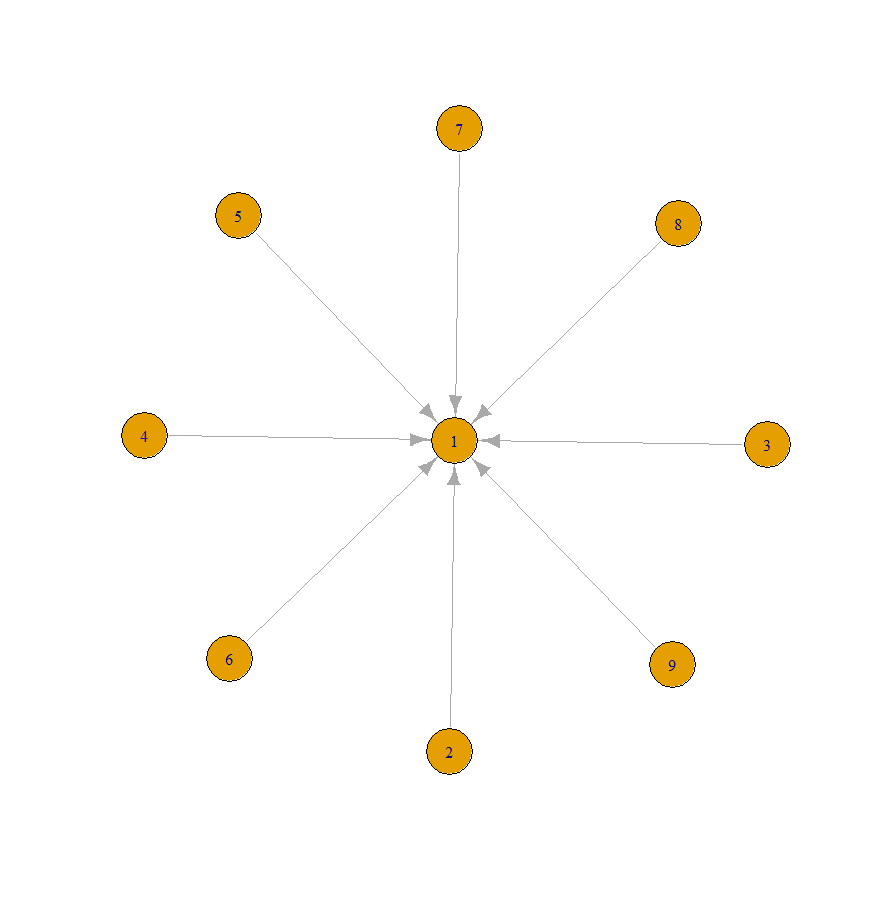
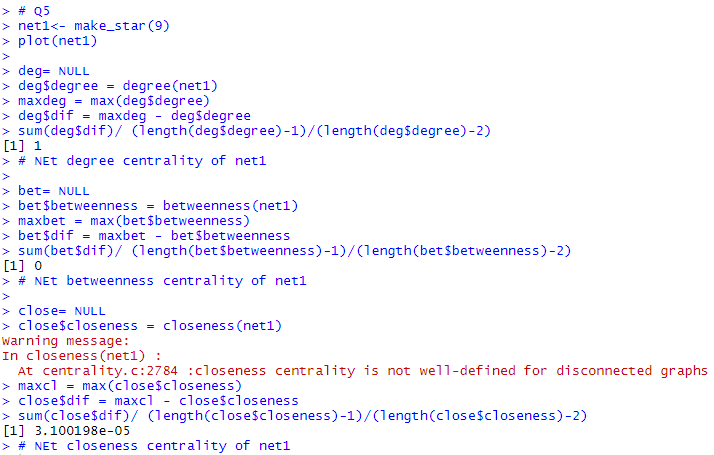
To begin, I combined the social and task ties into 1 tie by just indicating presence of a tie based on presence of either social or task tie. I then created an adjacency matrix and multiplied it to the 15th power using a for loop. I then did a quick count to see how many node pairs did not have walks. Below code output shows that there were 76 instances where a walk did not exist between a pair of nodes



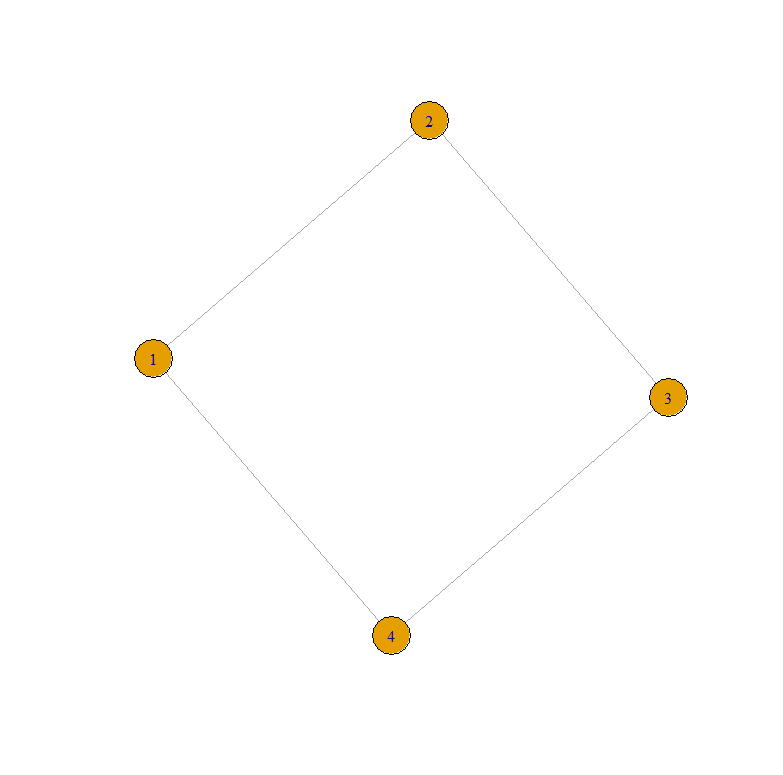
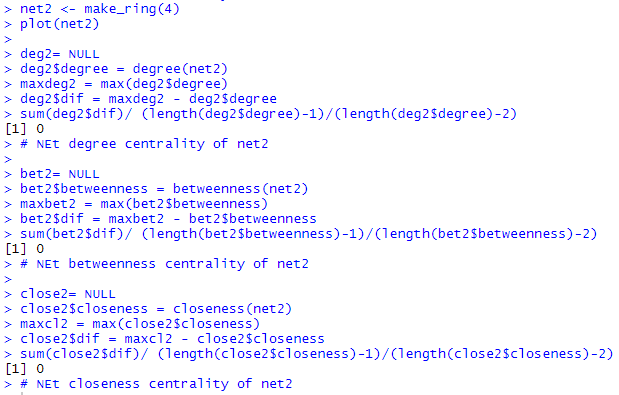


From the distance\_table function from igraph package we can verify the same. We see the same count of 76 in output above.

5. The network-level measure of degree centrality is a good indicator of the dispersion of the degree distribution in a network. Generate and plot a network in R in which the network-level measure of degree centrality, is equal to 1, and another where it is equal to 0. Would this relationship hold true for these networks for other measures of centrality, such as closeness or betweenness?

Above plot shows a star network with 8 nodes all connected to only one node at the center. This network will have a network level degree of centrality as 1. We can see the same from above code output. This network has a network level betweenness centrality of 0 and a network level closeness centrality of almost 0. We can see the values from above code output.

Above plot shows a ring network with 4 nodes, each connected to 2 adjacent nodes, forming a closed loop. This network will have a network level degree of centrality as 0. We can see the same from above code output. This network has a network level betweenness centrality of 0 and a network level closeness centrality of 0. We can see the values from above code output.

Based on the calculated values above, it seems that a network with network level degree centrality as 0 will have 0 for closeness and betweenness centrality as well. However, a network with network level degree centrality as 1 had a closeness of 0 and betweenness of almost 0. Therefore the relationship holds true for networks with network level degree centrality as 0 but not necessarily for networks with network level degree centrality as 1.