

NETWORK SCIENCE

Assignment 7

STUDENT | Erfan Raoofian

ID | 47315221

INSTRUCTOR | Prof. Ifeoma Adaji



IRVING K. BARBER SCHOOL OF SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
OKANAGAN CAMPUS

1 Question 1

I used this code snippet to create and plot the graph

```
1 edges<- read.csv("edgesList.csv",header=FALSE)
2 nodes<- read.csv("nodesList.csv",header=FALSE)
3
4 graph<-graph.data.frame(edges,directed=TRUE,vertices=nodes)
5 plot(graph,vertex.label=NA, edge.arrow.size=0.05, vertex.size=5)
```

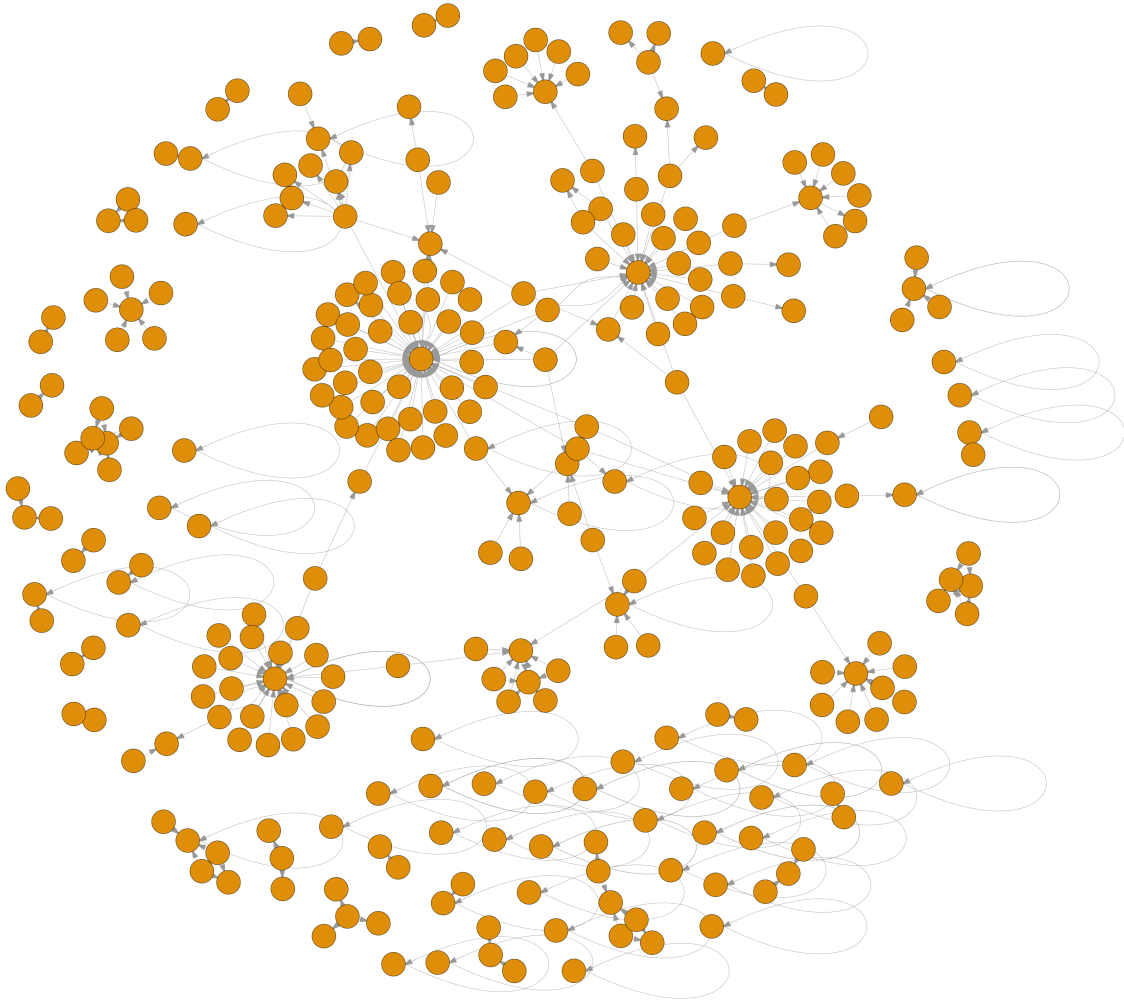


Figure 1: Network plot of Question 1

2 Question 2

I used the PageRank centrality measure to pick the most popular nodes. Because I think if node X mentions or replies to node Y, it shows the popularity of node Y and because in such a case there would be an edge from node X to node Y, the page rank centrality can exactly measure the popularity of the nodes. Other than PageRank, indegree and eigenvalue centrality can be good options as well, but I think pagerank is the superior option here. After calculating the pagerank I picked the

top 5 based on the pagerank centrality: "nocontextcanmnt", "onesoccer", "bensteiner00", "crate-soffrates", and "jimmybrennan11"

```
1 pageRank <- page_rank(graph)
2 nodes[,2] = pageRank$vector
3 colnames(nodes)[2] <-c("PageRank")
4 top5 = head(nodes[order(nodes$PageRank, decreasing = TRUE), ][,1],5)
```

3 Question 3

The betweenness centrality measure is very popular when it comes to dissemination of information in a social network as it shows how many times each node lies on the shortest path between nodes. It shows how many times each node was the reason the information was disseminated in the network. After calculating the betweenness scores and picking the top 5 nodes based on it, the answer would be: "onesoccer", "cplsoccer", "theroom442", "tv_jjd", and "brendan_dunlop"

```
1 Betweenness <- betweenness(graph,directed=TRUE)
2 nodes[,3] <-Betweenness
3 colnames(nodes)[3] <-"Betweenness"
4 top5 = head(nodes[order(nodes$Betweenness, decreasing = TRUE),
5 ][,1],5)
6
7 nodes$PageRank[which(nodes$V1=="onesoccer")] # output: 0.059
8 nodes$PageRank[which(nodes$V1=="fabrizioromano")] # output: 0.001
```

After looking at the other measures in the Nodes.xlsx file, I still think that the best centrality measure for finding the answer to this question would be the betweenness scores. For example, although the node with twitter id: "fabrizioromano" has the highest number of followers and much more than "onesoccer"(which is the node with highest betweenness score), if we look at its betweenness score its much less that "onesoccer". We can also see that the pagerank of node "fabrizioromano" is also very bad. This kind of analysis shows that the number of followers can not correctly tell us which node has the highest influence when it comes to dissemination of information. Also, comparing "onesoccer" and "fabrizioromano" shows that they have very similar number of tweets (21620 vs 21409 respectively), but the influence of "onesoccer" on the dissemination of information is much higher than "fabrizioromano".

4 Question 4

The Graph is shown in Figure 2.

```
1 CL <- cluster_edge_betweenness(graph,directed = TRUE)
2 nodes$community <- membership(CL)
3 V(graph)$community <- nodes$community
4 plot(graph, vertex.label=NA, vertex.color=V(graph)$community, edge.
5 arrow.size=0.05, vertex.size=5)
```

By running a simple code, it can be seen that community 2 has the largest size(181). This community is shown in light blue in 2.

```
1 CL$membership[which.max(sizes(CL))] # output: 2
2 max(sizes(CL)) # output: 181
```



By looking at different metrics and columns from the Nodes.xlsx file, I cannot find any meaningful relation between the number of followers or number of tweets and in which community the nodes are. Of course we did split the nodes based on edge betweenness and it means that they are much more connected to each other in the same community compared to other nodes outside of the community, but the question is what is the common thing among them that results in such communication between them. As we are checking the tweets related to FIFA world cup and Canada's national team, it shouldn't be a surprise if the nodes in the same community have the same taste in that regard. I can see in the description column in the excel file that the nodes in one particular have very related keywords like Canada's sports team, FIFA, and sports journalists.

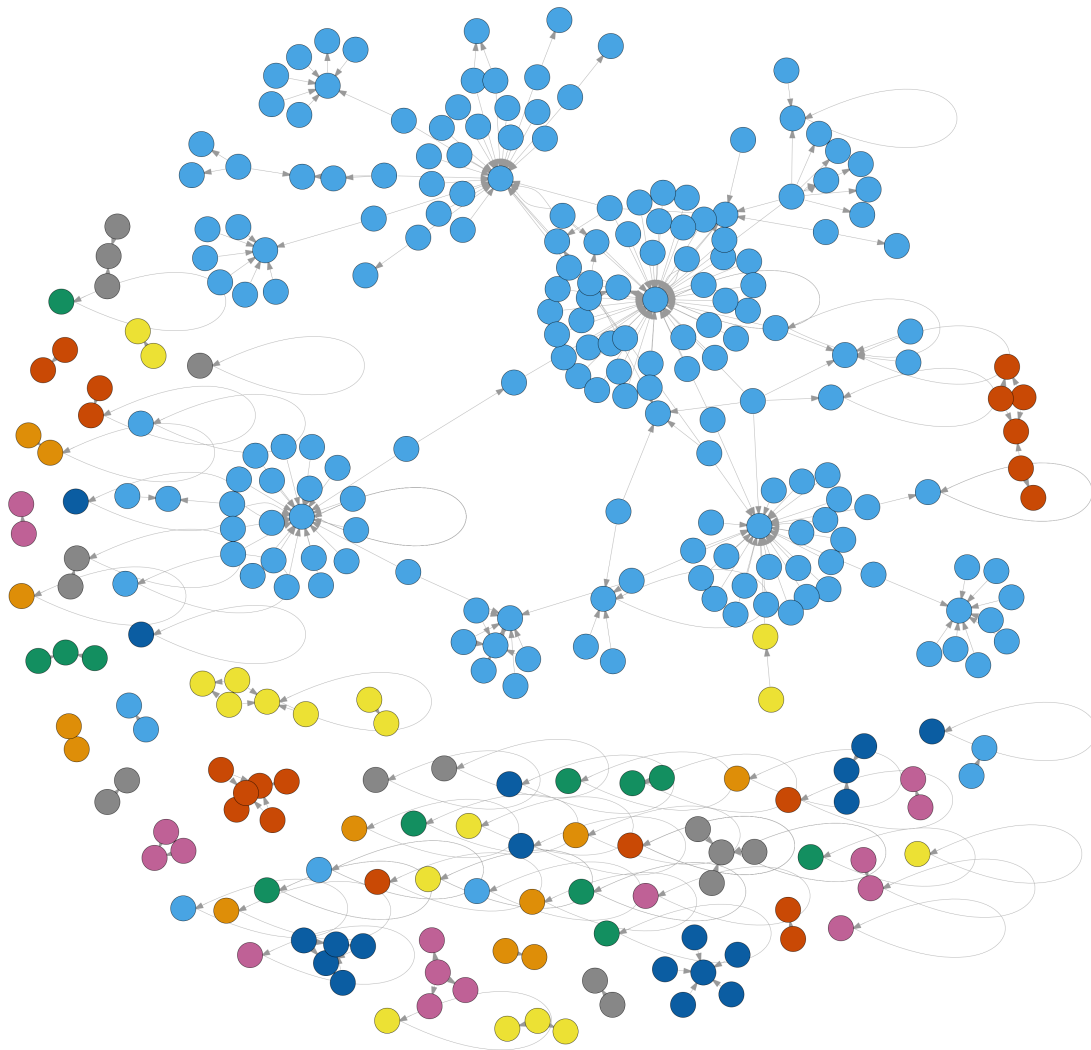


Figure 2: Network plot of Question 4