Homework 1

Alper Kocabiyik

April 11th 2018

Introduction

This network represens interacting pairs of protein in Humans (Homo sapiens).

```
data_nodes<-read.delim("protein_tab.txt", header = FALSE, sep = "\t")
colMax <- function(data) sapply(data, max, na.rm = TRUE)
colMax(data_nodes)

## Warning in FUN(X[[i]], ...): no non-missing arguments to max; returning -
## Inf

##   V1   V2   V3
## 1706 1706 -Inf

nodes <- data.frame(id=(1:1706),name = (1:1706))

edges <- data.frame(from=data_nodes[,1],to=data_nodes[,2])

g <- graph.data.frame(edges,vertices=nodes)
```

Questions

1. What is the number of nodes and links?

```
cat("Number of nodes:", length(V(g)))

## Number of nodes: 1706

cat("\nNumber of links:", length(E(g)))

##
## Number of links: 6207
```

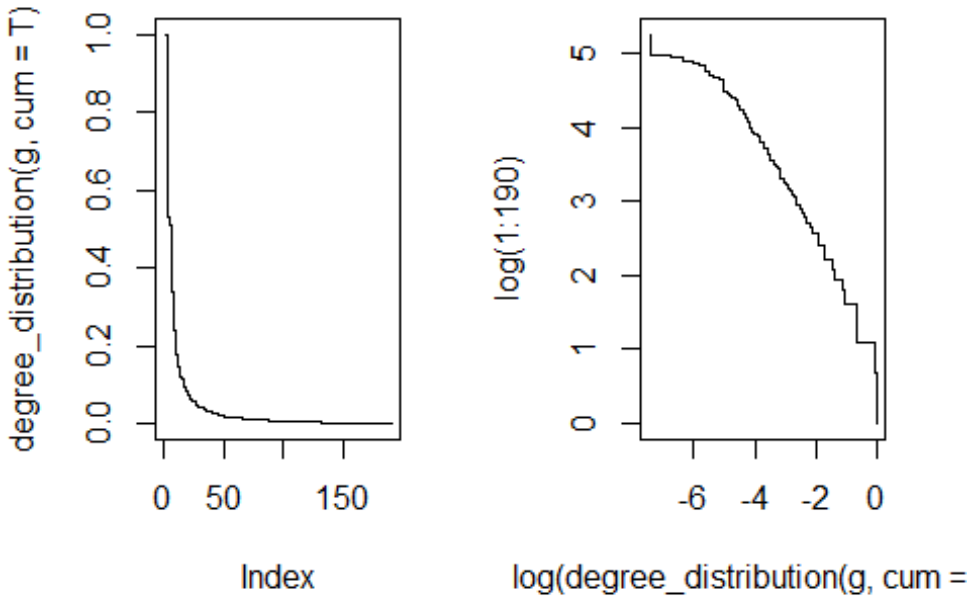
2. What is the average degree in the network? And the standard deviation of the degree?

```
cat("Average degree:", mean(degree(g)))

## Average degree: 7.276671

cat("\nStandart deviation of degree:", sd(degree(g)))
```

```
##
## Standart deviation of degree: 13.79084
```

3. Plot the degree distribution in linear-linear scale and in log-log-scale. What is the degree of the most connected node?

```
par(mfrow=c(1,2))
plot(degree_distribution(g,cum=T),type="s")
plot(log(degree_distribution(g,cum=T)),log(1:190),type="s")
```




```
#Get the ones which are most connected
cat("Degree of most connected:", max(degree(g)))

## Degree of most connected: 189
```

4. What is the clustering coefficient (transitivity) in the network?

```
cat("Clustering Coefficient:", transitivity(g))

## Clustering Coefficient: 0.005769231
```

5. What is the assortativity (degree) in the network?

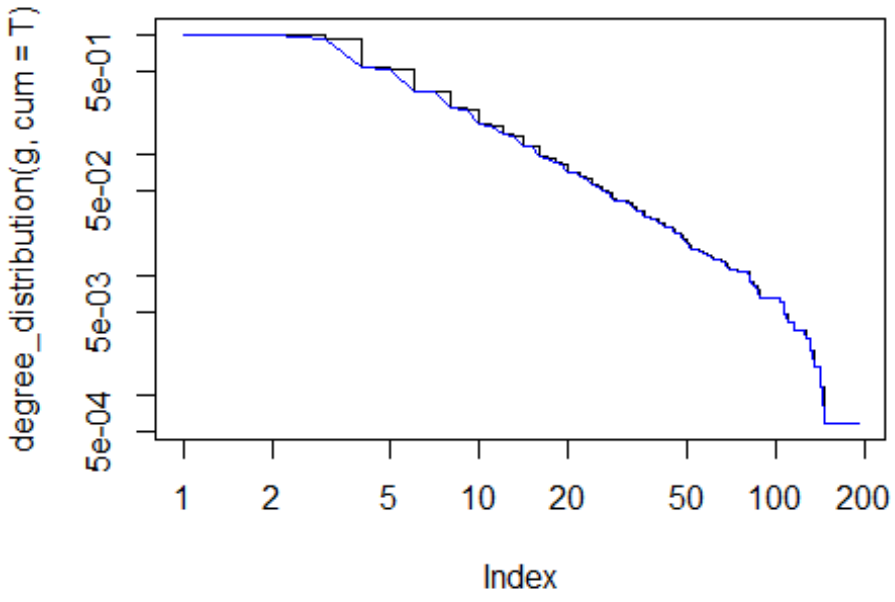
```
cat("Assortativity:", assortativity_degree(g))

## Assortativity: -0.1848108
```

6. Using the Louvain method, does the network have a community structure? If so, what is its modularity?

```
comm <- cluster_louvain(as.undirected(g, mode = "collapse"),weights=NULL)
cat("Number of groups:",length(comm))

## Number of groups: 69

cat("\nModularity:",modularity(comm))

##
## Modularity: 0.6149201
```

7. Test that the clustering coefficient in the network cannot be statistically explained by a configuration model in which the nodes have the same degree as the original.

```
#create model
model_new <- degree.sequence.game(degree(g))

#compare the main properties of the real network with the models
#Same
c(mean(degree(g)),mean(degree(model_new)))

## [1] 7.276671 7.276671

#Degree distribution is not same
plot(degree_distribution(g,cum=T),type="s",log="xy")
lines(degree_distribution(model_new,cum=T),col="blue")
```




```
#These metrics are different
c(diameter(g), diameter(model_new))
```

```
## [1] 13  9
```

```
c(transitivity(g), transitivity(model_new))
```

```
## [1] 0.005769231 0.057325894
```

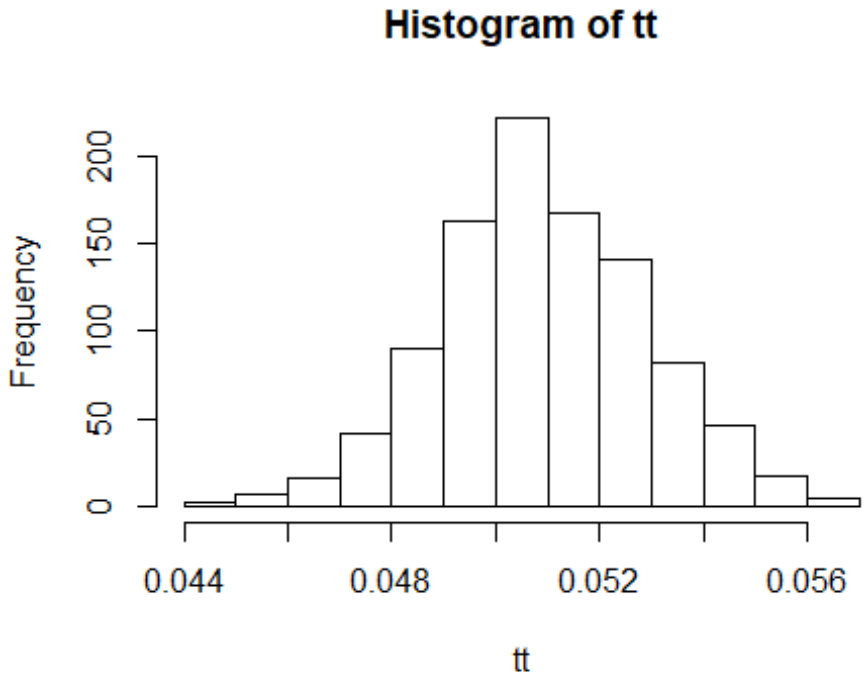
```
c(assortativity_degree(g), assortativity_degree(model_new))
```

```
## [1] -0.18481076  0.03384561
```

```
#compare transitivity
n <- 1000; tt <- rep(0,1000)
for(i in 1:1000) tt[i] <- transitivity(degree.sequence.game(degree(g)))
transitivity(g)
```

```
## [1] 0.005769231
```

```
hist(tt)
```



With the distribution above its almost impossible to obtain that value “0.005769231”.

8. Visualize the neighborhood of the node with the largest centrality (closeness)

```
require(visNetwork)
```

```
cc <- clusters(g)
```

```
table(cc$csize)
```

```
g1 <- induced_subgraph(g,which(cc$membership==which.max(cc$csize)))

is.connected(g1)

x<-closeness(g1)

maxClNames<-names(which.max(x))

data<-toVisNetworkData(g1)

visNetwork(data$nodes, data$edges, width="100%",height="400px")%>%

visOptions(highlightNearest = TRUE, nodesIdSelection = list(enabled = T,
selected=maxClNames))%>%

visPhysics(stabilization=T)
```