## Introduction:

People upload data to the web every day. With proper analysis of this data, we can examine and predict the future directions of our communities. While analysis of data objects provides us with useful information, examining the relationship between these objects provides us with a new dimension of analysis. For this project, we will be analysing Twitter account of Donald Trump, the President of United States of America. For assistance, Twitter API account and R statistical platform will be used.

### PART 1 – Friends of Trump

In the first part, using R we extract the number of people Trump is following. We then use coding in R to identify user IDs, screen name, number of followers, number of friends of these accounts; we sort this data according to no. of followers and pick the friends of Trump who have the greatest number of followers. We then examine each account individually and assess their type of relationship with the President of USA.

The top 20 friends of Trump are:

IvankaTrump – Daughter of Donald Trump, Ivanka is a businesswoman currently serving as a senior advisor to the President. Considering the given relationship, both have a strong positive link with each other.

DonaldJTrumpJr – Son of Donald Trump and the EVP of Development & Acquisitions at The Trump Organization. He is a republican and has a large following (5.1M) on Twitter. The relationship with Donald Trump is of positive nature.

EricTrump – Another son of Donald Trump and EVP at the Trump organisation. Eric has a twitter following of 3.6M members. Given his relationship with Donald Trump, the edge will be positive.

KellyannePolls – Kellyanne Conway is a pollster and a political consultant, serving as a counsler to the president of United States under administration of Donald Trump. As her current role predicts, she has a positive relationship with Trump.

Jim\_Jordan – is another republican politician currently serving as the US representative in the Ohio state. Since both Trump and Jim belong to Republican party, they have a positive relationship.

Foxandfriends – is a media morning show. Since it is not a person, it will be avoided. (a limitation and will be discussed at the end of this part.)

JesseBWatters – Jesse Watters is a conservative who hosts a TV show called “WattersWorld on Fox News Channel. Trump and Jesse seem to have a media - politician relationship. The background picture with Trump on Jesse’s twitter account shows that he is a Trump supporter.

greta - Greta Van Susteren is a lawyer and a TV host of “Full Court Press” show. She has 1.2M followers on her account and seems to have a strong voice in media. Considering she’s a media person, her relationship with Trump could be positive or negative.

Ericbolling – Eric Bolling is another TV host who has a presence on multiple media channels. He has a following of more than 1 million. In his twitter bio, Eric has proudly written that he is followed by the President which suggests his type of relationship with Trump.

MariaBartiromo – Another famous anchor from Fox news who is New York based and has a large number of followers of around 76k. Trump and Maria both are following each other but we do not have enough evidence to say whether their relationship is negative or positive.

GOPChairwoman – Ronna Mcdaniel is a republican who is currently serving as a chair of the Republican National Committee. With more than 68.9k followers on twitter, Mcdaniel is known for her prolific fundraising and staunch support for Trump. Hence it is clear that they have a positive relationship.

Scavino45 – Dan Scavino is an American political advisor who is currently serving as a White House deputy chief of Staff for communications and has previously worked as the director of social media for Trump’s previous election campaign. This suggests a strong positive link between Donald trump and Scavino.

KatrinaPierson – Katrina Pierson is an American activist and communications consultant. She was the national spokesperson of Donald Trump Election campaign 2016. Currently she is one of the senior advisors of Trump. We can say she has a positive relationship with Trump and is a close associate.

Garyplayer- Gary Player is the legendary South African golfer who now works as a philanthropist and a golf course architect. He has more than 400k followers on twitter. Trump and Gary both are following each other so it seems to be a positive relationship.

GeraldoRivera – Geraldo Rivera is an American reporter, author, and a talk show host. The New York based reporter is quite active on Twitter and posts tweets often. Donald Trump is one of his 357.2k followers. The relationship is of positive nature.

RealRomaDowney – Roma Downey is a popular actor, producer and the author of Box of Butterflies. With a following of 275k on twitter, she’s one of the rare Hollywood celebrities who are followed by the President of USA. This puts her in a special positive bond with Trump.

KatrinaCampins- Katrina campins is a real state expert and a TV personality. She has more than 231k followers on her twitter account. She seems to be a huge fan of Donald Trump and claims herself as a Trump’s surrogate on her Twitter bio. She’s another TV personality who Trump is following back. The two-way link makes it positive for both.

TrumpChicago – Official page of Trump Hotel Chicago

TrumpLasVegas – Official page of Trump Hotel Las Vegas

Trumpdoral – Trump national doral Miami is a gold resort owned by one of Trump companies.

From our sample data of top 20 friends of Trump with most followers, we can see that most of the people Trump is following are either his family members, his advisors, Media persons who are Trump supporters, or close associates of Donald Trump.

Limitations:

Limitations in this part were the 4 accounts that were not a person account due to which the quality of data was compromised. For a better and further analysis, a function could be created to drop specific and non-human accounts (for example selecting only personal accounts and dropping professional and company accounts) while extracting top 20 userIDs from the list.

### PART 2 – Followers of Trump

For part 2, we will use the same method as used in part 1. We extract the top 20 followers of Trump who have the greatest no. of following on twitter. We then examine each account individually and assess their type of relationship with the President of USA as below:

1yl – This account has a large follower base of 655k, and it seems like it is a Saudi based account operated by someone who is active in Saudi politics and world politics.

Clijsterskim – Kim Clijsters is a professional tennis player from Belgium who has 6 grand slam titles to her name and was ranked no. 1 in 2003. She doesn’t seem to have any political affiliation with Donald Trump.

Iranintl- Iran international is an official account of a 24/7 persian news Channel and is considered one of the mainstream news channels of Iran. This account is run in Persian language and has a following of more than 200k followers. Considering recent tensions between America and Iran, the relationship between Trump and this media account seems to be of negative nature.

DulaimAlna – This is an official account of Dulaim Al Nasher who is a popular businessman from Saudia and is founder and member of Board of Directors of “THE PROFESSIONAL EVENTS INDUSTRY ASSOCIATION”. Considering USA has a lot of investments in Saudi Arabia, the relationship between Dulaim and Trump could be termed as positive.

gezegen\_mehmet – Mehmet gezegen is a Turkish radio host who runs podcast on radio channels in Turkey. His twitter following of 169k followers suggests that he is quite popular in his people. Given the current sanctions on Turkey by USA, the relationship between Trump and Mehmet seems to be negative.

Nataliazardon – Natalia zardon seems to be an adult star as observed from her twitter bio. Therefore, this account seems irrelevant for our data analysis.

Dannyblu- This is another adult content account that is irrelevant for our analysis.

CriticaSinMiedo – From our observations, this seems to be an account of some Major General of US airforce and has a following of 134k on twitter. This person has done his PhD in Psychoacoustics. Since he is a part of US Airforce, the relationship between Trump and this person can be considered positive.

mariaamm123 – As the screen name suggests, this is an account of a female named Mariam who is based in Saudi Arabia. She has 133k followers and most of her tweets are in Arabic. Due to language barrier, it is hard to observe the nature of relationship between Trump and this account.

naief\_aloutaibi – Naeif Bin Hammad Al Outaibi is a Saudi citizen who is a founder of Touten App, built to help job seekers find jobs easily in Saudi Arabia. Since this account is also run in Arabic language, it is hard to observe the nature of relationship this account has with Donald Trump.

BurroSutil – Burro Sutil is a female, half Colombian half Venezolan who lives in Mexico. She has a good number of followers on twitter (89k) and is following more accounts than her number of followers. Since Mexico and USA are not on good terms with each other even though the two countries share borders, we will consider this relationship to be of negative nature.

Pueblaonline – It is an official account of online newspaper of Puebla and Mexico. The account posts news, videos and other services and is run in native language of Mexico. A total of 72.9 people follows this account. Given that it is a Mexico based News channel, the relationship with Trump will be considered negative.

iLLPeTiLL – This seems to be an account of someone who is passionate about music and produces music regularly. The large number of following of this account suggests that he/she is popular and posts good content on twitter. Unfortunately, there is not much information available on twitter regarding the person who runs this account. So we cannot predict if it is a positive or negative relationship with Trump.

AAltuwaim – A young doctor from Saudi Arabia who enjoys poetry and reading books. Due to his passion for literary, he has a large number of followers (53.4k). Again, since the account is run in arabic language, its difficult to predict the relationship between Trump and this user.

LaVieofaBlonde- This account is private so no information could be collected about this account.

jamelalsaleh0 – Jamel al Saleh is a Syrian resident whose tweets are mainly in Arabic. He seems to be an active twitter member and tweets very often. He also has large following of 40.6k. Since Syria is not on good terms with USA, the relationship with Donald Trump will be considered negative here.

HenryPacheco – Henry Pacheco is a US based chef who works as a Executive chef Sea Seahawk and has a following of 40k. From his retweets and shared posts, Henry seems to be a Trump supporter. Therefore, the relationship between Trump and Henry is positive.

Orblueduck – Katie Porter is an American politician who is currently a US house representative for CA-45. Considering her status in US politics, she has a strong following of 243.5k. Katie belongs to Democratic Party while Trump is a Republican, therefore the relationship between the two is strongly negative.

manal\_\_\_\_\_141- This is a female account named Manal with a following of almost 40k people. Since the account is run in Arabic, it can be said that the person is based in middle east. Due to language barrier, we cannot predict if the person has a positive relationship with Donald Trump.

Kwabs – This account is of a London based brown artist who sings. He is also a social worker who runs a local foodbank for needy people. His large following suggests that he is popular in his local community.

From our data sample of followers, we can see that it is a mixture of different type of people in the list. They range from international sports personalities to international media channels to radio hosts and some random followers. Since trump has one of the highest followings on twitter with more than 82.2M followers, it is too large a data for R application to extract top 20 followers from. Therefore, R chooses a random sample size of followers and picks top 20 from that sample based on sorting of no. of followers.

Limitations:

One of the limitations is related to R where when R is given a large number of users to collect a sample 20 from, it goes through first(most recent) N number of followers of a user account and then picks top 20 from that sample. This compromises the quality of data collected. Another limitation is that the sample data collected does not filter non-human or private accounts. Furthermore, if some account is not relevant for your analysis, you cannot drop it from your sample in R. Due to these limitations, our data for top20 followers is compromised and does not actually represents the top 20 followers of Trump on Twitter.

To improve the quality of analysis and to extract more accurate data from huge followers, a new command/function could be created in R which prefers to read the userIDs who followed Trump first. This will make our data more authentic and more interesting.

### PART 3 – BYPASSING TRUMP

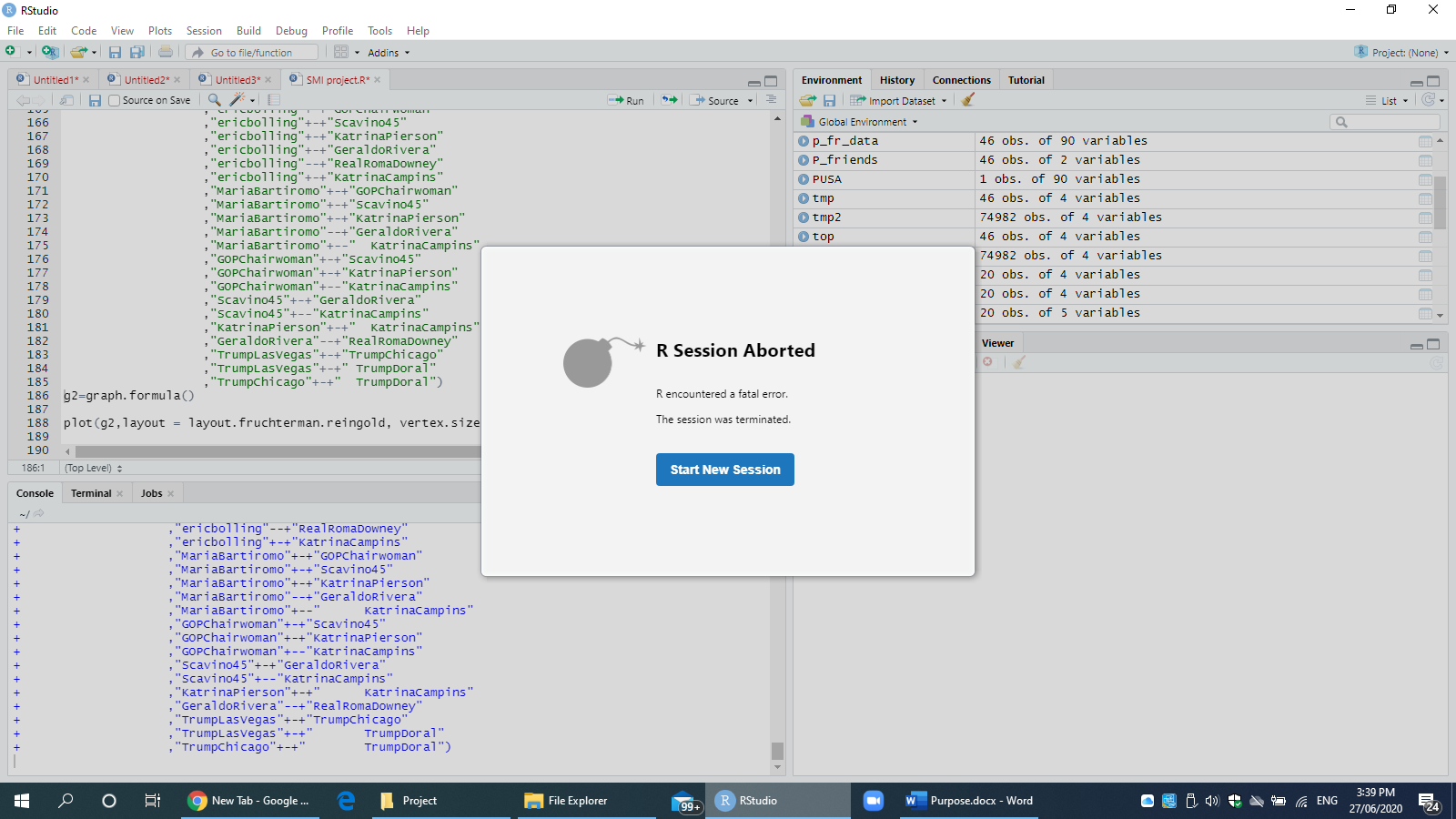
In part 3, we plot a graph of Trump’s 20 friends and 20 followers obtained in part 1 and 2.

A picture containing flower

Description automatically generated

From the graph above, we can see the 20 followers and 20 friends of Trump but none of the friends are friends with any of the followers of Trump. There are two group of nodes and there is no bridge between the two nodes.

We then observe the graph and see if there are any friends of Trump in the graph who are following each other, and any followers of Trump who are following each other.



Due to this continuous error, I was not able to build a graph for friends of Trump who are following each other and followers of Trump who are following each other if any.

Due to a continuous unexpected error in R application, we could not build graphs for friends of Trumps following each other, therefore density and diameter of graph 1 will be calculated for the remaining analysis.

Density:

The density of a graph is the ratio of the number of edges and the number of possible edges. Density of graph 1 is calculated as 0.04645761 which is quite low. This means that the group of networks are weak.

Diameter:

Diameter of a graph is the longest shortest path between two vertices. For graph 1, diameter is calculated as 2.

Neighborhood overlap:

Neighbourhood overlap is the ratio of nodes that are neighbors of i and j, and nodes that are neighbors of either i and j. The smaller the neighbourhood overlap, the closer the edge between i and j is to becoming a local bridge. For graph 1, NO is calculated as 1.

To calculate homophily, we named each node from graph1 as Supporter “S” and Non-supporter “NS” based on their relationship with Donald Trump. We then build a hypothesis to test the homophily for the given graph.

Hypothesis:

H0: Homophily does not exist in this network

H1: Homophily does exist for the given network

We add repetition and randomisation of labels for the edges by using method 1 which is loop method. We plot the histogram to show the cross-edge count for graph 1.

A screenshot of a cell phone

Description automatically generated

The histogram shows that cross edges are more than 30. Now we will cross check this count against graph1 to see if the randomised output is greater or less than the given graph. If the results are same, we cannot reject null hypothesis. If the results are different, we will reject null hypothesis and state that homophily exists.

To observe results, we will see if our p-value is small (pVal = mean(crossEdgeCount < dataCrossEdgeCount).

The p-value turns out to be 0.002 which is quite small. Therefore, we reject null hypothesis and say that homophily exists.

R COMMANDS USED:

library("igraph")

library("rtweet")

PUSA= lookup\_users("realDonaldTrump")

names(PUSA)

PUSA$friends\_count

#finding friends of Trump

P\_friends = get\_friends("realDonaldTrump") #46 friends

View(P\_friends)

p\_fr\_data <- lookup\_users(P\_friends$user\_id)

View(p\_fr\_data)

#top 20 friends of Trump

tmp <- (p\_fr\_data[,c('user\_id',"screen\_name",'friends\_count',"followers\_count")])

top = tmp[order(-tmp$friends\_count),]

head(top)

top20\_fr = top[1:20,]

View(top20\_fr)

write.csv(top20\_fr,file = "top20fr.csv")

trumpfr20 = read.csv("top20fr.csv")

#finding followers of Trump

PUSA$followers\_count

P\_followers = get\_followers(

"realDonaldTrump", n= 90000, retryonratelimit = TRUE

)

?get\_followers

View(P\_followers)

p\_fol\_data <- lookup\_users(P\_followers$user\_id)

View(p\_fol\_data)

#top 20 random followers of Trump

tmp2 <- (p\_fol\_data[,c('user\_id',"screen\_name",'friends\_count',"followers\_count")])

top2 = tmp2[order(-tmp2$followers\_count),]

head(top2)

top20\_fol= top2[1:20,]

View(top20\_fol)

write.csv(top20\_fol,file="top20fol.csv")

trumpfol20 = read.csv("top20fol.csv")

pfrIDs= P\_friends$user\_id

pfolIDs= P\_followers$user\_id

commonpeople = intersect(pfrIDs,pfolIDs)

View(commonpeople)

top20\_frN=top20\_fr$screen\_name

top20\_folN=top20\_fol$screen\_name

#graph1

all = as.data.frame(rbind(c("P\_friends","P\_followers", "black")))

names(all) = c('Type', 'Friend', 'Colour')

View(all)

relations1 = merge(data.frame(Type='P\_friends',Friend=top20\_fr$screen\_name,

Colour ="blue"),

data.frame(Type= "P\_followers", Friend=top20\_fol$screen\_name,

Colour = "red"), all=T)

g1 = graph.data.frame(relations1, directed = FALSE)

V(g1)$label = V(g1)$name

par(mar = c(0,0,0,0))

V(g1)$label.cex = 0.5

plot(g1)

plot(g1, layout = layout.fruchterman.reingold, vertex.size = 20, edge.color = E(g1)$Colour)

#finding friends who are followinf each other

frnfol=lookup\_friendships(top20\_fr$screen\_name,top20\_fol$screen\_name)

View(frnfol)

f=lookup\_friendships("TrumpChicago","TrumpDoral")

View(f)

#performed for all friends of trump against other 19 friends, one at a time.

#plotting th friends graph

g2 = graph.formula("IvankaTrump"+-+"DonaldJTrumpJr","IvankaTrump"+-+"EricTrump","IvankaTrump"+-+"Jim\_Jordan",

"IvankaTrump"+-+"JesseBWatters","IvankaTrump"+-+"KellyannePolls","IvankaTrump"+-+"foxandfriends"

,"IvankaTrump"+-+"greta","IvankaTrump"+-+"ericbolling","IvankaTrump"+-+"MariaBartiromo",

"IvankaTrump"+-+"GOPChairwoman","IvankaTrump"+-+"Scavino45","IvankaTrump"+-+"KatrinaPierson",

"IvankaTrump"+-+"garyplayer","IvankaTrump"+-+"GeraldoRivera","IvankaTrump"+--"RealRomaDowney",

"IvankaTrump"+--"KatrinaCampins","IvankaTrump"+-+"TrumpLasVegas","IvankaTrump"+-+"TrumpChicago",

"IvankaTrump"+-+"TrumpDoral"

,"DonaldJTrumpJr"+-+"EricTrump"

,"DonaldJTrumpJr"+-+"KellyannePolls"

,"DonaldJTrumpJr"+-+"Jim\_Jordan"

,"DonaldJTrumpJr"+-+"foxandfriends"

,"DonaldJTrumpJr"+-+"JesseBWatters"

,"DonaldJTrumpJr"+-+"greta"

,"DonaldJTrumpJr"+-+"ericbolling"

,"DonaldJTrumpJr"+-+"MariaBartiromo"

,"DonaldJTrumpJr"+-+"GOPChairwoman"

,"DonaldJTrumpJr"+-+"Scavino45"

,"DonaldJTrumpJr"+-+"KatrinaPierson"

,"DonaldJTrumpJr"+--"garyplayer"

,"DonaldJTrumpJr"+-+"GeraldoRivera"

,"DonaldJTrumpJr"+--"KatrinaCampins"

,"DonaldJTrumpJr"+-+"TrumpLasVegas"

,"DonaldJTrumpJr"+-+"TrumpChicago"

,"DonaldJTrumpJr"+-+"TrumpDoral"

,"EricTrump"+-+"KellyannePolls"

,"EricTrump"+-+" Jim\_Jordan"

,"EricTrump"+-+""

,"EricTrump"+-+""

,"EricTrump"+-+"foxandfriends"

,"EricTrump"+-+" JesseBWatters"

,"EricTrump"+-+"greta"

,"EricTrump"+-+"ericbolling"

,"EricTrump"+-+"MariaBartiromo"

,"EricTrump"+-+"GOPChairwoman"

,"EricTrump"+-+"Scavino45"

,"EricTrump"+-+"KatrinaPierson"

,"EricTrump"+-+"garyplayer"

,"EricTrump"+-+"GeraldoRivera"

,"EricTrump"+-+"KatrinaCampins"

,"EricTrump"+-+"TrumpLasVegas"

,"EricTrump"+-+"TrumpChicago"

,"EricTrump"+-+"TrumpDoral"

,"KellyannePolls"+-+"Jim\_Jordan"

,"KellyannePolls"+-+"foxandfriends"

,"KellyannePolls"+-+"JesseBWatters"

,"KellyannePolls"+-+"greta"

,"KellyannePolls"+-+"ericbolling"

,"KellyannePolls"+-+"MariaBartiromo"

,"KellyannePolls"+-+"GOPChairwoman"

,"KellyannePolls"+-+"Scavino45"

,"KellyannePolls"+-+"KatrinaPierson"

,"KellyannePolls"+--"GeraldoRivera"

,"KellyannePolls"+--"KatrinaCampins"

,"Jim\_Jordan"+-+"foxandfriends"

,"Jim\_Jordan"+-+"JesseBWatters"

,"Jim\_Jordan"+-+"greta"

,"Jim\_Jordan"+-+"ericbolling"

,"Jim\_Jordan"+-+"MariaBartiromo"

,"Jim\_Jordan"+-+"GOPChairwoman"

,"Jim\_Jordan"+--"Scavino45"

,"Jim\_Jordan"+--"KatrinaPierson"

,"Jim\_Jordan"+--" KatrinaCampins"

,"foxandfriends"+-+"JesseBWatters"

,"foxandfriends"--+"greta"

,"foxandfriends"+--"ericbolling"

,"foxandfriends"+-+"MariaBartiromo"

,"foxandfriends"+--"GOPChairwoman"

,"foxandfriends"+--"Scavino45"

,"foxandfriends"+-+"KatrinaPierson"

,"foxandfriends"+-+"GeraldoRivera"

,"foxandfriends"+--"RealRomaDowney"

,"foxandfriends"+--"KatrinaCampins"

,"JesseBWatters"--+"greta"

,"JesseBWatters"+-+"ericbolling"

,"JesseBWatters"+-+"MariaBartiromo"

,"JesseBWatters"+--"GOPChairwoman"

,"JesseBWatters"+-+"Scavino45"

,"JesseBWatters"+-+"KatrinaPierson"

,"JesseBWatters"+-+"GeraldoRivera"

,"JesseBWatters"+--"KatrinaCampins"

,"greta"+-+"MariaBartiromo"

,"greta"+-+"GOPChairwoman"

,"greta"+-+"Scavino45"

,"greta"+--"KatrinaPierson"

,"greta"+-+"GeraldoRivera"

,"greta"+-+"RealRomaDowney"

,"ericbolling"+-+"MariaBartiromo"

,"ericbolling"+-+"GOPChairwoman"

,"ericbolling"+-+"Scavino45"

,"ericbolling"+-+"KatrinaPierson"

,"ericbolling"+-+"GeraldoRivera"

,"ericbolling"--+"RealRomaDowney"

,"ericbolling"+-+"KatrinaCampins"

,"MariaBartiromo"+-+"GOPChairwoman"

,"MariaBartiromo"+-+"Scavino45"

,"MariaBartiromo"+-+"KatrinaPierson"

,"MariaBartiromo"--+"GeraldoRivera"

,"MariaBartiromo"+--" KatrinaCampins"

,"GOPChairwoman"+-+"Scavino45"

,"GOPChairwoman"+-+"KatrinaPierson"

,"GOPChairwoman"+--"KatrinaCampins"

,"Scavino45"+-+"GeraldoRivera"

,"Scavino45"+--"KatrinaCampins"

,"KatrinaPierson"+-+" KatrinaCampins"

,"GeraldoRivera"--+"RealRomaDowney"

,"TrumpLasVegas"+-+"TrumpChicago"

,"TrumpLasVegas"+-+" TrumpDoral"

,"TrumpChicago"+-+" TrumpDoral")

plot(g2,layout = layout.fruchterman.reingold, vertex.size = 20, edge.color = E(g2)$Colour,edge.arrow.size=0.4)

#graph density and diameter

graph.density(g1)

diameter(g1)

?diameter

## Computing the neighborhood overlap of each edge

## get the neighborhood graph of all nodes.

gn = neighborhood(g1, order = 1)

## get pair of nodes that are at the end of each edge.

g.ends = ends(g1, E(g1))

# number of edges

N = nrow(g.ends)

# make space for neighborhood overlap score

NO = rep(0, N)

for (a in 1:N) {

## for every edge

x = g.ends[a,1] # x is the node at one end of the edge

y = g.ends[a,2] # y is the node at the other end of the edge

## compute the intersection of the neighbourhoods of x and y

i = length(intersect(gn[[x]], gn[[y]])) - 2

## compute the union of the neighbourhoods of x and y

u = length(union(gn[[x]], gn[[y]])) - 2

## Note that we subtract 2 since each neighbourhood includes x and y.

## we don't want to include x and y in the counts.

## the neighbourhood overlap is the intersection/union

NO[a] = i/u

}

NO[a]

#Homophily

par(mar = c(0, 0, 0, 0))

V(g1)$label.cex = 0.5

V(g1)$label = c("S", "S", "S", "S", "S", "S", "S",

"NS", "S", "NS", "S", "NS", "S", "S", "S", "S", "S",

"S", "S", "S", "S", "S", "NS", "S",

"NS", "S", "NS", "NS", "NS", "S", "S", "S", "S", "S",

"NS", "NS", "S", "NS", "S", "NS", "S", "S")

plot(g1, layout = layout.fruchterman.reingold, vertex.size = 20)

V(g1)

class = c("S", "S", "S", "S", "S", "S", "S",

"NS", "S", "NS", "S", "NS", "S", "S", "S", "S", "S",

"S", "S", "S", "S", "S", "NS", "S",

"NS", "S", "NS", "NS", "NS", "S", "S", "S", "S", "S",

"NS", "NS", "S", "NS", "S", "NS", "S", "S")

## extract the adjacency matrix

A = get.adjacency(g1)

## repeat 1000 times

R = 1000

###----- Method 1: Using a for loop

crossEdgeCount = rep(0,R)

for (a in 1:R) {

### permute the row and column labels

permutedClass = sample(class)

### record the number of cross S-NS edges

xPos = which(permutedClass == "S")

yPos = which(permutedClass == "NS")

crossEdgeCount[a] = sum(A[xPos, yPos])

}

## examine the cross edge distribution

par(mar = c(2, 2, 3, 2))

hist(crossEdgeCount,40)

## compare the number of cross S-NS edges from the data to the

## random distribution

xPos = which(class == "S")

yPos = which(class == "NS")

dataCrossEdgeCount = sum(A[xPos, yPos])

## If the they look the same we cannot reject H0

## If they look different, we can reject, meaning there is homophily

pVal = mean(crossEdgeCount < dataCrossEdgeCount)

pVal

## p value is very small, so reject H0: there is no homophily,

## therefore there must be homophily