A NETWORK ANALYSIS OF TRUMP'S TWITTER FRIENDS AND FOLLOWERS

Introduction

Social media has become an important tool in political engagement and interaction over recent years (Getachew & Beshah, 2019). Nowhere is this more apparent than in Donald J Trump's tenure, as the 45th President of the United States. He has been described as the first 'social media' president, due to his prolific use of Twitter (Navarria, p 92, 2019). In conjunction with these developments and others, social media, and twitter specifically has been subject to scrutiny on political and sociological grounds. Twitter is a social networking website that allows users to post messages, known as 'tweets' within a 280-character limit (Balboa, *Design Critique: Twitter*, 2019).

The aim of this report is to analyse Trump's twitter relationship with his friends and followers and the impact he has on them via network analysis. The analysis will be divided into six parts as directed by FoT (Friends of Trump group), employing a combination of qualitative and quantitative methodologies to extract insights from Twitter. In each section the aim of the task will be stated and explained, the methodology used (which consists of also explaining the code written in R), and insights and conclusions found, and lastly, problems and issues with the analysis what could be done in future to improve it.

PROGRAMS USED IN ANALYSIS

The programming language R via R studio was used for the twitter analysis done in this report including all graphs. Specifically, the packages "rtweet", "base64enc" and "httpuv" were used to scrape data from Twitter. "Igraph" was used to build the network plots and graphs.

ANALYSIS

PART 1: Friends of Trump.

In this section we are concerned with finding the top 20 friends of Trump that have the most followers. We will then examine the twitter accounts and summarise the people.

Twitter defines friends and followers distinctly for the purposes of the social media network. Friends are Twitter users that a specific user follows (e.g. people that they are following). In contrast followers are Twitter users who are following a particular user. (Developer Twitter, 2020).

Given this, we are interested in finding the 20 friends of Trump that have the most followers. This means, out of the people that Trump follows, we want the top 20 who themselves have the most followers.

The function to use for this is **get_friends().** This function gives us a number of twitter ID's of users who follow trump.

```
trump_friends <- get_friends("@realDonaldTrump",
                              n = 5000)
```

In this case, the number of potential friends of trump we can retrieve is 5000. This can be problematic if Trump has more than 5000 friends, however this isn't an issue as Trump's follower count is less than this.

```
trump_friends_info <- lookup_users(trump_friends$user_id)

trump_friends_names <- trump_friends_info$screen_name

trump_friends_names
```

Using the **lookup_users()** function we can find the corresponding screen names of the user ID's gathered from the **get_friends()** function. This enables us to see the twitter names of the Trump's friends. By getting their names, we can filter out the Trump company twitter handles that are also following Trump, since these won't be pertinent to our analysis.

```
 [1] "JudgeJeanine"    "Jim_Jordan"      "MariaBartiromo"  "VP"              "GOPChairwoman"   "parscale"
 [7] "PressSec"        "TuckerCarlson"   "JesseBWatters"   "WhiteHouse"      "Scavino45"       "KellyannePolls"
[13] "Reince"          "RealRomaDowney"  "Trump"           "TrumpGolf"       "TiffanyATrump"   "IngrahamAngle"
[19] "Mike_Pence"      "TeamTrump"       "MrsVanessaTrump" "LaraLeaTrump"    "seanhannity"     "CLewandowski_"
[25] "DiamondandSilk"  "KatrinaCampins"  "KatrinaPierson"  "foxandfriends"   "MELANIATRUMP"    "GeraldoRivera"
[31] "ericbolling"     "MarkBurnettTV"   "garyplayer"      "VinceMcMahon"    "DanScavino"      "TrumpDoral"
[37] "TrumpCharlotte"  "TrumpLasVegas"   "TrumpChicago"    "TrumpGolfDC"     "TrumpGolfLA"     "EricTrump"
[43] "BillOReilly"     "greta"           "DonaldJTrumpJr"  "IvankaTrump"
```

Figure 1: List of Trump's friends.

In figure 1 we can see all of Trump's friends and that he has 46 of them. This is much lower than the 5000 max limit the **get_friends()** function can extract so there is no problem here. The next step is to therefore remove the Trump's twitter company handles.

```r
company_names <-  c("TrumpGolf", "TrumpDoral", "TrumpCharlotte", "TrumpLasVegas",
                                          "TrumpChicago", "TrumpGolfDC", "TrumpGolfLA")

N = length(company_names)
indi <- c()

for (i in 1:N){
  indi[i] <- which(trump_friends_names == company_names[i])
}

trumpfriends_filt <- trump_friends_names[-indi]
```

The code above looks for the indices in the **trump_friends_names** vector where his twitter handle company names are placed. They are then removed from the vector leaving the rest of his friends in the vector.

```r
# Count how many followers the friends of trump have and count the top 20 friends with the most.

followers_of_tfriends <- lookup_users(trumpfriends_filt)$followers_count

followers_table <- cbind(trumpfriends_filt, followers_of_tfriends)

followers_table <- as.data.frame(followers_table)

followers_table[ ,1] <- as.character(followers_table[, 1])
followers_table[ ,2] <- as.numeric(as.character(followers_table[, 2]))

str(followers_table)

top_20 <- order(followers_table$followers_of_tfriends, decreasing = TRUE)[1:20]
top_20

# top 20 friends of trump
fot_followers20 <- followers_table[top_20, ] #decided to keep whitehouse in it.
```

We can now find the top 20 friends of trump with the most followers. The code above shows the process of filtering out for the top 20 friends with the most followers. We use the **lookup_users()** function and focus on the **followers_count** column. We then use the **order()** function to find the top 20 followers counts of Trump's friends on twitter. The top 20 followers of trump are shown in table 1 with their corresponding followers count.

Summarizing the twitter accounts of Trump's top 20 friends can then be done by exploring the **description** column of the **lookup_users()** function. This is summarized in table 2 below. The people that are friends of trump who also have the most followers tend to be political commentators, senators, and those active in the American political landscape. Most, if not all of his friends with the most followers are either republicans and or advocates for the party and family members.

```
#info of followers
tfriends_info <- lookup_users(fot_followers20$trumpfriends_filt)
tfriends_info$description
followers_des <- cbind(fot_followers20, tfriends_info$description)
```

	trumpfriends_filt	followers_of_tfriends
10	WhiteHouse	22478912
4	VP	9082081
39	IvankaTrump	8825935
38	DonaldJTrumpJr	5058894
22	seanhannity	4905802
18	Mike_Pence	4750630
7	PressSec	4663623
35	EricTrump	3561908
17	IngrahamAngle	3438589
8	TuckerCarlson	3291266
12	KellyannePolls	3259156
36	BillOReilly	3218219
33	VinceMcMahon	2396570
1	JudgeJeanine	2067884
19	TeamTrump	1725972
24	DiamondandSilk	1440969
2	Jim_Jordan	1370191
27	foxandfriends	1362877
9	JesseBWatters	1224894
37	greta	1217499

Table 1: Top 20 friends of Trump with the most followers.

From the analysis shown above, we can see that all of Trump's friends are those who share similar political leanings and or family ties. This indicates that Trump does not follow too many people who are dissimilar to himself and supports the basic network thesis that people tend to be attracted to likeminded people, or similar to themselves in socially significant ways. (Park, p 1, Lecture Notes 3, 2020). This is also called Homophily. While this isn't a strict quantitative or technical statement of such a principle, it is a basic qualitative generalization we can pronounce from the findings in this section.

There are a few issues and possible ways to improve the analysis for the future. In removing Trump's corporate twitter handles from the friends list, one could ask whether or not the official White House twitter account should have also been removed. The decision to leave it in was purely one of taking a conservative approach. It was not a corporate Trump account and the FoT did not explicitly mention what to do with such an account. However, it would also be a reasonable assumption to choose to remove it from the analysis. Categorizing the friends of Trump into exact categories instead of a rough summary of who they were may also be a better way of analysis in the future.

trumpfriends_filt	followers_of_tfriends	tfriends_info$description
IvankaTrump	8825935	Wife, mother, sister, daughter. Advisor to POTUS on job ...
WhiteHouse	22478912	Welcome to @WhiteHouse! Follow for the latest from Pr...
VinceMcMahon	2396570	Vince McMahon, Chairman & CEO of WWE, Inc., is a thir...
Mike_Pence	4750630	Vice President of the United States
VP	9082081	Vice President Mike Pence. Husband, father, & honored ...
seanhannity	4905802	TV Host Fox News Channel 9 PM EST. Nationally Syndicat...
TeamTrump	1725972	The official Twitter for the Trump Campaign. Together, w...
BillOReilly	3218219	Simple man looking out for the folks. Watch the @NoSp...
Jim_Jordan	1370191	Proudly serving Ohio's beautiful Fourth District. Rankin...
DiamondandSilk	1440969	President Donald J Trump's Most Loyal Supporters. Host ...
KellyannePolls	3259156	Mom. Patriot. Catholic. Happy Warrior. Counselor. Surviv...
IngrahamAngle	3438589	Mom, author, host, The Ingraham Angle, 10p ET @FoxN...
JudgeJeanine	2067884	Judge Pirro is a highly respected District Attorney, Judge...
greta	1217499	Host of Full Court Press (Sunday show syndicated by Gra...
TuckerCarlson	3291266	Host of "Tucker Carlson Tonight", weeknights at 8 PM ET...
EricTrump	3561908	Executive Vice President of The @Trump Organization. H...
DonaldJTrumpJr	5058894	EVP of Development & Acquisitions The @Trump Organi...
JesseBWatters	1224894	Co-Host of @TheFive & Host of @WattersWorld on Fox ...
foxandfriends	1362877	America's #1 cable morning news show
PressSec	4663623	@WhiteHouse Press Secretary. Connecting the American...

Table 2: Description of Trump's top 20 friends.

PART 2: Followers of Trump.

In this section we will look at the top 20 followers of trump according to have the most followers and determine if they have a negative or positive relationship with trump based on their tweets.

As stated in the previous section. A follower is somebody who followers a particular account or user, in this case, we are interested in those that follow Donald Trump. Firstly, we use the **get_followers()** function to download a large list of people who follow Donald Trump.

```
followers_trump25k <- get_followers("realDonaldTrump", n = 25000)

tfollowers_followers <- as.data.frame(lookup_users(followers_trump25k$user_id))

trumpf_f <- tfollowers_followers[ , c("user_id","screen_name", "followers_count")]

top_20fol <- order(trumpf_f$followers_count, decreasing = TRUE)[1:20]
```

In order to get a list of the top 20 followers of Trump we use the **get_followers()** function and take a sample of 25,000 followers of trump. We then get the user_id of these 25,000 users and turn the output into a dataframe. Then we add **screen_name** and **followers_count** to the columns of the dataframe and find the top 20 followers of trump by how many followers they have. A table of the top 20 followers of trump is shown in table 3.

	user_id	screen_name	followers_count
1785	228794007	DonCheadle	837076
7965	4236806414	JohnKStahlUSA	102881
20832	1218558561061613574	queensavagedoll	88854
19257	299158459	Abdulmalik_fr	63350
8794	2836303298	rikomrnk	55018
4661	815824583911673856	Gideon__Lagat	40112
4001	277541083	thedrsec	35137
5678	1004589941328838656	VIP0552525000	34223
20942	299239037	RightWingLawMan	28871
18785	134588802	MintzGolf	27586
1367	736914891039072257	pinoyhotxxx	26357
11815	1058320140176158725	Khanpk122	22279
21996	1017532453337133056	rebeccabutlerm2	22065
16449	1144719727618658305	maturetsceline	20711
8328	196377029	maquialifraco	20599
9844	270721352	ceibonacional	15793
12599	2881781831	AshleyJohnsonh	15720
9565	1261717719231143938	InformedPartiot	15695
10984	321816956	KenzieKay69	15312
15763	1103068483380826112	alkrit2	14464

Table 3: Top 20 followers of Trump by followers count.

Don Cheadle the American actor, famous for his roles in movies such as Iron Man is the top follower of Donald J Trump with of a follower count of just over 800,000. John K Stahl is the person with the second most followers count after Don Cheadle who also follows Donald Trump. He is a radio commentator and Donald Trump supporter. Mintz Golf is a golf commentator, Dr Sec is a news commentator and Gideon Lagat is a footballer player in Europe. The rest of the followers for Donald Trump in the top 20 of follower counts are relatively unknown.

In order to summarize whether or not each individual had a negative or positive relationship to trump, each twitter account was examined qualitatively by going into each twitter account and looking through their tweets to determine how they felt about Trump using the search section of their twitter account page. The following table 4 is a summary given this qualitative analysis.

	Followers	Positive Relationship?
1	DonCheadle	No
2	JohnKStahlUSA	Yes
3	queensavagedoll	Yes
4	Abdulmalik_fr	Yes
5	rikomrnk	No
6	Gideon__Lagat	No
7	thedrsec	Yes
8	VIP0552525000	Yes
9	RightWingLawMan	Yes
10	MintzGolf	Yes
11	pinoyhotxxx	Yes
12	Khanpk122	Yes
13	rebeccabutlerm2	Yes
14	maturetsceline	No
15	maquialifraco	Yes
16	ceibonacional	No
17	AshleyJohnsonh	Yes
18	InformedPartiot	Yes
19	KenzieKay69	No
20	alkrit2	No

Table 4: Relationship with Trump

In summarising the top twitter followers of Trump, again we can use the **description** column from the **lookup_users()** function. As stated above, most of these people are not well known and are mostly random twitter accounts (some which may be considered "spam" accounts). The majority of the top 20 followers seem to be Trump followers. Only four of the top 20 followers have a 'negative' relationship with Trump.

From this section of the analysis the data seems to suggest that the majority of followers of Trump are advocates or have a positive relationship with Trump. This reinforces the idea that people who are likeminded tend to attract each other, and in the case of Trump's Twitter followers, they tend to be mostly supporters.

One limitations of this section in the analysis is with the sample size of the followers from **get_followers()** function. Since we could not extract all the followers of Trump since it would be too large, the decision to download a dataset of size 25,000 was made. It is possible that there are followers of trump with a larger follower count than was extracted from this 25,000-sample size.

screen_name	description
DonCheadle	OFFICIAL DON CHEADLE, OG Kung Fu Kenny <U+0001F9...
NavyFlyBoyUSA	Make America Great Again means we make our own stuf...
queensavagedoll	<U+0001F451>Adult STAR <U+0001F31F> Management...
Abdulmalik_fr	<U+0633><U+062A><U+062C><U+062F> <U+0645>...
rikomrnk	<U+533B><U+5E2B>·<U+30B8><U+30E3><U+30FC>...
Gideon__Lagat	#KOTLoyals Chelsea fan<U+0001F499>#KTBFFH #KOTwave
thedrsec	My radio program "The Waiting Room with Dr. SEC" airs ...
VIP0552525000	<U+200F><U+0627><U+0644><U+062F><U+0648><...
RightWingLawMan	Cowboy Action Shooter. God, Family, Freedom! <U+000...
MintzGolf	Golf analyst <U+0001F399> Golf Ambassador @exec_to...
pinoyhotxxx	all pinoy hottie and all pinoy hot vids plus straight porn!
Khanpk122	<U+200F><U+200F><U+200F><U+200F><U+200F><...
rebeccabutlerm2	Nothing much to see here. Nurse to barrister. @speechu...
maturetsceline	18 PLUS ADULT ONLY Very Amateur Mature transexual M...
maquialifraco	Abogada. Made in los 90's<U+2728>
ceibonacional	No hay esperanza donde la Corrupcion se instala,No ha...
AshleyJohnsonh	<U+0001F525>Sexy Single looking For casual Sex<U+0...
RealNickJericho	An 18 year old that knows the truth. I like exposing Libs ...
alkrit2	<U+200F><U+200F><U+200F><U+200F><U+200F><...

Table 5: Description of top 20 Trump Followers

Ways to improve the methodology for the future would be either (i) download the whole dataset of followers of Trump over a period of time or (ii) take many samples and pick the highest follower(s) from each and combine them. The problem with the first approach is that it is too time consuming and may take too much computing power. The second approach is much more efficient however, there is still an issue of not capturing the true top 20 followers of Trump with the most follower counts.

Another limitation of the study was the methodology employed for determining whether or not a follower had a positive or negative relationship with Trump. Using a more quantitative approach by text mining and using sentiment analysis of tweets about Trump from the top 20 followers may be a more robust approach for the future.

PART 3: Bypassing Trump

In this section of the report, we will construct a network graph of the friendships between Trump's friends and followers. Additionally, any friends or followers of Trump that are not friends with one another that we might infer should be will also be connected via edges in the network graph.

The first step in plotting the graph is to combine the friends and followers of Trump into one vector for ease of analysis.

```
friends_and_followers <- data.frame(fof = c(fot_followers20$trumpfriends_filt, top_20list$screen_name))

friends_and_followers <- as.vector(friends_and_followers)

friends_and_followers <- apply(friends_and_followers, 1, FUN = as.character)
```

The code above gives us a vector of both the top 20 friends and followers of Trump determined by their respective follower counts.

In the next step we want to determine whether or not the friends and followers of Trump are also friends with one another. To do this we use the following code:

```
#lookup_friendships

adj_val <- as.data.frame(matrix(data = 0, nrow=40, ncol=40, byrow = TRUE))

for(i in 1:40){
  for(j in 1:40){
    if(dim(lookup_friendships(friends_and_followers[i], friends_and_followers[j]))[1] < 1){
      adj_val[i ,j] <- "FALSE"
    }else{adj_val[i,j] <- lookup_friendships(friends_and_followers[i], friends_and_followers[j])[[11,4]]}
  }
}

adj_val
```

This gives us a matrix of the friendships between the followers and friends of Trump, where the TRUE values represent friendship edges, while FALSE represent no friendship edges between two people. To represent as an adjacency matrix we can use the following code:

```
adj_valx <- adj_val
for(i in 1:dim(adj_val)[1]){
  for(j in 1:dim(adj_val)[2]){
    if(adj_val[i,j] == "TRUE"){
      adj_valx[i, j] <- 1
    }else{ adj_valx[i, j] <- 0}
  }
}

adj_valx <- as.matrix(adj_valx)
adj_valx <- apply(adj_valx, 2, FUN = as.numeric)
rownames(adj_valx) <- friends_and_followers
colnames(adj_valx) <- friends_and_followers

adj_valx <- as(adj_valx, "dgCMatrix")

ga <-graph.adjacency(adj_valx, mode = "undirected")

plot(ga)
```

From looking at this adjacency matrix we can see which followers and friends of Trump are also friends and construct a preliminary graph.

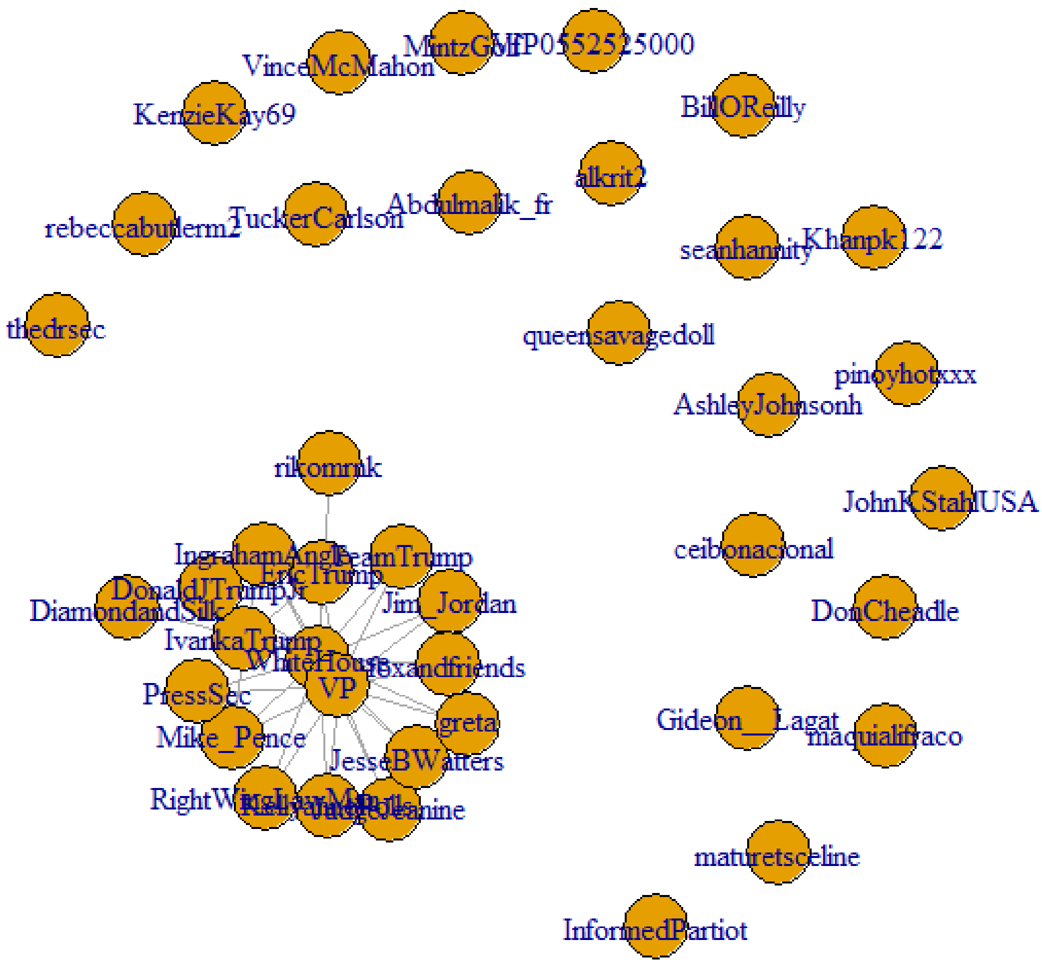

Figure 2: Preliminary friendship network of Trump's followers and friends.

As we can see, there is a cluster of connected people and many disconnected nodes with no edges at all. It seems from inspection of the network graph that it is mostly Trump's friends who are highly connected, while his followers are not friends with each other. This can be largely attributed to the large amount of followers Trump has (over 86 million) and the number of friends he has (only 46). It is much more likely that 46 people that Trump actively follows, know each other, compared to a random sample from 86 million people knowing each other.

Given the background of Trump's friends and followers, a more precise graph can be constructed. Using the information about the user's backgrounds from parts I and II, we create more edges between people who we think would be friends. The edges can be seen in the following code below which detail the possible friendships between the 40 nodes.

```
g = graph.formula("WhiteHouse" - "VP", "WhiteHouse" - "IvankaTrump", "seanhannity", "WhiteHouse" - "DonaldJTrumpJr",
                  "WhiteHouse" - "Mike_Pence", "WhiteHouse" - "PressSec", "WhiteHouse" - "EricTrump",
                  "WhiteHouse" - "IngrahamAngle", "TuckerCarlson", "WhiteHouse"-"KellyannePolls", "BillOReilly",
                  "VinceMcMahon", "WhiteHouse" - "JudgeJeanine", "WhiteHouse" - "TeamTrump",
                  "WhiteHouse" - "DiamondandSilk", "WhiteHouse" - "Jim_Jordan", "WhiteHouse" - "foxandfriends",
                  "WhiteHouse"- "JesseBWatters","WhiteHouse"-"greta", "DonCheadle", "JohnKStahlUSA", "queensavagedoll",
                  "Abdulmalik_fr", "WhiteHouse"- "rikomrnk", "Gideon_Lagat", "thedrsec", "VIP0552525000",
                  "WhiteHouse"-"RightWingLawMan", "MintzGolf", "pinoyhotxxx", "Khanpk122", "rebeccabutlerm2",
                  "maturetsceline", "maquialifraco", "ceibonacional", "AshleyJohnsonh", "InformedPatriot",
                  "KenzieKay69", "alkrit2",

                  "VP"-"WhiteHouse", "VP"-"IvankaTrump", "VP"-"DonaldJTrumpJr", "VP"-"Mike_Pence", "VP"-"PressSec",
                  "VP"-"EricTrump", "VP"-"IngrahamAngle", "VP"-"KellyannePolls", "VP"-"JudgeJeanine",
                  "VP"-"TeamTrump", "VP"-"Jim_Jordan", "VP"-"foxandfriends", "VP"-"JesseBWatters", "VP"-"greta",
                  "VP"- "RightWingLawMan",

                  "IvankaTrump"-"DonaldJTrumpJr", "IvankaTrump"-"Mike_Pence", "IvankaTrump"-"PressSec",
                  "IvankaTrump"-"EricTrump", "IvankaTrump"-"IngrahamAngle",

                  #what i added
                  "TuckerCarlson"-"foxandfriends", "TuckerCarlson"-"seanhannity", "TuckerCarlson"-"Mike_Pence",
                  "TuckerCarlson"-"IngrahamAngle", "TuckerCarlson"-"VP", "TuckerCarlson"-"BillOReilly",
                  "TuckerCarlson"-"EricTrump", "TuckerCarlson"-"DonaldJTrumpJr", "TuckerCarlson"-"IvankaTrump",

                  "MintzGolf"-"TuckerCarlson", "MintzGolf"-"seanhannity", "MintzGolf"-"BillOReilly",
                  "MintzGolf"-"foxandfriends", "MintzGolf"-"IngrahamAngle", "MintzGolf"-"KellyannePolls",



                  "DonCheadle"-"WhiteHouse", "DonCheadle"-"VinceMcMahon",

                  "VinceMcMahon"-"WhiteHouse", "VinceMcMahon"-"IvankaTrump", "VinceMcMahon"-"DonaldJTrumpJr",
                  "VinceMcMahon"-"Mike_Pence", "VinceMcMahon"-"PressSec",
                  "VinceMcMahon"-"EricTrump", "VinceMcMahon"-"IngrahamAngle", "VinceMcMahon"-"KellyannePolls",
                  "VinceMcMahon"-"JudgeJeanine",
                  "VinceMcMahon"-"TeamTrump", "VinceMcMahon"-"Jim_Jordan", "VinceMcMahon"-"foxandfriends",

                  "JohnKStahlUSA"-"JesseBWatters", "JohnKStahlUSA"-"greta",
                  "JohnKStahlUSA"- "RightWingLawMan", "JohnKStahlUSA"-"TuckerCarlson",

                  "rebeccabutlerm2"-"InformedPatriot",
                  "thedrsec"-"JohnKStahlUSA", "thedrsec"-"TuckerCarlson", "thedrsec"-"BillOReilly",
                  "thedrsec"-"seanhannity", "thedrsec"-"foxandfriends", "thedrsec"-"greta",
                  "thedrsec"-"KellyannePolls",

                  #sport
                  "Gideon_Lagat"-"thedrsec",

                  #strippers/porn/spam
                  "queensavagedoll"-"pinoyhotxxx", "queensavagedoll"-"AshleyJohnsonh",
                  "AshleyJohnsonh"-"pinoyhotxxx")
```

This gives us the final network graph for our analysis as shown in figure 3. Don Cheadle, Tucker Carlson, Vince McMahon, Mintz Golf and Dr Sec formed edges within the main component of the network. Don Cheadle being an actor and entertainer, may have a friendship with Vince McMahon who is also in the entertainment industry. It is also the case that Don Cheadle, given that he follows Donald Trump, should also follow the White House twitter account.

Vince McMahon forms edges with the main cluster given that his wife was the Administrator of the Small Business Administration from 2017 to 2019 for the Trump administration. (Jagoda, 2017). Tucker Carlson is one of the biggest political commentators for Fox News, a right-wing leaning news organization, so it would be natural to expect him to have ties to main cluster of the Trump's top friends and followers' network. Mintz Golf and Dr sec are both media figures so they too would have connections with other media personalities and political figures.

The insight we can take from this section of the report is that there is a main cluster of people who are highly connected and form the main component of the network, while the rest of the nodes are out in the peripheries with low connections and mostly isolated. These people who comprise of the main component of the network are either those followed by Trump himself (his friends) or important figures in entertainment, media, and politics. Network theory has presented the idea that there are usually giant connected components in a network and in this instance with Trump's followers and friends network graph, this seems to be the case (Park, p 14, lecture 1, 2020).

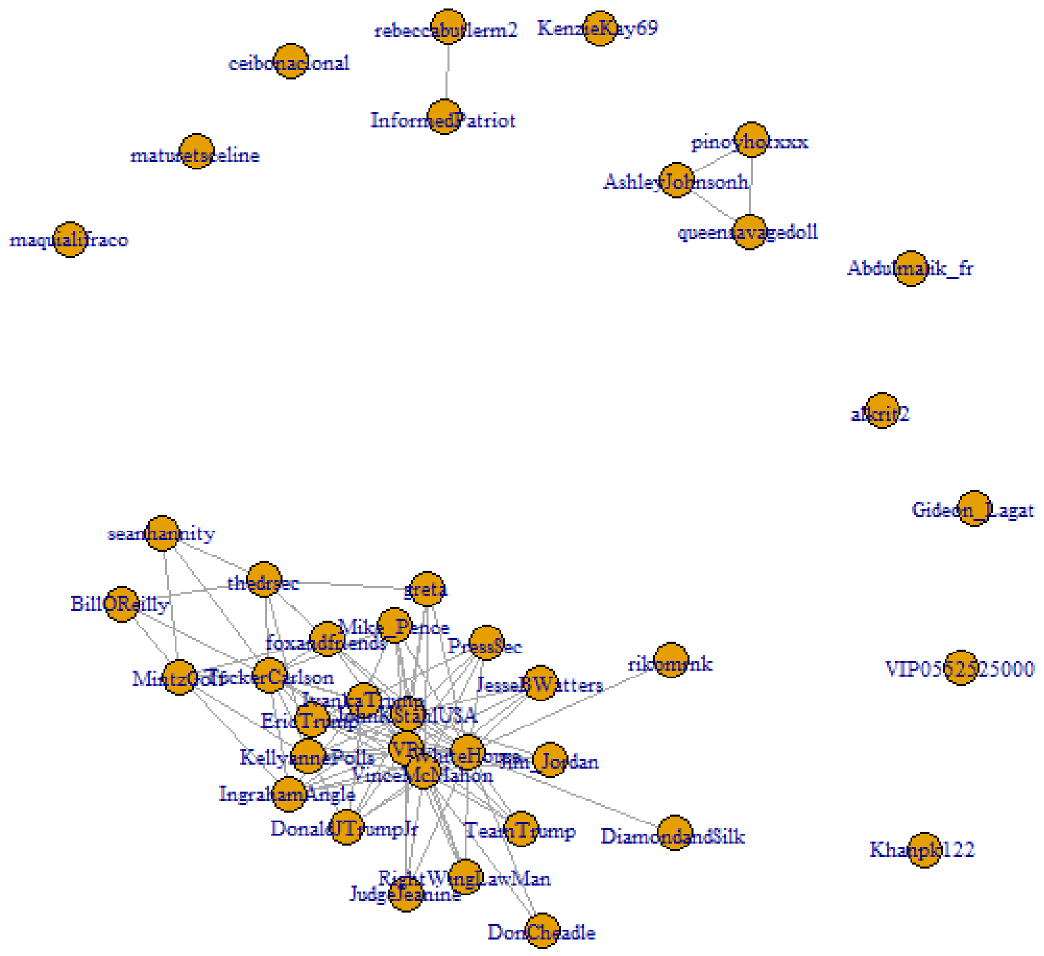

Figure 3: Network graph of Trumps followers and friends revised.

One way to improve the methodology of this section would be to develop a more quantitative way of determining whether or not friends and followers of Trump would be friends. Developing a more consistent approach that is automated, such as using keywords within users' tweets to indicate friendship or not would be one way.

PART IV: GRAPH STATISTICS

In this section of the report we will calculate the diameter and density of the graph and neighbourhood overlap and determine which of the nodes have the greatest social capital.

Firstly, we must define the concepts of diameter, density and neighbourhood overlap in the context of network theory. The diameter of a network can be defined as the shortest distance between the two most distant nodes in a network. In other words, of the pair of nodes which contain the longest paths in a network, it is the shortest.

The density of a graph can be defined as the ratio of number of edges and the number of possible edges (igraph R manual pages, 2020). It is the ratio of how many edges exist in a network and the number of edges that could exist potentially in a network.

Calculating these two metrics can be done using the igraph functions **diameter()** and **graph.density().** The following code below performs this calculation:

```
# DIAMETER
# It is the shortest distance between the two most distant nodes in the network.

dia.g <- diameter(g)

# Density
# Density is the number of actual edges as a ratio of possible maximal edges.

den.g <- graph.density(g)
```

Which give values 4 and 0.123 respectively. From the values we can see that the density is not very high from an absolute scale. Roughly, only 12% of the potential network edges have been realised in the network. Therefore, it seems like the network of Trump's top friends and followers isn't highly connected as it could potentially be.

The next metric we will measure is neighbourhood overlap. The concept of neighbourhood overlap is intimately tied up with the notion of bridges. A bridge is an edge that if deleted, would cause the node pairs that make up the endpoints of an edge to lie in two separate components of the network. (Park, p3, Lecture 2, 2020). It is rare to find a bridge in social networks due to the connectivity of the nodes (Park, p 3). However, there is a weaker version of the notion of a bridge called a 'local bridge'. This can be defined as the edge between two pairs of nodes that have no friends in common. The span on the local bridge is the shortest path between two pairs of nodes when it is not removed (Park, p 4). To generalize the notion of a local bridge, we can look at the concept of 'neighbourhood overlap'.

The neighbourhood overlap of an edge connecting A and B is the ratio:

$$\frac{\text{number of nodes who are neighbors of both A and B}}{\text{number of nodes who are neigbors of atleast one of A and B}}$$

In other words, we can determine to what degree an edge is a 'local' bridge. When the ratio equals 0, then we have a local bridge at minimum. (Easley et al., p 52). The numerator of the neighbourhood overlap is embeddedness.

For the Trump top friends and followers' network, we can calculate the neighbourhood overlap of the edges using the code shown:

```
#overlap

#neighbors of both A and B/neighbors of A or B.

#use ends function to get vertices of edge
# intersect and neighbors

n_pairs <- ends(g, E(g))
n_pairs <- as.data.frame(n_pairs)

neighborhood_overlap <- function(x){
  node_pairs <- ends(x, E(x))
  numerator <- c()
  denom <- c()
  no <- c()
  for(i in 1:length(E(x))){
    numerator[i] <- length(intersect(neighbors(x, v = node_pairs[i,1]),
                                     neighbors(x, v = node_pairs[i, 2])))
    denom[i] <- length(union(neighbors(x, v=node_pairs[i, 1]),
                             neighbors(x, v=node_pairs[i,2])))-2
    no[i] <- numerator[i]/denom[i]

  }
  return(no)
}

n_overlap <- neighborhood_overlap(g)
n_pairs[ , 3] <- n_overlap
no_edges <- n_pairs
```

By looping over all the node pairs with edges, we can calculate the neighbourhood overlap for all the edges to see how 'close' they are to being local bridges.

```
# Three Local bridges

ind_lb <- which(no_edges[, 3] == min(no_edges[, 3], na.rm = TRUE))
no_edges[ind_lb, ]

# Whitehouse - DiamondandSilk
# Whitehouse - rikomrnk
# KellyannePolls - MintzGolf
```

	V1	V2	V3
	WhiteHouse	DiamondandSilk	0
	WhiteHouse	rikomrnk	0
	KellyannePolls	MintzGolf	0

Table 6: Neighbourhood Overlap of 0 (Local Bridges)

From table 6 we can see that there are three edges which have a neighbourhood overlap of 0, indicating they are local bridges at minimum. These are the “WhiteHouse-DIamondandSilk”, “Whitehouse-rikomrnk” and “KellyannePolls-MintzGolf” edges.

From this, we can conclude which twitter users have the most social capital. Individuals that occupy bridges have higher social capital, since they can control the flow of information more since they ‘bridge’ different people together in the case of bridges, and more efficiently in the case of local bridges. Therefore, the users WhiteHouse, DiamondandSilk, rikomrnk, KellyannePolls and MintzGolf can be considered to have the highest social capital in the network as they have the lowest neighbourhood overlap scores. Therefore, these nodes play an important role in Trump’s network of friends and followers on Twitter. The results are relatively obvious from looking at the graph. This is because we can see how the edges don’t have any neighbours in common, but the nodes have friends that aren’t connected to the nodes in the other side of the local bridges directly.

The limitations of the methodology of this section is that structural holes were not considered in the analysis of social capital. Structural holes can be defined as nodes, when removed, would break the network into multiple components (Park, p 8). Nodes that occupy structural holes have power as a ‘gatekeeper of information’ (Park, p 8). Therefore, they are also important in the calculation of social capital. Any further analysis or report should include the analysis of structural holes.

PART V: Graph Homophily

This section of the report is concerned with Homophily in the graph. Homophily provides us with an illustration of how a network’s context can drive the formation of its edges and links (Easley et al., p 78). The concept of homophily can be defined as the principle that we tend to be similar to our friends (Park, Lecture 3, p 2).

In the case of networks, Homophily arises when nodes which have the same properties are connected to each other more so than they are connected with nodes in which they have different properties. In this section of the report, we will measure whether or not the Trump friends and followers graph displays the phenomena of Homophily by doing a statistical test. The friends and followers in the network will be assigned to be “supporters” or “non supporters” of Trump and then we will run statistical test to see if there is evidence of homophily.

The assignment of being a supporter or non-supporter of Trump is the same as whether or not one had a positive or negative relationship with Trump in the earlier sections of the report.

```
positive <- c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
              "Yes", "Yes", "Yes",  "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
              "No", "Yes", "Yes", "Yes", "No", "No", "Yes", "Yes", "Yes", "Yes", "Yes",
              "Yes", "Yes", "No", "Yes", "No", "Yes", "Yes", "No", "No")
```

A hypothesis testing will be performed to see if there is evidence of homophily. The hypotheses are:

Null Hypothesis: Homophily does not in the Trump social network data with respect to supporters and non-supporters.

Alternate Hypothesis: Homophily exists in the Trump social network data with respect to supporters and non-supporters.

To measure homophily for the attribute of 'supporter' we have to compute the frequency of cross-supporter edges for the sample data.[1] Then we generate a distribution of the null hypothesis and see whether the test statistic is within the p-values to accept or reject the null hypothesis.

This is a one-sided test. The test statistic is number of cross-supporter links (edges) from Trump's social network of friends and followers. The rejection of the Null hypothesis will occur if the p-value is less than 0.05, the standard convention.

In order to generate the null distribution, we use the permutation approach as outlined in the code:

```
relation_trump <- as.data.frame(friends_and_followers)
relation_trump[ , "Positive"] <- positive
relation_trump

g2 <- g
V(g2)$label <-c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
                "Yes", "Yes", "Yes", "Yes",
                "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes",
                "Yes", "No", "No", "Yes", "Yes",
                "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "No", "Yes", "Yes",
                "No", "No")
```

[1] By cross-supporter edges I mean edges which connect people of two different stances of support for Trump (supporters vs non-Supporters).

```
# create adjacency matrix
# no of edges 96

adj_mattrump <- get.adjacency(g2)
mat <- as.matrix(adj_mattrump)
class <- c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
           "Yes", "Yes", "Yes", "Yes",  "Yes", "Yes", "Yes", "Yes", "Yes",
           "Yes", "Yes", "No", "Yes", "Yes", "Yes", "No", "No", "Yes", "Yes",
           "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "No", "Yes", "Yes",
           "No", "No")

colnames(mat) <- class
rownames(mat) <- class

#indices of the cross-sentiment edges.
N_pos <- which(class == "No")
Y_pos <- which(class == "Yes")

mat[N_pos, Y_pos]
```

Firstly, an adjacency matrix of the nodes is made, then we create a matrix which gives us all the cross-link edges. The **sum()** function is used to calculate the number of cross-supporter edges in the network. There are 3 cross link edges. The null hypothesis is then generated:

```
#replicate 1000 times and make a histogram

set.seed(1)
z2 <- replicate(1000, {
  samp.class <- sample(class, 40, replace = FALSE)
  N_samp <- which(samp.class == "No")
  Y_samp <- which(samp.class == "Yes")
  sum(mat[N_samp, Y_samp])
})

hist(z2, breaks =20, col = 'lightblue', title = "Crosslink Distribution")
abline(v=mean(z2), col = 'green')
```

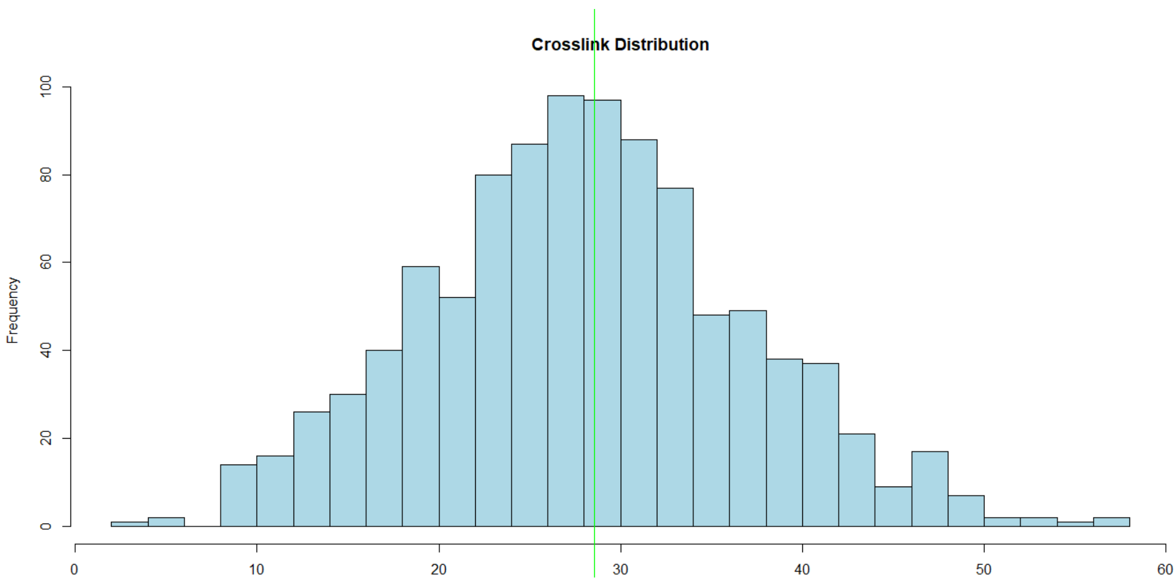

Figure 4: Null hypothesis distribution where there is NO homophily

Now P(H <= 3) = 0.001, therefore we reject the null hypothesis and accept the alternative, that there is homophily since the P value is less than 0.05.

```
# analytic method

e_crosslinks <- 2*(7/40)*(33/40)*96

# 27.72

# HYPOTHESIS TEST alpha = 0.05

mean(z2 <= network_cgl) #0.001
```

From this section of the report the test that we have performed suggests that there is Homophily in the network, meaning that those who are similar in a particular attribute or property (in this case whether they are a supporter or not) tend to form links with one another at a much greater rate than those who they do not share a particular attribute or property with.

The limitations with the test was that we had to use the top 20 followers and friends of Donald J Trump and the top 20 followers were taken from a random sample. It could be the case that the sample was an anomaly and that most other samples have followers of Trump who have negative sentiments towards him and don't support him. Then we might have seen more cross-edge links with supporters of trump with their accounts.

PART VI: Structural Balance

This section of the report will determine whether or not if the signed network is structurally weakly balanced. Firstly, we will give the edges of the network positive or negative signs given their relationship to trump in the earlier sections.

```
mat2 <- mat

mat2[1, 21] <- -1
mat2[1, 25] <- -1
mat2[13, 21] <- -1
mat2[21, 1] <- -1
mat2[25, 1] <- -1
mat2[21, 13] <- -1

matx <- mat2
rownames(matx) <- c(1:40)
colnames(matx) <- c(1:40)
```

The code has given all the edges of the network which contain at least one 'negative' node a negative edge relationship and all the ones containing all positive nodes a positive edge relationship. We can visualize the graph using hierarchical clustering.

```
D <- as.matrix(1-mat2) ## create a matrix of distances
# a distance of 1 if there is no edges, a distance of 2 if they are in another group, and a distance of
# 0 if there are not in the same group.
image(D) ## visualise the distances
## Notice the block structure

I <- as.matrix(mat2+1)
plot(graph_from_adjacency_matrix(I))

# look at components that are connected.
h <- hclust(as.dist(D), method = 'single')
plot(h)
image(D[h$order, h$order])
```

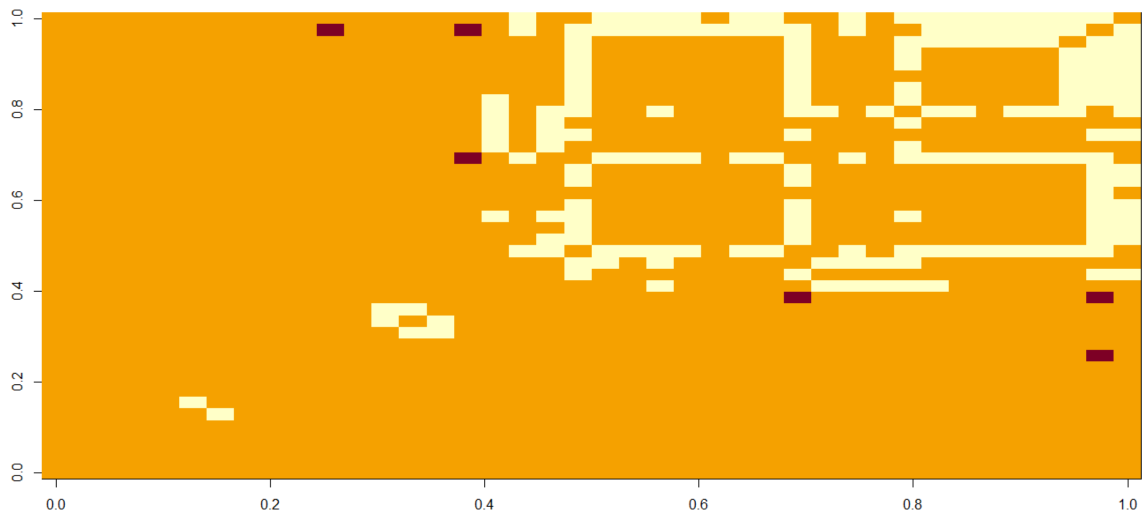

Figure 5: hierachical clustering of edges.

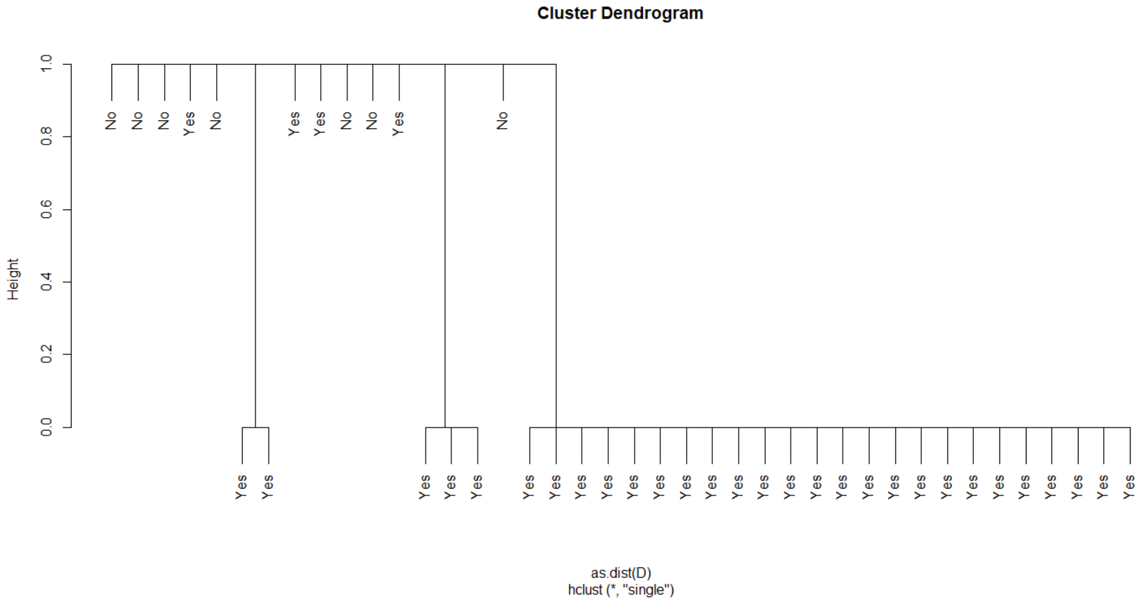

Figure 6: Dendrogram representation of the groups in the network split by hclust().

Unfortunately, since not all nodes are connected to the graph, the graph is not weakly balanced. However, we can see from the figure above how the **hcluster()** function tried to cluster the groups. Therefore, we can conclude that the followers and friends of Trump's social network is not weakly balanced.

The limitations of this analysis section was that not all nodes had connections to other nodes, which effectively ruled out the possibility of structural balance. One way to improve this for future results is to perhaps focus on the giant component of the network structure, and assume for all intents and purposes that it is the whole network structure and do an analysis of the structural balance of the giant/main component. This perhaps could leave to some further insight into the nature of Trump's followers and fans in the network.

CONCLUSION

In conclusion we performed an analysis on Trump's top followers and friends on the Twitter network which were determined via the follower counts of such friends and followers respectively. In the first section, we can see that all of Trump's friends shared similar political leanings and or familial ties. The second section of the analysis the data seems to suggest that the majority of followers of Trump are advocates or have favourable opinions towards the 45th president. The third section of the report showed that there is a main cluster of people who are highly connected and form the main component of the network, while the rest are relatively isolated. In the fourth section the users WhiteHouse, DiamondandSilk, rikomrnk, KellyannePolls and MintzGolf had the lowest neighbourhood overlap scores and thus were considered to be the nodes with the highest social capital. The fifth section suggested that homophily was present in the network structure as we rejected the null hypothesis. And lastly, the network was considered to not be weakly balanced because not all the nodes were connected in the network of Trump’s top friends and followers.

Bibliography

Balboa, P., 2020. *Design Critique: Twitter (Ios App) – IXD@Pratt*. [online] Ixd.prattsi.org. Available at: <http://ixd.prattsi.org/2019/09/design-critique-twitter-ios-app/> [Accessed 30 May 2020].

Developer.twitter.com. 2020. *Overview*. [online] Available at: <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/overview> [Accessed 31 May 2020].

Easley, D. and Kleinberg, J., 2010. *Networks, crowds, and markets* (Vol. 8). Cambridge: Cambridge university press.

Getachew, A. and Beshah, T., 2019, June. The Role of Social Media in Citizen's Political Participation. In *International Working Conference on Transfer and Diffusion of IT* (pp. 487-496). Springer, Cham.

Igraph.org. 2020. *Igraph R Manual Pages*. [online] Available at: <https://igraph.org/r/doc/edge_density.html> [Accessed 31 May 2020].

Jagoda, Naomi (January 31, 2017). Senate panel approves McMahon to lead Small Business Administration . *TheHill*. [Accessed 30 May 2020].

Navarria, G., 2019. The First Twitter President. In *The Networked Citizen* (pp. 91-138). Palgrave Macmillan, Singapore.

Park, L. (2020). Introduction to the Unit - Lecture 1*. Social Media Intelligence (301116)*

Park, L. (2020). Strong and Weak Ties – Lecture 2. *Social Media Intelligence (301116)*

Park, L. (2020). Networks in their surrounding context - Lecture 3. *Social Media Intelligence (301116)*

APPENDIX: R CODE

```
# ASSIGNMENT

key = "Wtn7cEgiT5okpz8qljQ1nbNzJ"
secret = "ariCoPDXMCov0SWacAjifw3zwsVE3tbkfmHkhJ5Dap7VHgSvto"

#API key:
#Wtn7cEgiT5okpz8qljQ1nbNzJ

#API secret key:
#ariCoPDXMCov0SWacAjifw3zwsVE3tbkfmHkhJ5Dap7VHgSvto



#===============#
# AUTHORISING R #
#===============#

library("rtweet")
library("base64enc")
library("httpuv")
library("gridExtra")
library("Matrix")



#Set up OAuth

create_token(
  app = "my_twitter_app",
  consumer_key = key,
  consumer_secret = secret)

# 1. Find 20 friends of trump that have the most followers. Use only Trump's friends,
# not corporate accounts. Examine the twitter and summarize these people.

trump_friends <- get_friends("@realDonaldTrump",
                              n = 5000)
```

```
# The function above gives us the number of friends that trump has (the people he follows).
# (Use twitter definition in report).

trump_friends_info <- lookup_users(trump_friends$user_id)

trump_friends_names <- trump_friends_info$screen_name

trump_friends_names



# the last two functions have given us the screen_names of his friends. This enables us to filter
# out Trump's company twitter handles.

# Get rid of "TrumpGolf", "TrumpDoral", "TrumpCharlotte", "TrumpLasVegas", "TrumpChicago",
# "TrumpGolfDC", "TrumpGolfLA",

# Get indices for these names

company_names <-  c("TrumpGolf", "TrumpDoral", "TrumpCharlotte", "TrumpLasVegas",
                    "TrumpChicago", "TrumpGolfDC", "TrumpGolfLA")

N = length(company_names)
indi <- c()

for (i in 1:N){
  indi[i] <- which(trump_friends_names == company_names[i])
}

trumpfriends_filt <- trump_friends_names[-indi]

# Count how many followers the friends of trump have and count the top 20 friends with the most.

followers_of_tfriends <- lookup_users(trumpfriends_filt)$followers_count

followers_table <- cbind(trumpfriends_filt, followers_of_tfriends)

followers_table <- as.data.frame(followers_table)
```

```
followers_table[ ,1] <- as.character(followers_table[, 1])
followers_table[ ,2] <- as.numeric(as.character(followers_table[, 2]))

str(followers_table)

top_20 <- order(followers_table$followers_of_tfriends, decreasing = TRUE)[1:20]
top_20

# top 20 friends of trump
fot_followers20 <- followers_table[top_20, ] #decided to keep whitehouse in it.

#info of followers

tfriends_info <- lookup_users(fot_followers20$trumpfriends_filt)

tfriends_info$description

followers_des <- cbind(fot_followers20, tfriends_info$description)




# Summary
# The peope that are friends of trump who also have the most followers tend to be political
# commentators, senators, and those active in the american political landscape. Most, if not all
# of his friends with the most followers are either republicans and or advocates for the party and
# family members.

#2. Find the 20 people who follow Trump and have the most followers. Examine if they have a positive
# or negative relationship with Trump based on their tweets. Download a dataset with a large number of
# followers and select the 20 that have the greatest number of followers.

# lookup_users()$followers_count
# get_followers()

followers_trump25k <- get_followers("realDonaldTrump", n = 25000)

tfollowers_followers <- as.data.frame(lookup_users(followers_trump25k$user_id))
```

```
trumpf_f <- tfollowers_followers[ , c("user_id","screen_name", "followers_count")]

top_20fol <- order(trumpf_f$followers_count, decreasing = TRUE)[1:20]

# top 20 followers of trump

top_20list <- trumpf_f[top_20fol, ]

followers_info <- lookup_users(top_20list$user_id)

followers_info$description



# negative or positive relationship

pos_follow <- c("No", "Yes", "Yes", "Yes", "No", "No", "Yes", "Yes",
"Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "No", "Yes", "Yes", "No", "No")

relation_fol_list <- cbind(top_20list[ ,2], pos_follow)

colnames(relation_fol_list) <- c("Followers", "Positive Relationship?")

followers_info[, c("screen_name", "description")]

# 3. Bypassing Trump
#Plot the graph containing Trump's 20 friends and 20 followers. Identify if any of the found friends
#or followers are friends with each other and add these edges to the graph. Then determine if
#any of the friends and followers should be friends, based on their background, and add those edges
#to the graph.

friends_and_followers <- data.frame(fof = c(fot_followers20$trumpfriends_filt, top_20list$screen_name))

friends_and_followers <- as.vector(friends_and_followers)

friends_and_followers <- apply(friends_and_followers, 1, FUN = as.character)



adj_mat <- matrix(data = 0, nrow = 40, ncol = 40)
```

```
adj_df <- as.data.frame(adj_mat)

colnames(adj_df) <- friends_and_followers
rownames(adj_df) <- friends_and_followers

adj_df

followers_pp<-
lookup_users(as.numeric(as.character(get_followers(friends_and_followers[1])$user_id)))$screen_name


# LOOK UP USERS

#lookup_friendships

adj_val <- as.data.frame(matrix(data = 0, nrow=40, ncol=40, byrow = TRUE))

for(i in 1:40){
  for(j in 1:40){
    if(dim(lookup_friendships(friends_and_followers[i], friends_and_followers[j]))[1] < 1){
      adj_val[i ,j] <- "FALSE"
    }else{adj_val[i,j] <- lookup_friendships(friends_and_followers[i], friends_and_followers[j])[[11,4]]}
  }
}

adj_val

adj_valx <- adj_val
for(i in 1:dim(adj_val)[1]){
  for(j in 1:dim(adj_val)[2]){
    if(adj_val[i,j] == "TRUE"){
      adj_valx[i, j] <- 1
    }else{ adj_valx[i, j] <- 0}
  }
}

adj_valx <- as.matrix(adj_valx)
adj_valx <- apply(adj_valx, 2, FUN = as.numeric)
```

```
rownames(adj_valx) <- friends_and_followers

colnames(adj_valx) <- friends_and_followers


adj_valx <- as(adj_valx, "dgCMatrix")


ga <-graph.adjacency(adj_valx, mode = "undirected")


plot(ga)


library("igraph")


g = graph.formula("WhiteHouse" - "VP", "WhiteHouse" - "IvankaTrump", "seanhannity", "WhiteHouse" -
"DonaldJTrumpJr",
        "WhiteHouse" - "Mike_Pence", "WhiteHouse" - "PressSec", "WhiteHouse" - "EricTrump",
        "WhiteHouse" - "IngrahamAngle", "TuckerCarlson", "WhiteHouse"-"KellyannePolls", "BillOReilly",
        "VinceMcMahon", "WhiteHouse" - "JudgeJeanine", "WhiteHouse" - "TeamTrump",
        "WhiteHouse" - "DiamondandSilk", "WhiteHouse" - "Jim_Jordan", "WhiteHouse" - "foxandfriends",
        "WhiteHouse"- "JesseBWatters","WhiteHouse"-"greta", "DonCheadle", "JohnKStahlUSA",
"queensavagedoll",
        "Abdulmalik_fr", "WhiteHouse"- "rikomrnk", "Gideon_Lagat", "thedrsec", "VIP0552525000",
        "WhiteHouse"-"RightWingLawMan", "MintzGolf", "pinoyhotxxx", "Khanpk122", "rebeccabutlerm2",
        "maturetsceline", "maquialifraco", "ceibonacional", "AshleyJohnsonh", "InformedPatriot",
        "KenzieKay69", "alkrit2",


        "VP"-"WhiteHouse", "VP"-"IvankaTrump", "VP"-"DonaldJTrumpJr", "VP"-"Mike_Pence", "VP"-"PressSec",
        "VP"-"EricTrump", "VP"-"IngrahamAngle", "VP"-"KellyannePolls", "VP"-"JudgeJeanine",
        "VP"-"TeamTrump", "VP"-"Jim_Jordan", "VP"-"foxandfriends", "VP"-"JesseBWatters", "VP"-"greta",
        "VP"- "RightWingLawMan",


        "IvankaTrump"-"DonaldJTrumpJr", "IvankaTrump"-"Mike_Pence", "IvankaTrump"-"PressSec",
        "IvankaTrump"-"EricTrump", "IvankaTrump"-"IngrahamAngle",


        #what i added
        "TuckerCarlson"-"foxandfriends", "TuckerCarlson"-"seanhannity", "TuckerCarlson"-"Mike_Pence",
        "TuckerCarlson"-"IngrahamAngle", "TuckerCarlson"-"VP", "TuckerCarlson"-"BillOReilly",
        "TuckerCarlson"-"EricTrump", "TuckerCarlson"-"DonaldJTrumpJr", "TuckerCarlson"-"IvankaTrump",


        "MintzGolf"-"TuckerCarlson", "MintzGolf"-"seanhannity", "MintzGolf"-"BillOReilly",
        "MintzGolf"-"foxandfriends", "MintzGolf"-"IngrahamAngle", "MintzGolf"-"KellyannePolls",
```

```
	"DonCheadle"-"WhiteHouse", "DonCheadle"-"VinceMcMahon",

	"VinceMcMahon"-"WhiteHouse", "VinceMcMahon"-"IvankaTrump", "VinceMcMahon"-
"DonaldJTrumpJr",
	"VinceMcMahon"-"Mike_Pence", "VinceMcMahon"-"PressSec",
	"VinceMcMahon"-"EricTrump", "VinceMcMahon"-"IngrahamAngle", "VinceMcMahon"-"KellyannePolls",
	"VinceMcMahon"-"JudgeJeanine",
	"VinceMcMahon"-"TeamTrump", "VinceMcMahon"-"Jim_Jordan", "VinceMcMahon"-"foxandfriends",
	"VinceMcMahon"-"JesseBWatters", "VinceMcMahon"-"greta",
	"VinceMcMahon"- "RightWingLawMan", "VinceMcMahon"-"TuckerCarlson",

	"JohnKStahlUSA"-"WhiteHouse", "JohnKStahlUSA"-"IvankaTrump", "JohnKStahlUSA"-
"DonaldJTrumpJr",
	"JohnKStahlUSA"-"Mike_Pence", "JohnKStahlUSA"-"PressSec",
	"JohnKStahlUSA"-"EricTrump", "JohnKStahlUSA"-"IngrahamAngle", "JohnKStahlUSA"-
"KellyannePolls",
	"JohnKStahlUSA"-"JudgeJeanine",
	"JohnKStahlUSA"-"TeamTrump", "JohnKStahlUSA"-"Jim_Jordan", "JohnKStahlUSA"-"foxandfriends",
	"JohnKStahlUSA"-"JesseBWatters", "JohnKStahlUSA"-"greta",
	"JohnKStahlUSA"- "RightWingLawMan", "JohnKStahlUSA"-"TuckerCarlson",

	"rebeccabutlerm2"-"InformedPatriot",
	"thedrsec"-"JohnKStahlUSA", "thedrsec"-"TuckerCarlson", "thedrsec"-"BillOReilly",
	"thedrsec"-"seanhannity", "thedrsec"-"foxandfriends", "thedrsec"-"greta",
	"thedrsec"-"KellyannePolls",

	#sport
	"Gideon_Lagat"-"thedrsec",

	#strippers/porn/spam
	"queensavagedoll"-"pinoyhotxxx", "queensavagedoll"-"AshleyJohnsonh",
	"AshleyJohnsonh"-"pinoyhotxxx")

V(g)$label.cex = 0.7

plot(g, layout = layout.fruchterman.reingold, vertex.size = 8)
```

WHO WOULD BE FRIENDS?

4. Graph Statistics

#Compute the diameter and density of the graph, and neighbourhood overlap of each edge and determine which

#nodes have the greatest social capital. State if the results are obvious from the graph structure and why.

DIAMETER

It is the shortest distance between the two most distant nodes in the network.

```
dia.g <- diameter(g)
```

Density

Density is the number of actual edges as a ratio of possible maximal edges.

```
den.g <- graph.density(g)

#overlap

#neighbors of both A and B/neighbors of A or B.

#use ends function to get vertices of edge
# intersect and neighbors

n_pairs <- ends(g, E(g))
n_pairs <- as.data.frame(n_pairs)

neighborhood_overlap <- function(x){
  node_pairs <- ends(x, E(x))
  numerator <- c()
  denom <- c()
  no <- c()
  for(i in 1:length(E(x))){
    numerator[i] <- length(intersect(neighbors(x, v = node_pairs[i,1]),
                                     neighbors(x, v = node_pairs[i, 2])))
    denom[i] <- length(union(neighbors(x, v=node_pairs[i, 1]),
                             neighbors(x, v=node_pairs[i,2])))-2
    no[i] <- numerator[i]/denom[i]
```

```
  }
  return(no)
}

n_overlap <- neighborhood_overlap(g)
n_pairs[ , 3] <- n_overlap
no_edges <- n_pairs


# Three Local bridges

ind_lb <- which(no_edges[, 3] == min(no_edges[, 3], na.rm = TRUE))
no_edges[ind_lb, ]

# Whitehouse - DiamondandSilk
# Whitehouse - rikomrnk
# KellyannePolls - MintzGolf

#Structural Holes
holes <- constraint(g, nodes = V(g), weights = NULL)

#5 HOMOPHILY

positive <- c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
              "Yes", "Yes", "Yes",  "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
              "No", "Yes", "Yes", "Yes", "No", "No", "Yes", "Yes", "Yes", "Yes", "Yes",
              "Yes", "Yes", "No", "Yes", "No", "Yes", "Yes", "No", "No")

relation_trump <- as.data.frame(friends_and_followers)
relation_trump[ , "Positive"] <- positive
relation_trump

g2 <- g
V(g2)$label <-c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
                "Yes", "Yes", "Yes", "Yes",
                "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes",
                "Yes", "No", "No", "Yes", "Yes",
                "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "No", "Yes", "Yes",
```

```
                "No", "No")

# create adjacency matrix
# no of edges 96

adj_mattrump <- get.adjacency(g2)
mat <- as.matrix(adj_mattrump)
class <- c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
           "Yes", "Yes", "Yes", "Yes",  "Yes", "Yes", "Yes", "Yes", "Yes",
           "Yes", "Yes", "No", "Yes", "Yes", "Yes", "No", "No", "Yes", "Yes",
           "Yes", "Yes", "Yes", "Yes", "Yes", "No", "Yes", "No", "Yes", "Yes",
           "No", "No")

colnames(mat) <- class
rownames(mat) <- class

#indices of the cross-sentiment edges.
N_pos <- which(class == "No")
Y_pos <- which(class == "Yes")

mat[N_pos, Y_pos]

#Total cross sentiment edges.
network_cgl <- sum(mat[N_pos, Y_pos])

#replicate 1000 times and make a histogram

set.seed(1)
z2 <- replicate(1000, {
  samp.class <- sample(class, 40, replace = FALSE)
  N_samp <- which(samp.class == "No")
  Y_samp <- which(samp.class == "Yes")
  sum(mat[N_samp, Y_samp])
})

hist(z2, breaks =20, col = 'lightblue', main = "Crosslink Distribution")
abline(v=mean(z2), col = 'green')
```

```
# The mean of 1000 experiments.

m.sim <- mean(z2)
sd.sim <- sd(z2)

m.sim
sd.sim

# analytic method

e_crosslinks <- 2*(7/40)*(33/40)*96

# 27.72

# HYPOTHESIS TEST alpha = 0.05

mean(z2 <= network_cgl) #0.001

# Reject the null hypothesis, accept the alternative.

#6. Structural Balance
#Finally, determine if the signed network is weakly balanced (using hierarchical clustering) and identify
#if any within or between signed relationships are not as expected. To perform this analysis, first label
#all existing edges as either positive or negative, based on their association to Trump.

#1, 21
#1, 25
#13, 21
#21, 1
#25, 1
#21, 13

mat2 <- mat

mat2[1, 21] <- -1
mat2[1, 25] <- -1
mat2[13, 21] <- -1
mat2[21, 1] <- -1
```

```
mat2[25, 1] <- -1
mat2[21, 13] <- -1

matx <- mat2
rownames(matx) <- c(1:40)
colnames(matx) <- c(1:40)

D <- as.matrix(1-mat2) ## create a matrix of distances
# a distance of 1 if there is no edges, a distance of 2 if they are in another
#group, and a distance of
# 0 if there are not in the same group.
image(D) ## visualise the distances
## Notice the block structure

I <- as.matrix(mat2+1)
plot(graph_from_adjacency_matrix(I))

# look at components that are connected.
h <- hclust(as.dist(D), method = 'single')
plot(h)
image(D[h$order, h$order])
```