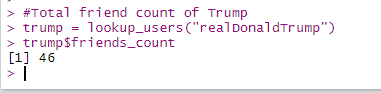
**Friends of Trump**

**Aim: - To examine the impact of Donald Trump’s recent statements on his supporters and examine the relationships of his Twitter friends and followers.**

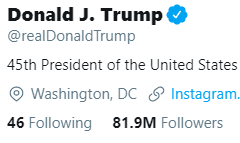
1. **Friends of Trump**

After making an authenticated session to twitter data API, let’s first find out the total no. of friends of trump using the rtweet library.



*Figure 1.1 Total no. of friends of trump.*

“Friends” is a twitter term used to reference the accounts followed by an account (displayed as “following” in the twitter UI:

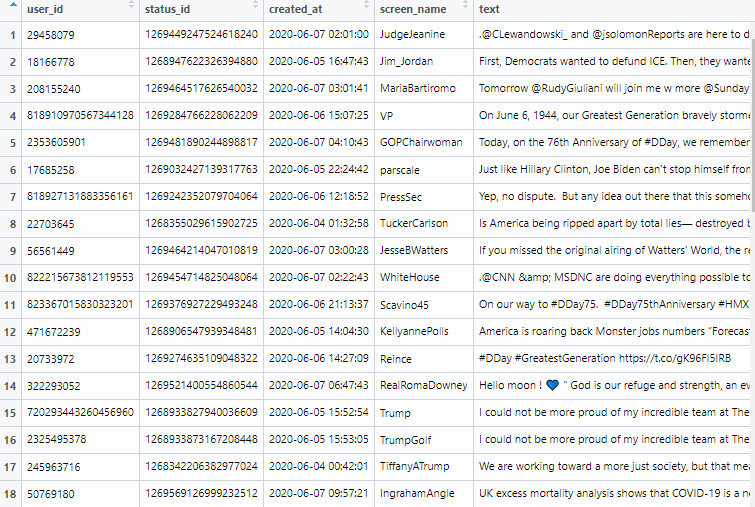


*Figure 1.2 Following count in twitter UI*

To examine the actual data of these users we have used the “get\_friends” function to get the user\_ids and “lookup\_users’ function to get more information about these accounts.



*Figure 1.3 collecting friends of Trump*



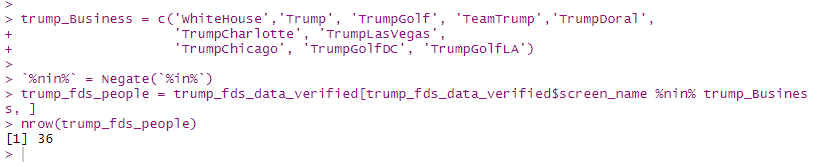
*Figure 1.4 subset of friend’s data collected*

As we can see we have both Trump’s company accounts (example @TrumpGolf at row 16) and other people accounts in Trump’s friends list. For proper analysis we have excluded the Trump’s company account to find the Top 20 friends.

First, we filtered out the accounts which are verified for a better analysis

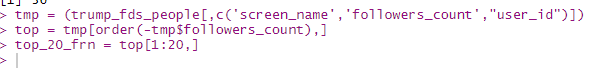


Turns out all the friends of Trump are verified. As a step two let’s exclude Trump’s business accounts.

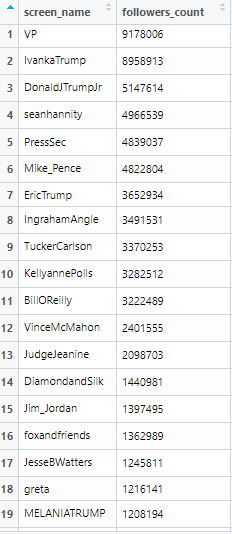


We excluded 10 business accounts and are left with 36 people accounts for analysis.

Next step we take out important properties screen\_name, user\_id and followers\_count from this data and arrange it in descending order of no. of followers and taking the top 20 from that list and storing it separately.



Hence, here are the top 20 friends of trump that have the most followers excluding Trump’s business handles.



*Figure 1.5 Top 20 friends of Trump*

By examining the “text” property of this list “trump\_fds\_people” we can the description of these accounts. On further examination we see that the top 20 list mostly contains portions of:

1. **Family of Trump** (Example: Ivanka Trump, Donald J Trump Jr, Eric Trump, Melania Trump)
2. **Republicans** (Example: Vince McMahon)
3. **People working in media and journalism industry** (Example: Sean Hannity, Kayleigh McEnany, Ingraham Angle, Tucker Carlson, Bill Reilly, Jeanine Pirro, Jesse Watters, Eric Bolling)
4. **Other government officials** (Example: VicePresident Mike Pence**,** Jim Jordan)

Let’s further analyze their relationship with Trump.

1.First we extract their recent 200 tweets using get\_timeline function

2.We find if “realDonaldTrump” is mentioned in each tweet

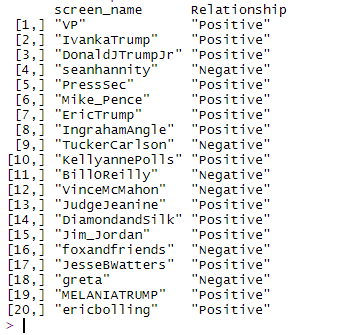
3. If he is mentioned we use **sentimentr** library for the sentimental analysis of the tweet. “**sentimentr”** is a library which gives us a score of average sentiment of a text.

4. If the average sentiment is “Positive” (more than zero) **or** the person has mentioned the word “Republican” in his tweet we consider it as a positive relationship

5. Otherwise if person has mentioned ‘Democrat” in his tweets **or** the average sentiment is “Negative” (less than zero) we consider it as a negative relationship.



Storing the results in a matrix we have: -



*Figure 1.6 Top 20 friend’s relation with Trump*

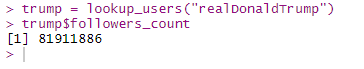
**Conclusion:** we can say that a **major portion of Trump friends contains accounts of people in Media and Journalism industry** some of which are Republicans. That concludes Trump’s twitter has a lot of connections with accounts of people in media and journalism.

**Suggestions for further analysis:** To examine the text analysis of tweets of top 20 friends beyond 200 tweets to conclude if they are in a supporter of Republican party of USA. Also we can analyze tweets containing #Trump and perform similar sentimental analysis.

**Limitations**: We have not considered refining other business accounts and accounts of organizations/groups which could add more value to this data. Also, we have considered the people who have not mentioned Trump in their tweets as a negative relationship while it should be considered neutral.

1. **Followers of Trump**

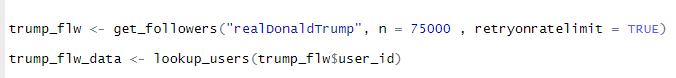
Let’s look at the no. of followers of Trump



*Figure 2.1 Followers count of Trump*

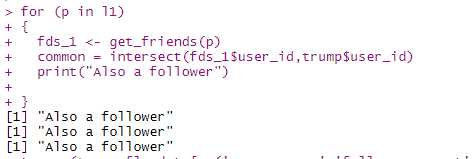
That turns out to be 8.91 million currently, but twitter data is very dynamic hence, the exact no. keeps on changing as more and more people keep following Trump. According to twitter documentation results are ordered with the most recent following first (in groups of 5,000 user IDs).

Starting with a subset of 75000 followers, we can extract user\_id from “get\_followers” function



Now that we have data of most recent 75000 followers, Let’s check first 3 friends of trump and check if any of them are his followers as well. For this we check the friends of “top 3 friends of trump”.

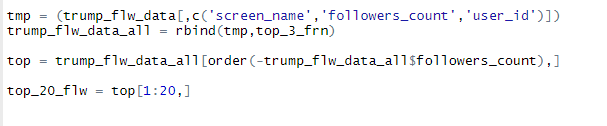




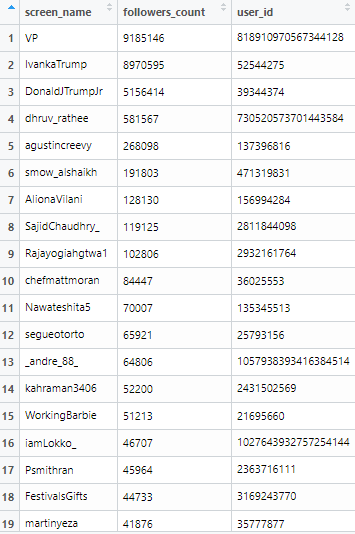
We take the user\_id for top 3 friends of Trump, extract their friends using get\_friends, finally we check the intersection between Top 3 friends and Trump’s user\_id to check if they follow Trump.

Turns out all 3 of them follow Trump as well. Hence, we add these 3 to trump\_flw\_data.

We then sort these in descending order of their no. of followers to find the top 20 followers of trump with maximum followers.



We finally have the Top\_20 followers of Trump:





*Figure 2.2 Top 20 followers of Trump*

By examining the twitter accounts of these Top 20 we can say that they have portions of:-

1. **Children of Trump**: (IvankaTrump, DonaldJTrumpJr)
2. **Politics Bloggers / Influencers** :(dhruv\_rathee, SajidChaudhry\_, Rajayogiahgtwa1, Nawateshita5, WorkingBarbie)
3. **People in Sports**: (agustincreevy)
4. **People in Entertainment industry** (AlionaVilani, Psmithran)
5. **Supporters of Trump** (thatNAchick)

Hence, we can say that a large portion of Trumps top 20 followers are politics bloggers/ Influencers.

We can find their relationship with Trump by analyzing their tweets.

1.First we extract their recent 200 tweets using get\_timeline function

2.We find if “realDonaldTrump” is mentioned in each tweet

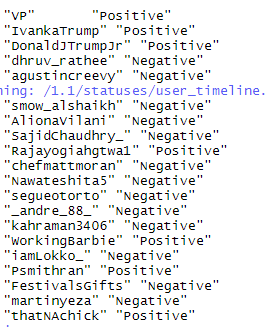
3. If he is mentioned we use **sentimentr** library for the sentimental analysis for the tweet. “**sentimentr”** is a library which gives us a score of average sentiment of a text.

4. If the average sentiment is “Positive” (more than zero) **or** the person has mentioned the word “Republican” in his tweet we consider it as a positive relationship

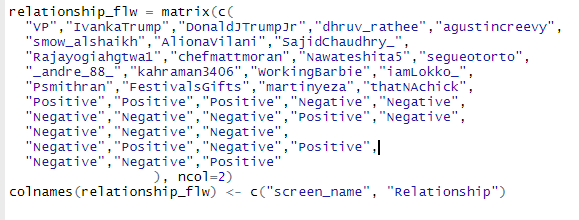
5. Otherwise if person has mentioned ‘Democrat” in his tweets **or** the average sentiment is “Negative” (less than zero) we consider it as a negative relationship.



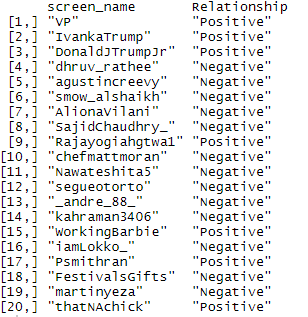
Example of the output:



We store these results in a matrix



Final result:



*Figure 2.2 Relationship of top 20 followers with Trump*

**Conclusion:** A major portion of Trump’s top followers are politics influencers and they have mixed relationship with Trump.

**Limitations:** these results are based on only recent 75000 followers for better analysis we should analyze all 8.9 million followers of them then we can have more details on what kind of most followed people follow Trump and their relationship with Trump.

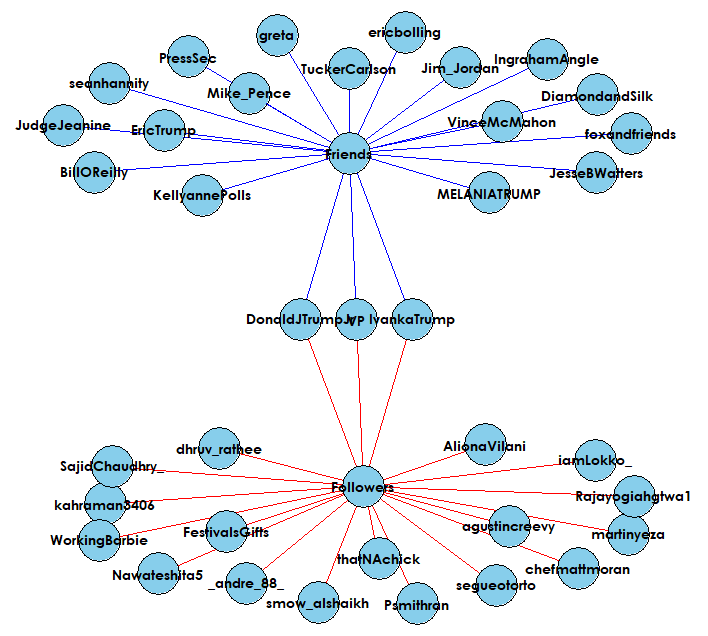
Also, we should only consider personal accounts not business accounts

**Suggestions for further analysis:** A better analysis to find relationship with Trump can be filtering out tweets of people where they mentioned Trump and **performing sentimental analysis for each of these tweets** and then counting a score for each sentiment and **averaging it out**. Hence, we can get a score how much pro and anti a follower posted against Trump and based on that we can conclude if he/she have a positive or negative relationship with Trump.

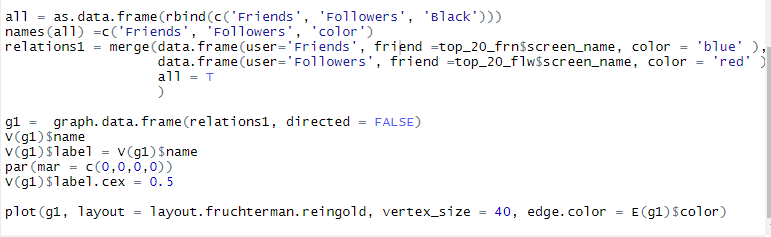
Combing results from previous step we can also find if the users are pro Republican but against Trump by finding the sentiment with which they mention Republican in their tweets.

1. **Bypassing Trump**

A graph between top 20 followers and friends of Trump:



*Figure 3.1 Graph between top followers and friends of Trump*

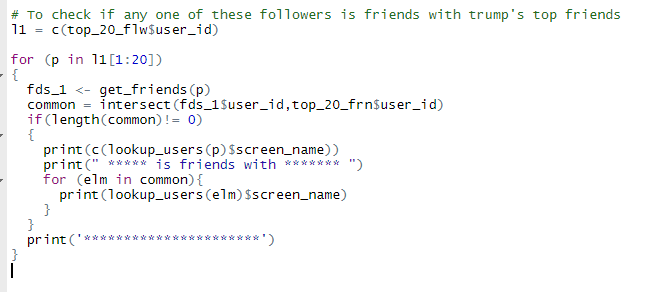


We add a relation between the two, denoting followers with red and friends with blue. We have found 3 common nodes between the two groups

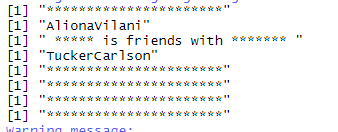


To find out if any of the found friends or followers are friends with each other we loop through top followers of trump and check if they follow any top friend of Trump if so the screen\_id is stored

We use get friends function to get the friends of these top followers and check if they follow any top friend.

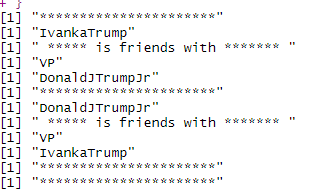


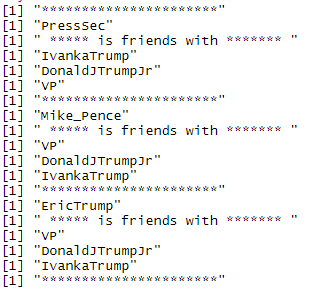
A subset of result:



We repeat the same for friends to check if they follow anyone from the top followers list and the result is stored.

A subset of result:



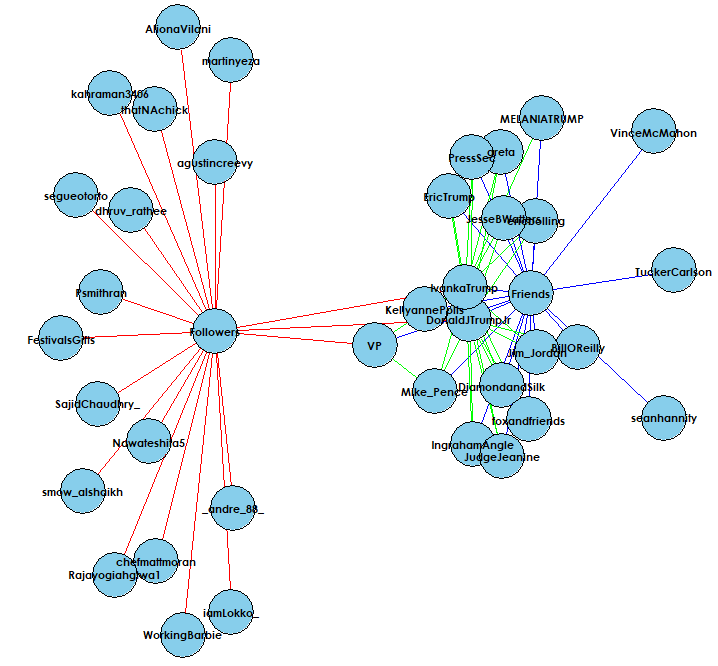


We then add these two results in a data frame and extract the accounts that follow each other **hence are friends with each other.**

Adding these edges to the graph and denoting in green color:



*Figure 3.2 Adding edges of top friends are followers of Trump which are friends which each other*

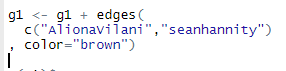


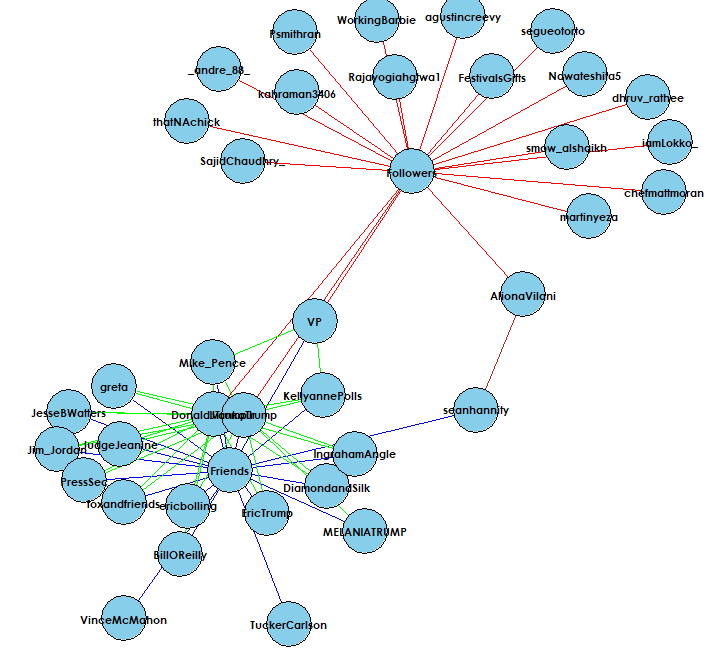
*Figure 3.3 Graph with connected edges*

We can see that " IvankaTrump " and "DonaldJTrumpJr" are very well connected with top friends of Trump.

Also, from top 20 friends and followers list we see that both AlionaVilani and Sean Hannity work in Television Entertainment Industry. Hence, they should be friends with each other.

Adding this edge to the graph





*Figure 3.6 Graph with all connected edges*

We can observe that VP, AlionaVilani, Sean Hannity, DonaldJTrumpJr, IvankaTrump are some nodes one of which will have the first access if there is information flow between the two clusters in the graph and the bridge between AlionaVilani and Sean Hannity is a local bridge.

**Conclusion**: The top friend’s network is very connected with each other.

DonaldJTrumpJr and IvankaTrump are most connected in the top friend’s circle.

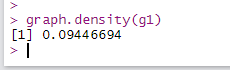
**Limitations**: As this is based on data from only 75000 followers the graph seems to be less connected, whereas in social media connections the networks are very connected. Also, as this is an undirected graph and twitter has a concept of “followers” we are only considering people who are friends of each other.

**Suggestions for further analysis:** For twitter data we should consider a directed graph as the main concept of twitter is “followers” to truly find how connected the network is.

**4.Graph Statistics**

The **density** of a graph is the ratio of the number of edges and the number of possible edges.

The densest graph will have density of 1 having all possible edges connected. Hence, it’s a measure of how well connected the graph is. The density of our graph:

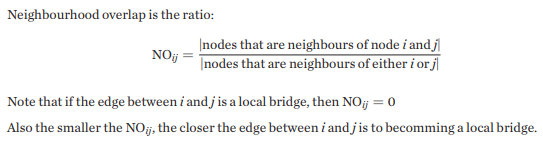


Which is not much hence our graph being not very well connected.

The **diameter** is the longest shortest path i.e. “greatest distance between any pair of vertices”

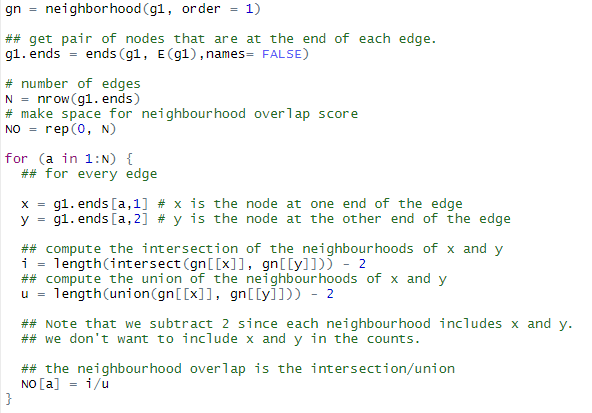


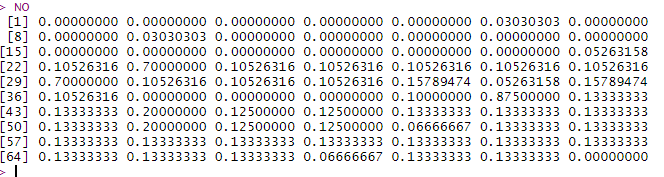
**Neighborhood overlap** lets us know how close an edge is to become a local bridge. The smaller the neighborhood overlap the closer the edge is to become a local bridge.



*Figure 4.1 Neighborhood overlap formula*

We compute the neighborhood overlap using the above formula for each edge:





*Figure 4.2 Neighborhood overlap*

We can see NO is 0 for many edges hence, they are local bridges. It shows our graph is not very connected.

By comparing NO and E(g1) we have the highest NO for edges 41,23,29 i.e. edges:

DonaldJTrumpJr—IvankaTrump (0.875)

Friends—DonaldJTrumpJr (0.70)

Friends—IvankaTrump (0.70)

Nodes with high neighborhood overlap have lower social capital. Social capital defines the importance of a presence of a node in graph in relation to connectivity.

AlionaVilani—seanhannity have lowest neighborhood overlap between these users.

Hence AlionaVilani and seanhannity will have highest social capital. Without these two nodes connected the graph have the tendency to be split into two components. They also act as gatekeeps and have early access to information

Yes, the results were obvious from graph structure we can see that our graph is not very connected i.e. why our density was low, and diameter was high. And we can see that IvankaTrump and DonaldJTrumpJr are very well connected with the “friends” group i.e. why they have high Neightbourhood overlap and AlionaVilani—seanhannity forms a local bridge hence, their social capital is high.

**Conclusion**: Our graph is not very much connected and AlionaVilani—seanhannity form a local bridge. IvankaTrump and DonaldJTrumpJr are most connected nodes.

**Limitations**: We only concluded a few parameters of graph for further analysis more measures need to be computed.

**Suggestions for further analysis:** We can also compute other parameters of graph such as degree, betweenness, closeness, degree\_distribution etc. for more insight

**5. Graph Homophily**

As we have already computed the relationship of Trump’s followers with Trump and Trump’s Friends with Trump, we label each follower having a positive relation as a supporter and Negative as a non-supporter.

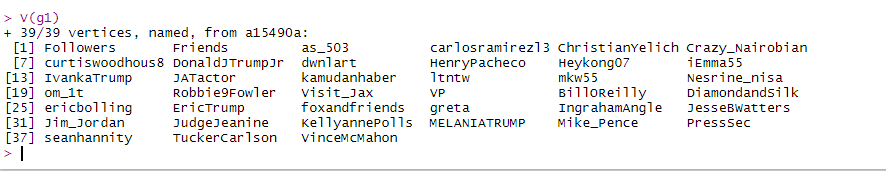
Let’s perform a hypothesis test.

H0 (Null hypothesis): homophily does not exist for the given attribute (supporter or non-supporter of Trump) in the network.

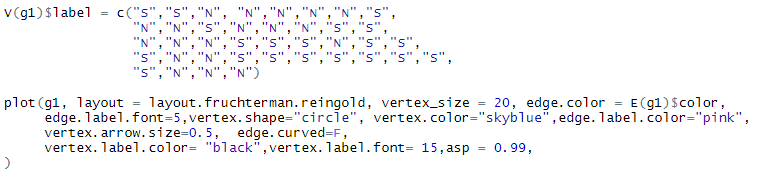
HA (Alternate hypothesis): homophily does exist for the given attribute (supporter or non-supporter of Trump) in the network.

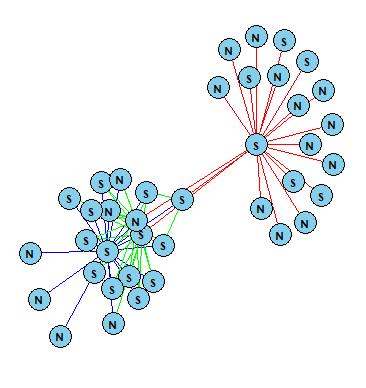
As for all hypothesis tests, we assume the Null hypothesis, then examine if our test statistic is likely to be a sample from the Null hypothesis distribution.

Lets first plot a graph replacing friend and follower names with whether they are a supporter or non-Supporter, we take S as a supported and N as a non-supporter and change the labels of the graph with information collected in part 2 and above.



Changing labels to supporter or non-supporter according to positive or negative relationship with Trump.

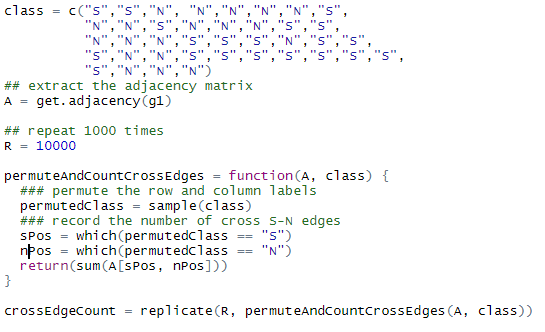




*Figure 5.1 Plot considering each node as either supporter or non-supporter of Trump*

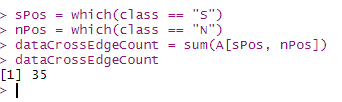
Step 1: Keeping the same nodes and edges of the graph, lets shuffle the labels randomly and calculate count of cross edges.

Step 2: Repeat it 1000 times to compute average



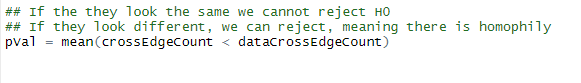
Now we have 1000 observational count of cross edges after shuffling the labels randomly using sample function and repeating it using replicate function.

Step 3: Compute no. of cross edges in actual graph



That comes out to be 35

Step 4: Lets compute p-value by calculating how many times the random distribution cross edges count was less than the actual cross edge count.



Step 5: p-value comes out to be 33%



As we are taking a significance level of 0.05, we don’t reject the null hypothesis.

**Hence, we state that there is homophily in the graph**.

Hence there is some relation between the cross edges as when they were randomly distributed the results were quite different from our actual graph.

**Conclusion**: There exists homophily between supports and non-supports of Trump which is what is excepted as they are somewhat link back to Trump thus have some common characteristic which is common in social networks.

**Suggestions for further analysis**: We can repeat the experiment more than 1000 times to receive better average results. We can also filter out tweets that contains #Trump instead of mentioning Trump directly and observe them to further analyze if a person is supporter or not.

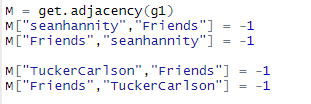
**Limitations**: We have only analyzed 200 tweets per individual. We can further analyze ideas of people and check if they match with Trumps for individuals who don’t mention Trump at all. Also, it’s very hard to say if a person if a supporter of Trump only based on some tweets, we need to analyze a good no. of tweets mentioning Trump to conclude if he/she is a supporter or not.

**6. Structural Balance**

First, we label each edge of same graph as positive or negative based on their relationship with Trump we found in part 1 and 2.

First, we label Top 20 followers and Top 20 friends relationship with Trump. We start by creating an Adjacency matrix and marking negative relationships using relations\_frn and relations\_flw computed in step 1 and 2 as “-1”.

Example: -



Lets further mark relationship of these top accounts which are connected to each other. As they are friends of each other and also have a common connection i.e. Trump and by analyzing their profiles we find that they have similar political views as most of them help Trump with his campaign also most of them follow accounts with support Trump’s cause, we mark them as a positive relationship (“1”) except edges ("greta","DonaldJTrumpJr")

("IvankaTrump","greta")

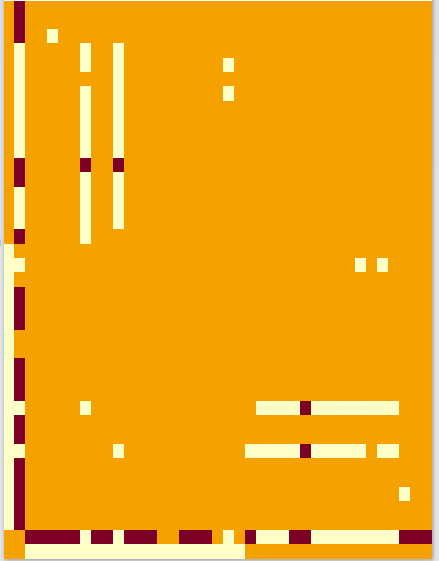
Which do not post similar posts and do not follow similar kind of accounts. So, we mark these edges as negative (“-1”).

Further, as ("AlionaVilani","seanhannity") both work in Television industry they seem to have a positive relationship so let’s keep the edge as “1”.

Then, lets create a matrix of distances and visualize it.



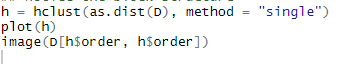
Result: -



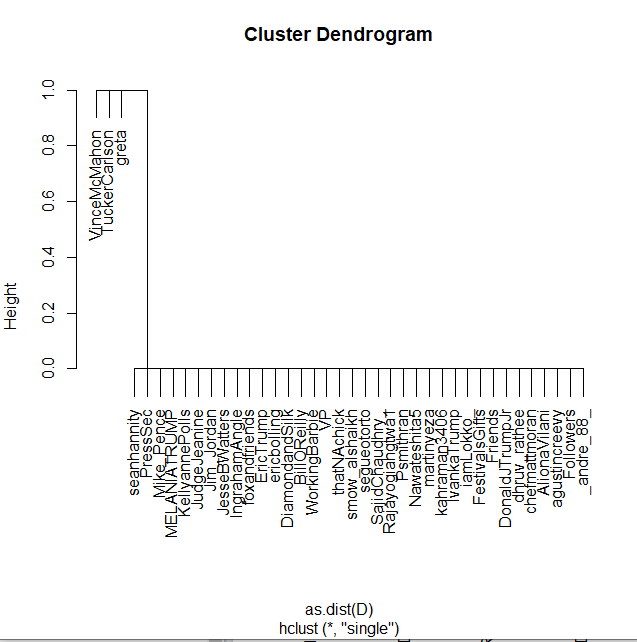
*Figure 6.1 Image of Adjacency matrix*.

Here, yellow represents that there is no edge or no relationship, red represents -1 or negative relationship and white represents +1 or positive relationship.

Let’s now use hierarchical clustering to find if the signed network is weakly balanced, for a weakly balanced graph we should expect the nodes to split into groups or clusters.

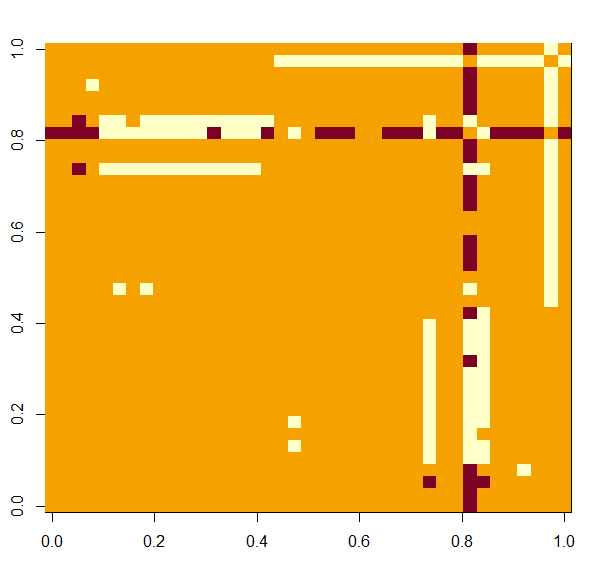


Result: -



*Figure 6.2 Plot of the cluster*

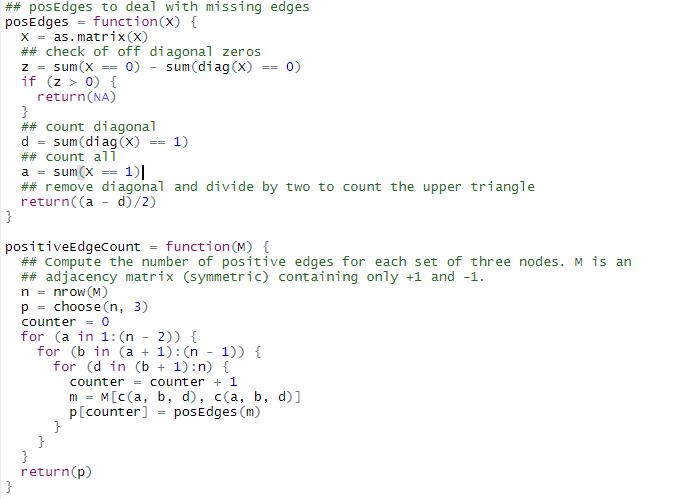
As we can see the nodes cannot be divided into separate groups where in each group have positive relationship with all nodes with same group while have negative relationship with all nodes outside the group.

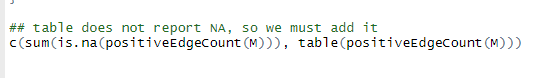


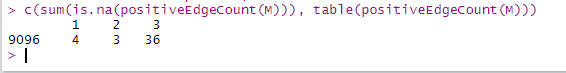
*Figure 6.2 Image of the cluster*

**We conclude that the signed network is not weakly balanced or Structurally Balanced.**

Further we compute the number of positive relationships in each triangle in the provided adjacency matrix.







Here we can see that we have 3 triangles with (++-) relationship, because of which our graph is not balanced. Also, we have 36 (+++) triangles and 4(+--) triangles. There is not (---) relationship/triangle in our graph.

These results are expected as its hard to find networks which is Structurally balanced also these correspond to few accounts taken into analysis. These 3 (++-) are expected to evolve over time and might become a (+++) relationship.

**Limitations**: - But as twitter data is very dynamic as this is analysis of only a snapshot of relationships in a time interval, we may find these relationships changing very often.

**Suggestions for further analysis:** We have considered only a few accounts to find if the network is Structurally balanced. If we perform the analysis on graph with more no. of nodes we might have better analysis as in a social media network nodes are generally very connected overall.

**References**

1. Lecture slides Social Media Intelligence
2. <https://twitter.com/realDonaldTrump> (Trump’s Twitter UI)
3. <https://rtweet.info/> (Retweet documentation)
4. <https://cran.r-project.org/web/packages/rtweet/vignettes/intro.html>
5. <https://towardsdatascience.com/sentiment-analysis-in-r-good-vs-not-good-handling-negations-2404ec9ff2ae>
6. <https://www.rdocumentation.org/packages/sentimentr/versions/2.7.1> (sentimentr documentation)