Drashti Patel

# *Social Media Analysis-*

# *United States Presidential Election 2016*

## 1. Project Introduction

### 1.1. Objective

There is no doubt that not only the Americans, but also the whole world wants to know who is the frontrunner in the election. The people also want to know if the ones they support are leading or not. Today, with the enormous amount of data on Social Media Networks I can certainly get to know which candidate is popular.

Therefore, my goal is to collect data from Twitter and get to know which candidate is trending so far. Moreover, which policies of theirs attract the people of America. This will also help, the presidential campaigners to strategize their campaigns.

### 1.2. **Social Networks**

Before choosing Twitter as my data source, I found out how effective could it be for to achieve my objective. Since the election buzz, there are an average of 16 million tweets about the presidential election. According to Miguel Rios, who is a Data Science Manager at Twitter, prior to Sanders primary win in Indiana, he was trending more on Twitter than Clinton. (see Figure 1)

Also with the time available, and with the Twitter data easily collectable, I decided to go forward with Twitter as my data source.

## 2. **Data Collection**

We used the Twitter API to collect data about the two major Presidential candidates: Donald Trump, Hillary Clinton. I used their names as the search word. The file DataCollection.R is used to collect the tweets. I were able to collect over 30000 tweets over a two-day period. The data was collected starting on June 26th, Tuesday morning and ran until June 27th Wednesday midnight. It was started again June 29th, Thursday morning and then again on July 1st, Friday morning.

### 2.1. **Parsing Data**

We first retrieved the data in the Twitter list format. To analyze the data in future I converted the data in CSV format. Then again based on my requirement, part of the data was converted into Twitter list, data frame, character etc.

The Original format:

{"Text";" favorited";" favoritedCount";" replytoSN";" created";" truncated";" replytoSID";"id";

” statusSource”;” screenname”;” retweetCount”;” isRetweet”;” retweeted”;” longitude”;” latitude”}

### 2.2. **Grabbing Data from Twitter**

We tried to augment my data by grabbing specific information about the users from the tweets I collected by mining data straight from the Twitter site. The only way to mine is using Twitter’s API. I create my own Twitter API: R_Mining_exp1 to get authorized access to grab the tweets.

## 3. **Data Analysis**

### 3.1. **Follower & Friend Network**

Follower count and Friend count are two important criteria to determine the influence of one user. I have used NodeXL to build a graph to show the comparison of Hillary Clinton’s followers vs. Donald Trump’s followers. (See Figure 2 to 10)

### 3.2. **Retweet per Day**

After analysis of the retweet data, I got a graph which indicates the Retweet count based on timeline. (See Figure 11 and 12). The code for retweet per day analysis is **Retweets per day.R .**

### 3.3. **Retweet per Tweet Count**

For retweet per tweet count analysis, I build a graph based on which twitter members have retweeted the tweets related to Hillary Clinton or Donald Trump. (See Figure 13 and 14) The code for retweet per count is **Retweets per count Trump And Hillary.R**

### 3.4. **Sentiment Analysis**

We have plotted the graph using Jeffrey Been’s algorithm to identify texts as either positive, neutral and negative words. (See Figure 15 to 17) The codes used for this analysis were **SentimentalAnalysisPart1.R, SentimentalAnalysisPart2.R and sentiment_new.R.**

### 3.5. **Word Cloud**

We did text mining with the tweets I extracted. I used a word cloud and tm package in **wordcloud.R** file(for mining) to display which words are mentioned by users frequently. In this way, I can know the feelings of users directly and obviously. The word 'Disaster' is the most frequent word for US Election analysis, the word 'Hillary Clinton' for Hillary Clinton analysis and the word 'Donald Trump' for Donald Trump analysis. (See Figure 18 to 20 ) .

### 3.6. **Country and Location**

We have used NodeXL software to retrieve the location, followers related data. For that, I used From Twitter Searched Network function to analyze this data. (See Figure 21)

## 4. **Summary**

We were able to get valuable information from Twitter about the ongoing US election, and this information can help us understand the popularity of the presidential candidates among people.

In this project, I worked as a team to collect, clean, visualize and analyze the data. I used tools like R, Gephi, NodeXL, and Excel for the previous mentioned processes. The most important thing that I found out was that nowadays Social Network Analysis can be crucial to understand people's mindset and this can be useful to predict the future trends.

## 5. Appendix

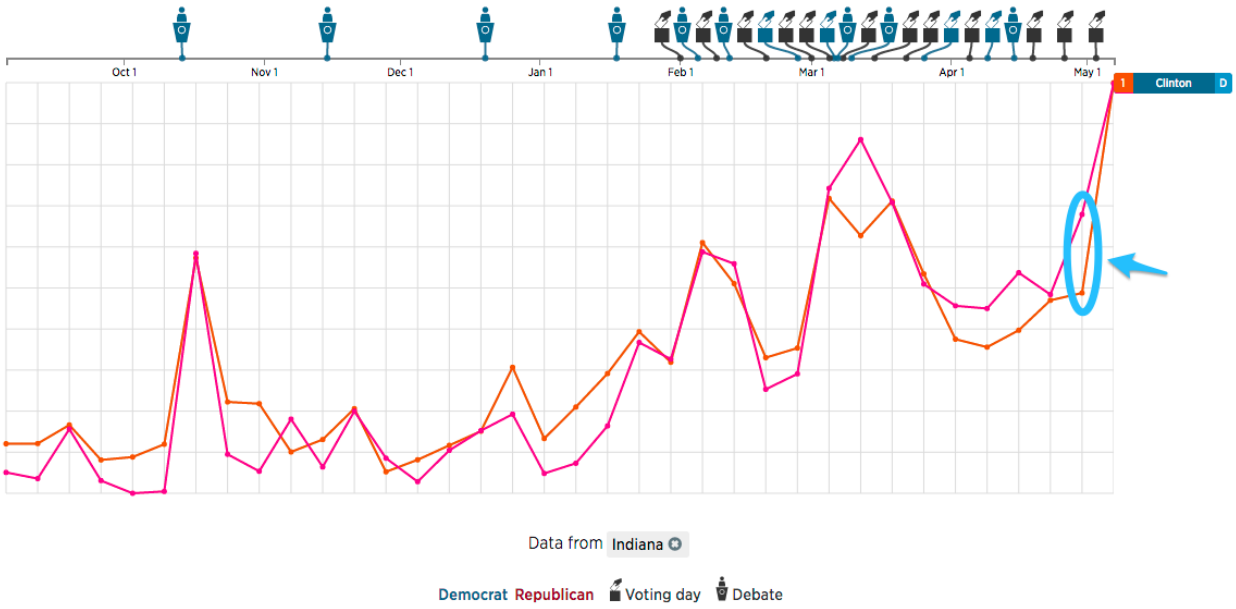


Figure 1: Sanders Vs. Clinton Twitter Analysis

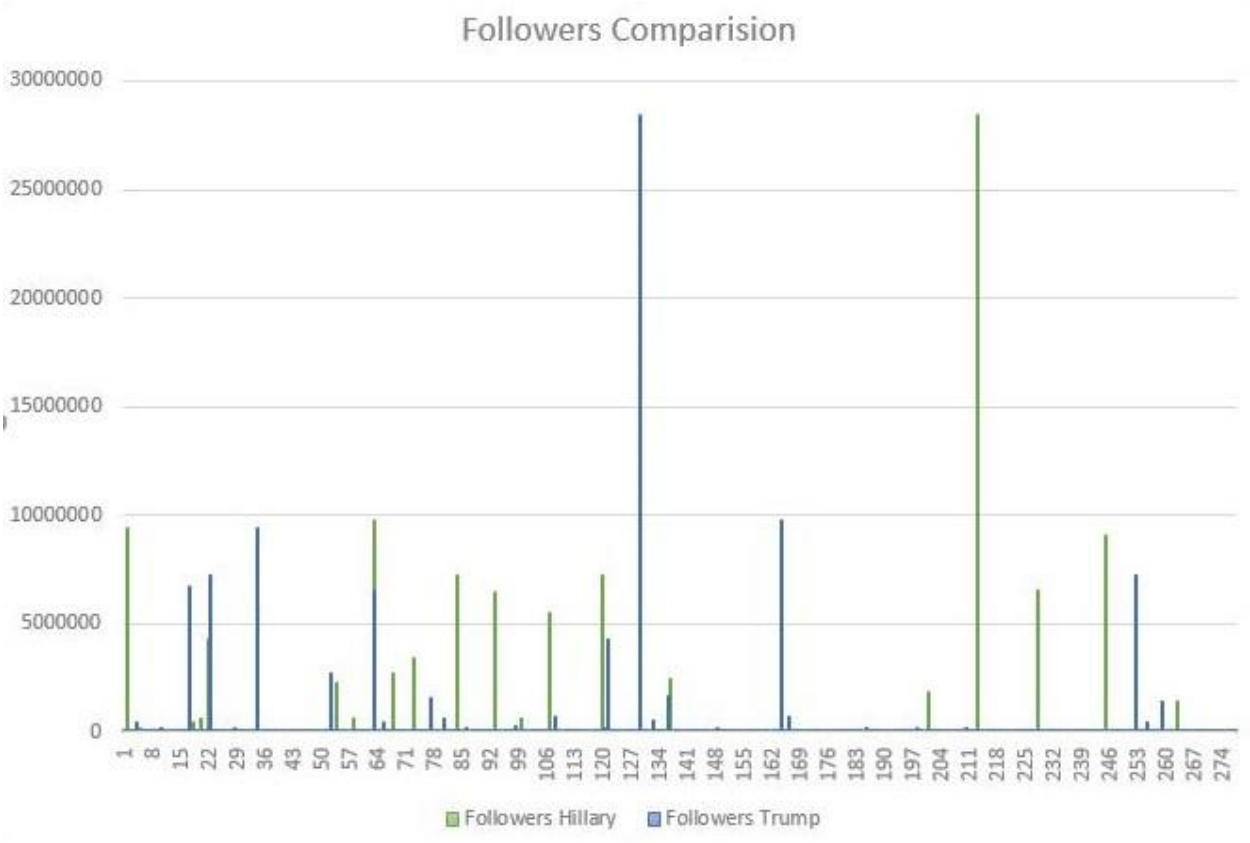


Figure 2: Comparison of Hillary Clinton and Donald Trump's followers

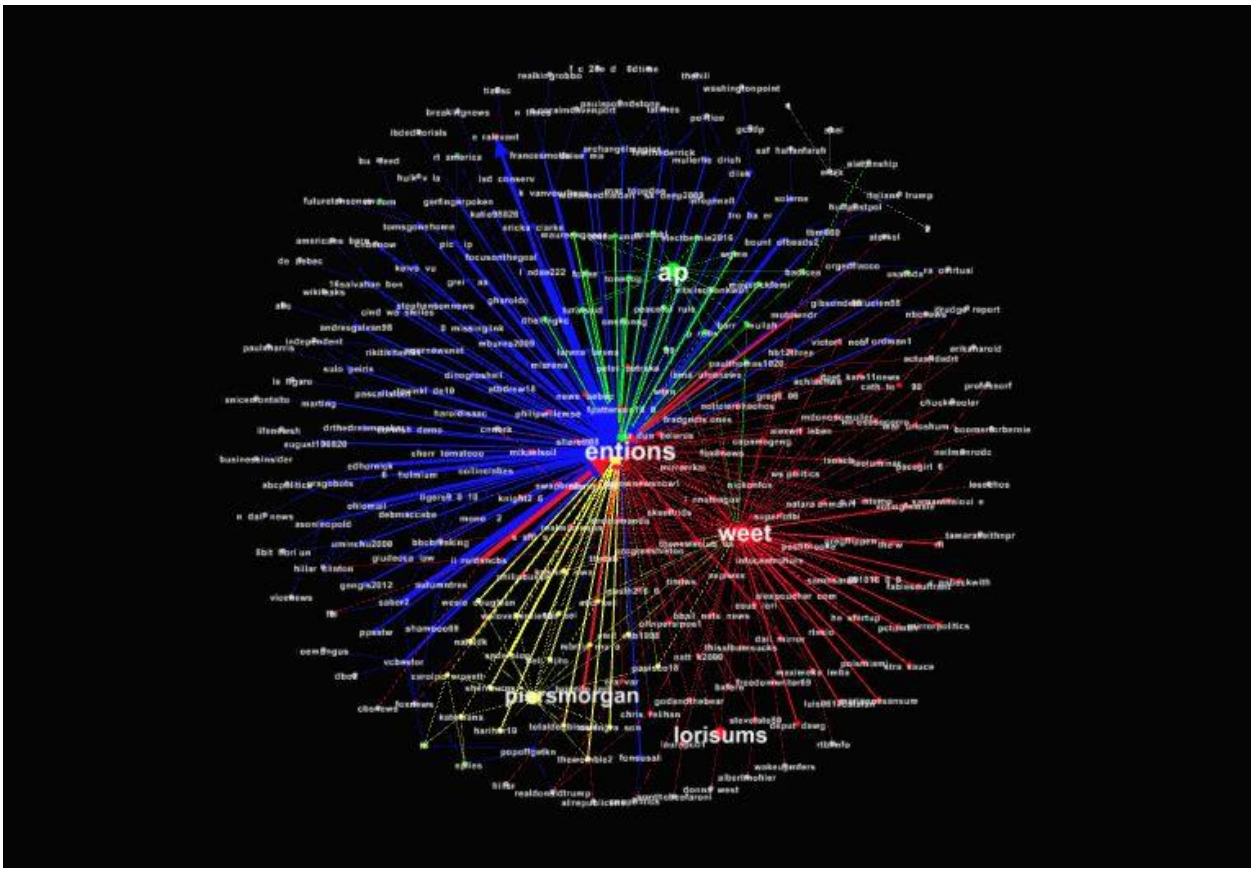


Figure 3: Network of friends and followers for Hillary Clinton

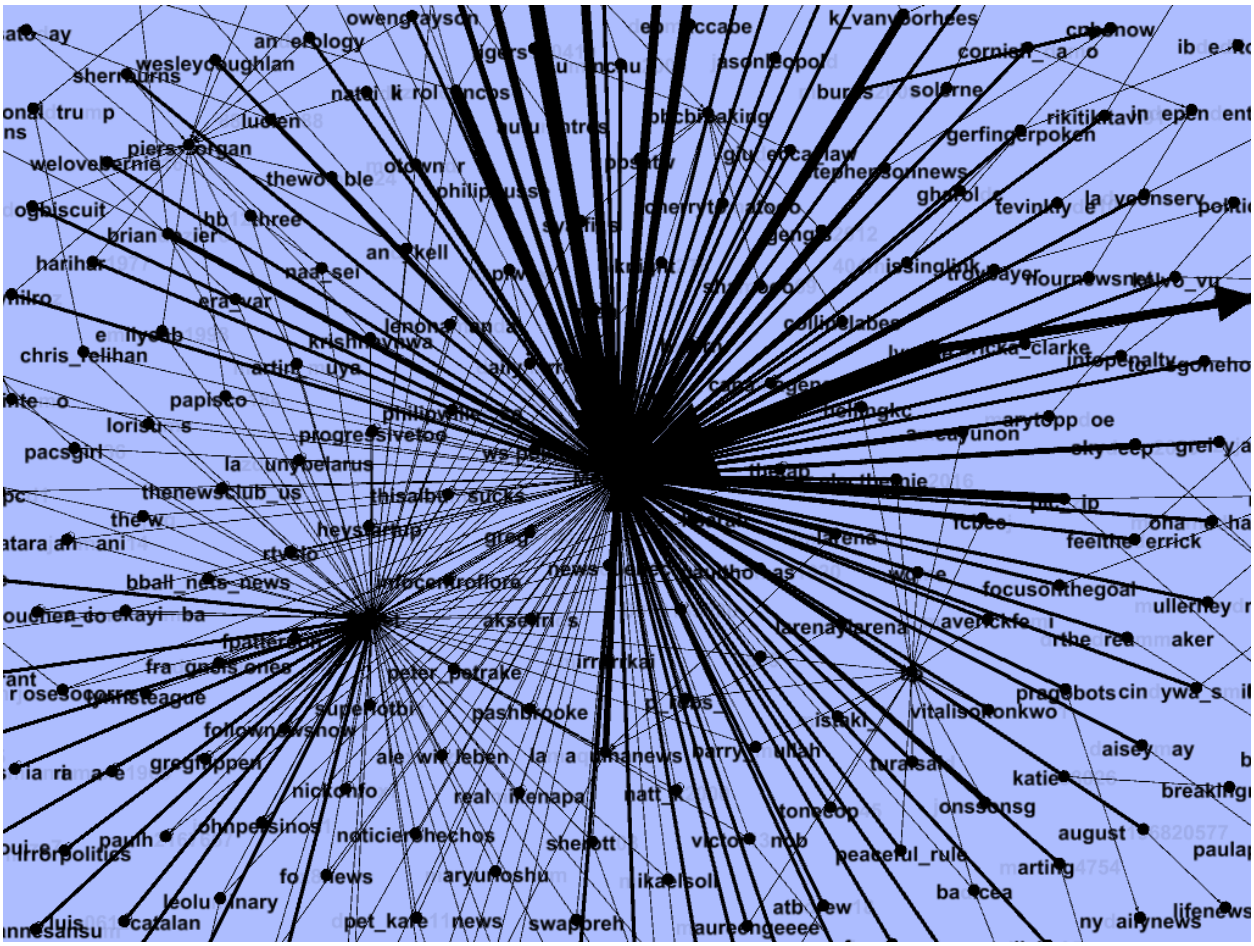


Figure 4: Closer view of friends and followers of Hillary Clinton

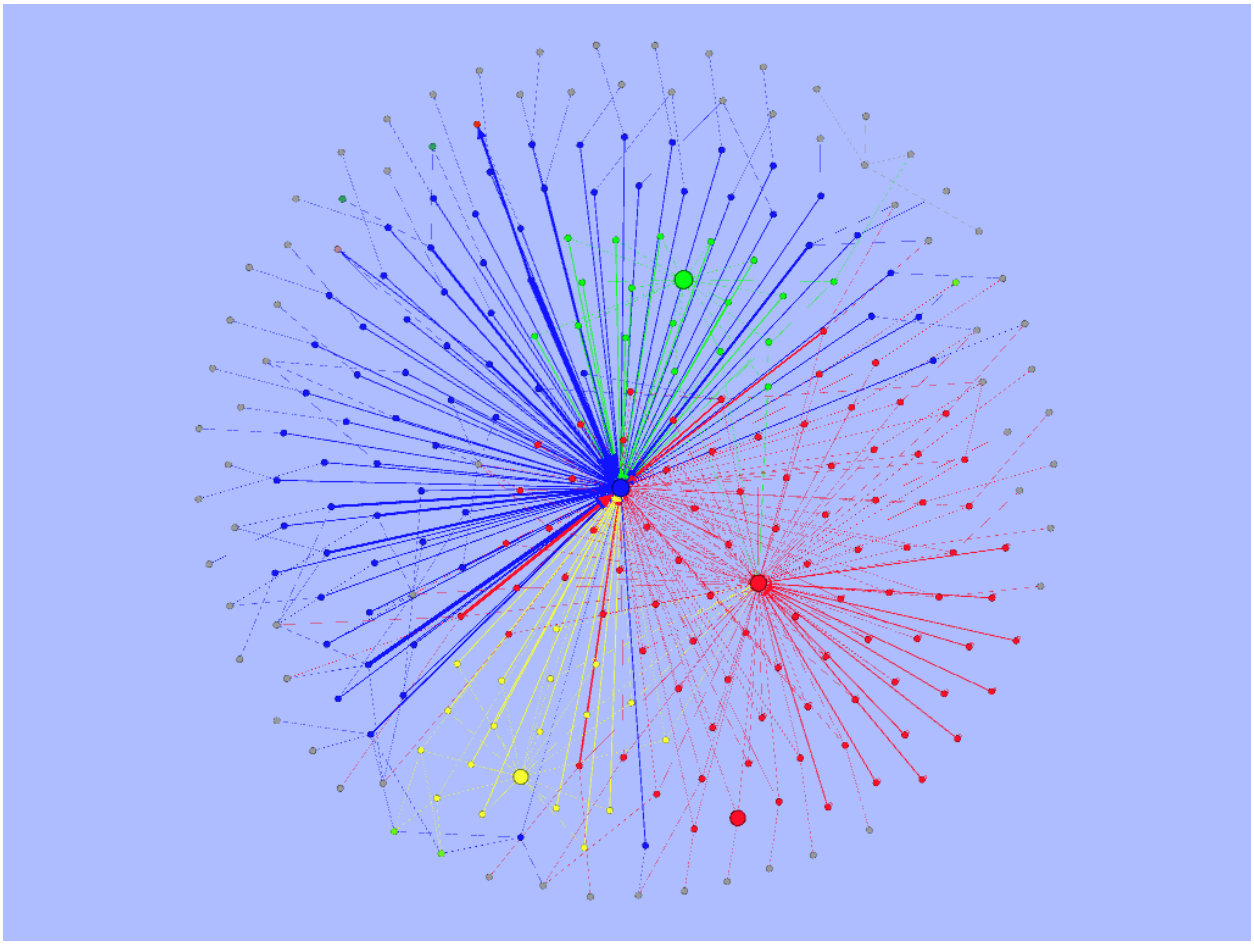

Figure 5: egocentric 2 network with neighbor for Hillary Clinton

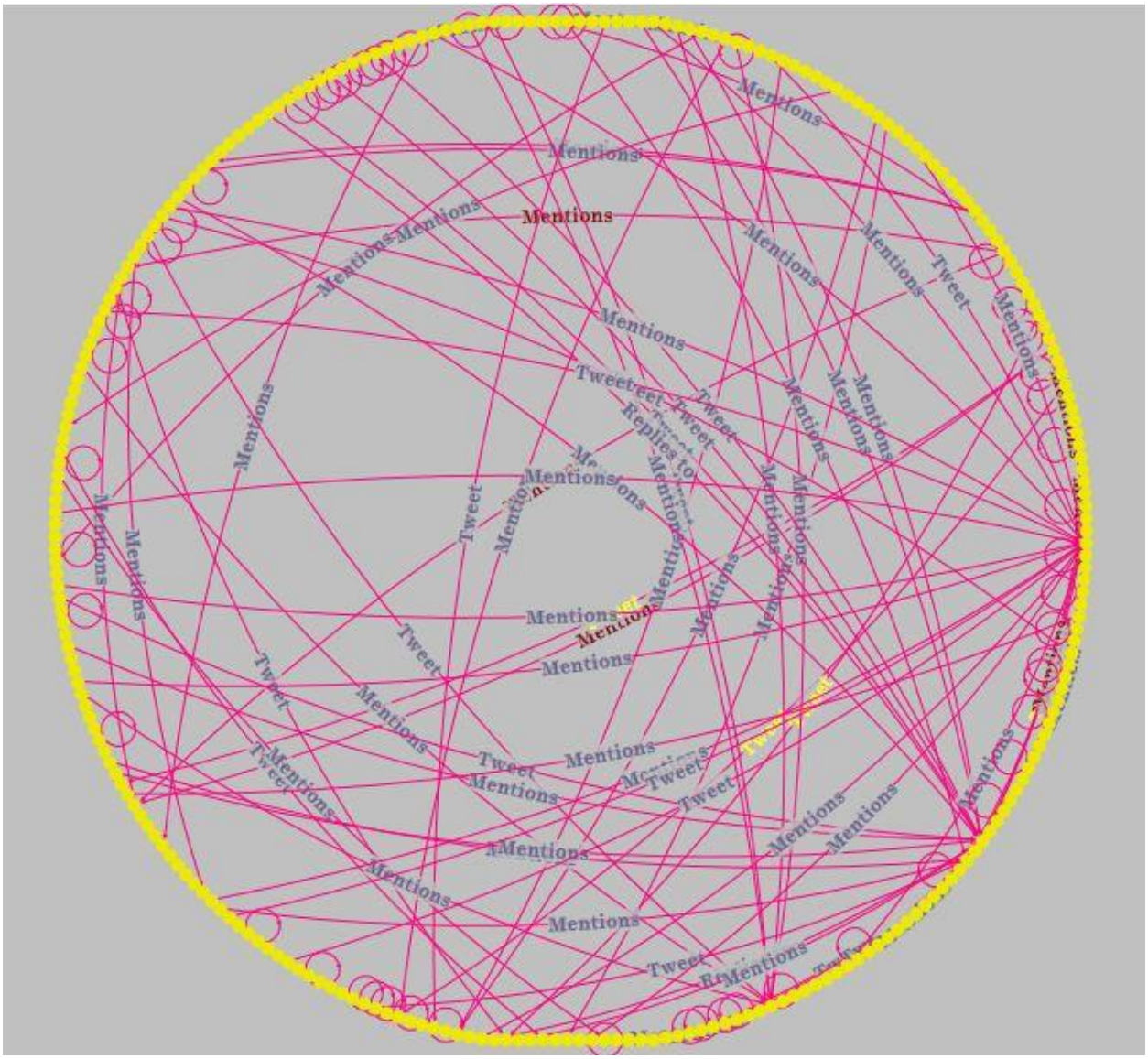


Figure 6: User Mentioned or tweeted about Hillary Clinton

Average Degree: 3.676

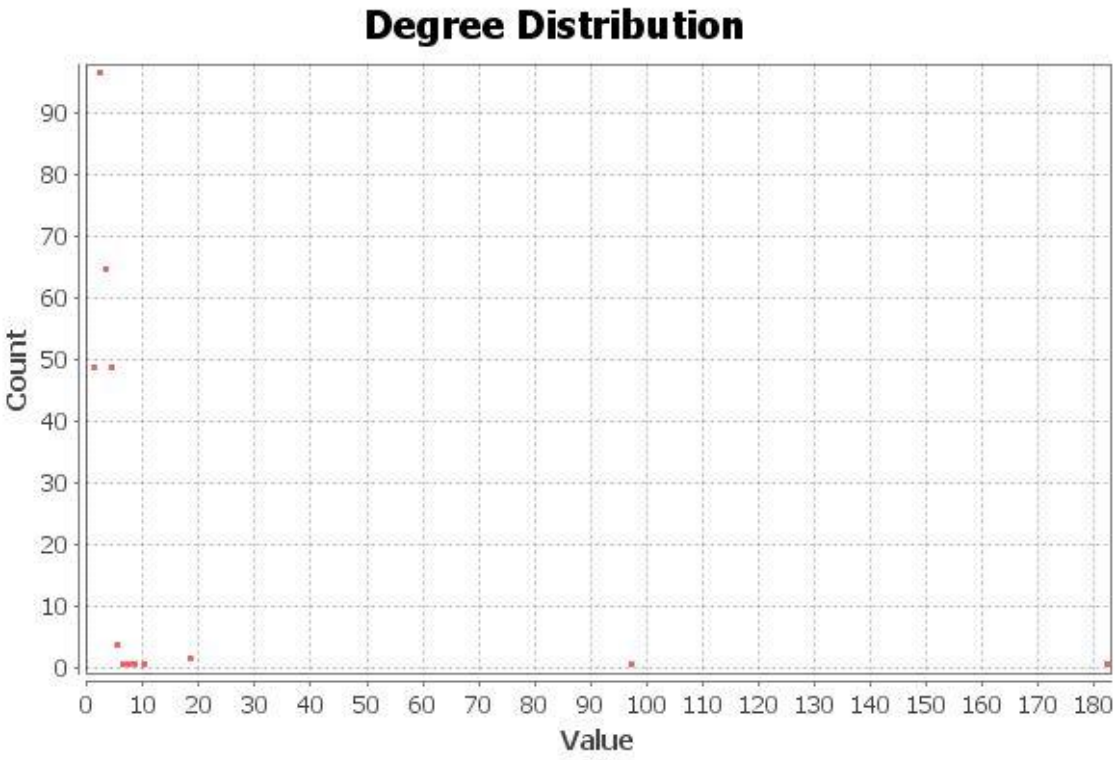


Figure 7: Average Degree Distribution for Hillary Clinton

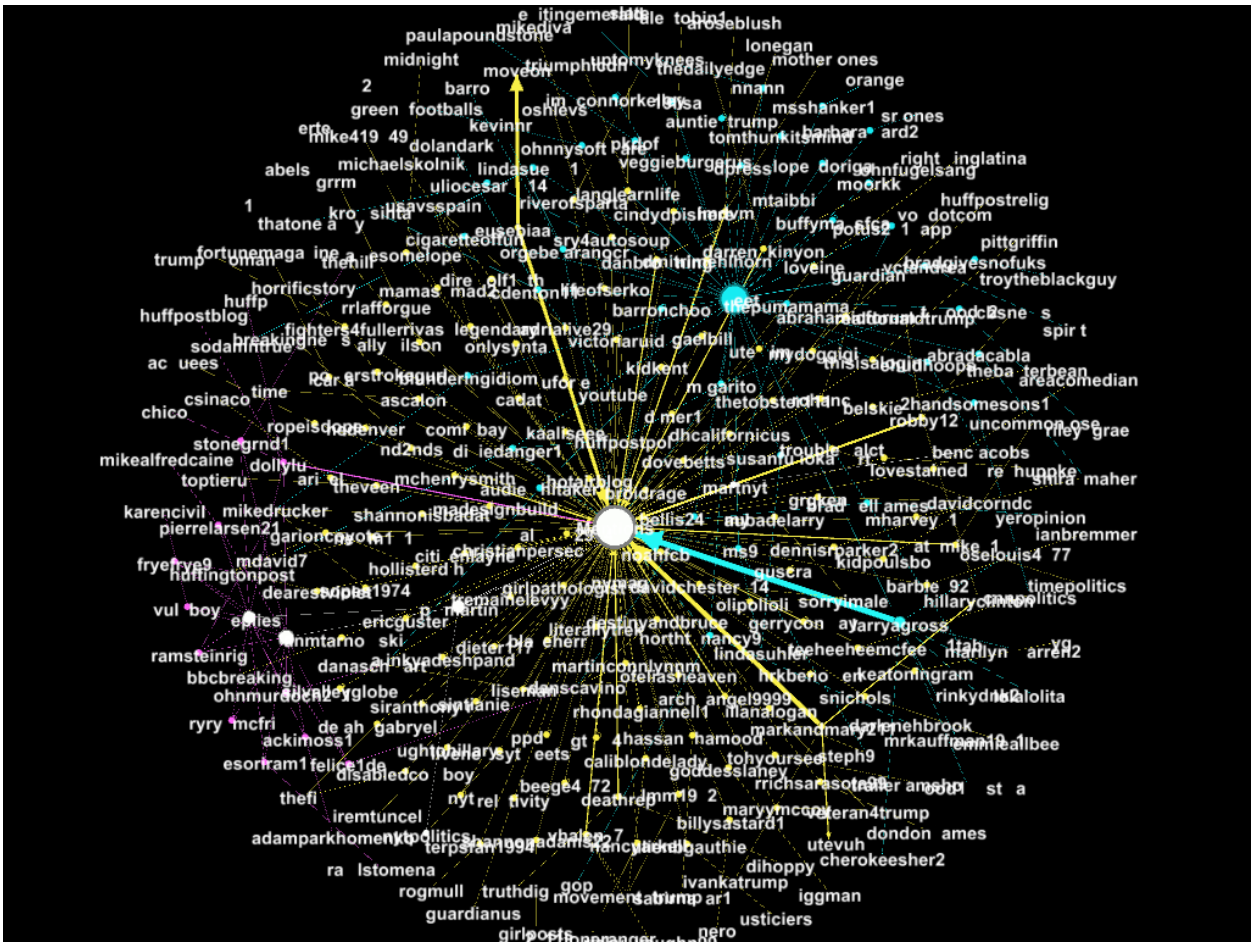

Figure 8: Network of friends and followers for Donald Trump

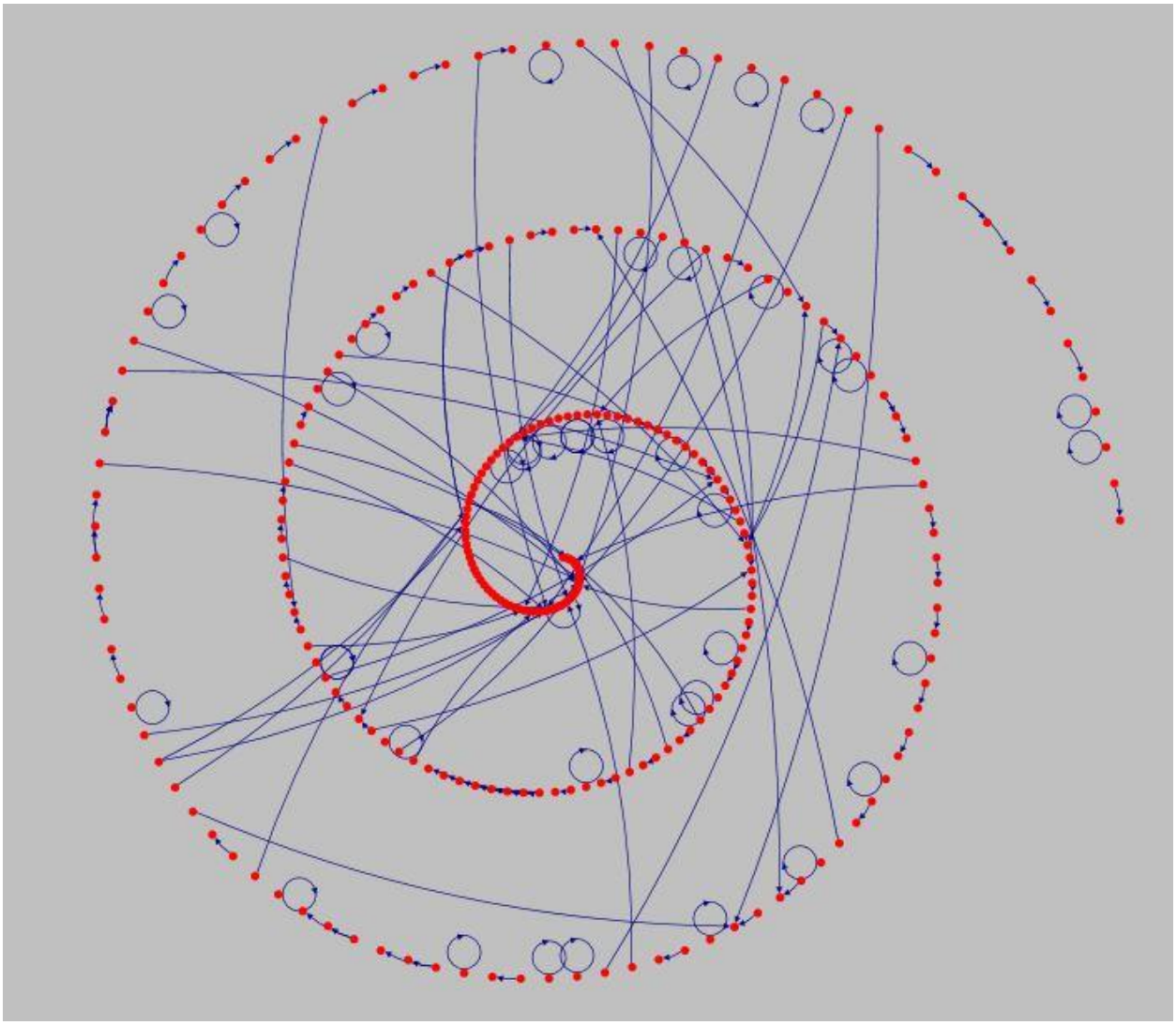

Figure 9: User Mentioned or tweeted about Donald Trump

**Results:**

Average Degree: 2.738

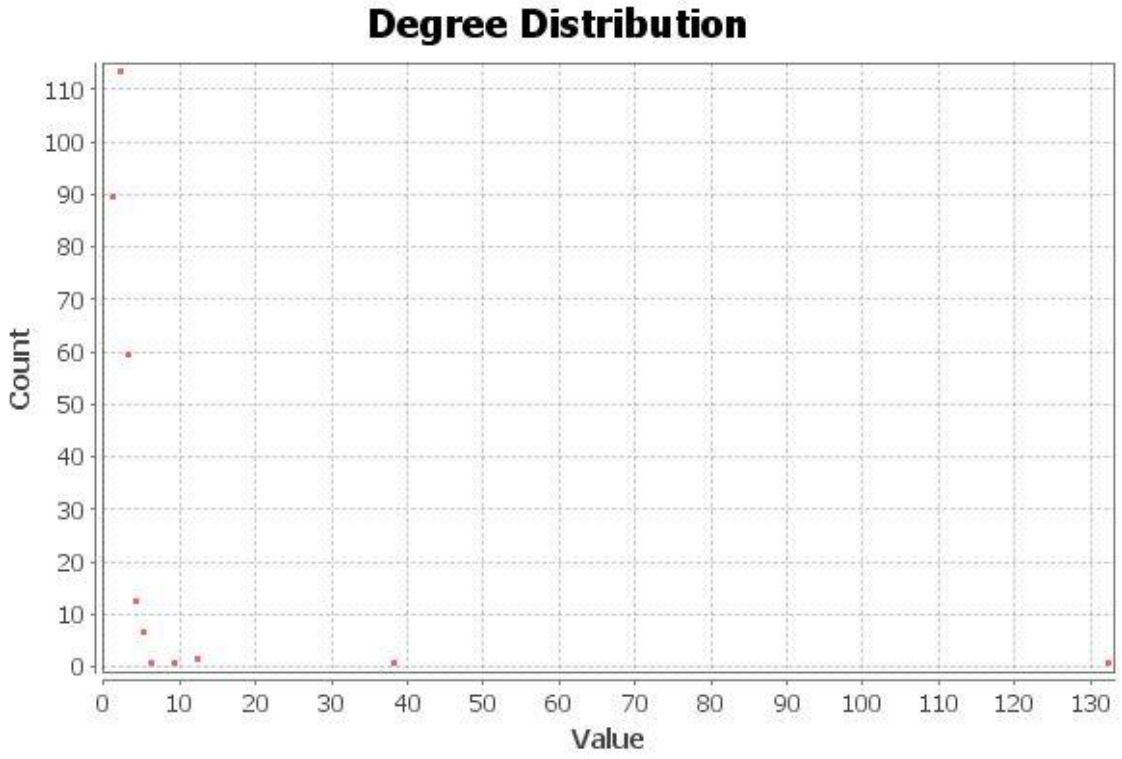


Figure 10: Average Degree Distribution for Donald Trump

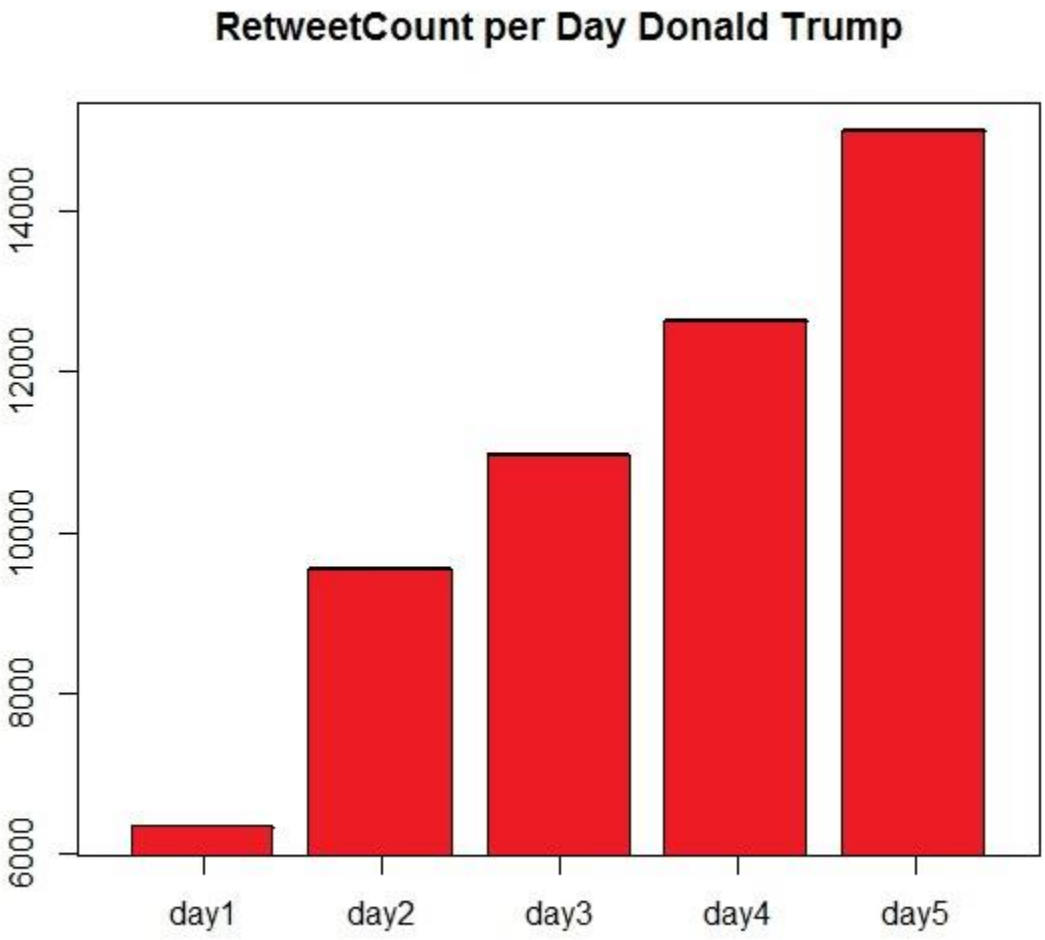


Figure 11: Retweet Count per Day for Donald Trump

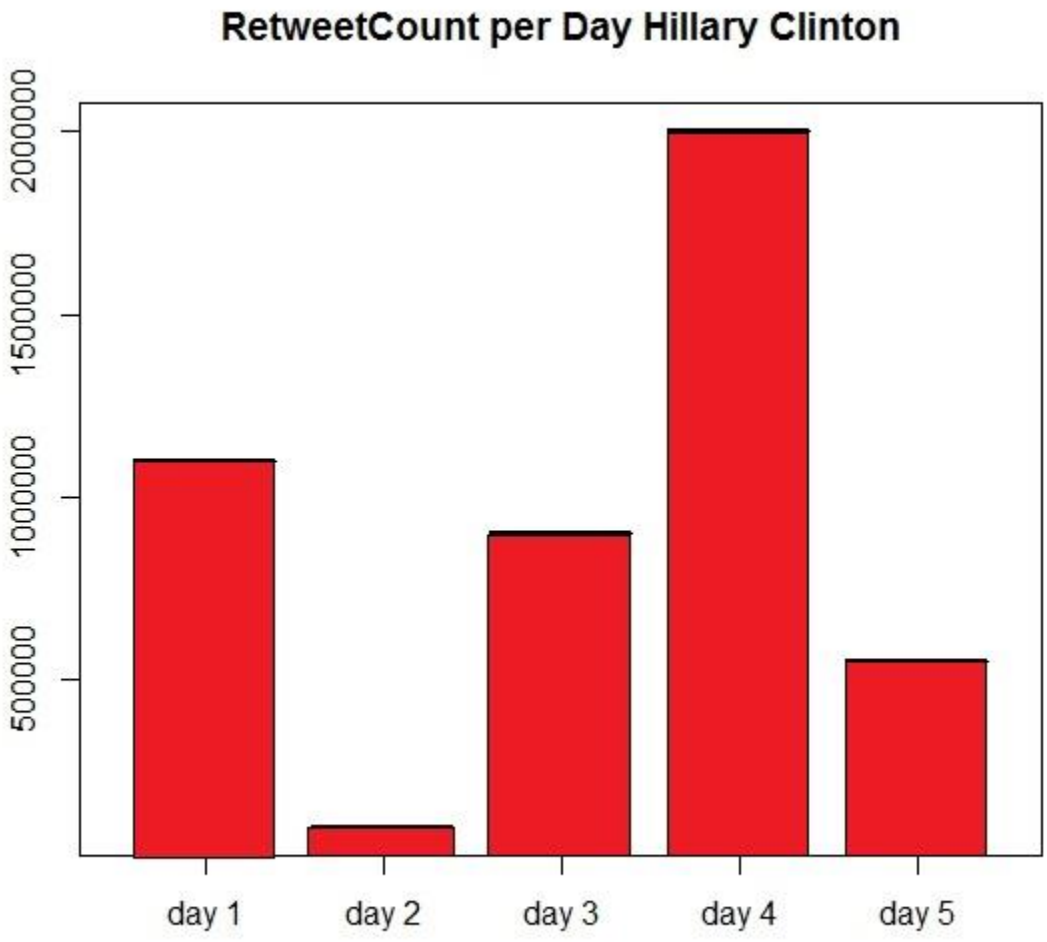


Figure 12: Retweet Count per Day for Hillary Clinton

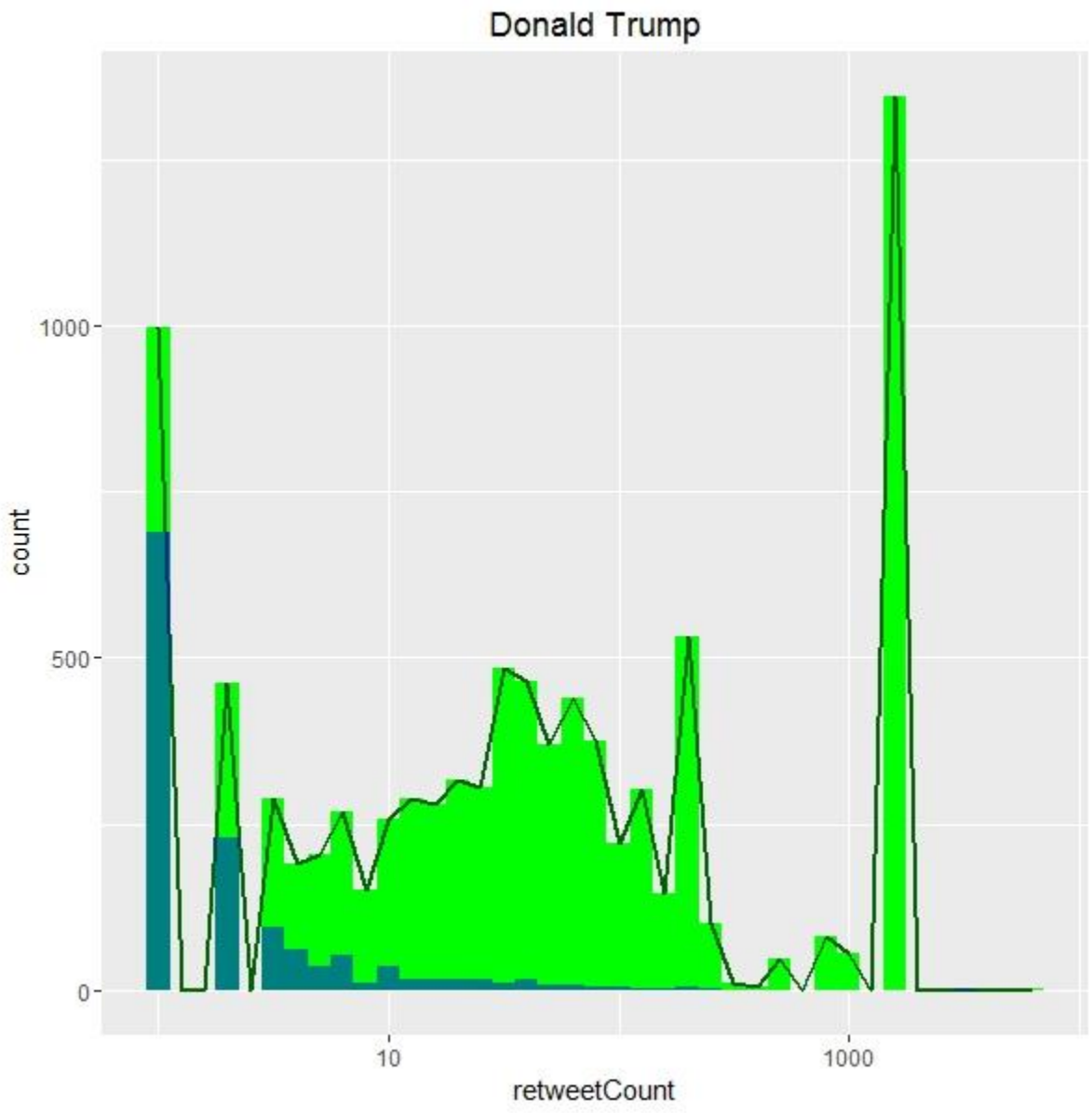


Figure 13: Retweet per Tweet Count Analysis for Donald Trump

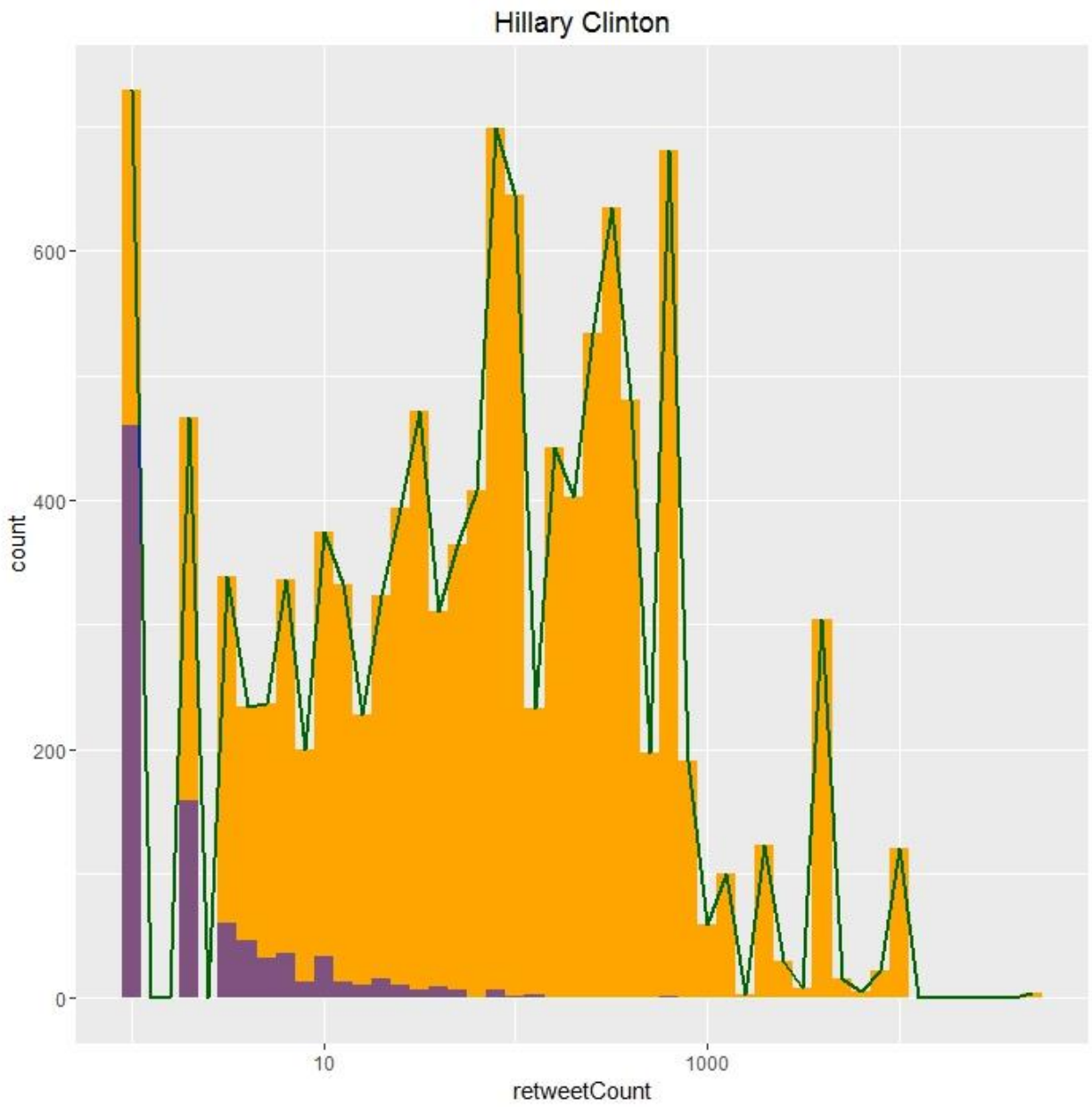


Figure 14: Retweet per tweet count analysis for Hillary Clinton

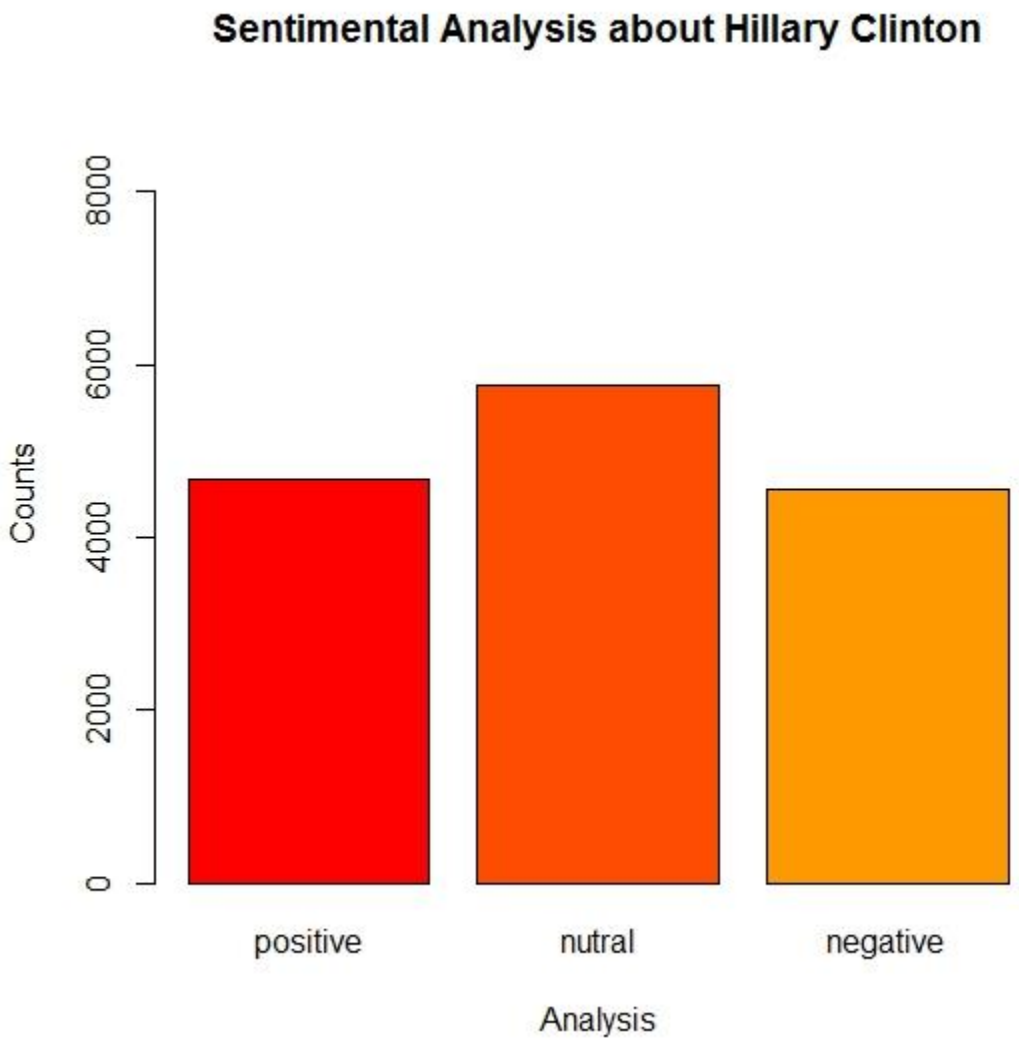


Figure 15: Sentimental analysis for Hillary Clinton

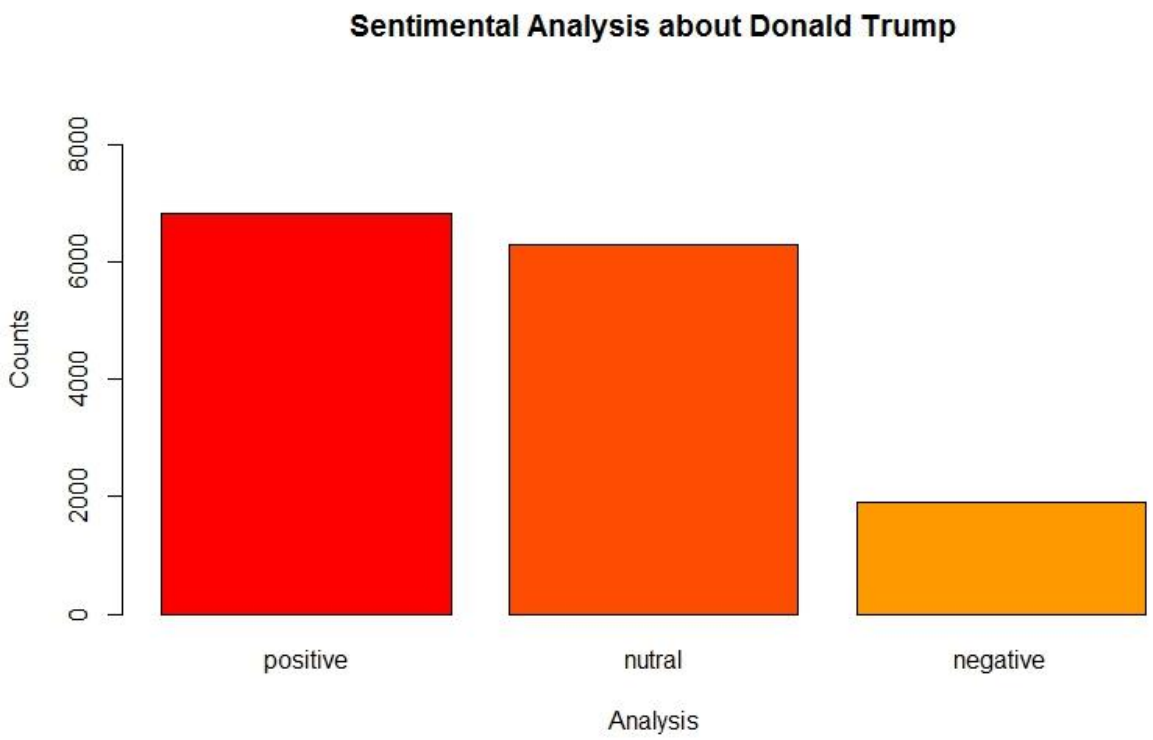


Figure 16: Sentimental analysis for Donald Trump

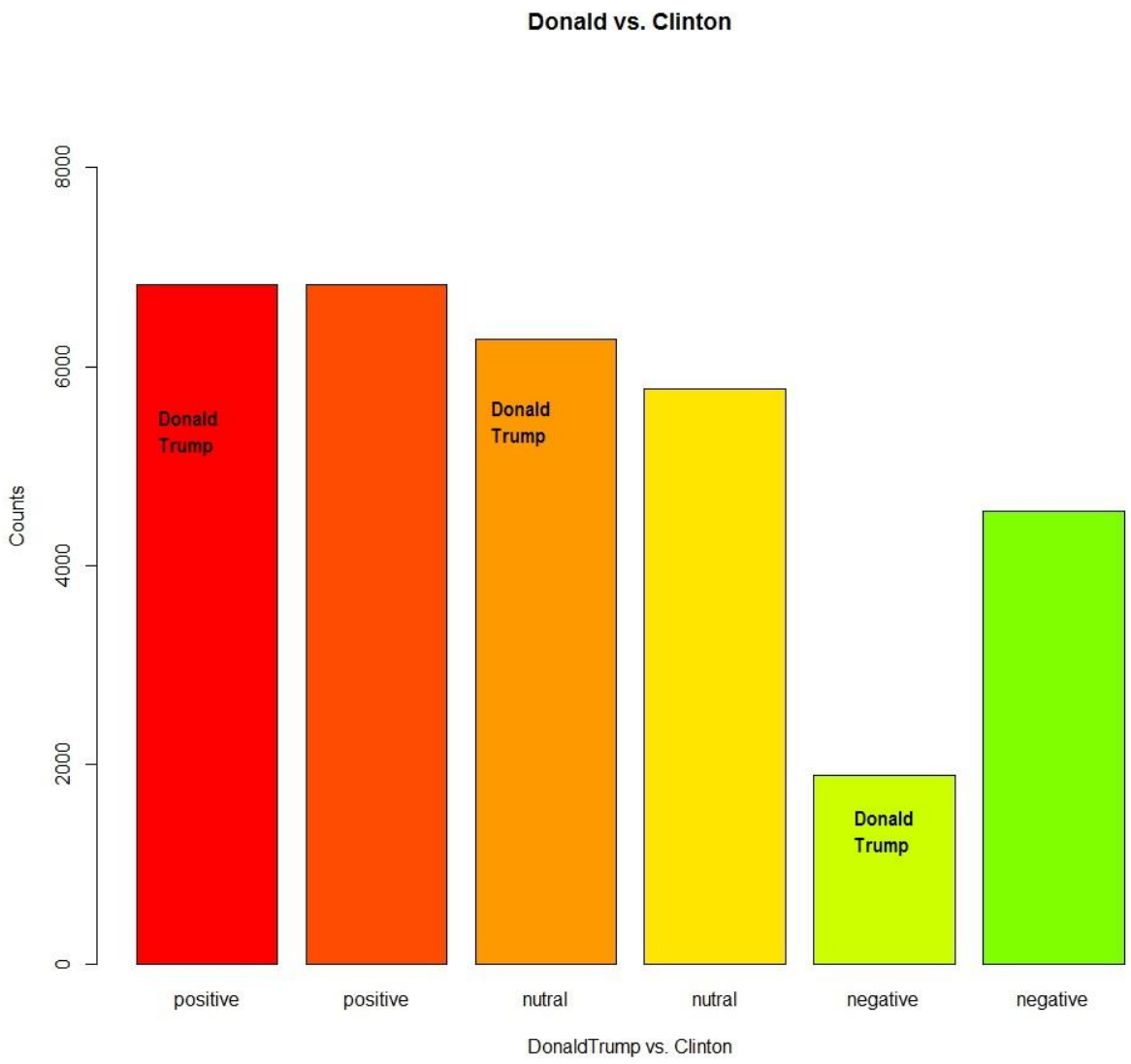


Figure 17: Sentimental Analysis between Hillary Clinton and Donald Trump


Figure 18: Word Cloud for US Election analysis

Figure 19: Word Cloud for Hillary Clinton analysis


Figure 20: Word Cloud for Donald Trump analysis

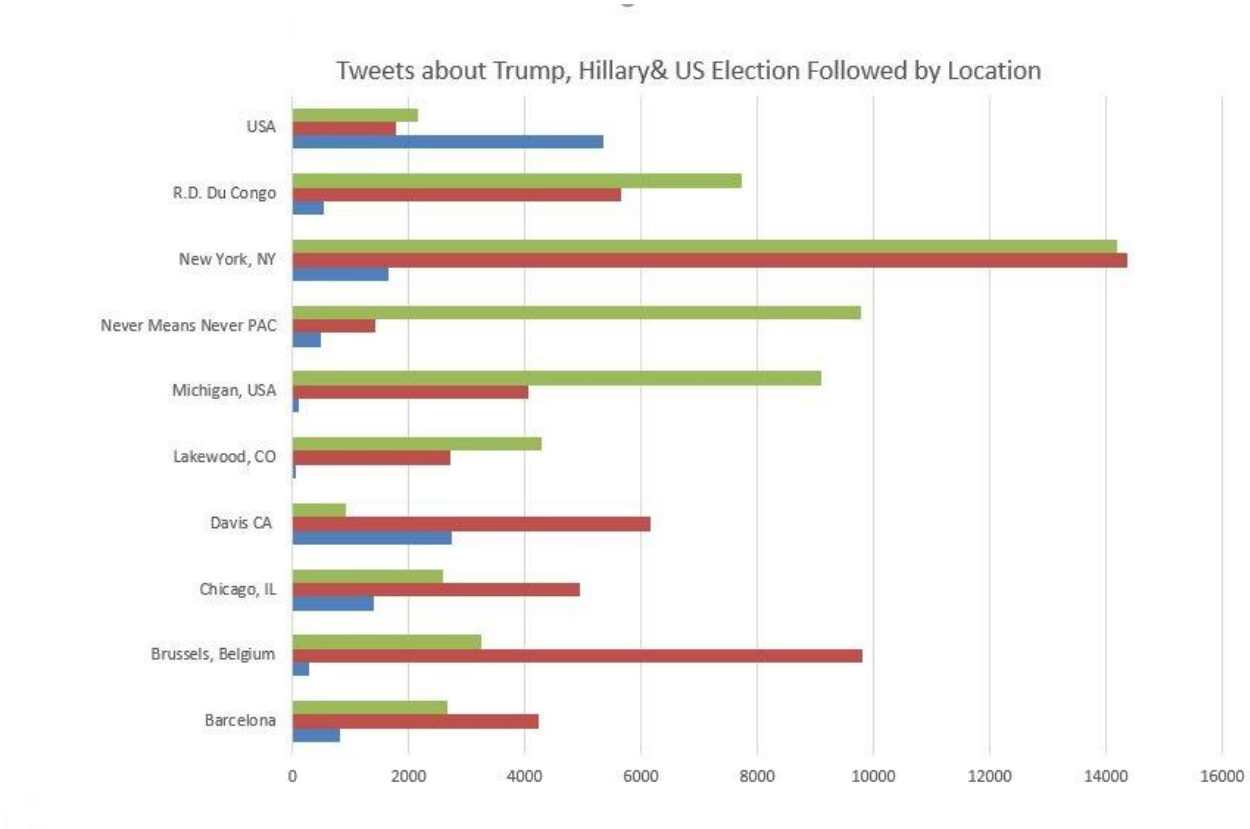


Figure 21: Country and Location wise analysis regarding tweets about Trump, Hillary and US Election.