Social Media Mining – Social Media Data Analysis

Praveenaa Kulandhaivel

September 23, 2022

I. Project Summary

To analyze Twitter data, the API method was used. Using Tweepy, tweets were crawled and scraped using the Twitter API based on vaccine sentiments i.e., Pro-Vaccine sentiments and Anti-Vaccine sentiments. The scraped tweets were stored in JSON format and then read from the file to build a network. The network was built as an Interaction network based on user mentions. From the network, network measures were calculated.

II. Data Collection

The data was collected using the API method and Tweepy library. The data was collected primarily based on hashtags as below:

i. Pro-vaccine tweets

#VaccinesSaveLives
#VaccinesWork
#ImVaccinated

The number of pro-vaccine tweets collected was 905.

ii. Anti-vaccine tweets

#VaccineSideEffects
#StopVaccination
#VaccineMandate
#NoVaccineForMe

The number of anti-vaccine tweets collected was 1256.

III. Network Visualization

Once the data was collected, an Interaction network was built where the nodes are the users, and the edges are user mentions, i.e., mentions of another user in a tweet or when a Twitter user interacted with another Twitter user within the tweets collected.

Two networks were built - one for pro-vaccine sentiment tweeters and another for anti-vaccine sentiment tweeters.

i. Pro-vaccine Network

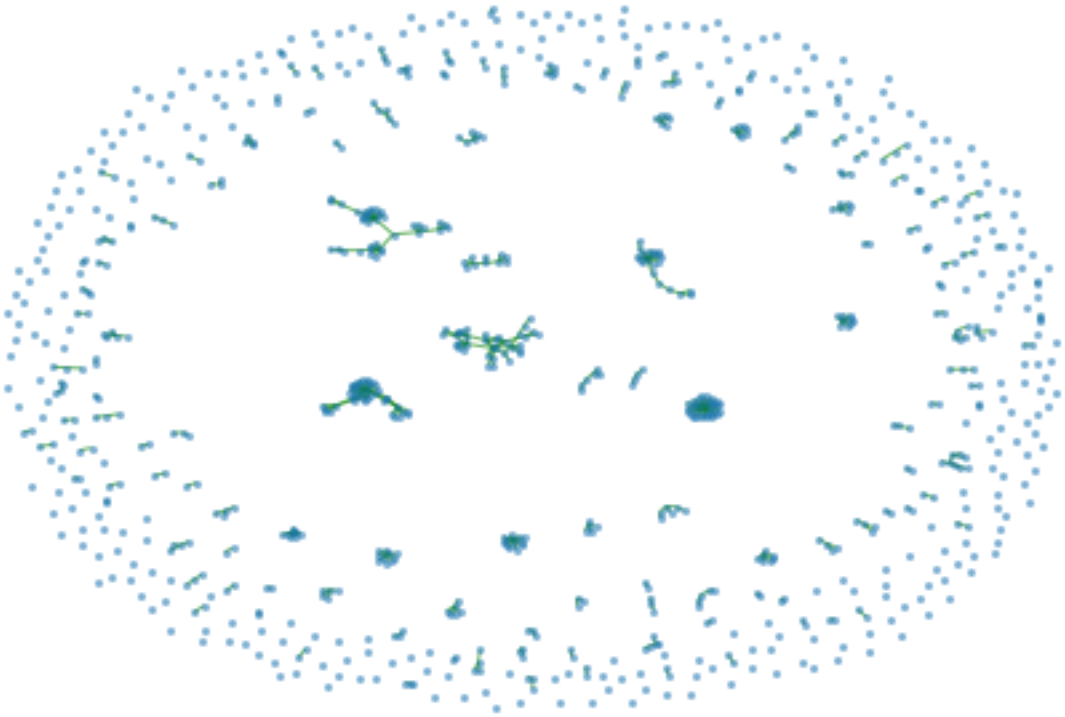

Figure 1 Network visualization of Pro-vaccine tweets' user and user mentions

Number of nodes in the Pro-Vaccine Interaction Graph:905

Number of Connected Components in the Pro-Vaccine Interaction Graph: 459

Figure 3 Largest sub-graph in the Pro-vaccine network graph

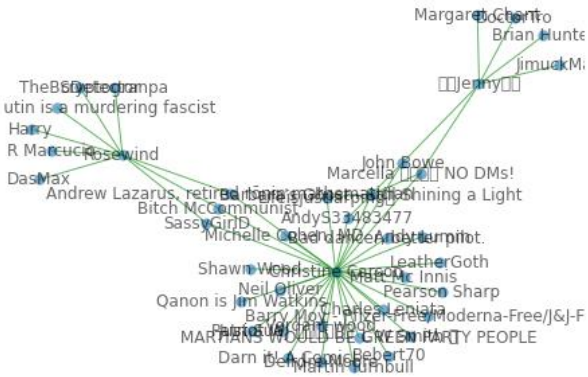

Figure 2 Largest sub-graph in Pro-vaccine network graph labelled

The largest subgraph in the pro-vaccine interaction Graph has 41 nodes and 47 edges

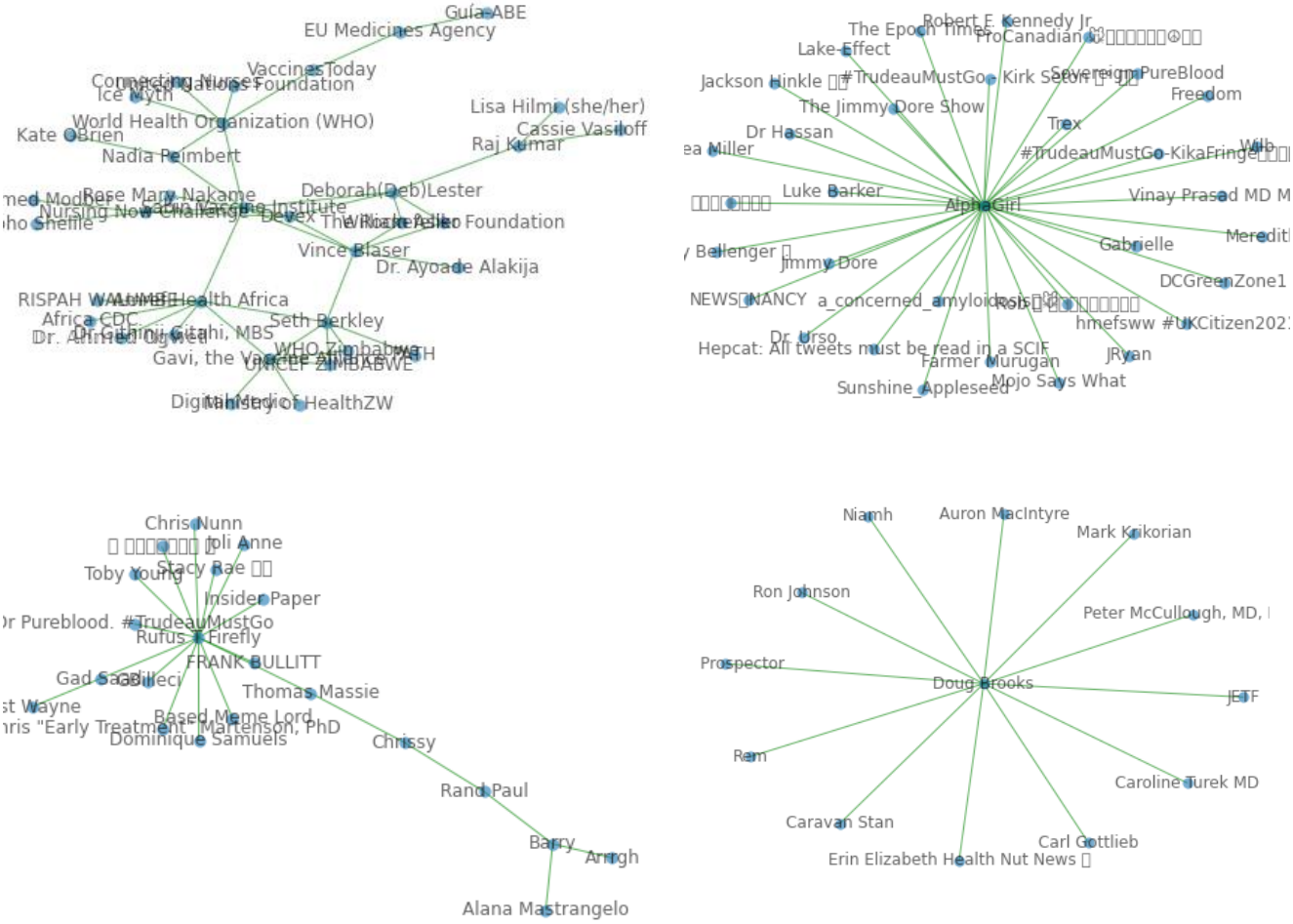

Figure 4 Oher connected sub graphs within the network

ii. Anti-Vaccine Network

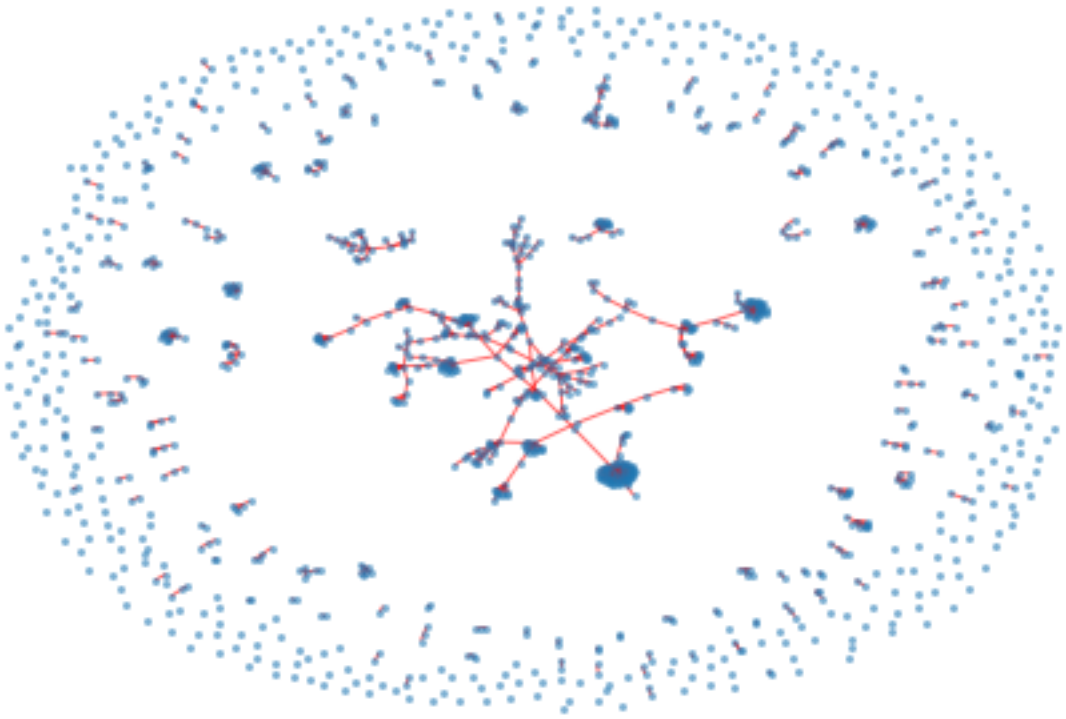

Figure 5 Network visualization of pro-vaccine tweets' users and user mentions

Number of nodes in the Anti-Vaccine Interaction Graph :1256

Number of Connected Components in the Anti-Vaccine Interaction Graph: 544

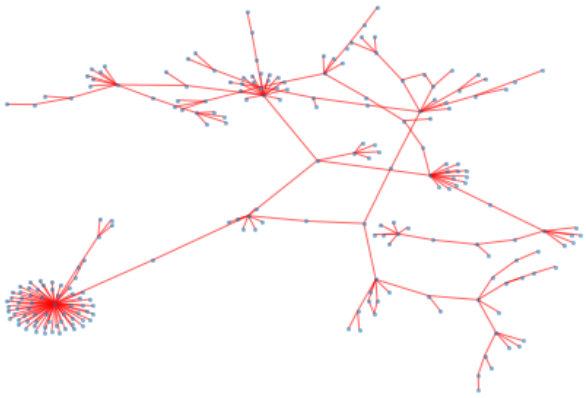

Figure 6 Largest connected subgraph in the Network

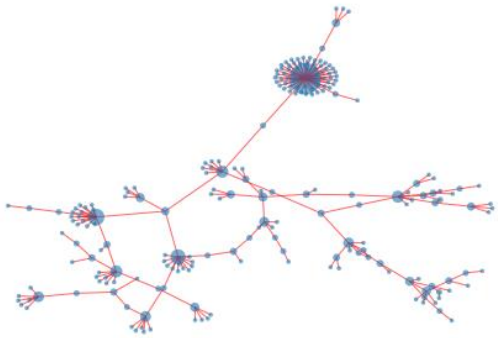

Figure 7 Largest Subgraph in the Network with node size varying according to the degree

The largest subgraph in the Pro-vaccine Interaction Graph has 223 nodes and 225 edges.

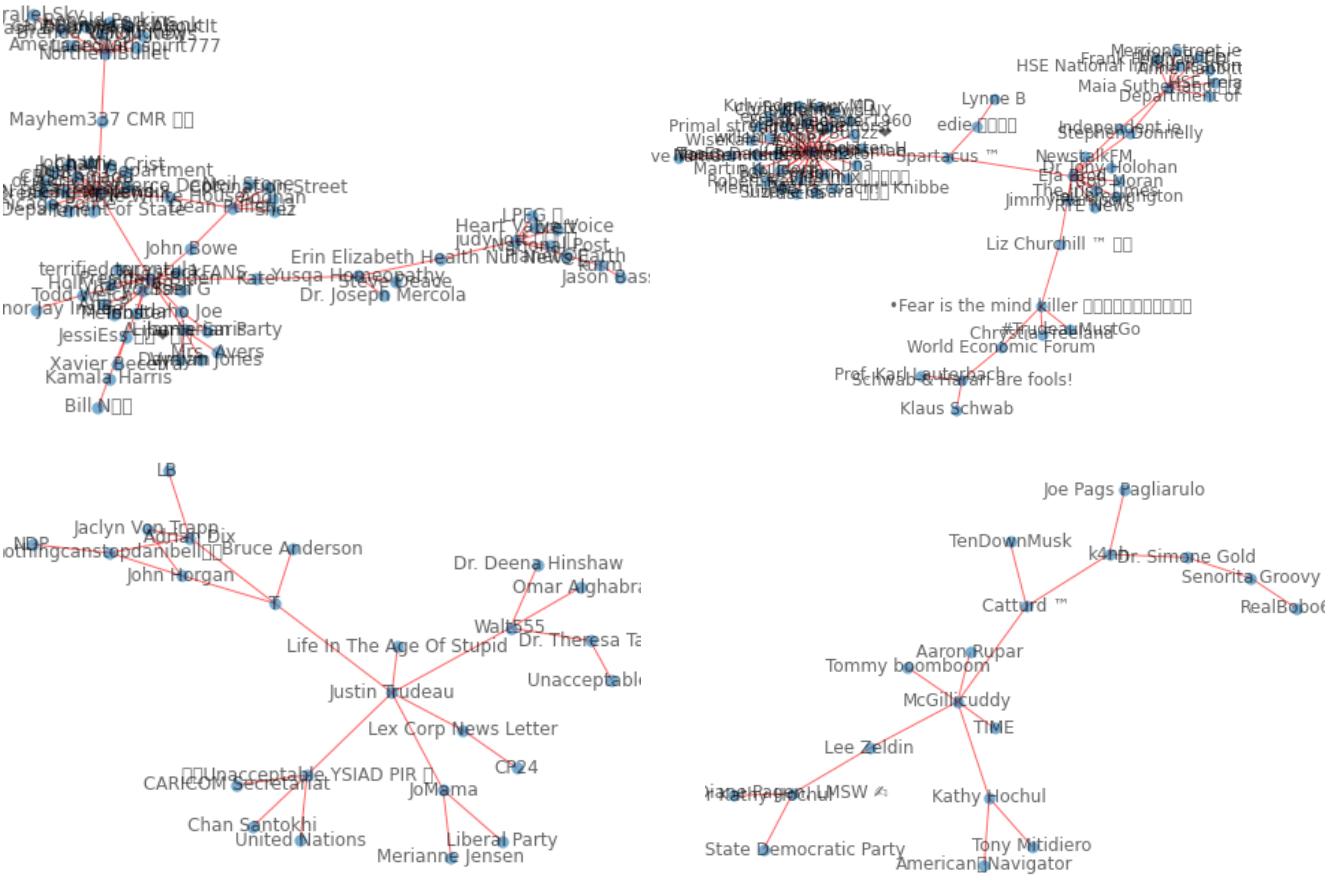

Figure 9 Other Connected Sub-graphs in the network

IV. Network Measures

i. Degree Distribution - Histogram

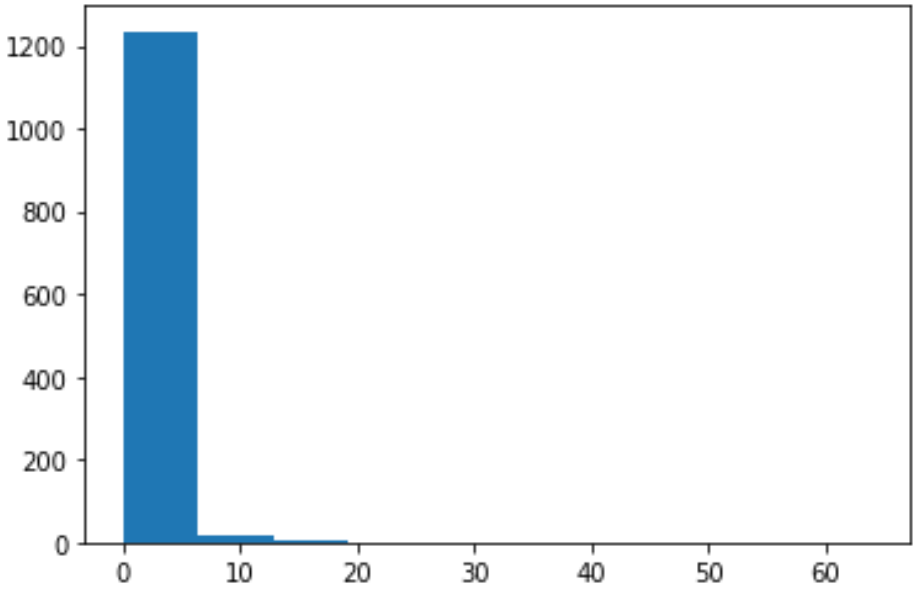

Figure 11 Degree Distribution Histogram for Anti-Vaccine Network

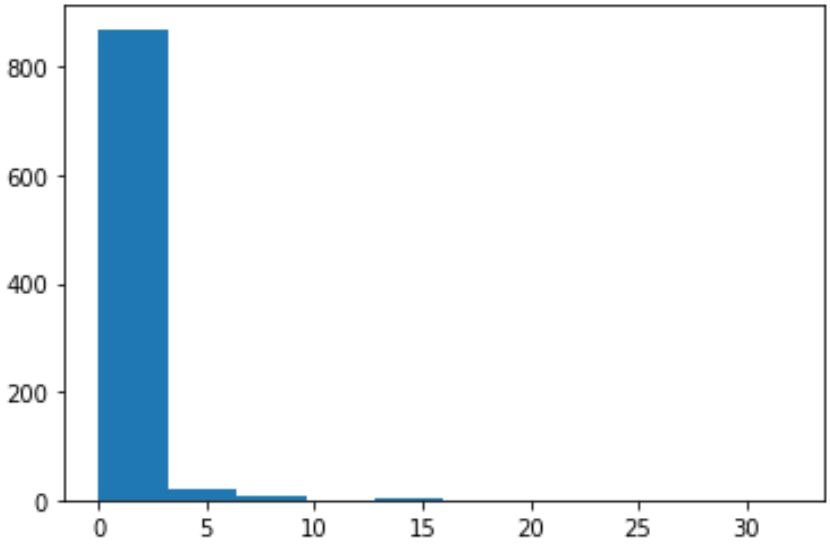

Figure 10 Degree Distribution for Pro-vaccine Network

The network measures below are calculated for the largest connected sub-graph, as the full graph is not completely connected.

ii. Diameter of the largest sub-graph

- Pro-vaccine Network - 6
- Anti-vaccine Network – 18

iii. Graph closeness

- Pro-vaccine Network - the node with id Christine Carson has a closeness centrality of 0.65 – the highest for the sub-graph
- Anti-vaccine Network - the node with id LaRoja has a closeness centrality of 0.22- highest for the sub-graph

iv. Graph Centrality

- Pro-vaccine network – the node with id 'Christine Carson' has a degree centrality of 0.70 – the highest for the subgraph
- Anti-vaccine Network – the node with id 'Rehmat' has a degree centrality of 0.29 -subgraph

V. References

- https://docs.tweepy.org/en/stable/
- https://networkx.org/documentation/stable/reference/functions.html
- https://github.com/ugis22/analysing_twitter/blob/master/Jupyter%20Notebook%20files/Interaction%20Network.ipynb
- https://matplotlib.org/stable/index.html
- https://docs.python.org/3/library/json.html