CSE 472 : SOCIAL MEDIA MINING

PROJECT 1 – Social Media Data Analysis

Name : Ayushi Nirmal
ASU ID : 1225466320

Prof : Dr. Huan Liu
TA : Zhen Tan

Email ID : ztan36@asu.edu

DECLARATION

I hereby declare that the project report for **Social Media Data Analysis**, is my original work and to my knowledge, has not been published before. The references and assumptions for the abstract have been defined in the report for each distribution.

Place : Tempe, Arizona.
Date : 22 September 2022

Name : Ayushi Nirmal
ASU ID : 1225466320

ABSTRACT

In this project, I have crawled twitter data to perform exploratory data analysis. The data is around Covid 19 Vaccine to deduce the biases towards two different attitudes i.e. ProVaccine and AntiVaccine. I have chosen API method using Tweepy library of Python in order to crawl data. First, I have crawled data for Provaccine and AntiVaccine queries respectively. Second, I have analyzed the data as a graph where users serve as nodes and their mentions and re-tweets from the mentions serve as an edge between the user and mentions and retweets respectively in the graph. After obtaining the graphs for the above two attitudes, I have performed Network Analysis over the graphs such as degree distribution, betweenness centrality, closeness centrality and clustering coefficient.

ACKNOWLEDGEMENT

I would like to express my gratitude to Dr. Huan Liu and Zhen Tan for providing us this project to work upon. The project has helped me understand social media network between people and the spread of biases among people. The project gave a real insight into the working of real social networking models and hands-on real time model. Through this project, I have had the experience to learn how to collect the data from Twitter and perform data cleaning, processing in order to extract useful data to perform data analysis.

TABLE OF CONTENTS

Step 1 : Collect the Data

1.1 : Get Elevated Accessfor Twitter Developer Account

- In order to collect the data, first we need to get an access to the twitter developer account through https://developer.twitter.com/ which usually takes 1-2 days.
- After we have got elevated developer access from Twitter, we can access API Key, Secret and Access Token and Secret.
- After we have got the keys, we now can access the twitter API using Tweepy library of python.

```
api_key = '8RQ5xPiDzAg1duMVpj9pWYEdX'
api_key_secret = 'BqfcElqH5G6yAFHafuMQWO5ljp8Bosa402DzldWgc3mzCOSrSB'
access_token = '1570146620653850629-erndFgwUuLXuK8prEhhrpkqlBUICih'
access_token_secret = 'hwjy8IzsfECPVje0SAzdq3YorFifjAZbpBBgmLxBxnGVc'
```

1.2 : Access the API Keys and get Twitter API authorization

Now that we have the api_key, api_key_secret, access_token, access_token_secret, we can access the API using Tweepy as below :

```
auth = tweepy.OAuth1UserHandler(api_key, api_key_secret)
auth.set_access_token(access_token, access_token_secret)
return tweepy.API(auth)
```

Step 2 : DATA CRAWLING

Now, we can go ahead with the data crawling using few hashtags for two different biases i.e. ProVaccine and AntiVaccine for **600 tweets** as below :

***Provaccine* :**

'#GetVaccinated', '#GetVaxxed', '#VaccinesWork', '#VaccinationWork', '#FullyVaccinated'

***AntiVaccine* :**

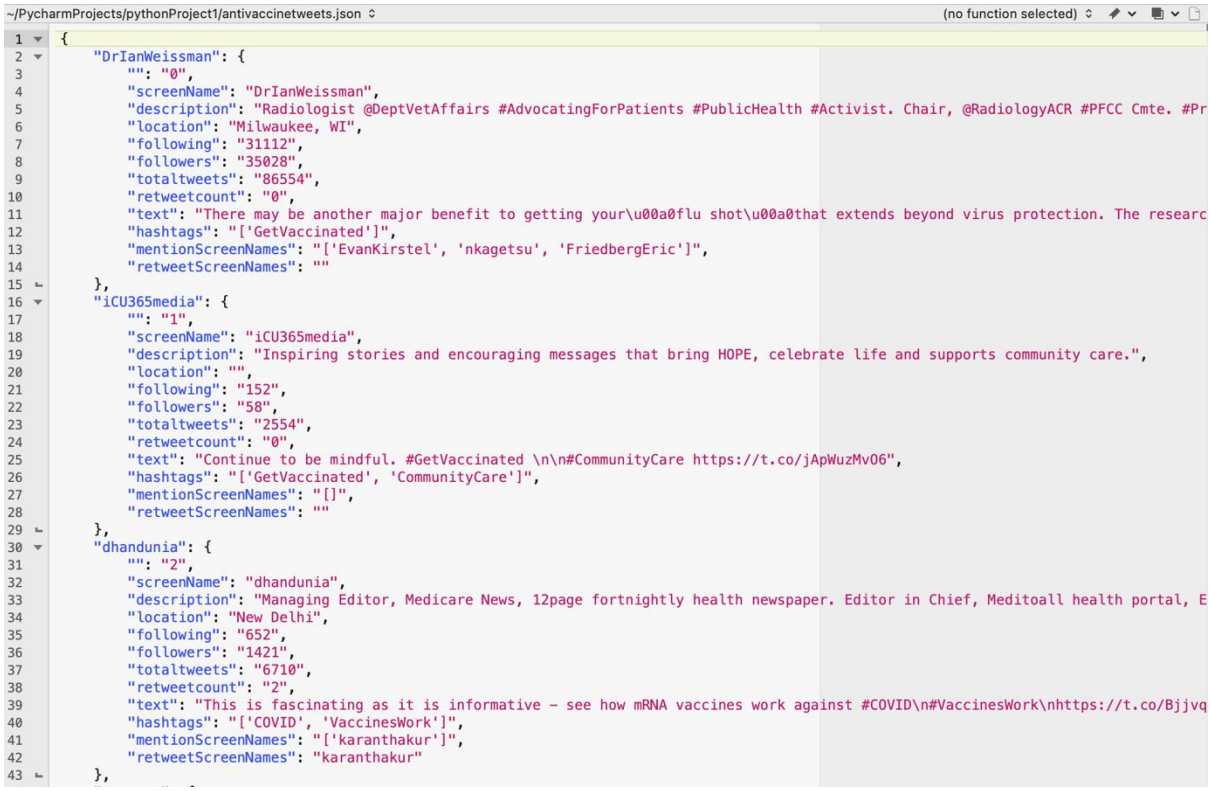
'#AntiVaccine', '#NoForcedVaccines', '#VaccinesKill', '#NoVaccineForMe', '#SayNoToVaccineMandate'

```
tweets = tweepy.Cursor(api.search_tweets,
                       words, lang="en",
                       since_id=date_since,
                       tweet_mode='extended').items(numtweet)
i = 1
retweet_username = ''
for tweet in tweets:
    username = tweet.user.screen_name
    description = tweet.user.description
    location = tweet.user.location
    following = tweet.user.friends_count
    followers = tweet.user.followers_count
    totaltweets = tweet.user.statuses_count
    retweetcount = tweet.retweet_count
    mentions = tweet.entities['user_mentions']
    hashtags = tweet.entities['hashtags']
    mention_usernames = getScreenNamesofMentions(mentions)
    try:
        text = tweet.retweeted_status.full_text
        retweet_username = tweet.retweeted_status.user.screen_name
    except AttributeError:
        text = tweet.full_text
    hashtext = list()
    for j in range(0, len(hashtags)):
        hashtext.append(hashtags[j]['text'])
    ith_tweet = [username, description,
                 location, following,
                 followers, totaltweets,
                 retweetcount, text, hashtext, mention_usernames, retweet_username]
    db.loc[len(db)] = ith_tweet
```

Pro Vaccine Json File


```
{
    "DrIanWeissman": {
        "": "0",
        "screenName": "DrIanWeissman",
        "description": "Radiologist @DeptVetAffairs #AdvocatingForPatients #PublicHealth #Activist. Chair, @RadiologyACR #PFCC Cmte. #Pr
        "location": "Milwaukee, WI",
        "following": "31112",
        "followers": "35028",
        "totaltweets": "86554",
        "retweetcount": "0",
        "text": "There may be another major benefit to getting your\u00a0flu shot\u00a0that extends beyond virus protection. The researc
        "hashtags": "['GetVaccinated']",
        "mentionScreenNames": "['EvanKirstel', 'nkagetsu', 'FriedbergEric']",
        "retweetScreenNames": ""
    },
    "iCU365media": {
        "": "1",
        "screenName": "iCU365media",
        "description": "Inspiring stories and encouraging messages that bring HOPE, celebrate life and supports community care.",
        "location": "",
        "following": "152",
        "followers": "58",
        "totaltweets": "2554",
        "retweetcount": "0",
        "text": "Continue to be mindful. #GetVaccinated \n\n#CommunityCare https://t.co/jApWuzMvO6",
        "hashtags": "['GetVaccinated', 'CommunityCare']",
        "mentionScreenNames": "[]",
        "retweetScreenNames": ""
    },
    "dhandunia": {
        "": "2",
        "screenName": "dhandunia",
        "description": "Managing Editor, Medicare News, 12page fortnightly health newspaper. Editor in Chief, Meditoall health portal, E
        "location": "New Delhi",
        "following": "652",
        "followers": "1421",
        "totaltweets": "6710",
        "retweetcount": "2",
        "text": "This is fascinating as it is informative - see how mRNA vaccines work against #COVID\n#VaccinesWork\nhttps://t.co/Bjjvq
        "hashtags": "['COVID', 'VaccinesWork']",
        "mentionScreenNames": "['karanthakur']",
        "retweetScreenNames": "karanthakur"
    },
```


Anti Vaccine Json File




```
{
    "DrIanWeissman": {
        "": "0",
        "screenName": "DrIanWeissman",
        "description": "Radiologist @DeptVetAffairs #AdvocatingForPatients #PublicHealth #Activist. Chair, @RadiologyACR #PFCC Cmte. #Pr
        "location": "Milwaukee, WI",
        "following": "31112",
        "followers": "35028",
        "totaltweets": "86554",
        "retweetcount": "0",
        "text": "There may be another major benefit to getting your\u00a0flu shot\u00a0that extends beyond virus protection. The researc
        "hashtags": "['GetVaccinated']",
        "mentionScreenNames": "['EvanKirstel', 'nkagetsu', 'FriedbergEric']",
        "retweetScreenNames": ""
    },
    "iCU365media": {
        "": "1",
        "screenName": "iCU365media",
        "description": "Inspiring stories and encouraging messages that bring HOPE, celebrate life and supports community care.",
        "location": "",
        "following": "152",
        "followers": "58",
        "totaltweets": "2554",
        "retweetcount": "0",
        "text": "Continue to be mindful. #GetVaccinated \n\n#CommunityCare https://t.co/jApWuzMvO6",
        "hashtags": "['GetVaccinated', 'CommunityCare']",
        "mentionScreenNames": "[]",
        "retweetScreenNames": ""
    },
    "dhandunia": {
        "": "2",
        "screenName": "dhandunia",
        "description": "Managing Editor, Medicare News, 12page fortnightly health newspaper. Editor in Chief, Meditoall health portal, E
        "location": "New Delhi",
        "following": "652",
        "followers": "1421",
        "totaltweets": "6710",
        "retweetcount": "2",
        "text": "This is fascinating as it is informative - see how mRNA vaccines work against #COVID\n#VaccinesWork\nhttps://t.co/Bjjvq
        "hashtags": "['COVID', 'VaccinesWork']",
        "mentionScreenNames": "['karanthakur']",
        "retweetScreenNames": "karanthakur"
    },
```

3. Building Graphs for ProVaccine and AntiVaccine

- In order to build a directed graph, I have taken users from tweets and have formed a connection from users to their mentions.
- If the tweet is a retweeted tweet, then form an edge between the original tweet user and the user from current tweet.
- Combining above two relations, I have formed a diffusion network graph and performed network analysis henceforth.

I have used **networkx** library in order to build graph and draw graph from the in-degree and out-degree of the graph.

```
def get_directedGraph(df):
    dg = nx.DiGraph()
    for index, r in df.iterrows():
        sn=r['screenName']
        rn=r['retweetScreenNames']
        mn=r['mentionScreenNames']
        if not mn and not rn:
            dg.add_node(r['screenName'])
        else:
            for i in mn:
                dg.add_edge(sn, i)
            if rn!='' and rn != sn:
                dg.add_edge(rn, sn)
    return dg
```

```
def degree_histogram_directed(G, in_degree=False, out_degree=False):
    nodes = G.nodes()
    if in_degree:
        in_degree = dict(G.in_degree())
        degseq=[in_degree.get(k,0) for k in nodes]
    elif out_degree:
        out_degree = dict(G.out_degree())
        degseq=[out_degree.get(k,0) for k in nodes]
    else:
        degseq=[v for k, v in G.degree()]
    dmax=max(degseq)+1
    freq= [0 for d in range(dmax)]
    for d in degseq:
        freq[d] += 1
    return freq
```

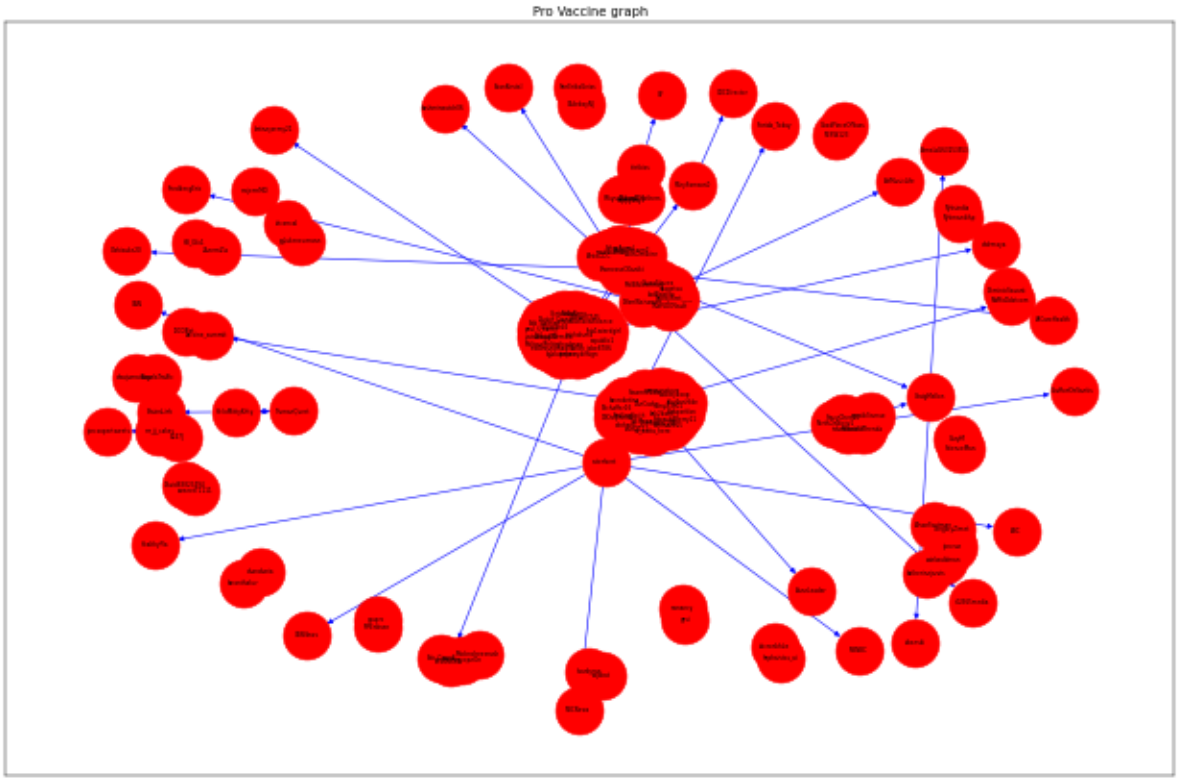
4. Drawing Graphs

Using networkx library's draw_networkx API, I have obtained directed graphs for ProVaccine and AntiVaccine as follows :

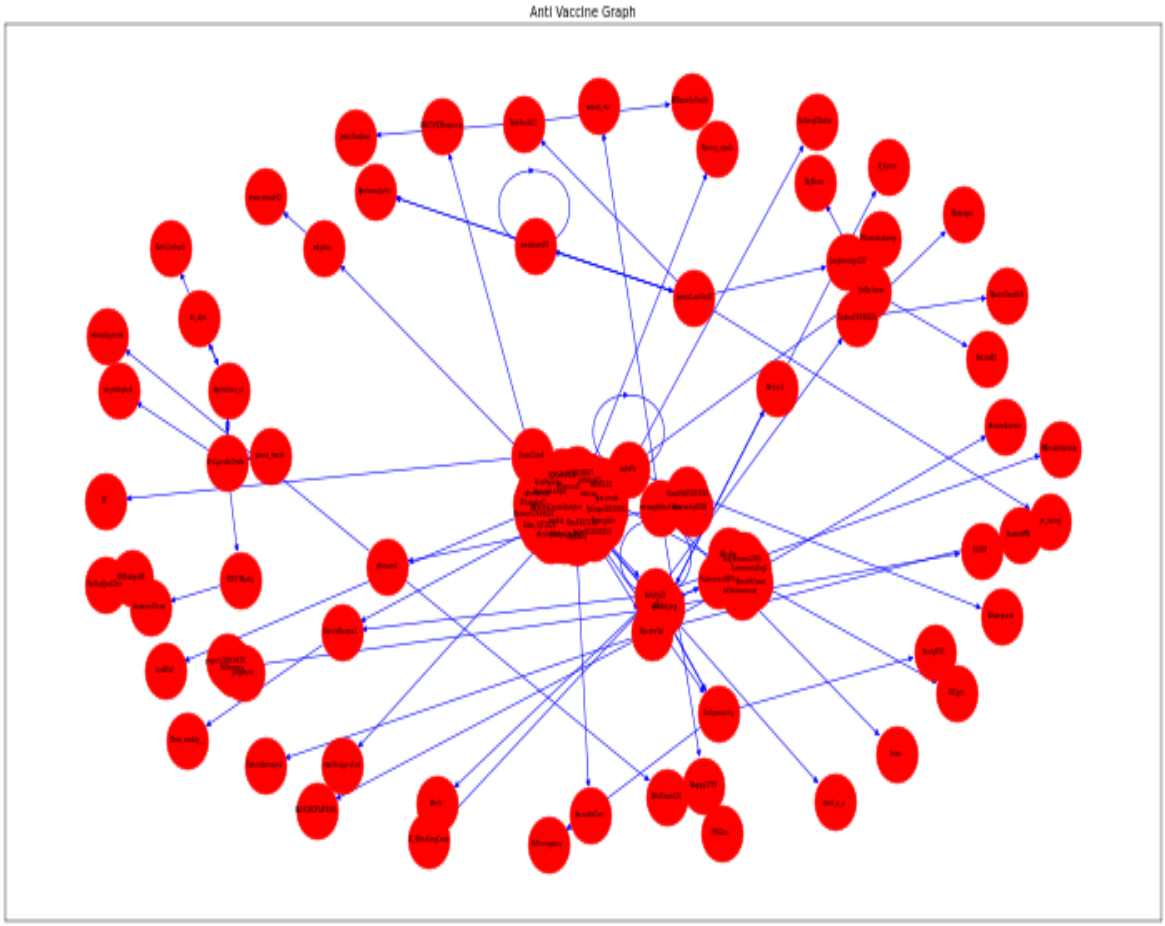
```
graphs = []
for i in range(len(pro_anti_queries)):
    df = tweet_sraping(pro_anti_queries[i], date_since, numTweets, api)
    df.to_csv(files[i]+".csv")
    graphs.append(get_directedGraph(df))
fig = plt.figure(1, figsize=(100, 80), dpi=40)
nx.draw_networkx(graphs[0], node_size=2000, node_color='r', edge_color='b', width=1, font_size=5)
plt.title('Pro Vaccine graph')
plt.show()

fig = plt.figure(2, figsize=(100, 80), dpi=40)
nx.draw_networkx(graphs[1], node_size=2000, node_color='r', edge_color='b', width=1, font_size=5)
plt.title('Anti Vaccine Graph')
plt.show()
```

Pro Vaccine Graph

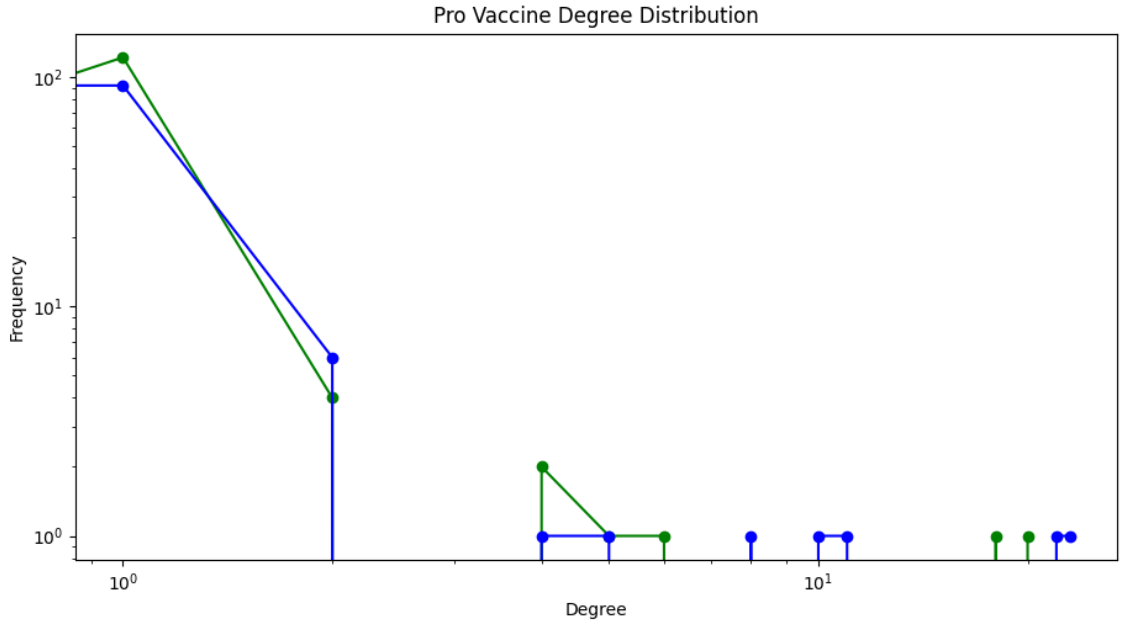


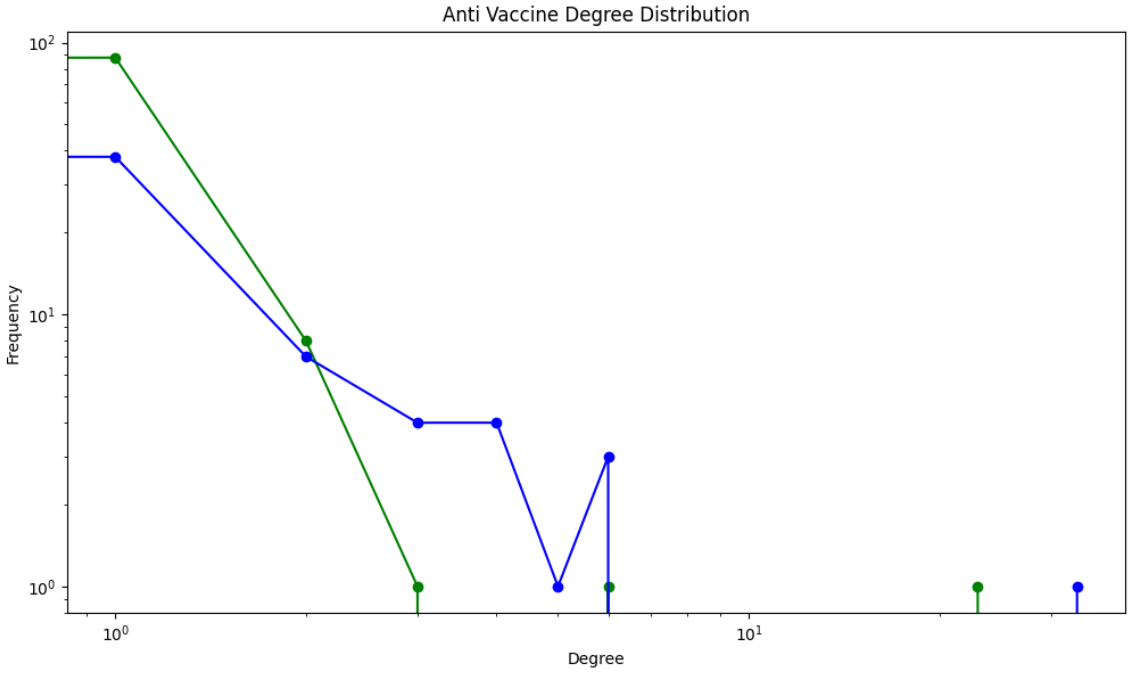
Anti Vaccine Graph



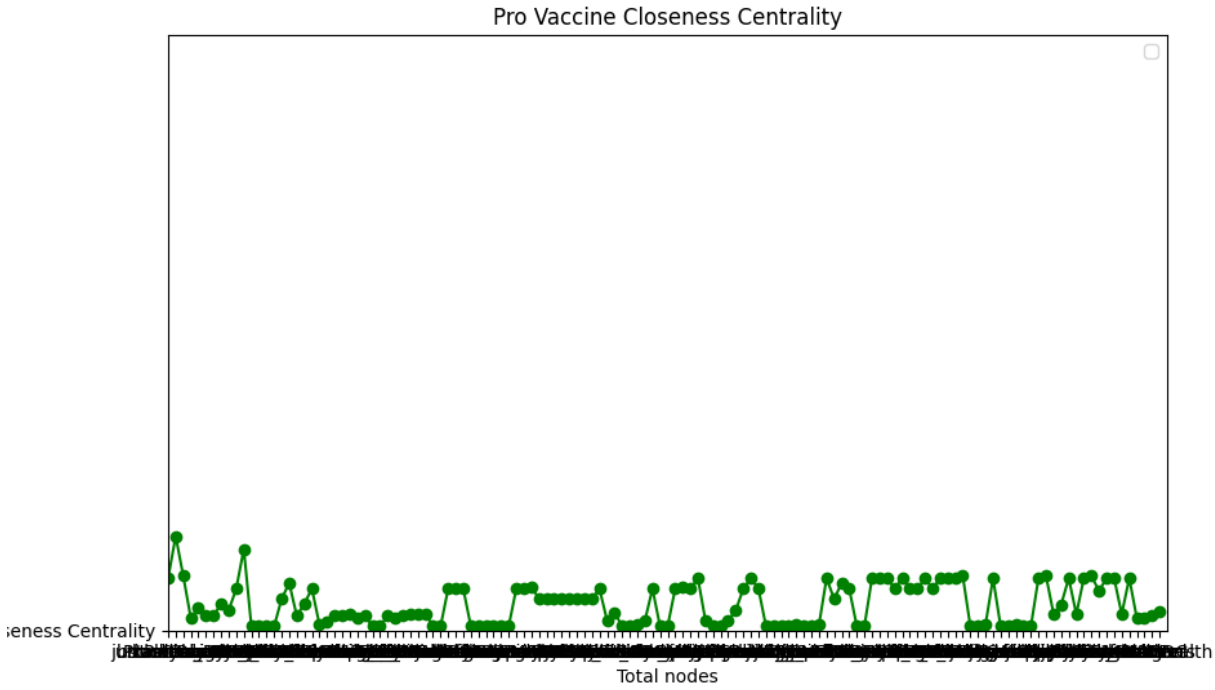
5. Network Measures from the graphs

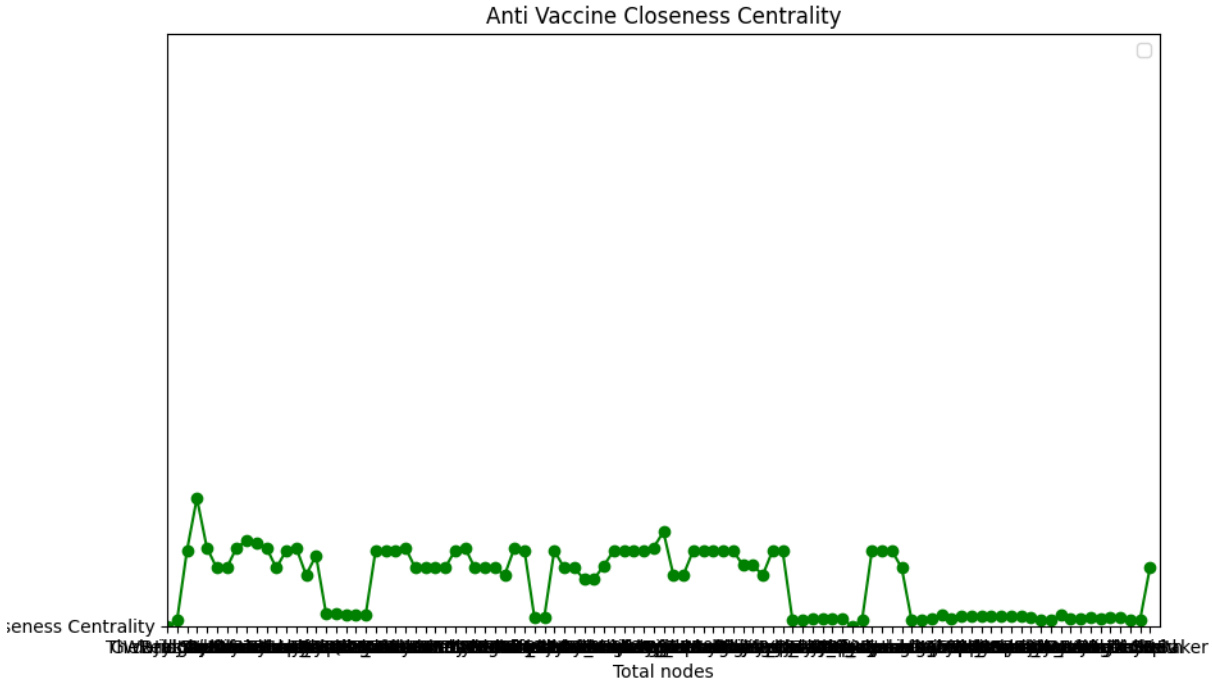
5.1 Degree Distribution



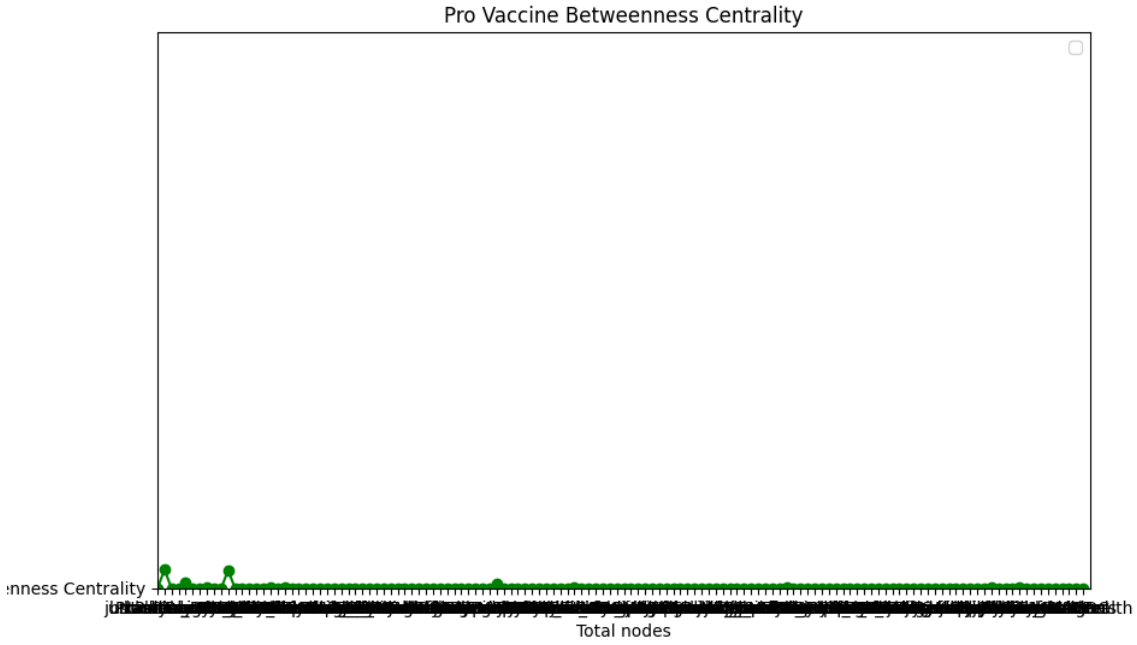


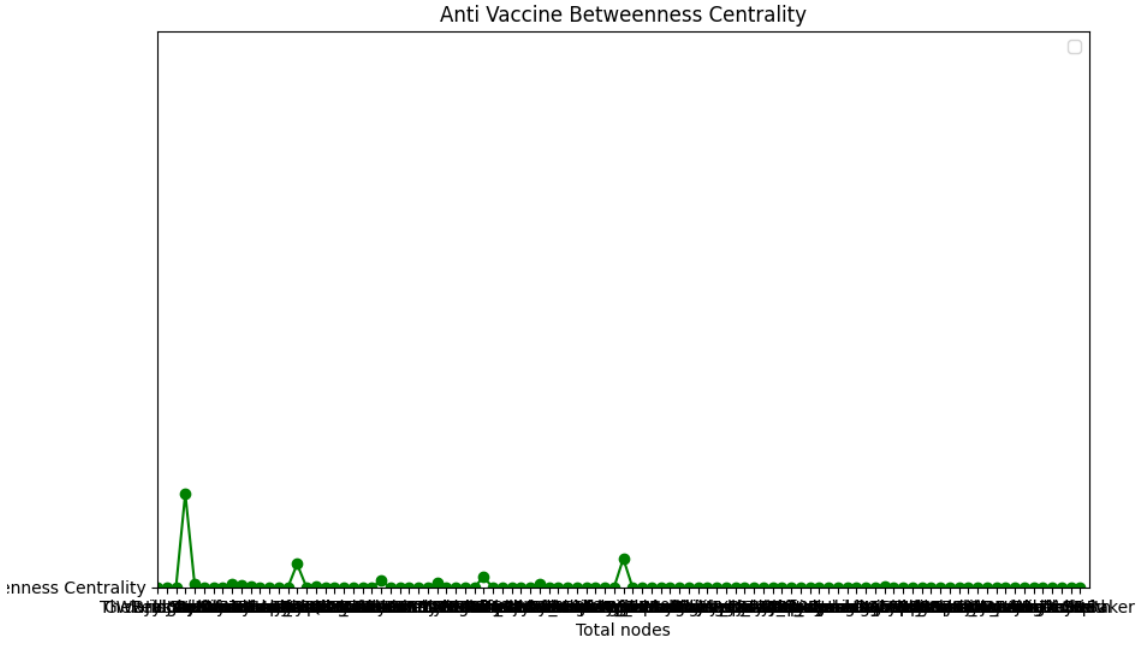
5.2 Closeness Centrality



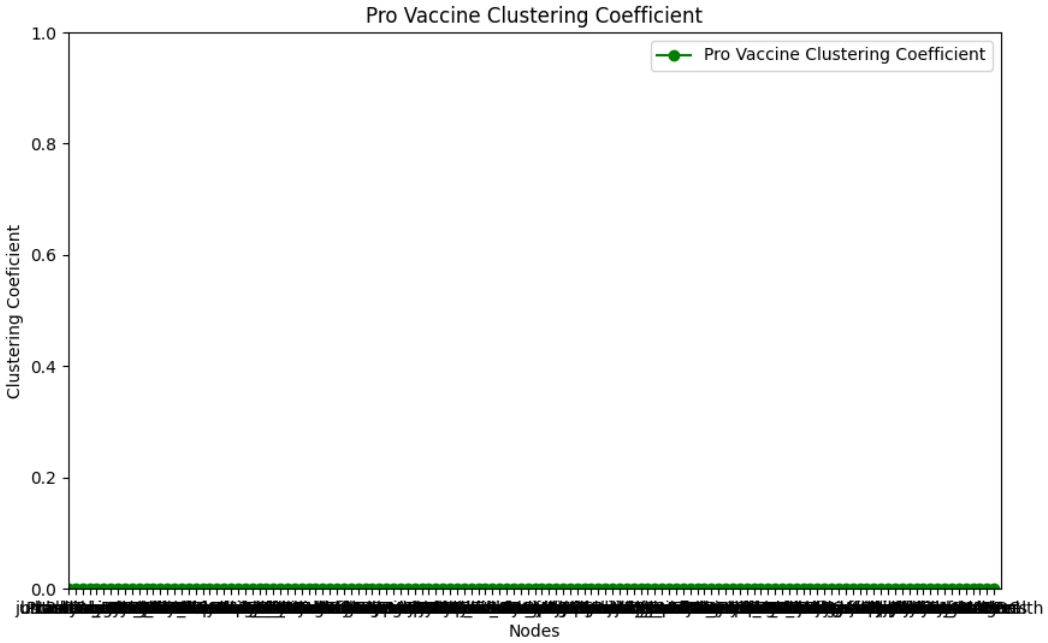


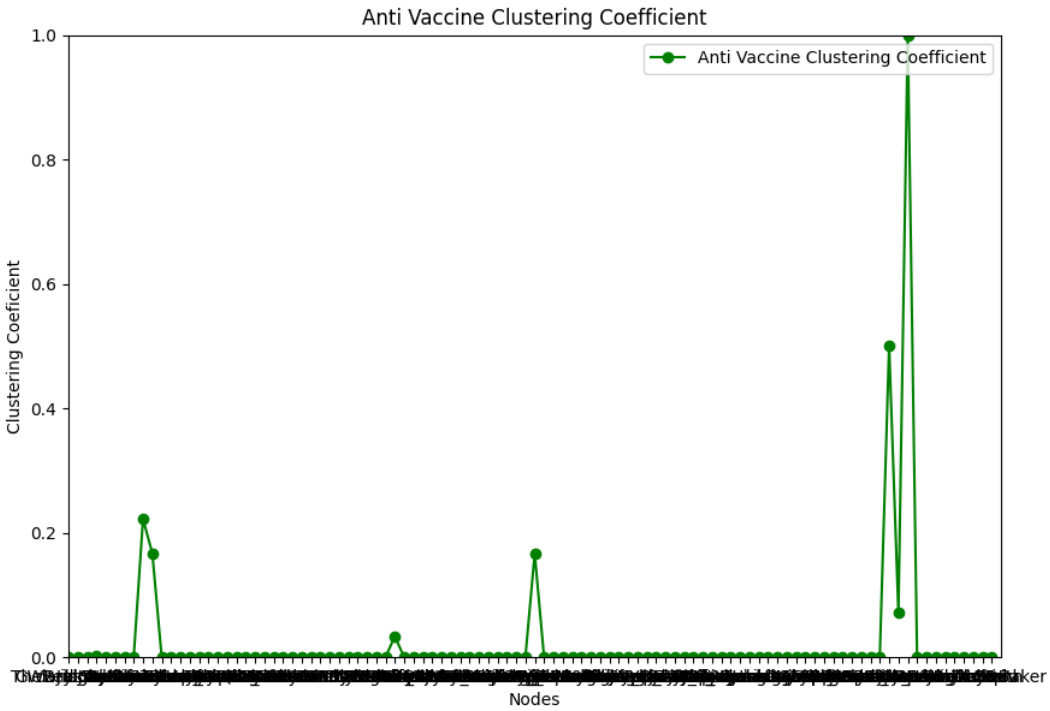
5.3 Betweenness Centrality





5.4 Clustering Coefficient





Conclusion

Through the above Social Media Data Analysis, it can be inferred that there are single source points for the ProVaccine attitude as seen in the graph whereas for AntiVaccine attitude, there are multiple origin points. Hence, a basic idea that negative reactions spread more quickly and people are more prone to express critics over appreciation can be analysed from the above two graphs and their behaviors.

References

- https://stackoverflow.com/questions/53958700/plotting-the-degree-distribution-of-a-graph-using-nx-degree-histogram
- https://docs.tweepy.org/en/stable/api.html
- https://www.geeksforgeeks.org/network-centrality-measures-in-a-graph-using-networkx-python/
- https://www.geeksforgeeks.org/extracting-tweets-containing-a-particular-hashtag-using-python/
- https://networkx.org/documentation/networkx-2.3/auto_examples/drawing/plot_degree_histogram.html