CSE 472 – Social Media Mining

Project I Report

Social Media Data Analysis

Haritha Athinarayanan – 1225396744 - hathinar@asu.edu

Abstract— The objective of this project is to understand social media data crawling and to perform exploratory analysis by processing the crawled data. For this project, tweets are crawled about a specific topic covid-19 vaccine, and these extracted data are then stored and retrieved in JSON format.

I. INTRODUCTION

The rise of social media in the last two decades has ignited tremendous growth and impact in understanding customer data and providing personalized content and targeted advertisements for consumers. As social media data are real-time data, we can extract the information on brand sentiments, demographic data, behavior, and share of voice of individuals.

Among the popular social media sites, Twitter can fetch insights into current social media trends, sentiments, and opinions with retweets and mentions. Twitter provides API to extract tweets based on a specific search query, username, etc., Twitter data can also be scrapped with tools like snscrape, Tweepy, Twint, etc.,

II. DESCRIPTION OF SOLUTION

Tweets from Twitter can be scraped using a scraping tool like **snscrape. snscrape** is a scraper for social networking services (SNS). I adopted the Non-Api method for crawling tweet information from Twitter. I chose two search texts for crawling data about two opinions on the COVID-19 vaccine. The search texts in my case are **'GetVaccinatedOrGetCovid'** to fetch tweets with Pro vaccine sentiment and '**SayNoToVaccines'** to pull out tweets with Anti-vaccine sentiment. Using **TwitterSearchScraper** details of the tweets are converted into a data frame and then to JSON. The required attributes for forming the networks of these two sentiments are appended and saved as two JSON files. Here the feature needed for building edges is the content of the tweets. Therefore, the content attribute of collected tweet data is stored in the JSON file.

Word Co-occurrence network, is a network formed by connecting pairs of words within a specified text. For Graphic visualization of gathered data, the content of two JSON files is read and the tweets are split as a list of words. Then stopwords in the list are filtered out, to give a valid co-occurrence words list in the tweets. Co-occurrence is identified using bigrams from nltk. The most redundant words are determined to form the Word Co-occurrence network.

NetworkX is a python package used for generating and visualizing graphs. Two directional graphs with edges connecting from the first co-occurring word pointing to the second one had been drawn with networkx and matplotlib.

Network Performance of social networks can be analyzed using the following concepts: Closeness, centrality, betweenness, and centralization. Calculated the Degree distribution of the two directional graphs and the histogram for the same is plotted. Additionally, the clustering coefficient, closeness centrality, and betweenness centrality for the two graphs are computed.

III. RESULTS

The networks generated for two different emotions are plotted. Each node consists of a word that is connected to the co-occurred word with directed edges. Fig 1 shows the graph for Anti-vaccine opinionated tweets. Likewise, Fig 2 shows the Pro vaccine word co-occurrence network.

Fig 1: Anti_Vaccine.png

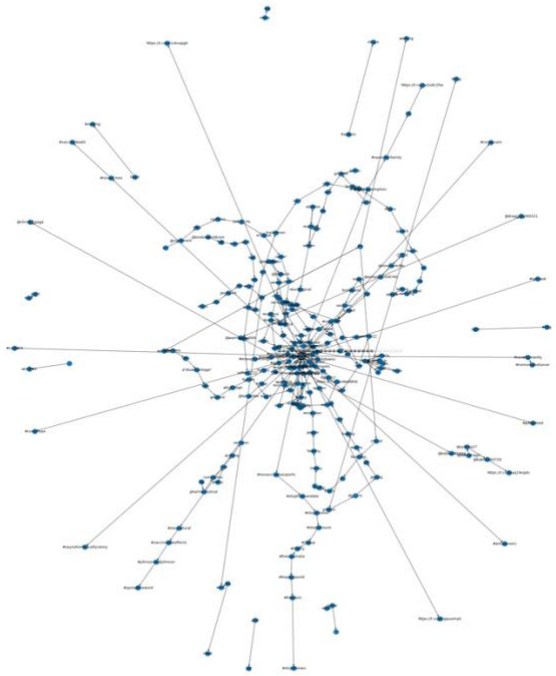
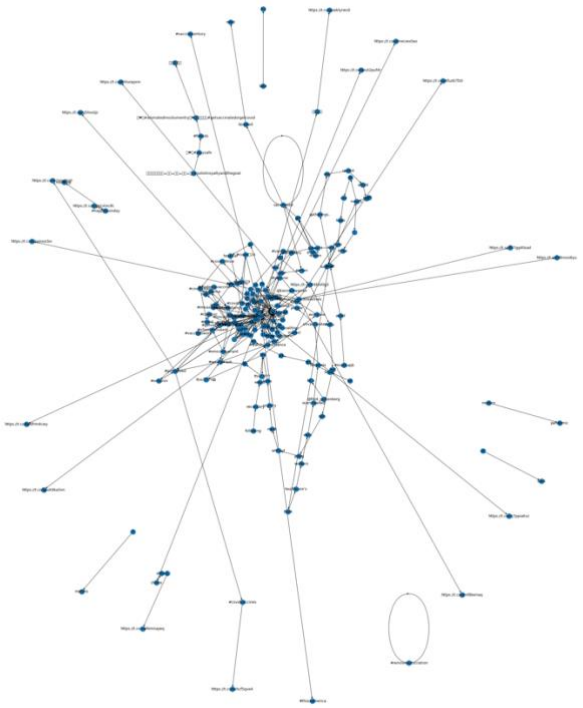


Fig 2: Pro_Vaccine.png



Degree distribution for Anti Vaccine and Pro Vaccine sentiment is assessed and the data is plotted as a histogram.

Fig 3: Anti_Vaccine_Histogram.png

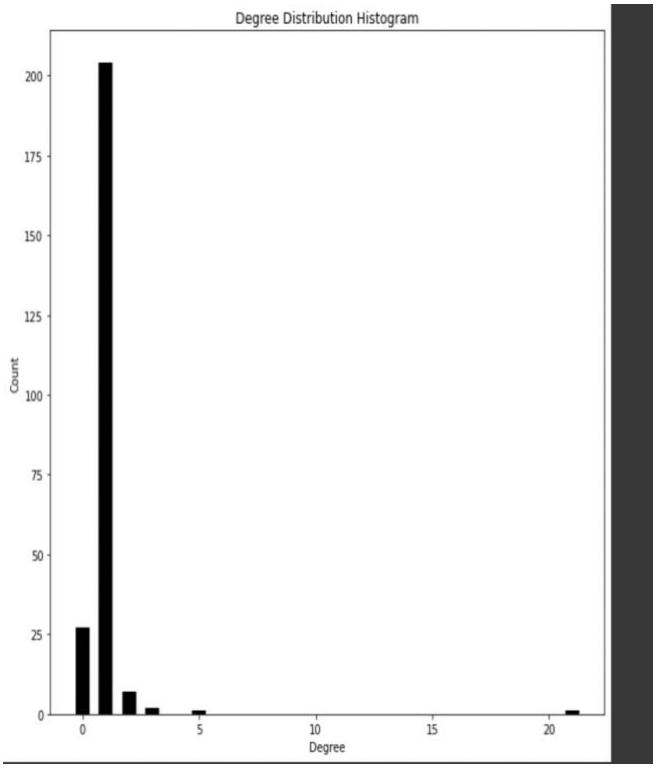
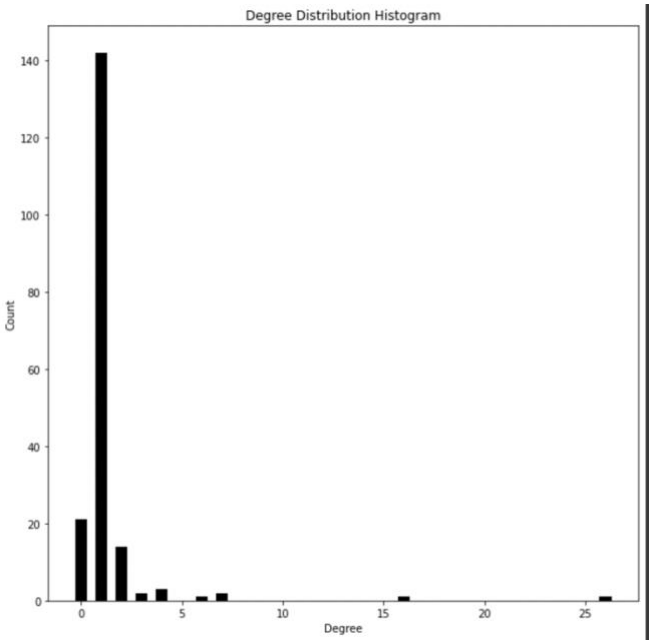


Fig 4: Pro_Vaccine_Histogram.png



Clustering coefficient, closeness and betweenness are computed for the two directional graphs and the result is attached. Among closeness centrality of the nodes, #saynotovaccines has the highest value.

IV. References

[1] https://medium.com/dataseries/how-to-scrape-millions-of-tweets-using-snscrape-195ee3594721

[2] https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/calculate-tweet-word-bigrams/

[3] https://networkx.org/documentation/stable/reference/introduction.html

[4] https://www.geeksforgeeks.org/removing-stop-words-nltk-python/

https://docs.python.org/3/library/itertools.html#itertools.chain

Closeness centrality:

```
Closeness                    centrality
{'#saynotovaccines':
0.09722950682304163,
'#novaccinepassports':
0.08677813839475235,
'#stopthemandate':
0.07848300322838234,    '#stoptheshot':
0.07173950588308232,    '#stopcensure':
0.06614976123411172,         '#riseup':
0.061441208494631035,       '#liberty':
0.0574208703405233,   '#freeaustralia':
0.05394828256560286,   '#freetheworld':
0.050918769499024776,        '#freedom':
0.048252728477283034,    '#stopthenwo':
0.04588854682550306,   '#saynotomasks':
0.011295527893038266,           'time':
0.00943040362127499,             'say':
0.008644536652835409,
'@innocentrifle': 0.07848300322838234,
'@anitasi41058740':
0.07173950588308232,           'world':
0.014522821576763486,  'indeed':  0.0,
'masks,':         0.007979572294924991,
'lockdowns':      0.007409602845287492,
'vaccines,':      0.006915629322268326,
'take':  0.012528313989616534,  'back':
0.012038537567004435,     'rights!!!!':
0.011634216406000639,
'#saynotolockdowns':
0.0110082856656675,
'#saynotomedicaltyranny':
0.08677813839475235,
'@warrior4ssr009':
0.06614976123411172,   '#savechildren':
0.08591926676960122,      '-':     0.0,
'windows':        0.004149377593360996,
'year':           0.0,           'old':
0.004149377593360996,           'free':
0.021339656194427976,     'australia,':
0.012448132780082987,            'uk,':
0.012448132780082987,         'mecca,':
0.012448132780082987,          'south':
0.013579781214635985,        'africa,':
0.009958506224066390,       'america,':
0.012448132780082987,            'tik':
0.011295527893038266,            'tok':
0.0110082856656675,             '😂🤣':
0.010762074449639471,
'pharmaceutical':
0.004149377593360996,      'companies':
0.0055325034578146606,       'pushing':
0.008644536652835409,       'vaccines':
0.00829875518672199,            'like':
0.017103532031170935,  '@christy1gaga':
0.061441208494631035,    'read':   0.0,
'🔻':              0.07030603062325941,
'@dragonlight9321':
0.06614976123411172,    'heart':   0.0,
'breaking':       0.004149377593360996,
'#nomask':         0.08677813839475235,
'#covidscam':      0.08677813839475235,
'every':          0.0,          'day':
0.004149377593360996,
'#mybodymychoice':                0.0,
'#saynotovaccinepassports':       0.0,
'it's':  0.01323103421279261,  'right':
0.0,   'thing':   0.004149377593360996,
'pretty':          0.0,          'sure':
0.009484291641967 99, 'architect': 0.0,
"sweden's": 0.004149377593360996, 'no-
lockdown':       0.0055325034578146606,
'covid-19':       0.011715889675372222,
'response':        0.01095695020746888,
'said':           0.010426641132035324,
'approach':       0.010038816758131442,
'basically':      0.009745327191746927,
'correct':        0.009517223210169267,
'https://t.co/vzv9ed18h8':
0.009336099585062 24,             'via':
0.009189725835603187,
'@businessinsider':
0.009069677799696165,  'freedom':  0.0,
'choice':         0.004149377593360996,
'talking':        0.0,         'people':
0.008839978351073427,           'dying':
0.00872532671501144,       'universal':
0.008786917256529168,          'deceit':
0.008471645919778701,                '—':
0.008298755186721992,        'telling':
0.008197550855176601,           'truth':
0.008136034496786265,  'revolutionary':
0.008097978851559363,             'act':
0.008074464505999775,           'work.':
0.008132780082987554,            'did,':
0.008047277756821326,            'need':
0.008002371072910491, 'vaccine.': 0.0,
'#crimesagainsthumanity':          0.0,
'#antivaxxers':    0.08677813839475235,
'working':        0.012138477735802317,
'#covid19':        0.08677813839475235,
'don't':          0.0,          'trust':
0.004149377593360996,   'sarah':   0.0,
'barnsey':        0.004149377593360996,
'johnson':         0.009336099585062 24,
'&amp;': 0.006224066390041493, 'april':
0.008298755186721992,            '9th':
0.007979572294924991,           '2021':
0.007861978597947149,         'severe':
0.011589640864215196,        'adverse':
0.009336099585062 24,       'reaction:':
0.007816269420052108,   'neurological':
0.010639429726566655,         'damage,':
0.01004149377593361,        'paralysis':
0.009637264087806184,             'age':
0.009349930843706777,              '33':
0.009137955149424216,   '#staynatural':
0.08677813839475235,
```

```
'#vaccinesideeffects':
0.07848300322838234,
'#johnsonandjohnson':
0.07173950588308232, '#spreadtheword':
0.06614976123411172,   'never':   0.0,
'potus':   0.004149377593360996,   '!':
0.0055325034578146606,   '@larryelder':
0.0,   'trump':   0.004149377593360996,
'elder':  0.0055325034578146606,  '24':
0.006224066390041493,  'really':  0.0,
'sick':          0.004149377593360996,
'tradesmen':     0.0055325034578146606,
'(and':  0.006224066390041493,  'men)':
0.006639004149377594,         'giving':
0.006915629322268326,          '#vaxx':
0.007113218731475992,         'speech':
0.007261410788381743,         'house,':
0.007376671277086215,            'far':
0.007468879668049793,           "i've":
0.007544322897019992,        'several':
0.007607192254495158,       'carrying':
0.007660389403127992,         'jabbed':
0.00872532671501144,       "etc...i'm":
0.00863610295853996,          'paying':
0.00856550088915234,            'job,':
0.008508850254740269,          'wanna':
0.008462854865555477,        'fucking':
0.008425132169260904,            'hear':
0.008393924397578896,              'ur':
0.008367911479944675,         'opinion':
0.008346086490068315,      'anything!':
0.008327670710020675,     '#covidhoax':
0.08677813839475235,
'#enoughisenough':                  0.0,
'#stoptheshots':  0.004149377593360996,
'#resistcovidvaccination':
0.0055325034578146606,
'https://t.co/40jppuomyb':
0.08677813839475235,          'listen':
0.07710311424816527,
'#vaccinationskill':
0.0642907273107847,     '🙏🤦\u200d♀':
0.08677813839475235,   '@katewolf1444':
0.061441208494631035,    '@axtromagic':
0.0574208703405233,       '@sylviaq17':
0.053948282565602 86,    '@bab69260728':
0.050918769499024776,
'https://t.co/skaq19oqdv':
0.048252728477283034,    'u.s.':   0.0,
'government':     0.004149377593360996,
'data':          0.0055325034578146606,
'confirms':       0.006224066390041493,
'143,233%':       0.006639004149377594,
'increase':       0.006915629322268326,
'cancer':         0.007113218731475992,
'cases':  0.007261410788381743,  'due':
0.007376671277086215,    'vaccination':
0.01095695020746888,     '#novaccines':
0.08677813839475235,   '#vaccinedeath':
0.07848300322838234,
'😂😂😂😂😂😂😂😂😂😂😂😂😂😂😂💥
💥😂😂😂😂😂😂☄☄☄☄☄☄☄
☄':          0.0,          '#vaccines':
0.004149377593360996,  '#unvaccinated':
0.0055325034578146606,     '#unvaxxed':
0.006224066390041493,
'https://t.co/ni1vbcapgh':
0.08677813839475235,   'africa':   0.0,
'rise':          0.004149377593360996,
'poisoned':      0.0055325034578146606,
'clot': 0.006224066390041493, 'shots':
0.006639004149377594,        'western':
0.014786872878159184,     'countries.':
0.013337285121517486,          'calls':
0.01235163562674901,
'a"disadvantage"':  0.0116424283930226,
'turn':          0.011110730084702172,
'greatest':      0.010699466508595138,
'advantage':      0.01037344398340249,
'human':         0.010109809495426513,
'history.':      0.009893097316190749,
'#goafrica':     0.009712481180920208,
'#stayunpoisoned':
0.009560165975103735,  '#savehumanity':
0.08677813839475235,
'#nomoreshotsever':
0.07848300322838234,
'@londonworldcom':    0.0,     'pfft.':
0.004149377593360996,          'polio':
0.0055325034578146606,         "ain't":
0.006224066390041493,           'bad.':
0.006639004149377594,           'kids':
0.006915629322268326,          'today':
0.007113218731475992,          'whiny':
0.007261410788381743,    'softies👶👶':
0.007376671277086215,        "they'll":
0.007468879668049793,           'stop':
0.007544322897019992,         'scared':
0.007607192254495158,        'realise':
0.007660389403127992,            'new':
0.007705986959098993,           'iron':
0.007745504840940526,          'lungs':
0.007780082987551867,            'arm':
0.007810593116914816,          'holes':
0.007837713231904103,          'still':
0.007861978597947149,     'selfies🤳':
0.012038537567004435,          'check':
0.011634216406000639,
'#religiousexemption':
0.08677813839475235,
'#saveyourfamily':
0.07848300322838234,           '🙏😑':
0.07173950588308232,
'https://t.co/2w2u9c1fiw':
0.06614976123411172,  '@des3650':  0.0,
'@jthetech':     0.004149377593360996,
'chick':        0.0055325034578146606,
'donating':      0.006224066390041493,
'single': 0.006639004149377594, 'red':
0.006915629322268326,          'cent.':
```

0.007113218731475992, '#pureblood': 0.007113218731475992, '#pureblood': 0.08677813839475235, 'big': 0.0, 'critical': 0.012138477735802317, 'business.': 0.01138548729885639, 'oh,': 0.010839190447963417, 'that’s': 0.010427566299837633, 'right!!': 0.010108258197360621, 'is.': 0.009854771784232365, 'money': 0.009649715333397665, 'healthy': 0.009481220304001031, 'people.': 0.009340924442728939, 'remember': 0.009222776247428432, 'that.': 0.00912229577777074, '#thinkforyourself': 0.009036101030838406, 'share!!': 0.06996547396254395}

Betweenness centrality

Betweenness centrality
{'#saynotovaccines': 0.1435338865836791, '#novaccinepassports': 0.027697095435684646, '#stopthemandate': 0.024757952973720607, '#stoptheshot': 0.021784232365145227, '#stopcensure': 0.018775933609958505, '#riseup': 0.015733056708160442, '#liberty': 0.012655601659751037, '#freeaustralia': 0.00954356846473029, '#freetheworld': 0.006396957123098202, '#freedom': 0.0032157676348547716, '#stopthenwo': 0.0, '#saynotomasks': 0.02372060857538036, 'time': 0.005791839557399724, 'say': 0.0001556016597510373 3, '@innocentrifle': 0.027852697095435683, '@anitasi41058740': 0.024896265560165973, 'world': 0.005809128630705394, 'indeed': 0.0, 'masks,': 0.0012621023513139696, 'lockdowns': 0.0023686030428769016, 'vaccines,': 0.003475103734439834, 'take': 0.024636929460580912, 'back': 0.0229253112033195, 'rights!!!!': 0.023340248962655602, '#saynotolockdowns': 0.0, '#saynotomedicaltyranny': 0.0, '@warrior4ssr009': 0.018775933609958505, '#savechildren': 0.012033195020746887, '-': 0.0, 'windows': 0.0, 'year': 0.0, 'old': 0.0, 'free': 0.005584370677731674, 'australia,': 0.0, 'uk,': 0.0, 'mecca,': 0.0, 'south': 8.644536652835407e-05, 'africa,': 0.000933609958506224, 'america,': 0.0, 'tik': 0.0026279391424619 64, 'tok': 0.002991009681881051, '😂🤣': 0.0033195020746887966, 'pharmaceutical': 0.0014004149377593361, 'companies': 0.0027662517289073303, 'pushing': 0.006742738589211618, 'vaccines': 0.007987551867219916, 'like': 0.015750345781466113, '@christy1gaga': 0.0, 'read': 0.0, '↓': 0.005376901798063624, '@dragonlight9321': 0.0, 'heart': 0.0, 'breaking': 0.0, '#nomask': 0.0, '#covidscam': 0.0, 'every': 0.0, 'day': 0.0, '#mybodymychoice': 0.0, '#saynotovaccinepassports': 0.0, 'it’s': 0.002921853388658368, 'right': 0.0, 'thing': 0.0, 'pretty': 0.0, 'sure': 0.0054633471645919775, 'architect': 0.0, "sweden's": 0.0010200553250345782, 'no-lockdown': 0.0020055325034578145, 'covid-19': 0.011825726141078838, 'response': 0.0015733056708160443, 'said': 0.002282157676348548, 'approach': 0.0029564315352697094, 'basically': 0.0035961272475795295, 'correct': 0.004201244813278008, 'https://t.co/vzv9edl8h8': 0.004771784232365145, 'via': 0.005307745504840941, '@businessinsider': 0.005809128630705394, 'freedom': 0.0, 'choice': 0.0, 'talking': 0.0, 'people': 0.014280774550484094, 'dying': 0.0, 'universal': 0.0006224066390041493, 'deceit': 0.0014349930843706776, '—': 0.0022130013831258644, 'telling': 0.0029564315352697094, 'truth': 0.003665283540802213, 'revolutionary': 0.004339557399723375, 'act': 0.004979253112033195, 'work.': 0.0024204702627939143, 'did,': 0.0001383125864453665 2, 'need': 0.0, 'vaccine.': 0.0, '#crimesagainsthumanity': 0.0, '#antivaxxers': 0.0, 'working': 0.0, '#covid19': 0.0, 'don’t': 0.0, 'trust': 0.0, 'sarah': 0.0, 'barnsey': 0.0010546334716459197, 'johnson': 0.003094744121715076, '&': 0.0, 'april': 0.004011065006915629, '9th': 0.004927385892116182, '2021': 0.005809128630705394, 'severe': 0.00850622406639004, 'adverse': 0.0001383125864453665 2, 'reaction:': 0.000933609958506224, 'neurological': 0.008990318118948824, 'damage,': 0.009699170124481328, 'paralysis': 0.010373443983402 49, 'age':

```
0.011013139695712309,           '33':
0.011618257261410789,   '#staynatural':
0.009232365145228216,
'#vaccinesideeffects':
0.006189488243430152,
'#johnsonandjohnson':
0.0031120331950207467,
'#spreadtheword': 0.0,  'never':  0.0,
'potus':  1.7289073305670816e-05,  '!':
0.0,   '@larryelder':   0.0,   'trump':
0.0008644536652835408,        'elder':
0.00169432918395574,            '24':
0.0024896265560165973,  'really':  0.0,
'sick':          0.001227524204702628,
'tradesmen':     0.0024204702627939143,
'(and':  0.0035788381742738 59,  'men)':
0.004702627939142462,         'giving':
0.005791839557399724,          '#vaxx':
0.006846473029045643,         'speech':
0.007866528354080221,         'house,':
0.008852005532503458,            'far':
0.009802904564315353,           "i've":
0.010719225449515906,        'several':
0.011600968188105118,       'carrying':
0.012448132780082987,         'jabbed':
0.014782157676348547,      "etc...i'm":
0.015491009681881052,         'paying':
0.01616528354080221,            'job,':
0.016804979253112032,          'wanna':
0.017410096818810512,        'fucking':
0.017980636237897647,           'hear':
0.018516597510373444,             'ur':
0.019017980636237896,        'opinion':
0.01948478561549101,       'anything!':
0.01991701244813278, '#covidhoax': 0.0,
'#enoughisenough':                0.0,
'#stoptheshots':   0.00084716459197787,
'#resistcovidvaccination':
0.0016597510373443983,
'https://t.co/40jppuomyb':        0.0,
'listen':        0.009042185338865836,
'#vaccinationskill':
0.0031293222683264174, '🙏🤦\u200d♀':
0.03077455048409405,   '@katewolf1444':
0.012586445366528354,    '@axtromagic':
0.009491701244813277,     '@sylviaq17':
0.00636237897648686,    '@bab69260728':
0.003198478561549101,
'https://t.co/skaq19oqdv': 0.0, 'u.s.':
0.0,                      'government':
0.001123789764868603,           'data':
0.0022130013831258644,       'confirms':
0.0032676348547717843,       '143,233%':
0.004287690179806362,        'increase':
0.005273167358229598,          'cancer':
0.006224066390041493,           'cases':
0.007140387275242047,             'due':
0.008022130013831258,     'vaccination':
0.010788381742738589,     '#novaccines':
0.003077455048409405,  '#vaccinedeath':
0.0,
'😂😂😂😂😂😂😂😂😂😂😂😂😂😂💥
💥😂😂😂😂😂😂😂☄☄☄☄☄☄☄
☄':          0.0,          '#vaccines':
0.0008644536652835408,
'#unvaccinated':   0.00169432918395574,
'#unvaxxed':     0.0024896265560165973,
'https://t.co/ni1vbcapgh':          0.0,
'africa':          0.0,          'rise':
0.0013312586445366529,       'poisoned':
0.002627939142461964,            'clot':
0.0038900414937759337,           'shots':
0.005117565698478562,         'western':
0.013623789764868603,      'countries.':
0.014367219917012447,            'calls':
0.015076071922544952,
'a"disadvantage"':
0.015750345781466113,            'turn':
0.016390041493775934,        'greatest':
0.01699515905947441,        'advantage':
0.017565698478561 55,            'human':
0.018101659751037343,        'history.':
0.018603042876901798,       '#goafrica':
0.01906984785615491,
'#stayunpoisoned':
0.01950207468879668,    '#savehumanity':
0.0030774550484094 05,
'#nomoreshotsever':                 0.0,
'@londonworldcom':     0.0,      'pfft.':
0.0012966804979253112,           'polio':
0.002558782849239281,            "ain't":
0.0037863070539419087,            'bad.':
0.004979253112033195,             'kids':
0.00613762102351314,             'today':
0.007261410788381743,             'whiny':
0.008350622406639004,    'softies👶👶':
0.009405255878284923,         "they'll":
0.010425311203319501,             'stop':
0.011410788381742738,           'scared':
0.012361687413554633,          'realise':
0.013278008298755186,              'new':
0.014159751037344398,             'iron':
0.015006915629322269,            'lungs':
0.015819502074688796,              'arm':
0.016597510373443983,            'holes':
0.017340940525587 83,            'still':
0.018049792531120 33,       'selfies🤳':
0.0017980636237897648,           'check':
0.002230290456431535,
'#religiousexemption':
0.009232365145228216,
'#saveyourfamily':
0.006189488243430152,            '🙏😑':
0.0031120331950207467,
'https://t.co/2w2u9c1fiw':          0.0,
'@des3650':      0.0,      '@jthetech':
0.0009163208852005532,           'chick':
0.0017980636237897648,       'donating':
0.002645228215767635,          'single':
0.003457814661134163,             'red':
```

```
0.00423582295988935,          'cent.':   0.006535269709543569,          'money':
0.004979253112033195,     '#pureblood':   0.007123098201936376,        'healthy':
0.0,     'big':     0.0,     'critical':   0.007676348547717842,        'people.':
0.003077455048409405,      'business.':   0.008195020746887967,       'remember':
0.003838174273858921,           'oh,':   0.00867911479944675,           'that.':
0.004564315352697096,         'that's':   0.009128630705394191,
0.005255878284923928,        'right!!':   '#thinkforyourself':
0.005912863070539419,           'is.':   0.00954356846473029, 'share!!': 0.0}
```