twitter

Social-Network Analysis of Anti-Vax & Anti-Mask Tweets

Search #trend

Our Team

Sebastian Salazar

Shivangi Soni

Steven Liang

Vivek Saahil

Business Case & Implications

- Social media serving as a medium for spreading conspiracy theories, misinformation, and confusion
- Anti-vaccination movements have been gaining momentum since the approval of COVID vaccines
- Individuals / communities refuse to wear masks:
  - People do not believe in the effectiveness of masks and perceive it as a threat to their rights
  - Belief that the pandemic is a hoax
- Analyze, gain insights / understand of the network of anti-vaccination and anti-masks movements can be helpful in determining a strategy to tackle these movements

Discover

#NetworkBuilding #LocationAnalysis #EvolutionOverTime/Spread #NetworkClustering #Influencers twitter

Hashtags Used*

Anti-Vax Top Hashtag: #IDoNotConsent

Other Hashtags:

#vaccineskill
#VaccinesHarm
#donttakethevaccine
#JUSTSAYNO
#MyBodyMyChoice
#remainingGatesfree
#vaccinehesitancy
#daretothink

Banned Hashtags:

#vaccinescauseautism
#vaccinesarepoison
#vaccinescauseaids

Anti-Mask Top Hashtag: #MasksDontWork

Other Hashtags:

#TakeOffYourMask
#antimaskers
#covidhoaxover
#endthelockdown
#scamdemic

* Based on analysis of ~36,000 tweets between March 27, 2021 – April 3, 2021

Discover

Twitter Scraping & Preprocessing

Tweepy		SNScrape
Fast, efficient, easy to setup / use, well documented	+	Fast, no tweet limits / time outs, no API keys required, able to source older tweets
Can't Source Tweets Older than 1 Week, has limit / timeouts, requires API key / Dev Account	−	Less tweets sourced, harder to setup / use, more preprocessing required post scrape

Preprocessing:

- Cleaning (Tokenization, Text Extraction, etc)
- Manual Excel Multi-Value Row Generation using Power Query
- Geopy Location Data Coordinates Sourcing API: Nominatim

```
df['mentioned_users'][8]
```

```
"[User(username='TheDoctors', displayname='The Doctors', id=1429
5156, description=None, rawDescription=None, descriptionUrls=Non
e, verified=None, created=None, followersCount=None, friendsCoun
t=None, statusesCount=None, favouritesCount=None, listedCount=No
ne, mediaCount=None, location=None, protected=None, linkUrl=Non
e, linkTcourl=None, profileImageUrl=None, profileBannerUrl=Non
e), User(username='IDRoadrunner', displayname='Ravina Kullar', i
d=2570388417, description=None, rawDescription=None, description
Urls=None, verified=None, created=None, followersCount=None, fri
endsCount=None, statusesCount=None, favouritesCount=None, listed
Count=None, mediaCount=None, location=None, protected=None, link
Url=None, linkTcourl=None, profileImageUrl=None, profileBannerUr
l=None), User(username='HighWireTalk', displayname='The HighWir
e', id=851985789072408576, description=None, rawDescription=Non
e, descriptionUrls=None, verified=None, created=None, followersC
ount=None, friendsCount=None, statusesCount=None, favouritesCoun
t=None, listedCount=None, mediaCount=None, location=None, protec
ted=None, linkUrl=None, linkTcourl=None, profileImageUrl=None, p
rofileBannerUrl=None)]"
```

Username	Mentioned Users
KimberlyKozik	[The Doctors, IDRoadrunner, HighWireTalk]

Username	Mentioned Users
KimberlyKozik	The Doctors
KimberlyKozik	IDRoadrunner
KimberlyKozik	HighWireTalk

```
df["Location"].value_counts()

USA                                 327
London                              222
United States                       206
England, United Kingdom             133
Scotland                            124
                                    ...
Where orders take me                  1
PLANET EARTH                          1
People'sRepublicofNovaScotia          1
UK + Worldwide                        1
united States of America              1
Name: Location, Length: 1243, dtype:
```

Geospatial Time Series Analysis

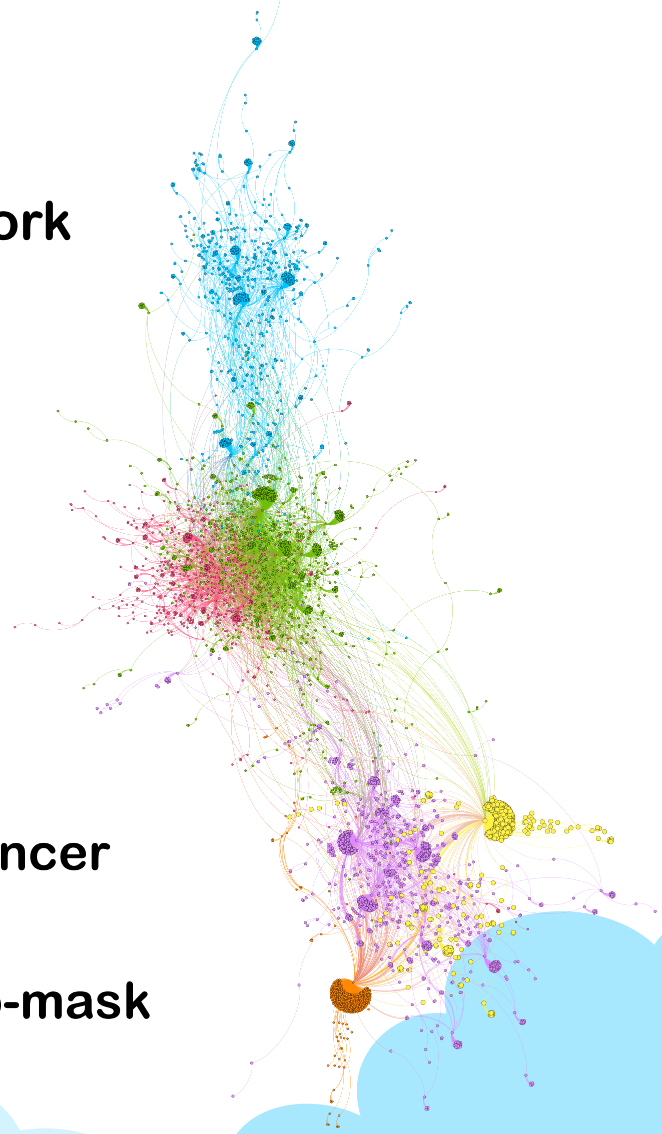
Enable analysis of trend spreads through time

Focus analysis on specific locations

Network Analysis

- # of Nodes = 17, 176 and # of Edges = 33,763
- NetworkX used to measure centrality values and create network graph
  - Network graph moved to Gephi for clustering
- Clustering Analysis
  - Used Gephi's modularity function to perform clustering
  - Around 2,361 clusters formed
  - ~50% of the users in 10 clusters
  - Average modularity of 0.583
  - Average clustering coefficient of 0.061
- Approx. 5% of the network concentrated around an author influencer who had tweeted a satirical post about #antimask
- Another 5% of network concentrated around a doctor whose pro-mask tweet reinforced the #TakeOffYourMask hashtag

*Please refer to the report for detailed explanation and images



Influencers

twitter

Anti – Mask Movement

Screen Name	Score
leew2030	0.08853
jennyrickson	0.07592
HagleyTom1234	0.07174
JujuliaGrace	0.06895
JamesMelville	0.06229
Seahorse1317	0.06167
tarynabell	0.06154
llamedos77	0.05798
northerness	0.05789
annaberu	0.05701

Anti – Vaccination Movement

Screen Name	Score
realDonaldTrump	0.01914
BorisJohnson	0.01272
MattHancock	0.01061
BillGates	0.00916
YouTube	0.009
10DowningStreet	0.00498
WHO	0.00476
simondolan	0.00474
POTUS	0.00469
RealCandaceO	0.00401

- Feature importance done using Random Forest
  - Sum of followers count and listed count treated as target variable
  - Betweenness centrality, degree centrality, and closeness centrality treated as independent features
  - Weights are equivalent to feature importance obtained from Random Forest
  - Score equations were different for both anti-mask tweets and anti-vaccination movement
  - Influencers – people with highest score

The top 10 users include mostly ordinary people, some famous people, some political figures and organizations

Insights

- Mixing of Politics with Science
  - A lot of tweets in England also had additional hashtag of #BorisJohnsonOut
  - Some studies showed that Donald Trump was the major source of misinformation last year
- Anti-vaccination and anti-mask supporters tweet less, but engage more in discussion (more retweets and replies)
- Pro-mask supporters can lead in inflation of the anti-mask hashtags

Recommendations

- Religious institutions, service organizations (national and local), celebrities spreading pro-vaccination and pro-mask sentiments so that the message spreads at a local and a personal level
- Political organizations giving clear guidelines
- Communicating benefits rather than threats
- Pro-mask people should refrain from using anti-mask and anti-vaccination hashtags
- Safe handling practices of mask should be clearly communicated

THANK YOU

Appendix:

Sample Data (After Preprocessing)

Datetime	Tweet Id	Text	Username	followersCount	listedCount	Location	mentioned_users	retweet	longitude	latitude	Index
2020-01-30 0:00	1.22E+18	#vaccines #vexit	mass4mf	968	6	Boston, MA			-71.0582912	42.3602534	0
2020-01-30 0:00	1.22E+18	#vaccinateyourd	Dontmesswmam	13	0	0			Not Found	Not Found	1
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	ceestave		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	PatriotsDontSlp		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	elenochle		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	IPOT1776		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	LisaMei62		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	X22Report		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	SGTreport		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	Cordicon		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	realDonaldTrump		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	okabaeri9111		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	Qanon76		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	jennajameson		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	gehrig38		Not Found	Not Found	2
2020-01-29 0:00	1.22E+18	@ceestave @Pa	zenmonster1	352	1	0	kate_awakening		Not Found	Not Found	2

Network Analysis - Metrics

Betweenness Centrality

Screen Name	Betweenness Centrality
HagleyTom1234	0.006997
VolvoMan17	0.003585
screenname87	0.002926
BillEHawthorne	0.002165
FunTimeFred1	0.00215
lorrain00414525	0.001993
MaizyDaizyZzzz	0.001944
MThebrexitparty	0.001818
GenuineBenny	0.00176
offshorebella	0.001619

Closeness Centrality

Screen Name	Closeness Centrality
leew2030	0.125711
jennyrickson	0.121001
WHO	0.111725
HagleyTom1234	0.109179
BorisJohnson	0.107343
northerness	0.106486
JamesMelville	0.105147
Seahorse1317	0.104192
MattHancock	0.101703
JujuliaGrace	0.100231

Degree Centrality

Screen Name	Degree Centrality
leew2030	0.080216
JujuliaGrace	0.05887
carolJhedges	0.052626
HagleyTom1234	0.049674
jennyrickson	0.048198
tarynabell	0.038831
JamesMelville	0.030372
Seahorse1317	0.029918
annaberu	0.02918
llamedos77	0.027704

The top 10 users include mostly ordinary people, some famous people, some political figures and organizations

Score Equations

The following score equations were used to find the influencers. The coefficients of these variables are the feature importance coefficients obtained from Random Forest

Anti-mask tweets:

Score = 0.18 betweenness centrality + 0.50* closeness centrality + 0.32* degree centrality*

Anti-vaccination tweets

Score = 0.14 betweenness centrality + 0.75* closeness centrality + 0.11* degree centrality*