Anti-Mask Twitter Network Analysis for the COVID Pandemic Era

Sebastian Salazar
Shivangi Soni
Steven Liang
Vivek Saahil

April 13th, 2021

Master of Management in Analytics
McGill University

Contents

1 Introduction and Problem Statement

COVID-19 emerged in China in December 2019 and has rapidly spread worldwide since then. To minimize its spread, various countries include Canada, UK, USA made it mandatory for people in some provinces or cities to wear masks. Additionally, to stop the transmission, various companies have developed vaccines. However, both measures that have been introduced to help people are actually taken in negative light by some people/groups. Social Media Platforms have been known as a medium that can be used to propagate misinformation and fake news. During the pandemic especially, there have been trending news about the Anti-Vax and Anti-Mask movements gaining momentum with the use social media to spread their conspiracy theories, confusion and influence the opinion of others to their cause. Due to their rhetoric, there's been a crackdown on major hashtags on platforms like Twitter and Instagram such as "**#vaccinescauseautism**", "**#vaccinesarepoison**" and "**#vaccinescauseaids**", for good reason.

Often, it is the vocal small minority groups that have a disproportionate impact on public opinions with the use of social media and using these platforms, they can widely and effectively disseminate a message that leads to real consequences impacting the wider society. At the end of the day, containing a pandemic has to do with people's behaviour and their willingness to cooperate such as staying home, adhering to quarantine, to not travel or take vacations. With messages that sway society to refuse vaccines, not wear masks, to gather en mass or general disregard for safety measures, we can see how these movements can sway legislation, and impede government implementation of efforts to contain the pandemic. Hence, it can be very useful to analyze these movements through a social media network perspective to gain insight and look at potential data driven actions to address these problems.

For the scope of this project, the tweets with Anti-Vax sentiments and Anti-Mask data were extracted. For Anti-Vax, the hashtag **#IDoNotConsent** was used and similarly, the hashtag **#MasksDontWork** was used for the Anti-Mask dataset. Then those tweets were used to construct and build a network by using NetworkX on Python and Gephi was used for visualization of the network. The location data was also sourced and analysis was conducted to determine where the hubs are, in addition to looking at changes over time from the beginning of the pandemic until present (i.e. Where are these messages the most prevalent and how did it spread to other locations?). Additionally, network clustering was also performed on Gephi along with identification of influencers to gather meaningful insights. These insights were then used to suggest recommendations to political leaders and various health organizations to address the Anti-Vax and Anti-Mask movements.

2 Data

2.1 Data Source

In terms of data sourcing, tweets were scraped from Twitter, using both Tweepy and snscrape. In its final state, the data looked as follows:

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

Figure 1: Data Sample

Tweepy: was fast, efficient, easy to use and setup. Given previous experience with the tool, the API developer account was already setup. The data sourced from Tweepy was of satisfactory quality and did not require extensive pre-processing.A large data set was sourced to start off the network building and analysis on anti-mask tweets and it was noticed that the tweets sourced, though of good quality, spanned only a short period of 7 days. One of the major limitations of Tweepy, which surfaced, was the inability to source tweets older than a week old. Some other limitations were the tweet cap and timeouts which was an inconvenience. In the interest of time, the network analysis was performed on past week's scraped data from Tweepy.

snscrape: Eventually, a new scraper called **snscrape** was discovered from GitHub, which can potentially source tweets older than a week. It does not require an API key or a developer account and with this new scraper, the older tweets were sourced. Though snscrape is able to source older tweets, it is difficult to setup and not as intuitive as Tweepy. For instance, it requires git and installation using a reference repository to run the scraper. In addition, snscrape also does not work on notebook, but only on Spyder and Python 3.8 environment. The scrape was also fairly complex in terms of modifying the code to suit the format of data required for analysis.

Finally, around 36,000 tweets were scraped using Tweepy to build social network and find influencers for the period between March 29, 2021- April 3, 2021 for anti-mask movement. The data obtained using snscrape did not have any retweeted tweets and included only tweets (including mentions), and thus was only used to conduct time-series analysis for anti-mask movement from April 2020 to March 2021.

2.2 Data Pre-Processing

Target/User Mentions: The target/mention user data, which was used to build the network, was retrieved in a non-traditional format, making data extraction difficult. Sourcing the hashtags also required additional efforts as it was only included in the body of the tweet. Standard NLP techniques such as tokenization were used and special search filter functions were applied to the column to extract the user information, followed by additional data pre-processing and cleaning steps to extract the user target/mentions data (Refer code for more details).

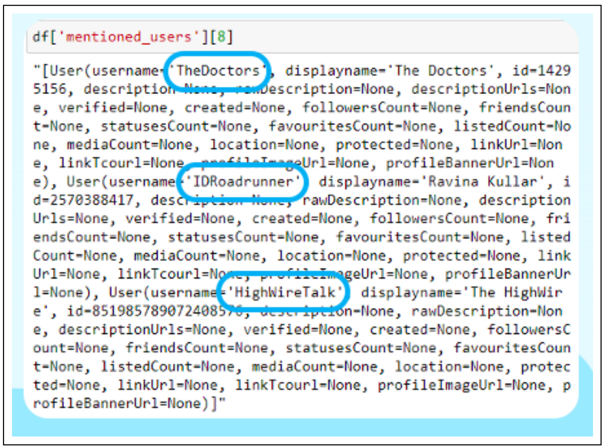


Figure 2: Raw Users Target/Mentions from snscrape

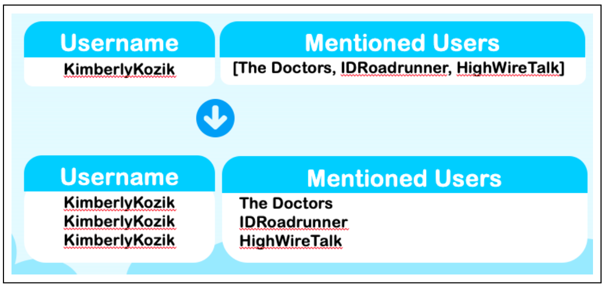


Figure 3: Multi-Valued Attribute Normalization Example

Excel Power Query: After sourcing the data, Excel Power Query was leveraged to generate rows that were not multi-valued rows from a list of potential target/mentioned users. To construct the network, one to one relationships are usually used, whereas the data sourced contained one user to multiple relation with other users. Thus, it needed to be simplified to have 1 row per target/mentioned user. This was a general step done for both the anti-vax and anti-mask datasets after scraping and website [1] was used as a reference to create these new rows.

Geocoding Location Data: The location analysis was conducted to determine where did the tweets originate from and how these locations changed overtime during the pandemic. The location data on twitter however is not refined as users can write arbitrary locations (such as Earth, Middle of Desert etc.) in the location field. The field was imprecise, unstructured and a majority of times didn't make sense or refer to a real location. A solution was found using a package called Geopy, that allows us to feed location data into a 3rd party API and returns geo-coordinates. The 3rd party API is robust enough to handle unstructured and imprecise data pretty well and with this code, clean location data was generated. This data was geocoded and along with the longitude and latitude, visualizations needed for the analysis were generated with Plotly.

3 Methodology

3.1 Detecting trend spreading through geographic time-series visualization

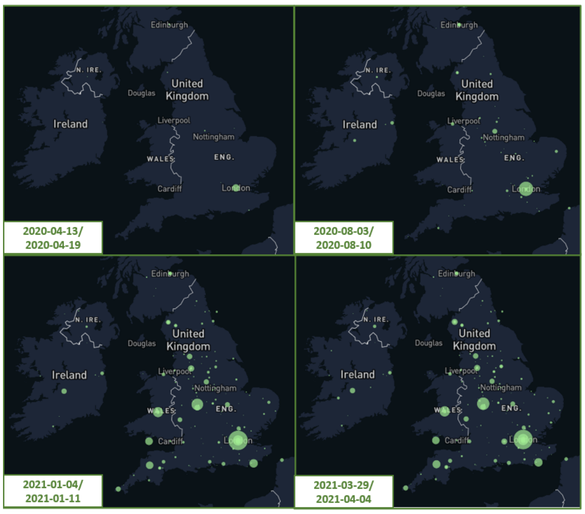
As previously stated, the goal of our project is to correctly assess the spread of Anti-vaxx or Anti-mask sentiment so government institutions can counter any fake news could result in huge COVID spikes. However, assessing public sentiment is an extremely complicated and costly task for governmental institutions as polling, focus groups, or other opinion studies must be done specially considering COVID restrictions. Given this, social media is one of the most efficient way to assess public sentiment and spread information, and through text analytics enable real time updatable sentiment for certain topics like the ones that interest our project. Hence, with this information we developed a Plotly Scatter Mapbox dashboard to track Anti-vax or Anti-mask sentiment around the world.

To enable this type of dashboards, our pre-processed data must be pre-processed again following these steps. First, tweets of a certain topic are grouped by creation date and location creating summary table the count the number of tweets on each month per topic. Once this is done, an additional column is created that tracks the cumulative count of tweets per topic as dates progress. This enables the graph to track which places have the highest number of tweets through time (January 2020-April 2021), which is a critical insight for policy makers. For this project two dashboards were develop to enable different insights.

3.1.1 Dashboard types

The first dashboard is a topic-specific weekly dashboard. The objective of this dashboard is to enable the user to track how a specific trend grows on a weekly basis. Plotly geographs have multiple functionalities that allow the users to get very specific information per location (city). Figure 4 demos how the dashboard displays cumulative values in London for the anti-mask trend, where the size of each circle is representative of the frequency of the tweets.

The second dashboard is a multi-topic monthly dashboard. The objective of this dashboard is to enable the user to track and compare different trends per location. Data timeline is monthly due to processing limitations, but this could potential be applied in a weekly-basis as the topic-specific dashboard. Figure 5 demos how the dashboard displays cumulative values in United States for the anti-vax and anti-mask trends.
It can be seen from both Figure 4 and 5 that the trends for both movements have been growing. This makes sense considering more and more people have started getting frustrated over masks, leading to the widespread of the hashtag. Additionally, in terms of anti-vaccination movement, it makes sense for the hashtag to gain more momentum since November 2020 since that was when the first COVID-19 vaccines results were in news.



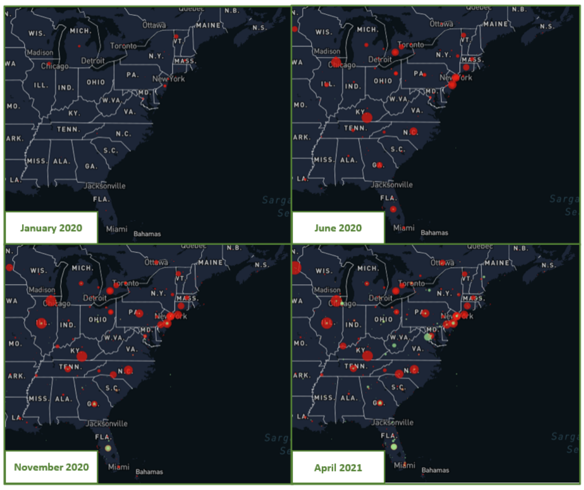


Figure 4: Anti-mask trend spread in UK through time

Figure 5: Multi-trend spread through time in US East Coast

3.1.2 Potential use of Dashboard

The provided dashboards have a business intelligence functionality to enable tracking capabilities for policymakers as, by knowing where spreads are growing rapidly, can lead them to take action and prevent any health-threatening situations. However, analyzing trend's spread may not be enough to stop the spread of anti-mask or anti-vax sentiment so the dashboards must be combined with other analysis. This project proposes using these dashboards to:

- Assess spread of a particular topic/theme through time
- Compare the growth of trends to prioritize them based on health impact for society and spreading rate
- Focus on geographical areas to downsize network analysis by geographical clusters

3.2 Determining influencers based on tweets

There are certain people on social media, generally referred to as influencers, who have established credibility in a specific industry [2]. These people usually have a large network, which makes it easier for them to influence others by virtue of their connections and reach. Since SNScraper was not able to retrieve any information on the retweeted tweets and low quality of data (lot of pre-processing required) and the data quality for , the tweets that were scraped using Tweepy for anti-mask hashtag during the time span of March 29-April 3, 2021 were used to build network. This was done so that the influencers from these networks can be determined and the tweets of those influencers can be analyzed in depth to determine specific recommendations related to anti-mask moment. In this case, the sum of followers count (how many people are following the user) and listed count (how many people have added you to a list that they follow) was taken as the proxy for the social influence. So the target variable was the sum of these metrics and the independent features included between centrality, degree centrality and closeness centrality. Random Forest was used to determine the feature importance, which was then used as weight for each of the independent features. The equation below was used to calculate the *social influence score* for each user and the top ten users with the highest scores. The results can be seen in Figure 1 in Appendix.

$$Score = 0.18 * \text{ betweenness centrality} + 0.50 * \text{ closeness centrality} + 0.32 * \text{ degree centrality} \tag{1}$$

Since the anti-vaccination tweets have been getting removed by twitter, anti-vax hashtag related tweets were hard to find. Hence, for analysis of anti-vaccination tweets, tweets obtained by using SNScarper were used. Similar procedure was followed to determine the feature importance of the centrality values to find scores of the influencers. Since the data was different, the weights obtained for features were different as well. The equation used to find score is shown below and the influencers for anti-vax movement are shown in Figure 2 in Appendix.

$$Score = 0.14 * \text{ betweenness centrality} + 0.75 * \text{ closeness centrality} + 0.11 * \text{ degree centrality} \tag{2}$$

3.3 Clustering using Gephi

Please refer to the appendix for individual cluster images.

Cluster (Modularity Partition)	%age of users	Top influencers	Cluster Attributes	Tweets Observation	Major Location	Recommendation
1	9%	1. andrewspoooner - Actor/Director 2. JolyOnRubs - Actor/Director 3. mrnickharvey - Former MP/Politician 4. happyjack1981 - local influencer 5. llamedos77 - Local Influencer	UK Famous People (Actors/Politicians/Local Influencers)	Tweeted against growing popularity of antimask hashtags, but strengthened it instead	UK	Ask to not promote the hashtag as it reinforces the trend
2	8.84%	1. jennyrickson - Local Influencer 2. annaberu - Local Influencer 3. blackswan_t2 - Local Influencer 4. SchwabBum - Parody Account	UK Antimask Influencers	Strongly vocal - antimask and antilockdown	UK	Notify local governing body for proactive action
3	5.73%	1. Milhouse_Van_Ho - Local Influencer 2. Sharon75571311 - Local Influencer 3. Chillie2 - Local Influencer	Canada Antimask Influencers	Strongly vocal - antimask and antilockdown	Canada	Notify local governing body for proactive action
4	5.09%	1. CarolJHedges - Author 2. dontbrexitfixit - Political Satire	Popular person/group	Sarcastic tone against the hashtag; Anti to anything the govt says	UK	Ask to not promote the hashtag as it reinforces the trend
5	4.96%	1. JujuliaGrace - Doctor	Doctor	Tweeted against growing popularity of antimask hashtags, but strengthened it instead	UK	Ask to not promote the hashtag as it reinforces the trend
6	4.22%	1. VolvoMan17 2. Andy_In_The_UK	Random People	Random usage of hashtag	USA/UK	No action needed

*Local Influencers - where follower:following ratio is greater than 1.5:1
*The last cluster made it evident that it was not worth exploring further clusters

Figure 6: Clustering Results and Inference

4 Insights and Recommendations

As can be seen from Figure 1 in Appendix, when some of the tweets of these influencers were analyzed in detail, it was observed that all of these people opposed wearing masks because they believed that masks violated their freedom of expression and liberty and they were not effective in the curbing the virus. Their reasoning was that had masks been effective, there would not have been seeing second or even third wave of the virus. These people did not back up their claim their claim by using any scientific factor so most of these negative sentiments were found to be frustration against the government, lack of trust in masks' effectiveness, violation of their right to make choices for their own body. Some of these tweets are shown in Figure 3 in the Appendix. Additionally, as mentioned earlier as well, the health debate was also translated into political debate in some tweets. For example, some of the tweets had the hashtag of #BorisJohnsonOut. This shows that people in UK associated this health debate with their lack of trust in their Prime Minister, Boris Johnson and expressed their discontent towards his leadership. This is also where the dashboard idea could be extremely helpful since there could be a sudden jump in these sentiments due to a political leader of some country making a statement or passing a law, that can infuriate people and further promote these negative sentiments. Although this was not observed as part of our tweets, some studies have shown that some of the anti-mask tweets have been about political leaders themselves not wearing masks, hence, going against their own messages[5]. Another interesting observation that was made was that not all influencers shared negative sentiments for masks. Some of the influencers, such as Jujulia Grace from Table 1 is actually a Doctor, who was more concerned about the presence of the hashtag as shown in Figure 5 in Appendix. Another common notion that was seen in the tweets was that pandemic does not exist and is a hoax or is not as lethal as is portrayed by media. It must be noted that insights obtained from anti-masks tweets are limited to that week and are subjected to change if another time period is considered.

It was interesting to find that the influencer for anti-vaccination movements were mostly influential known people including former and current President of United States, Donald Trump and Joe Biden, respectively and current Prime Minister of United Kingdom, Boris Johnson as shown in Figure 2 in Appendix. Additionally, the official pages of these political leaders also had higher scores. Some of the other influential known parties included Bill Gates, co-founder of Microsoft, YouTube and WHO (World Health Organization). This could be due to the fact that scores for influencers for anti-vaccination movements highly depended on the closeness centrality, which a way of detecting nodes that can spread information very efficiently through a graph. Hence, it is reasonable to have such political parties as influencers as these people and organizations will be in the best place to spread information efficiently to their network and so forth. Not only that it was seen that the anti-vaccination tweets were regarding people opposing the idea of having vaccines mandatory or having vaccines passports. As with anti-masks, people did not have scientific evidence to back up their claims that vaccines kill and that manufacturers should be held liable. A continuous notion of people being concerned about their children getting vaccination was seen as well. Additionally, some of the tweets included conspiracy theory where they considered vaccines as "Bill Gates' death serum", with some of these tweets shown in Figure 4 in Appendix.

4.1 Recommendations

After a year and half of the virus continuously affecting the lives of people worldwide, it is very clear that the world including ordinary citizens, political leaders and various organizations need to come together to fight it. However, the insights from these tweets have also showed that some people/communities do not believe in joining forces. Based on our analysis,we suggest the following recommendations the appropriate authorities and organization to deal with these anti-vaccination and anti-mask movements:

- The first general recommendation to health organization and political parties is that they need to understand the sentiments and the beliefs of these people and try to empathize with than rather than threatening them. If the people will feel heard, they are most likely to listen to authorities.
- As mentioned earlier, a lot of anti-mask tweets were associated with more of a political agenda. In this scenario, we suggest that different techniques should be tried to spread positive message regarding masks and vaccinations. This includes having more neutral organizations such as health organizations or even religious organizations trying to spread these positive sentiments. The attachments of masks with politics should fade away if the source of these recommendations are not only political authorities.
- People with positive sentiments should also be asked to refrain from using these hashtags as the inflation of these hashtags give a false impression that a lot of people share their anti-sentiment views.
- People are most likely to react if a certain level of personal fear is induced. In this case, it can be done by having more individuals who have gone through COVID-19 experiences sharing their stories. If people can relate on personal level to any of these individuals, they are more likely to follow the guidelines [3].
- A lot of people do not believe in the effectiveness of the masks and as mentioned earlier one of their reasoning is that there are still growth in case even after everyone wearing masks. We suggest that in order fight with this notion, health organizations should really focus on the idea of how a multiple factors contribute when we are talking about masks such as handling and usage of masks. Misusage of masks will definitely a counter-intuitive impact.
- We have seen that a lot of industries are currently using personalization techniques to attract more consumers. In this scenario, we can use the use the dashboard idea to focus on specific geographical areas and analyze the clusters/communities in those areas to understand the motivating factor behind these sentiments. The authorities can then analyze these factor and have specific recommendations for that certain geographic areas. For example, if one sees that people are against vaccines because they believe that it is against the idea of their religion or some local organizations are behind this negative messaging, then it is easier for authorities to take actions those organizations[4].
- Political authorities should start with giving clear guidelines. For example, in the beginning of pandemic, it was suggested to not wear masks because they could actually harm people who did

not have the virus and then after few months it was mandated to wear masks. Hence, it is very important for the authority to support their claims with scientific evidence. If that is not possible, then authorities should be open and transparent to people in terms of what they know vs what they do not.

References

[1] Harkins, S. (2018, September 20). A super easy way to generate new records from Multi-value columns using Excel Power Query. https://www.techrepublic.com/article/a-super-easy-way-to-generate-new-records-from-multi-value-columns-using-excels-power-query/.

[2] Definition: What is a social media influencer? Pixlee. (n.d.). https://www.pixlee.com/definitions/definition-social-media-influencer.

[3] Mattern, A. (2020, October 26). The psychology of wearing masks: Why some people are complying and others are not — CBC News. CBCnews. https://www.cbc.ca/news/canada/saskatoon/covid-19-mask-compliance-psychology-1.5777729.

[4] Dwoskin, E. (2021, February 16). On social media, vaccine misinformation mixes with extreme faith. The Washington Post. https://www.washingtonpost.com/technology/2021/02/16/covid-vaccine-misinformation-evangelical-mark-beast/.

[5] Ahmed,W.,Khan L., Malik, A.(2021). A Social Network Analysis of Pro-maskers and Antimaskers on Twitter during the COVID-19 Pandemic. shared through pro-masking and anti-masking accounts. https://doi.org/10.2196/preprints.27716

5 Appendix

Screen Name	Score	Description
leew2030	0.08853	cry later but for now lets enjoy the laughter
jennyrickson	0.07592	questioning the narrative, life is for living not just existing! my symbol for freedom
HagleyTom1234	0.07174	Fighting for freedom and normality.
JujuliaGrace	0.06895	Doctor, campaigner, forager, magpie. Chief Executive of @EveryDoctorUK . All views my own.
JamesMelville	0.06229	Made in Fife, Scotland. Communications, Sponsorship and Social Media Consultant. Freelance writer. Haverings are my own. Liberal james@eastpointswest.co.uk
Seahorse1317	0.06167	my other account @Seahorse was suspended by Twitter, without explanation. Against Lockdown and coerced vaccinations.
tarynabell	0.06154	A Holland Parker in the 80s (when the school was cool). Founder of the #showyoursmile hashtag as you're never fully dressed without one Grinning face with smiling eyes Same handle on Gab.
llamedos77	0.05798	Hi I'm rob, i'm a sci-fi fan. I tweet alot about mental health positivity + autism awareness. Father of 4 children and husband to the love of my life Claire :)
northerness	0.05789	"I did what I could, what I had to do, what my conscience told me I must do. That's all there is to it. Really, nothing more."
annaberu	0.05701	#KBF #endlockdown #NoMasks Flag of Scotland

Figure 1: Influencers for anti-mask movement

Screen Name	Score	Description
realDonaldTrump	0.01914	**Banned from Twitter so no description was obtained
BorisJohnson	0.01272	Prime Minister of the United Kingdom and @Conservatives leader. Member of Parliament for Uxbridge and South Ruislip.
MattHancock	0.01061	Secretary of State for Health & Social Care and MP for West Suffolk
BillGates	0.00916	Sharing things I'm learning through my foundation work and other interests.
YouTube	0.009	Like and Subscribe.
10DowningStreet	0.00498	Official page for Prime Minister @BorisJohnson 's office, based at 10 Downing Street
WHO	0.00476	We are the #UnitedNations' health agency - #HealthForAll. Right-pointing triangle Always check our latest tweets on #COVID19 for updated advice/information
simondolan	0.00474	**Account does not exist
POTUS	0.00469	46th President of the United States, husband to @FLOTUS , proud dad & pop. Tweets may be archived: http://whitehouse.gov/privacy
RealCandaceO	0.00401	New York Times best-selling author. Founder of @BLEXIT organization. Black people don't have to be Democrats— still.

Figure 2: Influencers for anti-mask movement

user	tweet
jennyrickson	b'I have so little trust in the Government at the moment, I wouldn\xe2\x80\x99t believe them if they said the sky was blue\xe2\x80\xa6 https://t.co/S5cXs5aPPY'
HagleyTom1234	b"@Revegelance Masks do nothing.\n\nI'm sorry you've been abused by your Government.\n\n#TakeOffYourMask"

Figure 3: Tweets from two influencers from anti-mask movement

@BillGates Stop with the murder campaign #vaccineskill #BillGatesmurderer #BillGatesEvil #BillGatesBioTerrorist #BillGatesIsNotADoctor
@BorisJohnson @CMO_England and @MattHancock are following the @BillGates and @gatesfoundation edict by killing the pensioners who survived the virus, with the #vaccine, why are the likes of #skynews #bbcnews ignoring all the vaccine casualties?? #COVID #vaccineskill #coronavirus
@BillGates In this NOW, I cancel all previous/now/future permissions & contracts granted knowingly/unknowingly to the Dark agenda. #IDoNotConsent to being used in the Dark agenda. I return your energies to the Love and Light of the One Infinite Creator. #Ascension #TheGreatAwakening

Figure 4: Tweets including mention for Bill Gates

user	tweet
JujuliaGrace	b'It\xe2\x80\x99s extremely worrying to see #TakeOffYourMask trending\n\nWe must not abandon the measures keeping us safe. Masks a\xe2\x80\xa6 https://t.co/mGXa55Qjpw'

Figure 5: Tweet from a pro-mask person

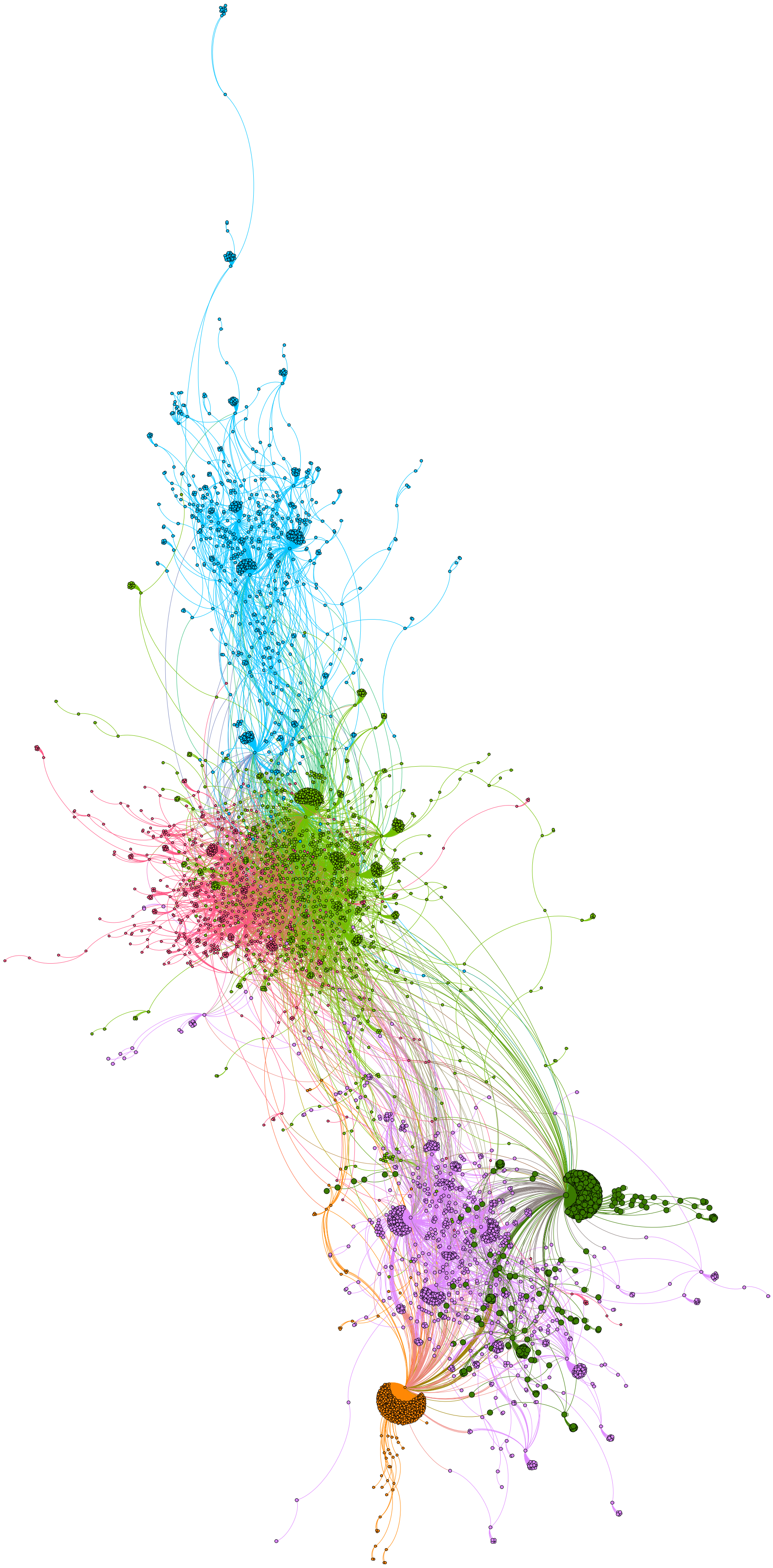
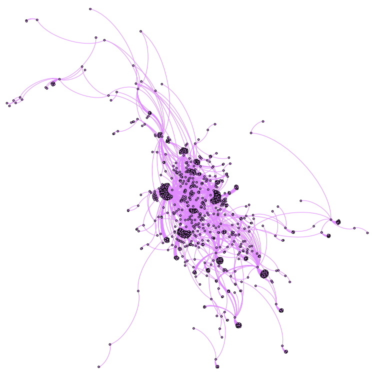


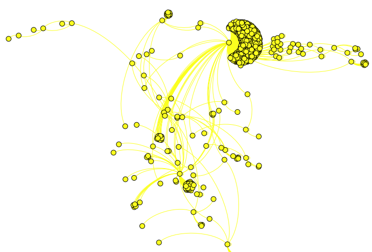
Figure 6: Clustering using Gephi



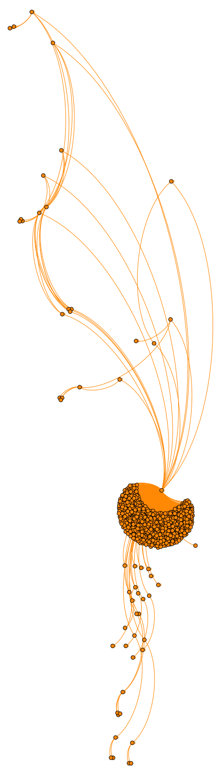
Cluster 1

Cluster 2

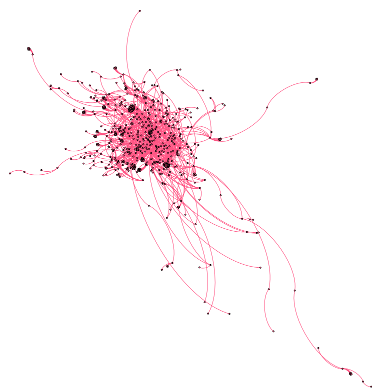
Cluster 3



Cluster 4



Cluster 5



Cluster 6

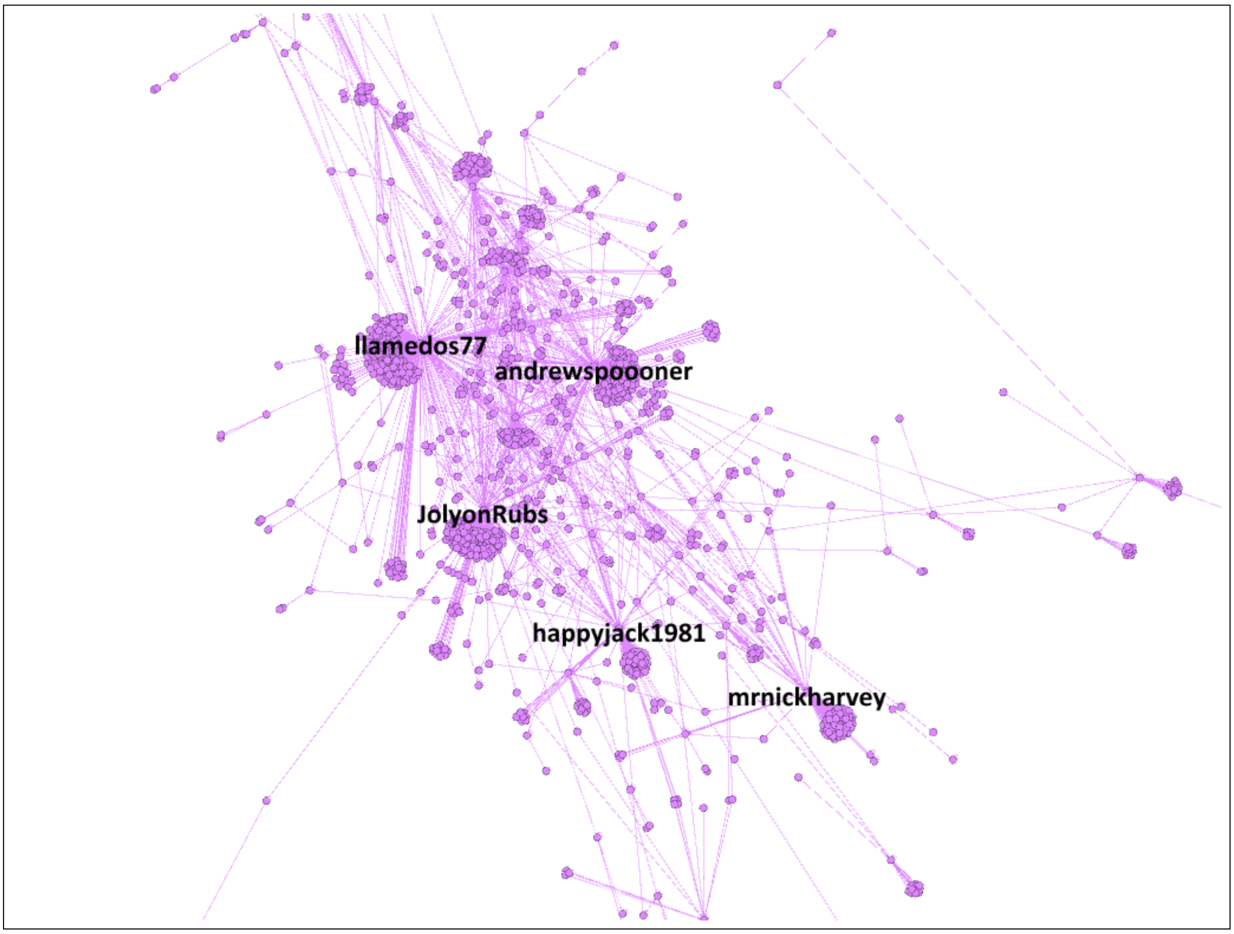


Figure 7: Cluster 1

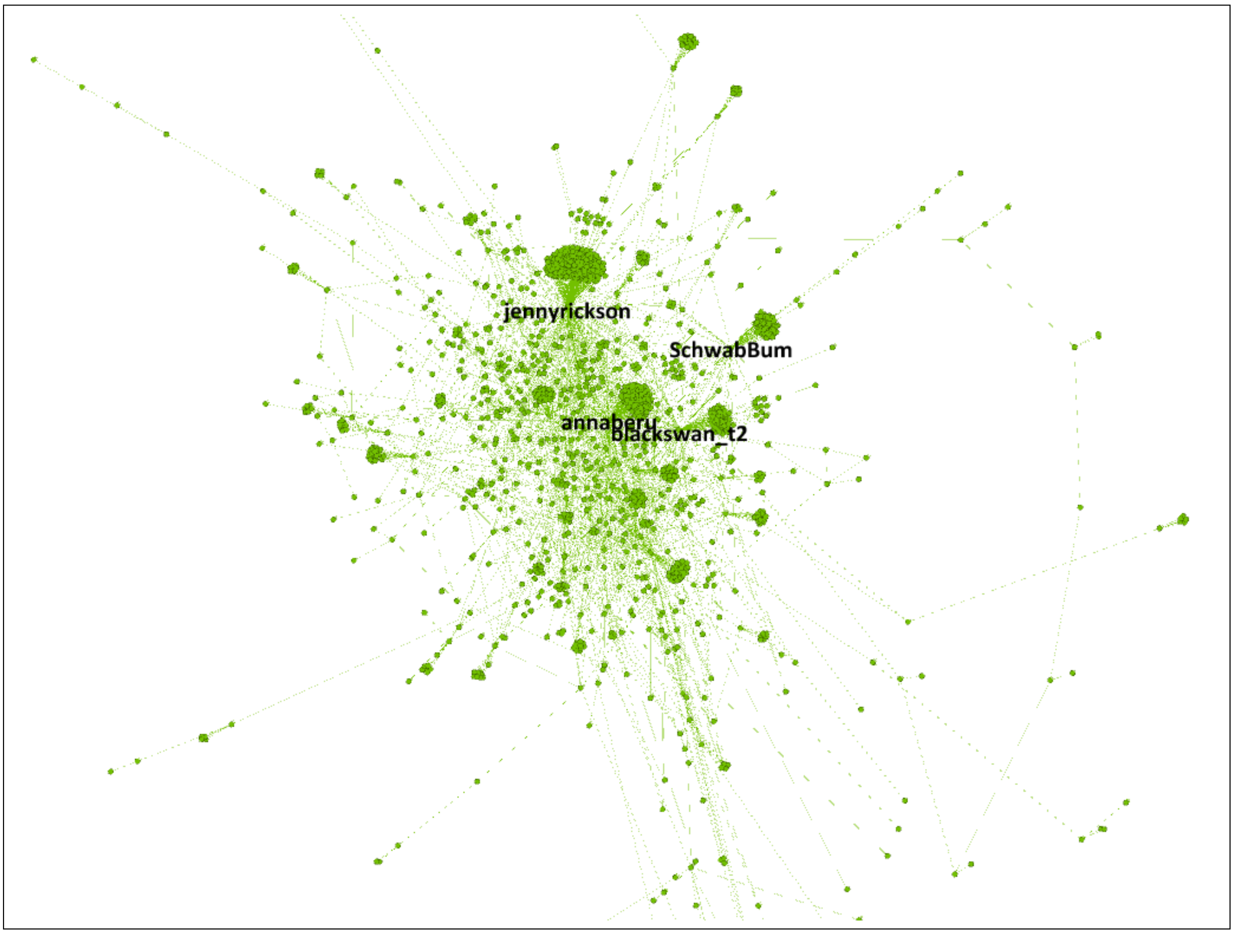


Figure 8: Cluster 2

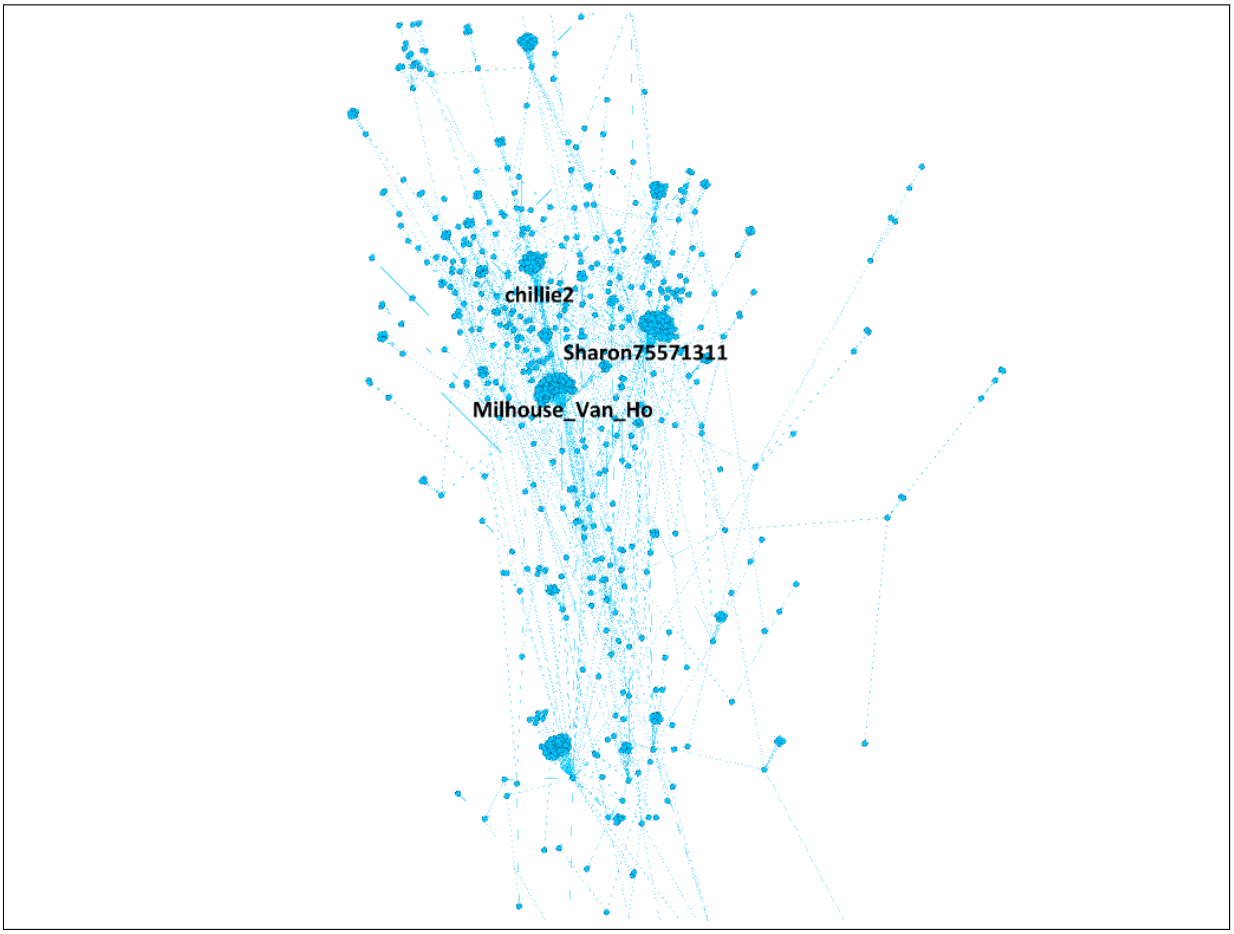


Figure 9: Cluster 3

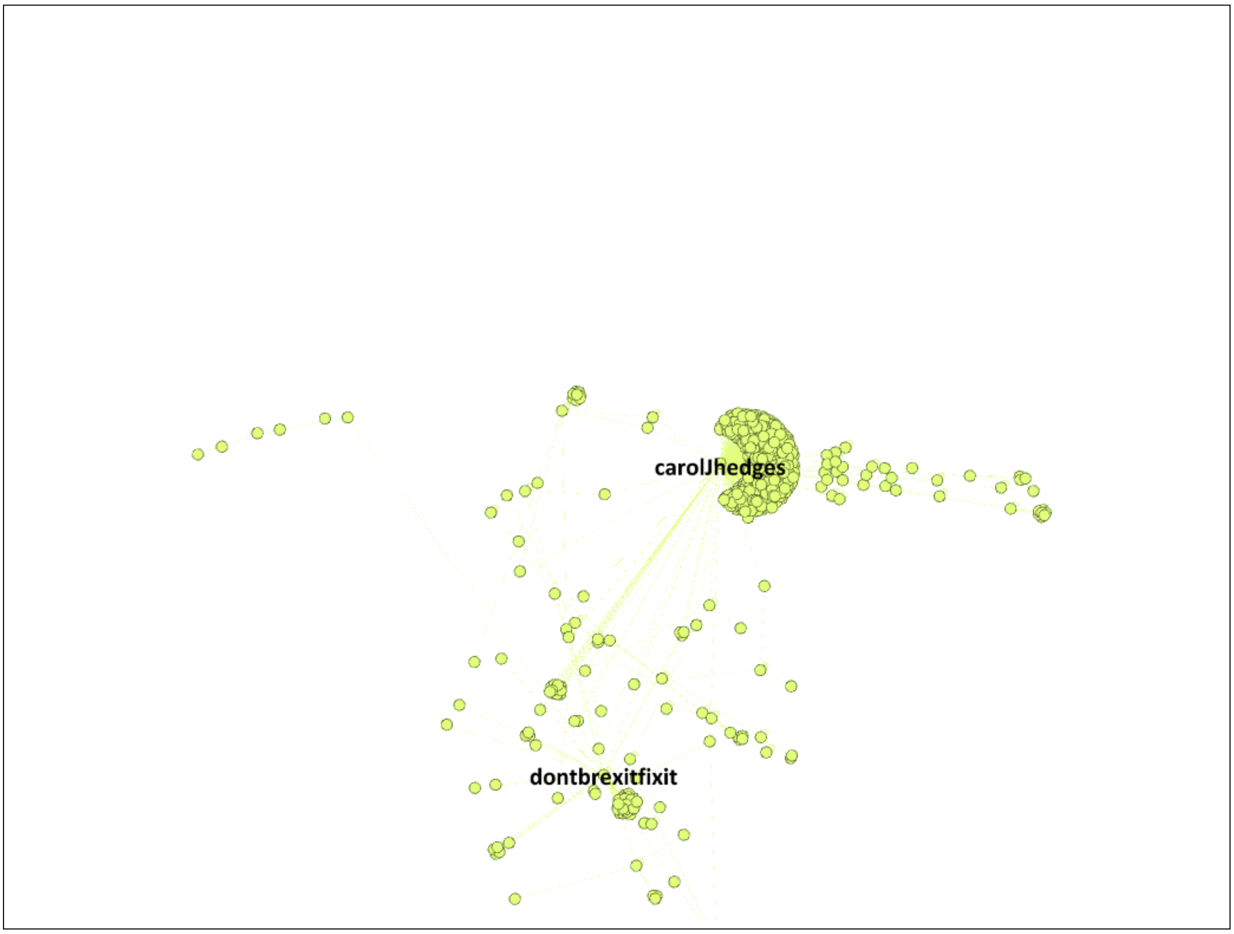


Figure 10: Cluster 4

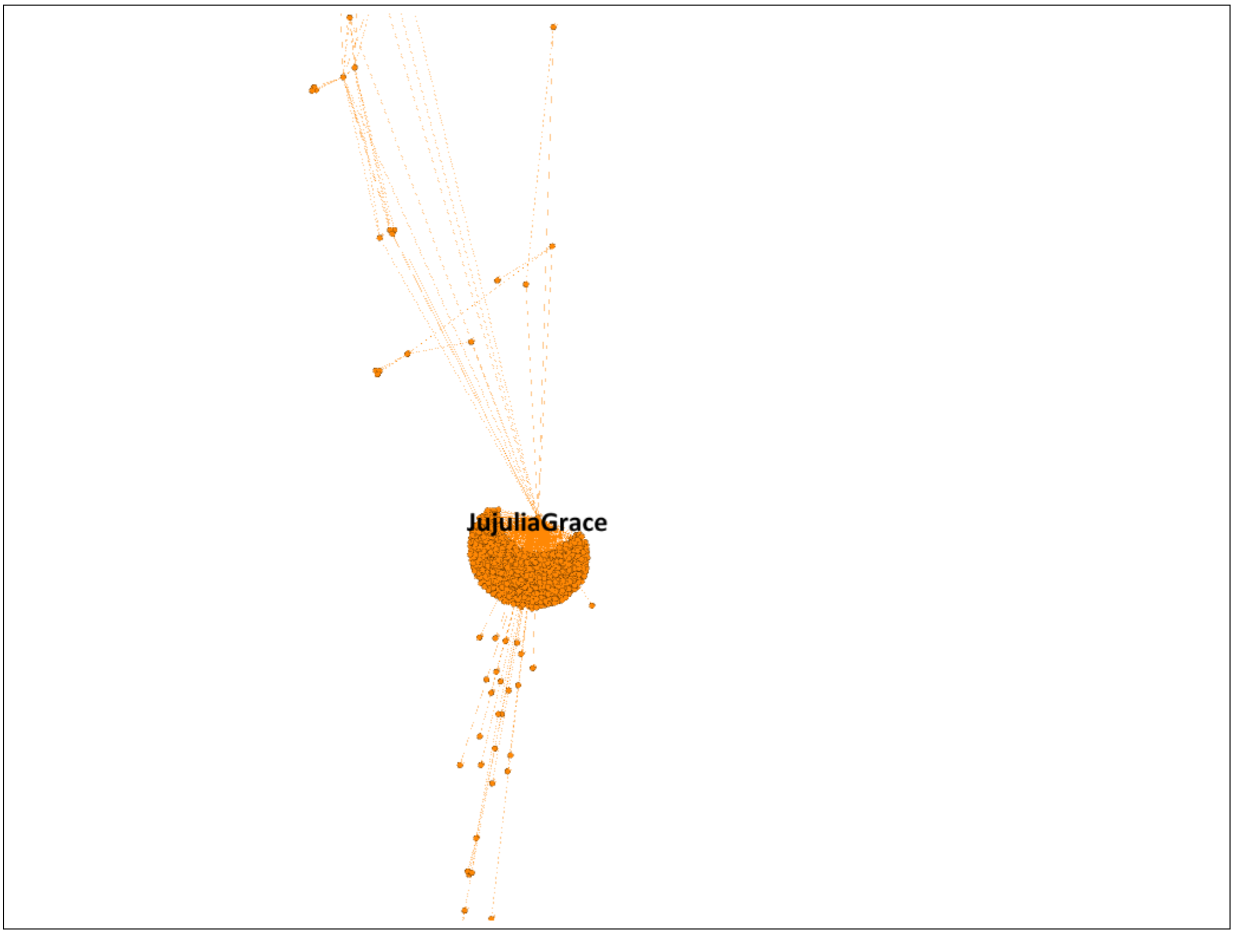


Figure 11: Cluster 5

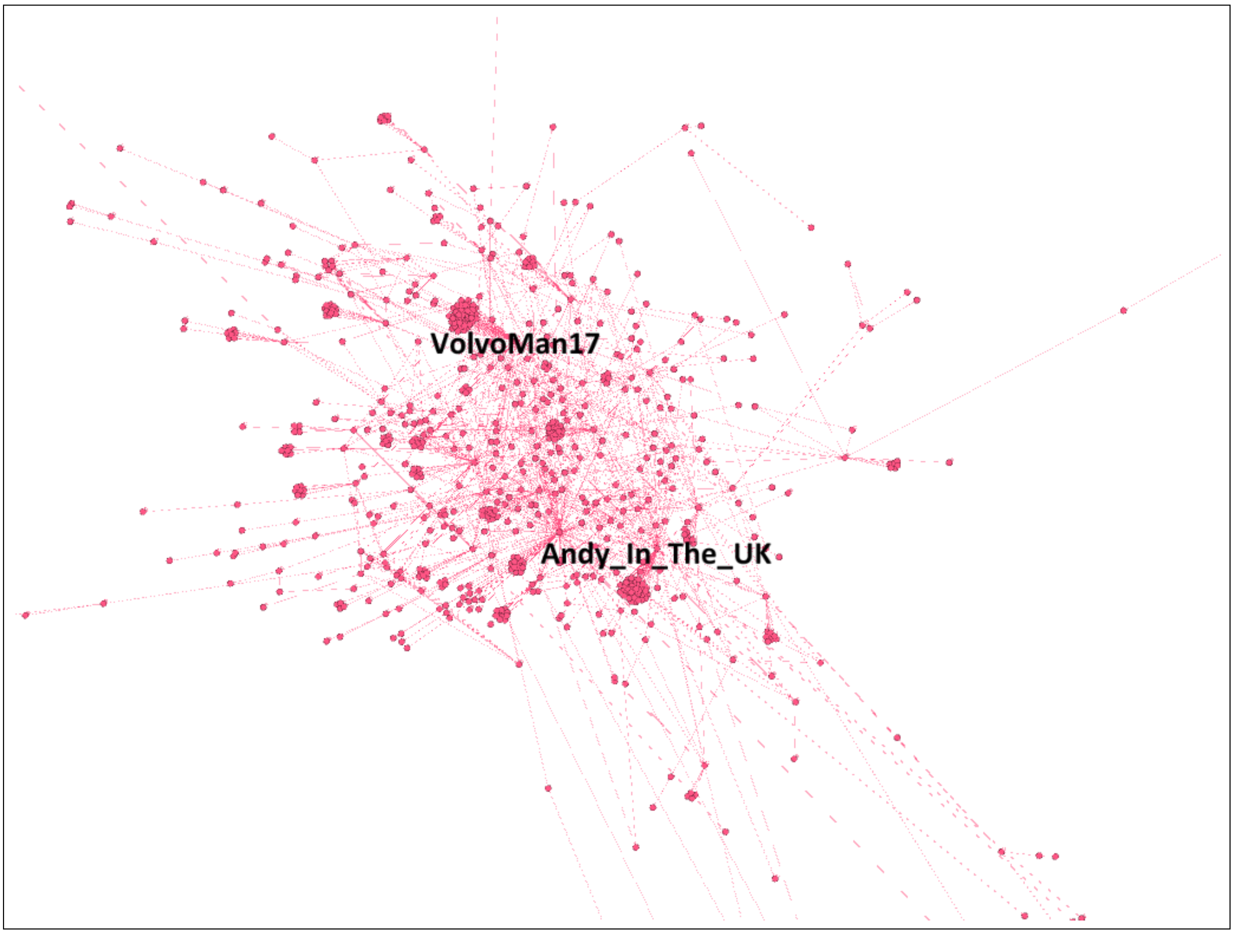


Figure 12: Cluster 6