An Analysis of Twitter Discussions Regarding COVID-19

Aaron Lohner, Nicholas Kiriazis, Mehdi El Jaafari
McGill University
aaron.lohner@mail.mcgill.ca, nicholas.kiriazis@mail.mcgill.ca, mehdi.eljaafari@mail.mcgill.ca

Introduction

Since the start of the pandemic in March 2020, the world has changed. Countries all over the world implemented health measures to mitigate the spread of the virus: lockdowns, masks mandates, and, once available, vaccines. In this report, we attempt to make sense of the current global sentiment regarding the pandemic and health measures. In particular, we focused on analyzing public opinion toward vaccines. We did this by collecting Twitter data (tweets) with certain filters that allowed us to peer into the current discussions surrounding COVID.

In brief, this was done by fetching 3,000 tweets related to COVID-19 during a 3-day period using the Twitter API. After an open coding of a small sub-sample, we classified 1,000 tweets into five categories: public health measures that are not vaccine-related, commentary on how groups and individuals have handled the pandemic, medical news and information about COVID-19, personal experiences with vaccination or COVID-19, and societal-level vaccination policy. We also annotated tweets in order to determine whether the sentiment was negative, positive or neutral.

We then computed the tf-idf scores of the tweets for each category, and saved the ten highest scoring words in each. We subsequently measured two more variables for every tweet: the length of the tweet and its relevance (both of these terms are precisely defined in the Results section).

Using these measures, we drew the following conclusions: (1) there is little positive sentiment among any category regarding the COVID-19 discussion, (2) there is strong negative sentiment directed at specific groups and individuals–such as politicians, people associated with certain political parties, and world leaders–regarding their pandemic response, and (3) tweets containing facts and news are generally short in length, contain one of a handful of characterizing words, and are neutral in sentiment, whereas emotionally-charged tweets tend to be longer, use a wider vocabulary, and have a negative (or occasionally positive) sentiment. Regarding the last point, we found that tweets expressing some degree of vaccine hesitancy tend to be more emotional and long, however they also typically employ one of a small set of characterizing words.

Data

We initially fetched 3,000 tweets over a 3-day period using 36 non-case sensitive keywords. The keywords we selected fell into three categories: (1) variations of the term ”COVID” (”covid”, ”coronavirus”, ”covid19”), twelve words relating to vaccines (e.g. ”vaccination”, ”vaxxed”, ”antivax”), and commonly-referred names of three different COVID-19 vaccine brands (”pfizer”, ”moderna”, ”astrazeneca”). We also queried the Twitter API with the hashtagged version of these 18 terms by adding the # symbol before each word. It should be noted that hyphenated words are considered separate words by the Twitter API, so tweets with terms such as ”covid-19” were included in our dataset since the word ”covid” is one of our keywords.

Our preliminary analysis on this set of tweets revealed that 6% of the time, the keyword(s) present in each filtered tweet was (were) contained in its URL rather than in the content of the tweet itself. Nevertheless, the content of these tweets was consistently relevant to the discussion of COVID, so we decided to keep them for our main analysis. Furthermore, it was interesting to observe that approximately 70% of the tweets mentioned one of the variations of the term ”COVID” (i.e. contained at least one of the keywords from the first category mentioned above), while only 36% had at least one mention of any of the other keywords.

In addition to filtering the data by keywords, we also chose to exclude retweets and replies from our dataset. From a practical perspective, this made sense to do because for these types of tweets fetched over the API, we noticed that the content was truncated beyond 140 characters. This design decision also helped us with our analysis. By removing retweets and replies, we ensured that we had a wide array of tweets to study, rather than several retweets and replies stemming from a single tweet.

Methods

In this section, we describe our methods for the three principal tasks of this project: data collection, data annotation, and word frequency analysis.

Data Collection

All our tweets were collected from November 28 to November 30, 2021. On each of the three days in this time period, we executed a Python script to fetch 1,000 tweets. The script fetched the most recently posted tweets that met the filtering conditions in batches of approximately 10-20 at a time, and looped until 1,000 were collected and entered into a CSV file. A 1.5-second pause was added to the script between fetching each batch to avoid exceeding the Twitter API's rate limit.

We used the Tweepy library to access the Twitter API. Although Tweepy also allows users to access v2.0 of the Twitter API, we elected to use version 1.1 because we found it slightly easier to enter the language, keyword, and other filtering restrictions into Tweepy's `search_tweets` function, which accesses the earlier API version. The string used to query the API contained the 36 keywords joined together by the "OR" operator. We chose this many words because the string it formed was nearest to the maximal length we were able to send for a query.

Data Annotation

Once all the tweets were collected, we conducted our open coding on a sample of approximately 200 of them. We first briefly read through the tweets and designed rough categories. We then repeated the process of attempting to categorize the 200 tweets in our categories followed by editing our label definitions based on the ambiguities we encountered. After three rounds of this iterative process, we began to annotate the entire dataset. Overall, we analyzed approximately 1,100 tweets (in roughly equal proportion from each day's collection of tweets). We occasionally encountered tweets that could be placed in zero or more than one our categories. When this occurred, we chose to discard these tweets from the annotation process. Of the roughly 1,100 tweets, we annotated 1,000 of them and indicated their sentiment (positive, negative or neutral).

Word Frequency Analysis

In order to extract salient topics from our dataset, we computed the tf-idf values for the words in each category. We computed term frequency and inverse document frequency as follows:

$$tf(w,c) = \text{number of times word } w \text{ appears in category } c \quad (1)$$

$$idf(w, dataset) = \ln\left(\frac{\text{number of categories}}{\text{number of categories containing word } w}\right) \quad (2)$$

However, given the unstructured and messy nature of tweet content, counting words was not straightforward. Therefore, before computing word frequency statistics, the following pre-processing procedure was used:

1. Replace every non-alphanumeric character with a space character.
2. Split each string on its spaces to find the words in the given string. A word is defined as a sequence of alphanumeric characters surrounded by whitespace.
3. Require a word to appear at least 5 times throughout the whole dataset in order to be considered in tf-idf calculations.
4. Remove stopwords and numerics.

Results

In this section, we describe the resulting categories that were decided through the open coding process and we present a table and figures that illustrate the results of our analysis.

Open Coding

The following is the list of categories we designed as well as descriptions of the types of tweets that are classified in each category:

1. **Public health measures that are not vaccine-related**
   Tweets that discuss mask mandates, travel bans, border restrictions, re-opening measures, and testing requirements. No distinction was made on who was issuing the policies, whether it be federal governments, local governments or individual venues.
   We opted to distinguish between general policies and vaccine-specific policies to better understand whether vaccine hesitancy was due to a general unwillingness to comply with measures, or whether vaccine hesitancy is due to the vaccines themselves.
2. **Commentary on how groups/individuals are handling the pandemic**
   Tweets that discuss the effectiveness of politician responses, the reliability of news outlets, and comments about how specific groups have handled the pandemic.
   Tweets that focus on discussing a specific policy but also mention the group/individual who implemented the policy were omitted from this category, and rather were included in categories 1 or 5 accordingly.
3. **Medical news and information about COVID-19**
   Updates on case/hospitalization counts, announcements about outbreaks, announcements of new variants, and summaries of symptoms and/or the health effects of COVID-19.
4. **Personal experiences with vaccination and/or COVID-19**
   Updates about one's health after contracting COVID-19, announcements of personal vaccination status, concerns about the effectiveness/side-effects of vaccination, and general questions about vaccine acquisition.
   This category focuses on understanding the medical questions and personal reasons that could potentially be an important part of understanding vaccine-hesitancy.
5. **Public vaccination policy**
   Tweets that discuss government ordered vaccination mandates, employer vaccination mandates, comments on personal liberty in regard to mandates, and vaccine allocation policy.
   This category is meant to encapsulate the discussion of civil liberties associated with vaccines, rather than medical questions about vaccination.

Measured Variables

In addition to manually classified tweets using the five aforementioned categories and determining their sentiment, two more variables were measured for each tweet:

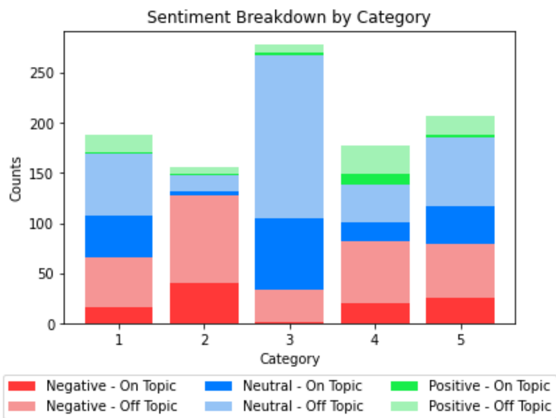

Figure 1: Distribution of the categories within the dataset. Within each category we distinguish negative tweets (red), neutral tweets (blue), and positive tweets (green). On-topic tweets are more saturated than their off-topic counterparts.

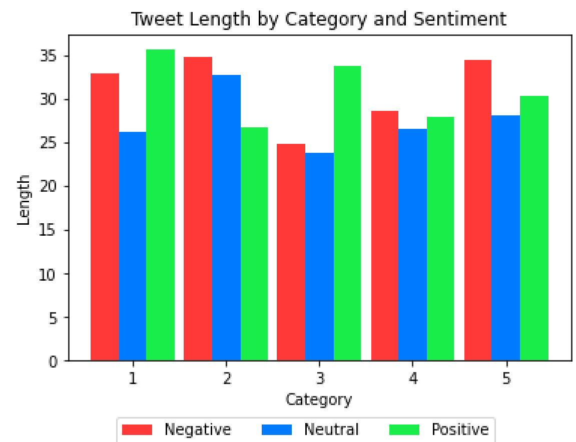

Figure 2: Average length of the tweets for each category/sentiment grouping.

1. **Length**
   In this context, tweet length refers to the number of words in its content, using the definition of word given in the Methods section.
2. **Relevance**
   We assigned a Boolean variable to each tweet. This variable was `True` for tweets containing at least one of the ten most popular words (based on tf-idf score) in their category, and `False` otherwise. For simplicity, from hereon Tweets with a `True` value for this variable will be referred to as being "on-topic," and tweets with a `False` will be referred to as being "off-topic."

Quantitative Results

The relative popularity of each category, the distribution of tweets among these categories, sentiment, and relevance is captured graphically in Figure 1. The length of tweets is illustrated by Figure 2. Furthermore, the ten most popular words (based on tf-idf) for each category are included in Table 1, which illustrates some of the key themes being addressed within each category.

Discussion

The results from our analysis reveal several interesting trends regarding the discussions of COVID-19 on Twitter.

Understanding the Distribution of Topics and Sentiment

We first observe that, with the exception of category 3, the number of tweets classified in each of the five categories ranges from approximately 150 to 200 tweets. This indicates that the relative engagement of Twitter users among these four topics is roughly evenly distributed. A possible reason for which category 3 has a significantly larger number of tweets may be that it mainly consists of posts disseminated by news outlets and health agencies rather than comments made by individuals. Figure 1 highlights the difference in the quality of information from this category. Here, we see that category 3 has the largest share of its tweets categorized as having neutral sentiment, as would be expected for these types of tweets.

Another trend that we notice from Figure 1 is that, among all categories, there is little positive discussion regarding COVID-19 and significant negative sentiment. However, the proportion of tweets that are negative or neutral varies by category. Category 3 has the lowest proportion of negative tweets, which is understandable for this category based on the points mentioned in the above paragraph.

In terms of topic scope, the proportion of positive tweets that are off-topic is noticeably larger than the proportion of off-topic negative or neutral tweets. This is likely due to the fact that positive discussion is much less common than negative or neutral discussion. Among all categories, neutral tweets in category 3 are the most likely to be on-topic.

We can also observe that the overall sentiment proportions are fairly similar among categories 1 and 5, with a large share (about 45%) of each categories' tweets being negative, and a fraction (roughly 15%) being positive. There are many possible implications from this roughly even distribution. For instance, one may interpret this as equal disdain of COVID policies and of vaccines regulations, and a general belief that COVID is not so severe and need not be the cause of many restrictions. Alternatively, people may be equally against vaccine mandates as well as other COVID policies because they are against government regulations in general. This is further suggested by the strong negative sentiment toward politicians and other leading figures as seen in category 2: assuming the outlook on regulations is negative, one would also expect negative views of the people implementing the policies, which is in fact the case here. In contrast, the complete opposite may instead be the case: it is possible that the negative sentiment is directed at those disrespecting the COVID- and vaccine-related rules, and that the negativity directed at those in charge (i.e. the negative sentiment in category 2) is anger about a lack of restrictions. Further in-

Rank	Category 1	Category 2	Category 3	Category 4	Category 5
1	indoor (8.05)	trump (8.68)	cases (11.60)	booster (10.73)	mandate (9.19)
2	bans (7.33)	rep (8.05)	november(9.66)	never (5.11)	booster(8.17)
3	rules (7.33)	jackson (8.05)	analytics (9.66)	okay (4.58)	companies (5.5)
4	travel (6.92)	democrats (5.5)	patients (7.33)	stay (4.58)	states (5.5)
5	wake (3.67)	gop (4.6)	confirmed (5.5)	moderna (4.09)	doses (4.58)
6	israel (3.67)	texas (4.58)	detected (5.5)	shouldn (3.67)	thousands (4.58)
7	event (3.67)	lazy (4.58)	ontario (5.5)	mind (3.67)	misinformation (4.58)
8	spreads (3.67)	political (4.58)	positive (4.6)	1st (3.67)	business (4.58)
9	guidelines (3.67)	american (4.58)	update (4.6)	dose (3.06)	moderna (4.09)
10	announces (3.67)	wrong (4.58)	ottawa (4.58)	booked (2.75)	appointment (3.67)
Mean	5.16	5.78	6.85	4.59	5.44
Variance	3.43	2.74	5.94	4.65	2.95

Table 1: Ten most popular words in each category by tf-idf score (with their tf-idf score in parentheses), as well as the mean and variance of the tf-idf scores for each category.

vestigation would be required to better under the subtleties of this negative sentiment, given that depending on the interpretation, one would consider opposite courses of action.

Dissecting Tweet Content

We now look at the implications of the results in Table 1. For our analysis, we interpreted the mean tf-idf score of the top ten words in each category as a measure of the overall characterizability of a category. A high mean among the top ten implies that the featured words are very important for the given category, whereas a low mean suggests that many different terms and topics are discussed. Variance also measures characterizability, but exclusively among the top ten words; theoretically, low variance would imply that the ten words listed are of roughly equal importance when characterizing a category, but in practice, low variance may simply be an artifact of truncating at ten words. One should therefore also take the mean into account when interpreting the variance: the larger the mean, the "better" the category is characterized by the retained words.

For example, the high mean and variance of scores for category 3 may imply that there are very few key words that characterize this class, such as “cases” and “analytics”. This would align with the intuition that news headlines use a small collection of “buzzwords” to capture the attention of readers. It also gives a similar conclusion that we drew from Figure 1 when we noted that tweets in category 3 are the most likely to be on-topic.

Table 1 provides insight on the vaccine discussion and vaccine hesitancy in the population. “Booster” is the highest scoring word for category 4 and second-highest for category 5, suggesting that people are discussing the third dose of vaccine from both a personal perspective (such as in regard to sign effects) as well as a mandate perspective. Another interesting point to note is that the top ten scores in category 4 have the lowest mean among all the categories. From our interpretation of this statistic mentioned above, this suggests that the “personal” COVID discussion covers a range of experiences that require the use of many different words to properly capture.

Category 5 contains several important words that relate to the discussion of vaccine policy, such as “mandate”, “companies”, and “misinformation”. Considering that very few of these tweets express positive sentiment, there is a significant chance that people are discussing these terms in a negative light. This gives several potential reasons why people may be hesitant to get vaccinated. For example, people may be discontent with newly-instituted requirements of some companies to require their employees to be vaccinated. Alternatively, many may be citing the plethora of misinformation being spread as a reason to avoid vaccines.

Interpreting Tweet Length

Looking at tweet lengths in Figure 2, we note that neutral tweets are consistently the shortest across all categories. We can hypothesize from this that tweets expressing non-neutral sentiment may typically involve more storytelling and explanation, in contrast to neutral, fact-based tweets. The shorter average length of tweets in category 3, along with the higher proportion of neutral sentiment, could be used to argue that Twitter is a powerful took for the rapid sharing of information, which is vital with a constantly evolving pandemic.

The length of negative tweets in category 5 is significantly larger than neutral and positive tweets, meaning that users expressing their discontent toward vaccine policies have a lot to say. Given that neutral, fact-based tweets such as those in category 3 tend to be shorter, it is plausible that the negative tweets in category 5 rely more heavily on emotion rather than fact. Furthermore, given the relative ease with which anecdotes can be understood as opposed to the scientific literacy required to analyze experimental findings or study results, an argument could be made that these longer and emotionally-charged tweets are key factors driving vaccine hesitancy.

Conclusion

This project allowed us to analyze the salient topics on Twitter regarding COVID-19. We deduced that the key areas of discussion are in regard to COVID policies, the handling of the pandemic by different groups/individuals, medical information and news, personal experience with the virus or vaccines, and vaccination policy. We also uncovered several interesting trends, such as strong negative sentiment toward groups and individuals handling the pandemic, overall similar sentiments toward COVID policy as toward vaccine policy, and longer tweets are more likely to express non-neutral sentiment. Furthermore, we outlined how these and other trends may be linked with vaccine hesitancy. For instance, the pervasive negative sentiment of tweets in category 5 coupled with its discussion of mandates, companies (potentially in relation to their vaccine requirements), and misinformation suggest that people may turn to these topics when looking for reasons to avoid getting vaccinated. In the future, there are several ways that this analysis can be built upon. With more tweets, we could potentially generate more precise categories and pinpoint more specific topics that are linked to vaccine hesitancy. We could also collect more data on the users of the tweets that can provide more insight, such as user location.