Identifying Sources of COVID-19 Vaccine Hesitancy Through Tweets

University of Chicago

CAPP 30254: Machine Learning for Public Policy

Link to Git Repository

Shashank Bharadwaj, shashab
Natalie Ayers, nayers
Fumi Yoshikoshi, fumiy

June 3, 2021

Table of Contents

1 Executive Summary

One of the most challenging aspects of COVID-19 recovery for policy makers has been addressing the widespread hesitancy surrounding the newly-developed vaccinations. In order to address these concerns, however, they must first be understood. Social media can often present more accurate representations of a user's beliefs and feelings than they might share with officials or pollsters, so we propose a Machine Learning-based approach to identify trends and common themes running throughout Twitter users' discussion of the COVID vaccine from January 2021 to present. Utilizing an open-source dataset of all COVID-related tweets, we leverage sentiment analysis to label vaccination-related tweets as positive, negative, or neutral and find that neutral and positive tweets predominate. A consideration of the sentiment distribution over time finds the discussion surrounding vaccines peaked early in the year and that positive and negative sentiment tweets followed a roughly parallel trend. We subsequently use the topic modeling technique Latent Dirichlet Allocation to identify common topics within our Twitter corpus.

While we do find some topics which reference various claims and disparagements of the vaccine's effectiveness, safety, and necessity, our final topics also include a number of themes unrelated to vaccine hesitancy, and occasionally unrelated to the vaccine as well. This is representative of the difficulty in working with Tweets, given their minimal content and words which are often incomplete or combinations of others. Nevertheless, this analysis allowed us to identify previously less-visible trends in particular claims, individuals, and corporations which had significant impact in the Twitter sphere yet whose prevalence was difficult to discern as a casual Twitter user. Similar analyses which further refine the data preparation and model-building could prove useful to policy makers as means of identifying the less-publicized beliefs which are motivating their bases, particularly by distinguishing these popular, persistent beliefs from the numerous others which are shared and forgotten on Twitter, in order to best meet and respond to the concerns of the constituents they serve.

2 Background and Overview of Solution: COVID Vaccine Hesitancy

The COVID-19 pandemic has been a massive-scale challenge for countries across the globe. Vaccines represent a path forward, as a vaccinated population allows countries to reopen without significant worry of risk. Children can go back to school, people can go back to work, and economies can begin to bounce back. In order for countries to return to normalcy, citizens must actually get the vaccine, which is no guarantee. In the US, for example, as of May 27th, 2021, only half of adults are fully vaccinated[1]. Unvaccinated citizens are likely to lead to huge spikes in healthcare costs down the road[2]. While there is evidence that overall vaccine hesitancy is dropping, various estimates show that 15-25% of American adults are still deciding whether to get the vaccine or have decided not to. For governments worldwide, whether at federal or local levels, it is hugely important to understand what the reasons are for vaccine hesitancy. Officials can use the information to target messaging to hesitant citizens based on these reasons to try to improve uptake. There are various polls aimed at understanding the reasons behind vaccine hesitancy, but polls often pre-select choices from which respondents must choose, or lag behind more real-time indicators. It's critical for policymakers to understand citizens' thinking as it evolves,

[1] "Half Of All U.S. Adults Are Now Fully Vaccinated Against COVID-19." n.d. NPR.Org. Accessed June 3, 2021. https://www.npr.org/sections/coronavirus-live-updates/2021/05/25/1000171685/half-of-all-u-s-adults-will-be-fully-vaccinated-against-covid-19-as-of-tuesday.

[2] Dovere, Edward-Isaac. 2021. "Vaccine Refusal Will Come at a Cost—For All of Us." The Atlantic. April 10, 2021. https://www.theatlantic.com/politics/archive/2021/04/vaccine-refusal-hesitancy-economic-costs/618528/.

such that they can react expeditiously and adjust to latest trends. People may also not report their feelings accurately to pollsters, in fear of being judged[3].

To that end, we propose that Twitter can be mined for more real-time information and a forum by which people may be more likely to share their true feelings. Our solution is to identify COVID vaccine related tweets and use the text from these tweets to identify sources of vaccine hesitancy. We will use machine learning for two primary text-mining solutions: sentiment analysis and topic modeling. We will use sentiment analysis to identify tweets that have negative sentiment regarding the vaccine, as these are tweets that are more likely to express hesitancy. We will then run topic modeling algorithms like Latent-Dirichlet Allocation to identify common topics in these tweets and manually analyze the topics to identify sources of hesitancy. We will also analyze how sentiment for vaccine-related tweets trends over time, to provide further insight.

As noted earlier, the intended audience for our analysis is government agencies in English-speaking countries[4] worldwide that are invested in getting people vaccinated. The information in the topics we find can be used to understand common sources of vaccine hesitancy, and especially ones people are sharing with their friends/followers online. In the US, for example, the information can be used in campaigns by the CDC or state and local governments that target these common sources and aim to address them. Topics can also be used as jumping off points for further analysis or focus group study. Sentiment analysis over time can provide guidance on how public opinion is moving and how much progress is being made, as a secondary indicator to overall vaccine uptake.

3 Data

Our initial idea was to use the Twitter API to pull tweets using keyword searches. However, Twitter's primary API limits searches to the last seven days, and we wanted to go back further. We were able to identify an existing open-source dataset[5] that pulled and stored tweet identifiers for tweets that included COVID identifiers. These tweet identifiers, along with code provided, could be used to pull the full text of the tweets. Their dataset began in March 2020 and is still being updated as of writing, but we limited our pull from the beginning of 2021 to early May. Additionally, as rehydrating the tweets was time-consuming, we further limited to pulling tweets every four days. We decided that this was a reasonable balance between acquiring sufficient sample size and also not spending too much time and effort in the data acquisition phase. The tweets we pulled are absent retweets (as one way to limit "spam") and are all in the English language. Beyond that, there is no location restriction as very few tweets are publicly geotagged. Restricting to a specific location would have made our dataset too small.

Because the dataset we acquired targeted COVID-related tweets generally, we needed to do significant pre-processing to focus the dataset on vaccine related tweets, as well as prepare it for text-mining algorithms. First, we filtered to tweets that included the text string 'vacc' to try to capture tweets related to the vaccine. Then we cleaned the text of components such as mentions, punctuation, emojis, and emoticons[6]. Next, we removed stopwords and other very common substrings. Finally each tweet was tokenized, both into single tokens as well as bigrams. Thus each row in our dataset is a single tweet,

---

[3] "Social-Desirability Bias." 2020. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Social-desirability_bias&oldid=992112847.
[4] Our dataset is limited to English-language tweets
[5] Banda, Juan M., Tekumalla, Ramya, Wang, Guanyu, Yu, Jingyuan, Liu, Tuo, Ding, Yuning, … Chowell, Gerardo. (2021). A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration (Version 64) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4876538
[6] See here for details

with the tokenized forms of the tweet being used for sentiment analysis and topic modeling. As our modeling approach is an unsupervised approach, there is no outcome label. We stored our cleaned data in Google Drive and used the Google API to pull it to local.

Our overall data pull included 5,560,000 tweets, which we whittled down to 1,170,000 tweets that we found to be vaccine related, or roughly 20%. On average, our dataset includes 35,500 vaccine-related tweets per day of data we pulled. On average, there were 11 single-word tokens per tweet after cleaning.

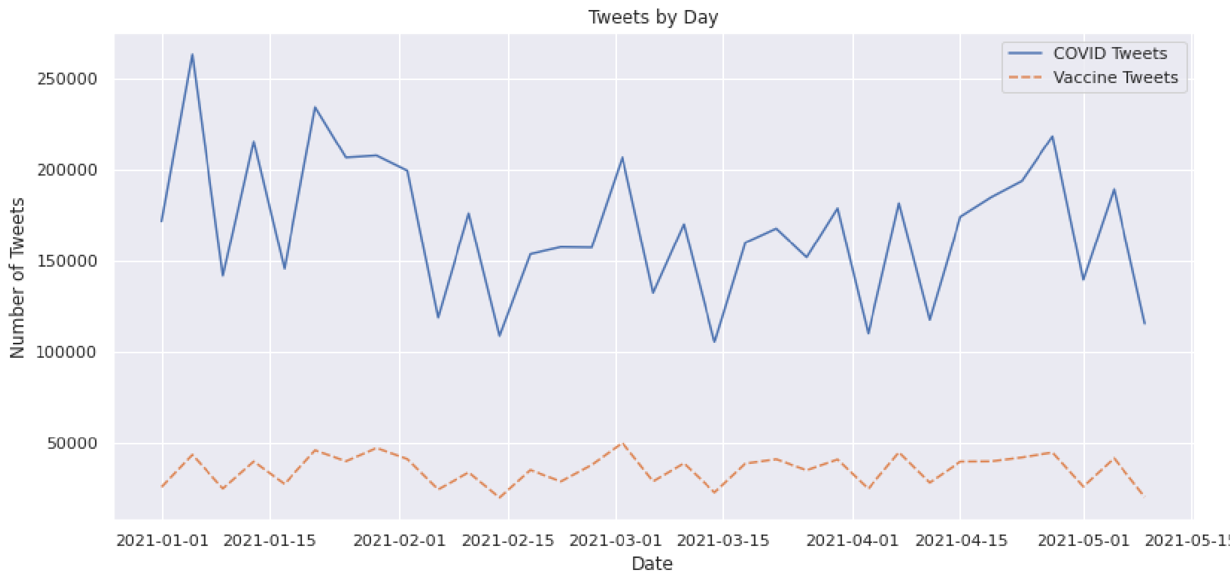


Figure 1. Count of COVID and vaccine-related tweets in 2021

4 Machine Learning and Solution Details

4.1: Sentiment Analysis

We have used python's package, Vader (Valence Aware Dictionary and sEntiment Reasoner), to label each cleaned tweet. VADER is a "lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media." The lexicon is sensitive to "both the polarity and the intensity of sentiments expressed in social media contexts" and can deal with more complex syntax such as negation. Using Vader, we assigned positive, negative, and neutral scores according to the words. These scores represent the percentage of either positive, negative, or neutral words in the text. An example: *VADER is smart, handsome, and funny.*

This sentence is assigned the scores:
{'pos': 0.746, 'compound': 0.8316, 'neu': 0.254, 'neg': 0.0}
This means that the sentence is 74.6% positive, 0% negative and 25.4% neutral. We can confirm that 74.6 + 25.4 = 100. [7] The compound score is calculated by summing up valence scores of the words, adjust them with rules such as negation and by normalizing it to be between -1 and 1, using following equation: $x = \frac{x}{\sqrt{x^2 + \alpha}}$, where x is sum of valence scores and α is a normalization constant, which is set

[7] Hutto, C., & Gilbert, E. (n.d.). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Retrieved from http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

to 15 by default. [8] We flagged the tweets, by convention, positive if the 'compound' score is equal or higher than 0.05, negative if the compound score is equal or less than -0.05 as negative, and neutral otherwise.

4.2: Latent Dirichlet Allocation

For the second component of our analysis, determining common causes of vaccine hesitancy, we employ the unsupervised technique Latent Dirichlet Allocation to identify the most common topics represented by the tweets. Given a set of documents - which, in the case of our project, is a set of tweets - LDA generates a user-specified number of topics which it determines were likely used to create the documents. Each topic defined by LDA is a grouping of tokens - individual words or n-grams - assigned a probability of occurring in documents represented by the topic. The tokens of a topic together identify the topic's content as a whole, and thus LDA is a useful tool for identifying and linking common content over a range of sources.

We determined LDA would be a useful tool to attempt to identify common sources of vaccine hesitancy as it should enable us to summarize the most common themes across our tweet dataset from January, 2021 to the present. Each topic identified should be a belief, feeling, or shared event which was held in common among a large proportion of English-speaking Twitter users. Once these topics are defined by the LDA model, as human interpreters we can subsequently identify and clarify the corresponding theme. We felt this would be particularly useful after performing sentiment analysis on this twitter corpus to single out the tweets relating to vaccines which are negative: once we'd narrowed down our corpus to negative tweets, we hoped to have the best chance of identifying common causes of this negativity, some of which may translate to reasons for hesitancy.

For the actual implementation of our LDA models in Python, we primarily used Gensim's LDA implementation (gensim.models.LdaModel)[9],[10] and Gensim's wrapper for MALLET's LDA (gensim.models.wrappers.LdaMallet).[11] Gensim is a popular LDA implementation which uses Variational Bayes in order to determine the probabilities of tokens belonging to a topic, while MALLET's LDA is a Java-based implementation which uses Gibb's sampling to determine topic assignment instead. Generally, Gibb's sampling provides full consideration of the entire corpus in classifying each token, which requires holding the entire dataset in-memory and can be computationally expensive, particularly on larger datasets. Variational Bayes is faster, as it considers subsets of the data at a time, but it can become stuck at local minima (rather than the true minima which would optimize the allocation to topics) in some scenarios.

The implementation of LDA first requires the creation of a dictionary which maps each token to a unique identifying number. The resulting data structure is of the form {'hundreds doses': 0, 'million people': 1, …}, where each key is a token and each value is its id. The dictionary can be utilized in this full state, where each token in every document is included, or the dictionary can be filtered to remove the least

---

[8] Source code for nltk.sentiment.vader. (n.d.). Retrieved June 02, 2021, from https://www.nltk.org/_modules/nltk/sentiment/vader.html#SentimentIntensityAnalyzer.score_valence

[9] Řehůřek, Radim, and Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA.

[10] Hoffman, Matthew, Francis Bach, and David Blei. 2010. "Online Learning for Latent Dirichlet Allocation." In Advances in Neural Information Processing Systems, edited by J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta. Vol. 23. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2010/file/71f6278d140af599e06ad9bf1ba03cb0-Paper.pdf.

[11] McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

common words, remove the most common words, and use only a pre-specified number of words. This is to improve the targeting of important words and can significantly decrease the time required to model as well. While our pre-processing included removing stopwords and other common words which we felt would add noise rather than clarity, this ability to filter provides another layer of control in determining the content to model. The dictionary is then used to build a corpus which identifies the frequency of each term in a document. The corpus is structured as follows:

```
[
[(0,2),(22, 1), ...],
[(1,1),(503,2),...],...
]
```

Each list within the overall corpus represents a single document, and each tuple within a document's list contains the id of a token and the count of times the token appears in the document. In the example above, the first document has two occurrences of the token 'hundred doses' (with id 0), one occurrence of the token with id 22, and so on.

A corpus can either be used as a set of counts, as in the above example, or it can be refined further by turning it into a Term Frequency-Inverse Document Frequency (TF-IDF) matrix. This causes common words to be weighted less heavily and amplifies the presence of more uncommon words, which acts as another means of attempting to isolate the most important components of each tweet.

We created our dictionaries and corpuses from our entire dataset, every 4th day's tweets from January to May 2021. Once created, they can be used in an LDA model along with other hyperparameters to build out the topics present in the set of documents. We conducted multiple grid searches to select the optimal model inputs, models, and corresponding hyperparameters.

Overall, we tested models using variations on model type (Gensim vs MALLET), token (single word vs digram), dictionary filters (no filters, removing bottom 200 or 400, removing top 20%), and corpus type (count vs TF-IDF). Given the short length of tweets, we primarily considered single words and digrams, as opposed to any longer n-gram options, in order to allow for a greater volume of tokens to be available for allocation within each tweet. Many LDA users found success when filtering their dictionary, so we tested a few filtering options as well on both smaller sample datasets and on our complete set of tweets. The filtering is particularly useful when using a corpus containing pure counts of occurrences, as it allows for an adjustment by token popularity and fulfils a similar purpose as TF-IDF. Other ML practitioners also saw benefit in using a dictionary filter as well as TF-IDF corpuses, however, so we did also test this combination.

Each combination above was tested in a separate grid search, as the runtime for training these models (between 5-20 hours per limited grid search) made it infeasible to cycle through all combinations of the above as well as through the hyperparameters of the model itself. In particular, within each grid search we varied the number of topics and the alpha and beta hyperparameters. The alpha hyperparameter represents the likely distribution of topics within a single document - ie, whether a document is more likely to contain only a few of the topics or whether it is likely to contain elements of most of the topics. The beta hyperparameter - which is only used in Gensim's LDA as MALLET's Gibbs Sampling method does not require it - is the distribution of words per topic, or whether each topic representation is likely to contain only a few of its representative words or most of its words. While we initially believed that our specific Twitter dataset would likely have a low alpha, given the concise nature of tweets, each of our grid searches showed that different combinations of both low and high alpha and beta parameters could produce optimal results, so we didn't feel we should narrow down the range of alpha or beta hyperparameters in training (as initially we hoped to do for performance). The number of topics also demonstrated success at multiple values, depending on the other parameters, so we felt we should keep this relatively broad as well. While it would have been ideal to test a full range of topic numbers in each grid search, again due to performance concerns we limited the possibilities to 5, 10, or 15. We felt over 15 would be difficult to interpret and severely hamper performance, while below 5 would likely be inaccurate given the large dataset we were modeling.

5 Evaluation and Results

5.1: Sentiment Analysis

We plotted graphs that show the number of tweets of any sentiment, positive sentiment, neutral sentiment, and negative sentiment (Appendix 1). We have two data points each week and aggregated them by week for this visualization. Overall, there are more neutral or positive tweets than negative tweets. In addition, the number of tweets increased rapidly in the first two weeks, showing that people's interest in vaccines rapidly increased as vaccine distribution accelerated. In contrast, the number of tweets has constantly been decreasing since mid-April. It is consistent with the search trend on Google of the word 'covid vaccine.'

Polarity scores per week (Appendix 2) show the intensity of peoples' sentiment. We have plotted the mean value of positive, negative, neutral, and compound scores assigned to each tweet. From the fact that neutral scores consistently exceed positive or neutral scores, it can be inferred that the tweets are mostly neutral during the period. On the other hand, the compound scores used to flag positivity go up in January and come down slowly towards mid-April. As the compound score below -0.05 is interpreted as negative text, negativity is dominant from mid-march to mid-April. Then the sentiment improved rapidly to mid-May in a month.

The sentiment analysis is a heuristic process, as tweets are not 'clean' and requires background understanding to interpret. The main takeaways from sentiment analysis are as follows.
The data included many repeated tweets. It suggests that the sentiment on Twitter can be volatile, as a specific tweet quickly dominates them, and the same tweet is assigned the same polarity. In Particular, people can retweet to express their agreement or disagreement, which does not appear in the text unless people add some comments when quoting. The understanding of sentiment of a tweet is high-contextual; we consider a user's background and past tweets into consideration...

- Many tweets referencing the fight against the vaccine and calling on people to come together to defeat it are understandably classified as negative yet would likely be considered positive by human observers. Sentiment labeling on Vader is helpful for processing large-scale data but may contradict human sentiment.
- Many tweets were labeled as neutral, which makes the result more difficult to interpret. One solution would be to customize the dictionary by adding some words and scores on our own. The alternative is to manually label some data and label other tweets using a machine learning method, but this approach requires those who label not to be partial.

5.2: Latent Dirichlet Allocation

In evaluating the LDA models in each of our grid searches, we primarily considered the Coherence score to identify the best models of the varieties considered. Coherence provides a value between 0 and 1 which is intended to score how well the topics form a comprehensive, understandable grouping, and the

Coherence score is widely-used in evaluating topic models. Another popular metric, Perplexity, aims to evaluate the success of the model in predicting a sample, but from the recommendations of TAs and others online, as well as our own evaluation of the usefulness of both scores, we determined to use Coherence as our primary evaluation metric. This is in line with a general consensus in topic modeling that Perplexity produces scores which are less aligned with human judgement of topic successes. While each grid search produced models scored by Coherence, we found that the top topic model by Coherence was usually not the most meaningful for human observers. Instead, we used the Coherence score to narrow the set of topics produced by each grid search to the top 10-15 and subsequently evaluated these manually to determine the topic model which allowed for the most meaningful interpretation.

The models which ultimately performed the best (by Coherence and manual inspection) utilized digrams rather than single words and utilized the Gensim LDA rather than MALLET's. While MALLET performed better than Gensim with the default parameters, the tailoring of parameters which occurred during the grid searches enabled Gensim's LDA to outperform MALLET in most configurations. Filtering out the top- and bottom-most frequently occurring tokens also proved to generally be more effective than models using the full set of tokens, even when TF-IDF was employed. The models which were both most informative and most relevant to the study of sources of vaccine hesitancy tended to utilize filters.

The model we ultimately determined was the most aligned with our priorities came from a grid search utilizing filters (no words occurring fewer than 200 times or in more than 80% of tweets) and TF-IDF. The best model among the group produced a Coherence score of 0.6279 and included 5 topics (Appendix A3). However, while this produced the highest coherence score and there were 2 topics which referenced sources of vaccine hesitancy (primarily adverse reactions and blood clots), overall this model produced topics with too broad a scope to enable successful analysis of the topics as well-defined, self-contained entities. We also felt it did not cover enough of the range of sentiments people have expressed surrounding the vaccines.

Instead, we determined a model with a lower Coherence score of 0.5973, yet higher human-comprehensibility and with 10 topics, was the best representation of the tweets over time. While this model is still fairly messy - with the majority of tweets being assigned to Topic 1, which appears to be the catch-all cluster, and the others being grouped into topics which each contain a few distinct themes - this model did enable us to gain some insight into the trends which have dominated the English-speaking Twitter sphere since January. We ultimately classified 249,710 tweets labeled as having negative sentiment with this model.

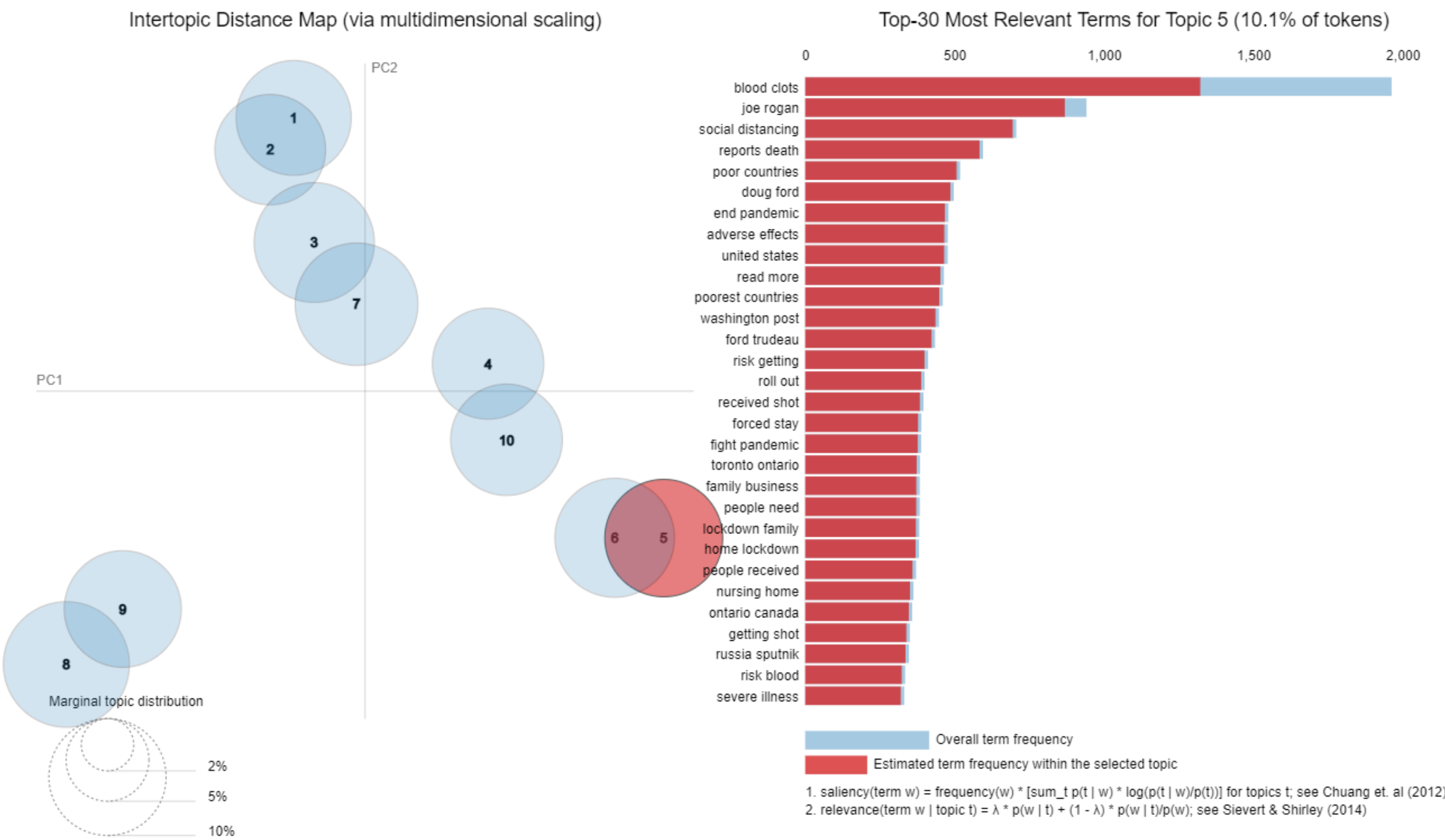


Figure 2. Topic Model of Negative Covid Vaccine Tweets. Model derives 10 topics and represents a sample of tweets spanning Jan 01, 2021 to May 9, 2021.

Figure 2 presents a visual representation of the topic breakdown by distance, while the official topics as defined by keywords are provided in Appendix A4, Table A4.1. Descriptions of the common trends within each topic, as determined by the topic keywords and a manual inspection of the top 200 tweets in each topic, is presented below (excluding Topic 1, the catch-all topic, which is too varied to permit identification of distinct themes):

Common Themes per Topic:

2: People are dying from the Covid vaccine; India ramping up vaccination and fighting vaccine hesitancy/misinformation; young people are dying from Covid; young people don't need the vaccine (Joe Rogan); there will always be new variants; people are selling fake vaccine cards to travel

3: Vaccines are weapons of mass destruction (Dr. Coleman); J&J effective against severe disease; discussion around delaying second dose; various commentaries on Biden (don't trust rush job on vaccine/Biden shouldn’t get credit/Biden giving vaccines to Guantanamo detainees/convincing Biden to prioritize cancer patients)

4: Drop vaccine patents (governments, big pharma); making vaccines affordable/accessible; comparing covid vaccine to flu - long term solution, people who get flu shots more likely to get covid shots; description of vaccination's mechanisms (spike protein)

5: Canada/Doug Ford not approving Russian vaccine when the situation's bad in Ontario; discussion around blood clots - vaccines cause clots/risk of blood clots higher with Covid; poor countries getting/not getting vaccine; Joe Rogan not a good source

6: Vaccines linked to blood clots vs low risk of clots; cases spreading - appeal to wear masks, wash hands, distance; doubt of reaching herd immunity; question of long-term vaccine effects

7: Advances in cancer research; adverse reactions to Covid vaccine (esp. Europe, Israel, US); New York's vaccination successes and challenges, variants in NY; doubting needing vaccine when it has '99% survival rate while vaccines are only 60%+ effective'

8: Tallies of cases and deaths; reports of emergency approval worldwide; lower income countries especially suffering/will suffer, low income countries not receiving vaccines or some being sent vaccines

9: Take/trust Ennaid Therapeutics Company's antiviral drug; docs/nurses giving vaccines are war criminals, vaccine is an experiment/don't trust; US/Europe export ban on raw materials for developing virus abroad

10: Millions won't have vaccines unless Biden reverses blockage of WTO emergency measures; Ron Johnson saying vaccine not needed/not safe; blood clots with vaccines vs vaccines are safer from blood clots; allergic reactions to Pfizer/Moderna; Florida news (people eligible, won't allow vaccine passports); debate over waiving patents

The distribution of topics in Figure 2 and the themes identified within each topic do support commonalities and overlap between topics, which is indicative both of the difficulty the model had in distinguishing precise boundaries between topics and the overall lack of cohesion in tweets more generally. Nevertheless, general trends are elucidated by this analysis. Topics 3 and 7, which display an overlap in distribution, both contain significant concerns with the vaccine and doubts about Covid. One interesting note is that Topic 3 - which otherwise contains concerns about the vaccine and the government's handling of the virus - also is assigned the bulk of the tweets surrounding Johnson & Johnson's trial results. This is likely driving the most significant spike in our dataset: Topic 3 in late January through early February (Appendix A4, Figure A4.1). This coincided not only with the release of the Johnson & Johnson vaccine results, but also with decisions by the US and other governments to limit vaccine exports due to concerns about supply, and with general concerns about the South African variant which the Johnson & Johnson vaccine helped combat. Another rise in Topic 3 occurred in mid-March, when a man who identified himself as Dr. Vernon Coleman produced a video claiming the vaccines were weapons of mass destruction.[12] While this video and surrounding conversation were unknown to our team members until this analysis, the prominence of this weapons of mass destruction claim is indicative of the reach and difficulties in eradicating such claims which drive many cases of vaccine hesitancy. Topic 7 contains similarly negative tweets about covid and the vaccine, with a particular focus on databases of adverse reactions which were circulating and doubts about the need for vaccines at all. Such anti-vaccine sentiment was also significantly present in Topic 2, which overlaps with Topic 7 as well. Claims of the vaccine causing death are most prevalent in this topic, as well as other

[12] Washington, District of Columbia 1100 Connecticut Ave NW Suite 1300B, and Dc 20036. n.d. "PolitiFact - No, the COVID-19 Vaccines Are Not Weapons of Mass Destruction." @politifact. Accessed June 3, 2021. https://www.politifact.com/factchecks/2021/mar/31/facebook-posts/no-covid-19-vaccines-are-not-weapons-mass-destruct/.

misleading ideas such as Joe Rogan's proclamation that young people don't need the vaccine.[13] Likely linked to this misinformation, which was called out as such in many tweets, this topic also discusses India's campaign against vaccine misinformation.

Topics 4 and 10 share some overlap which seems primarily driven by the issue of waiving patents for covid vaccines in order to enable more widespread manufacturing. However, the model finds a distinction between the conversation which was happening around patents earlier in the year (Topic 10), which was also linked to conversations around the US allowing emergency manufacturing measures, and the patent discussion which occurred more recently at the beginning of May (Topic 4). While the other themes linked in Topic 4 appear to be less cohesive, the distinction which the model finds between these two strands of patent discussions does enable an interesting analysis of the different contexts in which both took place. Topics 5 and 6 also share some overlap, primarily around the issue of blood clots. There are a range of opinions represented within these topics, from distrust of the vaccines due to clots to statements that the clotting risk is very low. In the chart of these vaccines over time, the extensive public discussion around these issues is clearly represented in the rise of Topic 5 toward the beginning of April and the spike in Topic 6 at the middle of April, when the Johnson and Johnson vaccine was put on hold due to the clotting investigation.

Topics 8 and 9 share overlap both in expressing a lack of trust in the vaccine and in discussions of world powers sharing or refusing to share vaccines and materials. Another interesting trend which appears in Topic 9 was Ennaid Therepeutic's announcement of their development of ENU200, an antiviral, oral medication for Covid-19.[14] This occurred in early April, and while it did not get much notice at least from our team members, it appears to have been influential in Twitter conversations around the vaccine. The spike in early May of Topic 8 was also more influential than we might have expected, with the likely primary driver being the WHO's emergency use approval of Moderna as well as other emergency approvals which were happening worldwide at that time.

Overall, in addition to providing a greater understanding of the difficulties present in topic modeling, especially on documents (tweets) which are very limited in size and content, this model did allow us to appreciate some of the ways ideas, claims, and themes can spread and gain enormous influence on social media while failing to obtain prominence elsewhere. The influence of individuals was particularly on display, with large proportions of multiple topics being devoted to praise (and disparagement, fortunately) for the ideas of certain individuals: Dr. Coleman equating vaccines to weapons of mass destruction in Topic 3, Joe Rogan's questioning of the vaccine for young people in Topic 2, and Senator Ron Johnson's doubts of the vaccine's safety and necessity in Topic 10.[15]

The broad influence of these individuals' claims, which can be easy to dismiss when not immersed in this environment due to the low-to-moderate prominence of the individuals themselves, demonstrates both the risks and opportunities available through social media and analysis like topic modeling. While such influence can have a frightening reach, these individuals and their claims do present a common

[13] Andrews, T. (2021, April 29). Joe Rogan is using his wildly popular podcast to QUESTION Vaccines. experts are fighting back. Retrieved June 02, 2021, from https://www.washingtonpost.com/technology/2021/04/28/joe-rogan-podcast-vaccine-coronavirus/

[14] "Ennaid Therapeutics Announces Development of ENU200, A New Antiviral Therapeutic for the Treatment of COVID-19." n.d. Accessed June 3, 2021. https://www.prnewswire.com/news-releases/ennaid-therapeutics-announces-development-of-enu200-a-new-antiviral-therapeutic-for-the-treatment-of-covid-19-301034232.html.

[15] The Vicki McKenna show - Are americans SOVIET? - THE Vicki McKenna Show. (n.d.). Retrieved June 02, 2021, from https://www.iheart.com/podcast/139-vicki-mckenna-27246267/episode/the-vicki-mckenna-show-are-82266303/

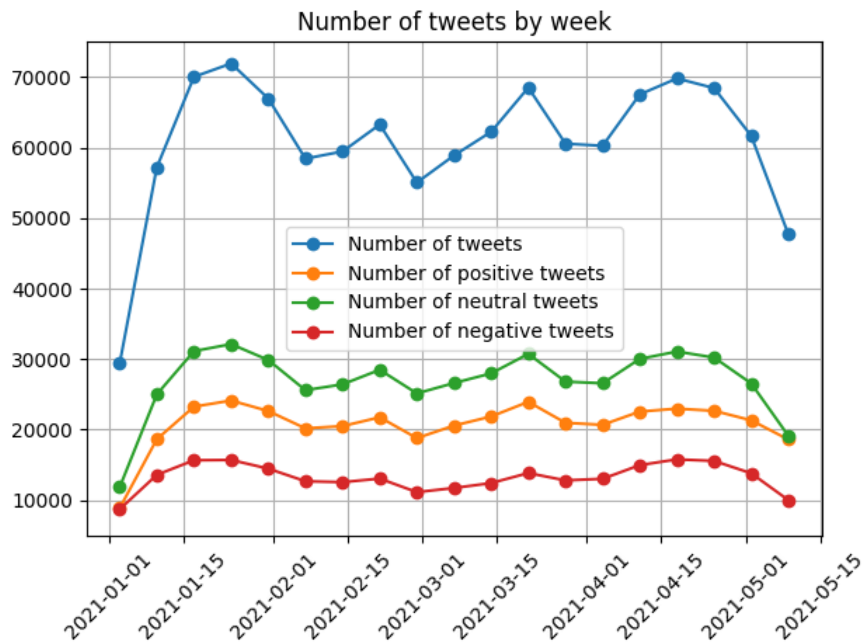
source of untruths which can be addressed and combated by policy makers and other influencers working to promote vaccinations and Covid safety. While the ethics and legality of social media regulation are also heavily debated, in theory the early identification of individuals or claims such as these could also be targeted by regulators at social media companies or in government for suppression. It's also possible - though perhaps more idealistic than realistic - that identifying these individual influencers early could allow for persuasion campaigns targeted at them directly, to try to convince them to moderate their own content.

Another factor which was illuminated by these trends is the scale of civic engagement around these issues: the people fighting against misinformation claims or those attempting to persuade the government to take certain actions. This is actually extremely encouraging and presents opportunities for policy makers to develop greater engagement with their constituents. At the basic level, policy makers could adopt this kind of social media topic modeling analysis within their offices to identify less-obvious, yet still relevant, trends or issues which they might pursue or use to connect with their base. There also seems to be an opportunity to use this sort of topic modeling to expand the methods of civic engagement available to citizens, potentially by developing a separate platform or environment which seeks citizen input on topics which are identified through analysis of conversation trends. This would provide a chance for citizens to channel their energy - which is currently heavily expended in responses to social media debates - towards relevant, productive policy challenges of particular interest to them by providing input or suggesting actions for decision makers.

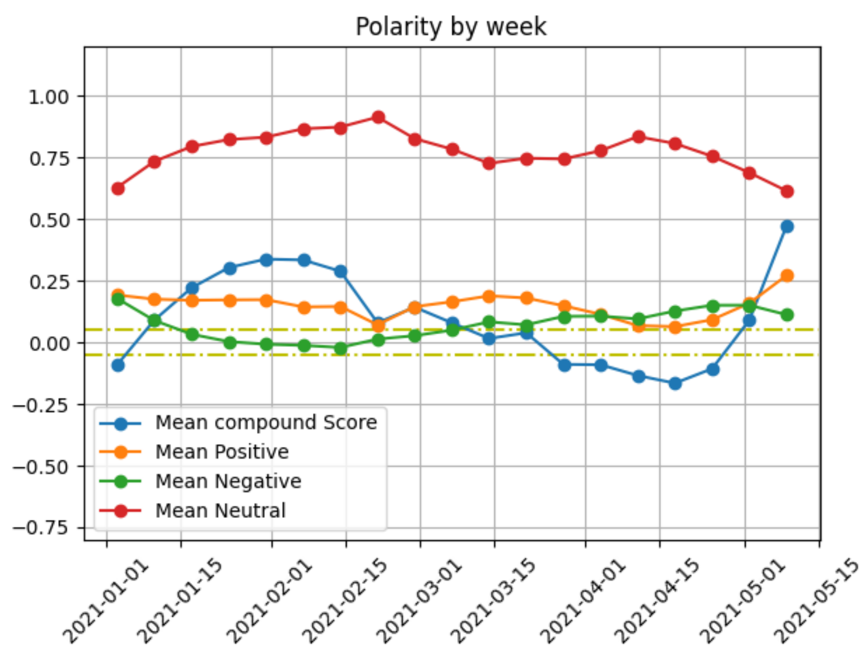
In order to expand this research on vaccine hesitancy further, for policy makers or other interested parties, it might be beneficial to significantly expand the number of topics under consideration and manually identify the few among them which relate precisely to vaccine hesitancy. While we limited our consideration of topics to a maximum of 15 due to concerns of increasing the complexity of the model and due to performance limitations, one potential avenue for exploration could be increasing the number of topics (perhaps considering 30-80), which would allow for much greater cohesion of individual topics. The topics most relevant to vaccine hesitancy could then be manually identified and traced over time and across sources.

Appendix

Appendix A1



Appendix A2



Appendix A3

The top-Coherence model from our chosen grid search

Topics	Terms per Topic
Topic1	private school, blood clot, wear mask, fake cards, health officials, pfizer biontech, adverse reactions, people dying, fully people, sars cov
Topic2	emergency use, people died, raw materials, young people, new variants, use listing, use authorization, social media, age group, second shot
Topic3	johnson johnson, blood clots, second dose, herd immunity, high risk, woman died, rare blood, people die, wearing mask, blood clotting
Topic4	patent protections, big pharma, long term, ennaid therapeutics, therapeutics company, survival rate, stay home, cases deaths, suspend patents, spike protein
Topic5	new cases, public health, second wave, anti vax, year old, death toll, biden administration, million doses, pfizer moderna, cases total

Appendix A4

Table A4.1. Our selected best model

Topics	Count of Tweets (Percent)	Terms per Topic
Topic1	155,787 (0.6239)	johnson johnson, second wave, people dying, fully people, wear masks, export ban, million people, joe biden, right now, people fully, cdc says, moderna pfizer, human rights, world leaders, real world, kill people, number people, dose pfizer, slow rollout, president joe
Topic2	9,149 (0.0366)	fake cards, people died, young people, new variants, million doses, sars cov, selling fake, people don, single dose, age group, adverse events, people getting, people infected, health minister, years age, clinical trials, adverse event, life saving, getting sick, pfizer astrazeneca
Topic3	13,454 (0.0539)	second dose, anti vax, high risk, biden administration, negative test, anti vaxxers, severe disease, low risk, fact check, killing people, doses administered, spread virus, health crisis, rest world, south africa, health experts, surge cases, cancer patients, know people, death rates

Topic4	8,940 (0.0358)	private school, big pharma, suspend patents, spike protein, death rate, stop spread, quickly possible, drop patents, flu shot, governments big, pharma drop, bbc news, affordable everyone, patents affordable, everyone quickly, gene therapy, don want, wearing masks, stay home, save lives
Topic5	12,055 (0.0483)	blood clots, joe rogan, social distancing, reports death, poor countries, doug ford, end pandemic, adverse effects, united states, read more, poorest countries, washington post, ford trudeau, risk getting, roll out, received shot, forced stay, fight pandemic, toronto ontario, family business
Topic6	11,664 (0.0467)	wear mask, herd immunity, year old, long term, rare blood, dont want, second shot, wash hands, pregnant women, prime minister, blood clots, disease control, breaking news, risk dying, people got, essential workers, fox news, social distance, term effects, sore arm
Topic7	10,789 (0.0432)	pfizer biontech, adverse reactions, survival rate, people die, federal government, new york, got shot, fake news, adverse drug, global pandemic, drug reactions, database adverse, dead injuries, european database, getting infected, injuries european, age groups, president biden, dont need, looks like
Topic8	11,982 (0.0480)	emergency use, new cases, use listing, cases total, cases deaths, income countries, use authorization, social media, blood clotting, higher risk, low income, poorer countries, active cases, book appointment, total cases, health ministry, world health, fatality rate, reports new, recoveries death
Topic9	7,650 (0.0306)	raw materials, health officials, ennaid therapeutics, therapeutics company, woman died, wearing mask, medicine cure, cure produced, die you, produced ennaid, prevailed love, company die, usa humanity, ego prevailed, racism ego, warned usa, fake providing, humanity racism, listen warned, immunity medicine
Topic10	8,240 (0.0330)	patent protections, public health, blood clot, pfizer moderna, ron johnson, dont know, adverse reaction, millions people, people world, new strains, department health, clot risk, allergic reactions, sen ron, donald trump, state health, trump administration, rich countries, people risk, live news

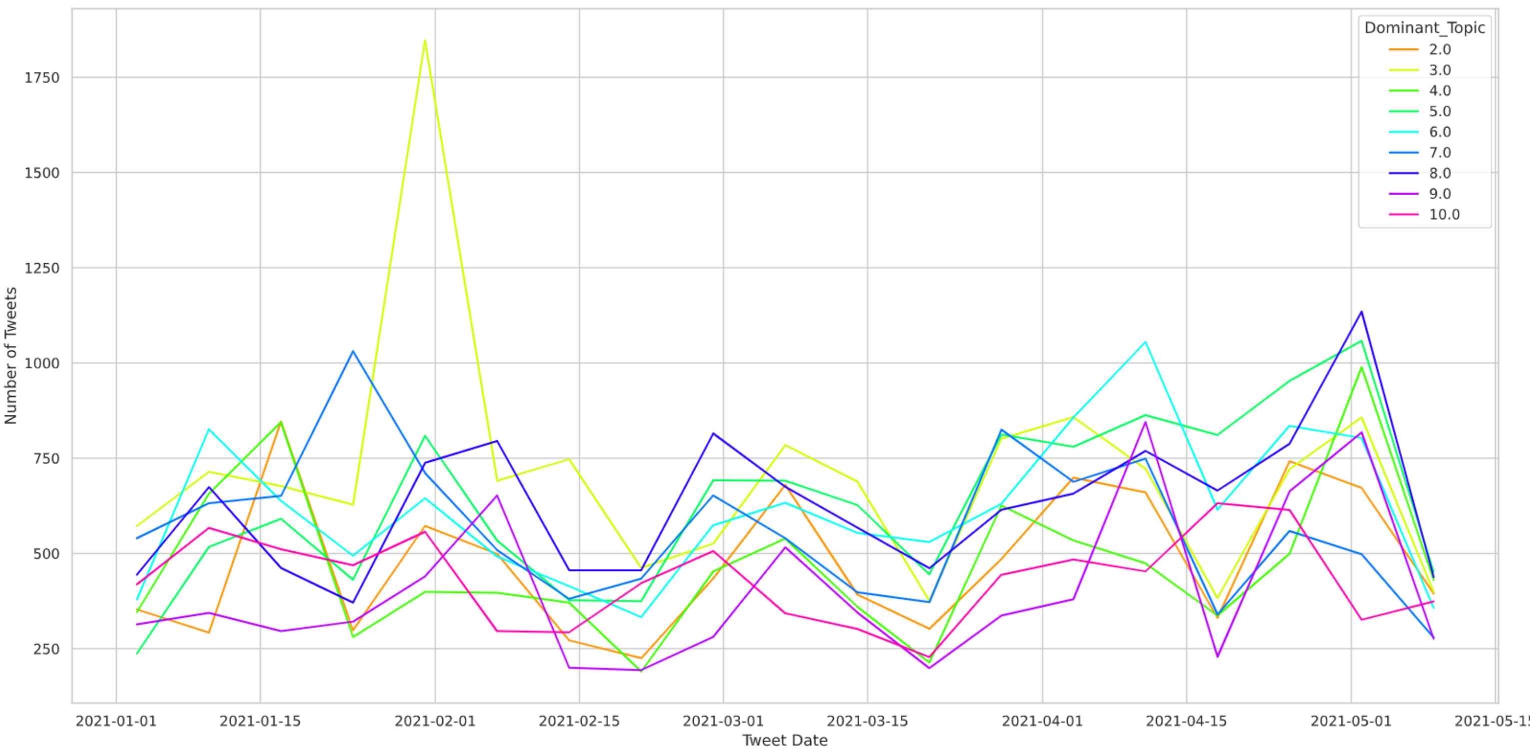


Figure A4.1. Tweet Counts by Topic. Excludes Topic 1, which acts as the catch-all for the majority of tweets and is at a scale that dwarfs the others.

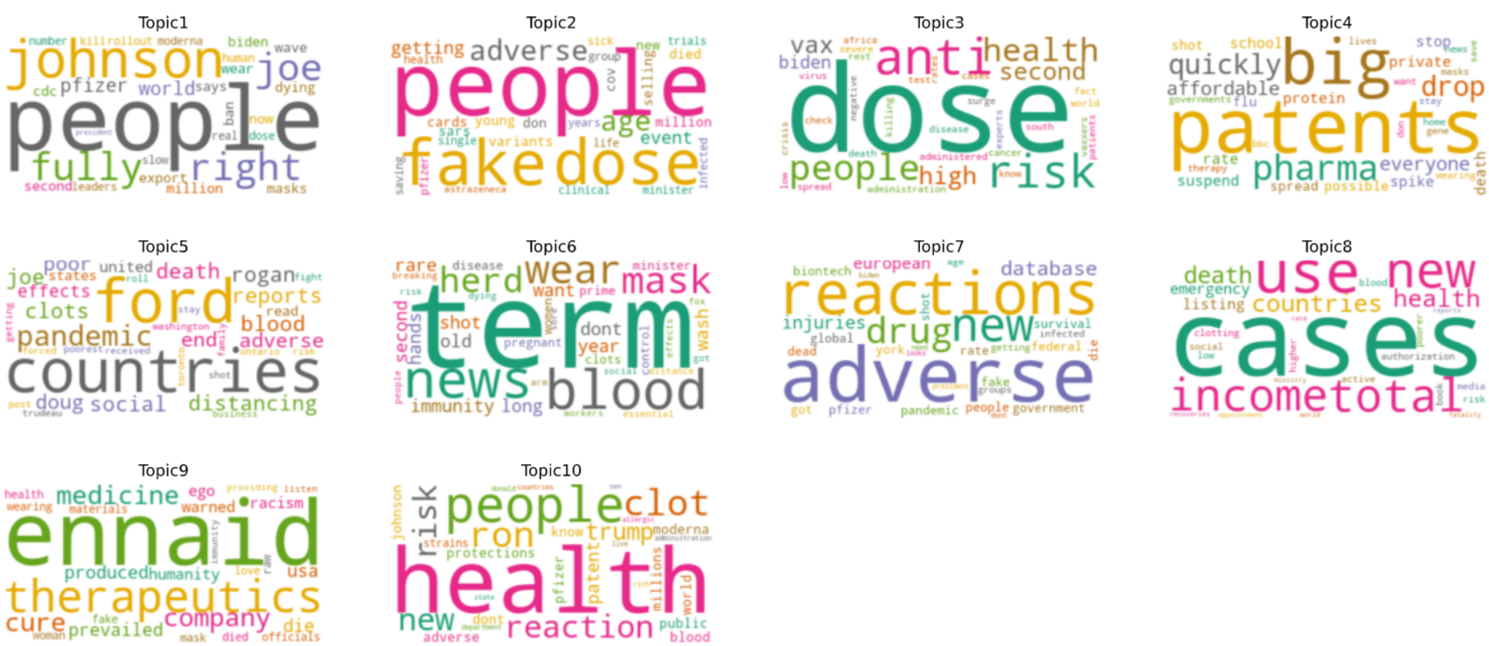


Figure A4.2. Wordcloud representations of topics in our chosen LDA model.

References

Andrews, T. (2021, April 29). Joe Rogan is using his wildly popular podcast to QUESTION Vaccines. experts are fighting back. Retrieved June 02, 2021, from https://www.washingtonpost.com/technology/2021/04/28/joe-rogan-podcast-vaccine-coronavirus/

Banda, Juan M., Tekumalla, Ramya, Wang, Guanyu, Yu, Jingyuan, Liu, Tuo, Ding, Yuning, … Chowell, Gerardo. (2021). A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration (Version 64) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4876538

Dovere, Edward-Isaac. 2021. "Vaccine Refusal Will Come at a Cost—For All of Us." The Atlantic. April 10, 2021. https://www.theatlantic.com/politics/archive/2021/04/vaccine-refusal-hesitancy-economic-costs/618528/.

"Ennaid Therapeutics Announces Development of ENU200, A New Antiviral Therapeutic for the Treatment of COVID-19." n.d. Accessed June 3, 2021. https://www.prnewswire.com/news-releases/ennaid-therapeutics-announces-development-of-enu200-a-new-antiviral-therapeutic-for-the-treatment-of-covid-19-301034232.html.

"Half Of All U.S. Adults Are Now Fully Vaccinated Against COVID-19." n.d. NPR.Org. Accessed June 3, 2021. https://www.npr.org/sections/coronavirus-live-updates/2021/05/25/1000171685/half-of-all-u-s-adults-will-be-fully-vaccinated-against-covid-19-as-of-tuesday.

Hoffman, Matthew, Francis Bach, and David Blei. 2010. "Online Learning for Latent Dirichlet Allocation." In Advances in Neural Information Processing Systems, edited by J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta. Vol. 23. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2010/file/71f6278d140af599e06ad9bf1ba03cb0-Paper.pdf.

Hutto, C., & Gilbert, E. (n.d.). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Retrieved from http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

Kurt, Senol. 2020. "Topic Modeling — LDA Mallet Implementation in Python — Part 2." Medium. July 1, 2020. https://medium.com/swlh/topic-modeling-lda-mallet-implementation-in-python-part-2-602ffb38d396.

Li, Susan. 2018. "Topic Modeling and Latent Dirichlet Allocation (LDA) in Python." Medium. June 1, 2018. https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24.

M, Sahil. n.d. "NLP-A Complete Guide for Topic Modeling- Latent Dirichlet Allocation (LDA) Using Gensim! | LinkedIn." Accessed June 3, 2021. https://www.linkedin.com/pulse/nlp-a-complete-guide-topic-modeling-latent-dirichlet-sahil-m/.

McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

Naskar, Anindya. 2019. "Guide to Build Best LDA Model Using Gensim Python." ThinkInfi. August 15, 2019. https://thinkinfi.com/guide-to-build-best-lda-model-using-gensim-python/.

Prabhakaran, Selva. 2018. "Topic Modeling in Python with Gensim." ML+ (blog). March 26, 2018. https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/.

"Python | Convert a String Representation of List into List." 2019. GeeksforGeeks (blog). February 18, 2019. https://www.geeksforgeeks.org/python-convert-a-string-representation-of-list-into-list/.

"Python - No Module Named PyLDAvis." n.d. Stack Overflow. Accessed June 3, 2021. https://stackoverflow.com/questions/66759852/no-module-named-pyldavis.

"Quality Control for Banking Using LDA and LDA Mallet." n.d. Accessed June 3, 2021. https://mickzhang.com/quality-control-for-banking-using-lda-and-lda-mallet/.

Rafferty, Greg. 2019. "LDA on the Texts of Harry Potter." Medium. May 22, 2019. https://towardsdatascience.com/basic-nlp-on-the-texts-of-harry-potter-topic-modeling-with-latent-dirichlet-allocation-f3c00f77b0f5.

Řehůřek, Radim, and Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA.

Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 399–408. WSDM '15. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2684822.2685324.

"Social-Desirability Bias." 2020. In Wikipedia. https://en.wikipedia.org/w/index.php?title=Social-desirability_bias&oldid=992112847.

Source code for nltk.sentiment.vader. (n.d.). Retrieved June 02, 2021, from https://www.nltk.org/_modules/nltk/sentiment/vader.html#SentimentIntensityAnalyzer.score_valence

The Vicki McKenna show - Are americans SOVIET? - THE Vicki McKenna Show. (n.d.). Retrieved June 02, 2021, from https://www.iheart.com/podcast/139-vicki-mckenna-27246267/episode/the-vicki-mckenna-show-are-82266303/

Washington, District of Columbia 1100 Connecticut Ave NW Suite 1300B, and Dc 20036. n.d. "PolitiFact - No, the COVID-19 Vaccines Are Not Weapons of Mass Destruction." @politifact. Accessed June 3, 2021. https://www.politifact.com/factchecks/2021/mar/31/facebook-posts/no-covid-19-vaccines-are-not-weapons-mass-destruct/.