RESEARCH ARTICLE

Characterizing polarization in online vaccine discourse—A large-scale study

Bjarke Mønsted[1], **Sune Lehmann**[1,2]*

1 Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark, **2** Center for Social Data Science, University of Copenhagen, Copenhagen, Denmark

* sljo@dtu.dk

Abstract

Vaccine hesitancy is currently recognized by the WHO as a major threat to global health. Recently, especially during the COVID-19 pandemic, there has been a growing interest in the role of social media in the propagation of false information and fringe narratives regarding vaccination. Using a sample of approximately 60 billion tweets, we conduct a large-scale analysis of the vaccine discourse on Twitter. We use methods from deep learning and transfer learning to estimate the vaccine sentiments expressed in tweets, then categorize individual-level user attitude towards vaccines. Drawing on an interaction graph representing mutual interactions between users, we analyze the interplay between vaccine stances, interaction network, and the information sources shared by users in vaccine-related contexts. We find that strongly anti-vaccine users frequently share content from sources of a commercial nature; typically sources which sell alternative health products for profit. An interesting aspect of this finding is that concerns regarding commercial conflicts of interests are often cited as one of the major factors in vaccine hesitancy. Further, we show that the debate is highly polarized, in the sense that users with similar stances on vaccination interact preferentially with one another. Extending this insight, we provide evidence of an epistemic echo chamber effect, where users are exposed to highly dissimilar sources of vaccine information, depending the vaccination stance of their contacts. Our findings highlight the importance of understanding and addressing vaccine mis- and dis-information in the context in which they are disseminated in social networks.

Introduction

Vaccine hesitancy, defined as the reluctance or refusal to vaccinate [1], is a growing threat to global health, and is believed to be driven mainly by the 'three C's': Confidence, Complacency, and Convenience [2]. Social media platforms may potentially influence vaccine hesitancy through the former two, for example by enabling easy and wide-spread sharing of content that exaggerates the risks of vaccination and/or understating the risk of vaccine-preventable diseases [3]. Vaccine hesitancy exist on a continuous spectrum [4], where the extreme positions of rejecting or accepting all vaccines tends to be overrepresented in online settings [5].

While vaccine hesitancy is a nuanced and context-dependent phenomenon, some general factors influencing hesitancy have been identified in the literature [6]. Key among those factors are the availability of information regarding vaccines [4, 7], the accuracy of beliefs about the risks and benefits of vaccines and vaccine-preventable diseases [8, 9], social norms regarding vaccination, i.e. whether or not vaccinating is perceived as a 'normal' thing to do [5, 10], and trust in health authorities and/or the pharmaceutical industry, particularly concerns regarding commercial conflicts of interest [4, 7, 8]. The list above is by no means exhaustive, rather a few central factors which are well-described in the literature and of particular relevance for this paper.

These factors are strongly connected to the topic of social networks and online misinformation: Anti-vaccine messages on Twitter typically aim to alter the reader's perception of risks and benefits regarding vaccination, often drawing on conspiracy theories [9]. In addition to an inaccurate risk picture, anti-vaccine content on Twitter, especially during the COVID-19 pandemic, has focused on commercial interests in the pharmaceutical sector [11], and often rely on conspiracy theories in doing so [12].

The detrimental effects of reduced vaccine uptakes on public health are well described in the literature [13–18]. Somewhat paradoxically, vaccination rates have declined in part due to the success of vaccines in preventing disease, leading to complacency [19, 20]. However, online misinformation has also been linked to decreasing vaccine uptake [21–24], and outbreaks of vaccine-preventable diseases have been observed in areas where anti-vaccine activists have organized disinformation campaigns [25]. In this sense, the growing amount of online misinformation [26, 27] can be characterized as a threat to public health [28].

Countering medical misinformation in online systems is no easy task. While the scientific literature is rife with evidence which disproves the narratives outlined above [29, 30], individuals at the 'rejection' extreme of the vaccine hesitancy continuum often have a strong sense of identity regarding their stance on vaccines [31]. Individuals tend to reinterpret or disregard information if it conflicts with a stance that they strongly identify [32], an effect which has been demonstrated in numerous contexts [33] including vaccination [34].

The challenges of countering misinformation is compounded by the fact that strongly anti-vaccine individuals often form tightly knit communities in large social networks, such as Twitter [35, 36] and Facebook [37]. In such environments, evidence challenging in-group beliefs is dismissed as untrustworthy [8], and often ends up only reinforcing said beliefs [37, 38].

Therefore, study of the interplay between vaccination attitudes and vaccine-related online (mis)information is essential to inform policy [39–41], also at the community level [42].

We utilize two large datasets to study this interplay. The first (Dataset 1) is a large, random sample consisting of of approximately 60 billion tweets. The second (Dataset 2) consists of 6.75 million tweets obtained via Twitter's search API for tweets containing vaccination-related terms. Both datasets are discussed in detail in the Methods section. Using these datasets, we construct a large network which captures interactions on Twitter, and use machine learning methods to identify Twitter profiles with vaccine stances at the 'rejection' and 'acceptance' extremes of the hesitancy continuum, known colloquially as anti- and pro-vaxxers, respectively.

Based on the data and methods outlined above (which are elaborated upon in the materials and methods section), the remainder of the paper presents a number of analyses on the interplay between strong vaccination stances, social network structure, and online information.

Anti- and pro-vaccine profiles distinct types of URLs

We use a deep neural network to classify vaccine sentiments expressed in tweets from dataset 2, and identify 'antivaxx' and 'provaxx' profiles which consistently express highly negative and positive attitudes toward vaccination, respectively. Full details regarding the stance detection

methods are presented in the materials and methods section. In the following, we assess the degree to which anti- and provaxx users tend to rely on distinct types of outside sources, what characterizes these sources, and whether interactions occur disproportionally between profiles with similar stances. After estimating vaccine sentiment in the individual tweets, we estimate the vaccine *stance* of each individual profile. We define a profile's stance as provaxx or antivaxx if at least half of vaccine-related tweets posted by the profile are assigned a probability of at least 50% of expressing pro- and anti-vaccine sentiment, respectively. Approximately 48% of profiles were assigned pro- and antivaxx stances, as some profiles had posted only a few tweets regarding vaccination, or did not unambiguously express the same attitudes toward vaccination. In terms of the vaccine hesitancy continuum [4], these labels correspond loosely to the extremes, which tend to be over represented in online discussions [5]. Note that the correspondence is not exact, because hesitance is defined in terms of *accepting* or *rejecting* vaccines, whereas the labels used here refer to *attitudes* regarding human vaccination. However, intention to vaccinate is strongly influenced by attitudes regarding vaccination [43].

The distributions of tweet sentiments and profiles stances are shown in Fig 1. Of the vaccine-related tweets in English, approximately 35% could be identified as originating from the United States. An analysis at the state level of the vaccine sentiments expressed in tweets is provided in S2 Appendix. Details on data gathering, classification, and geolocation are provided in the materials and methods section.

Of those tweets, approximately 2.65 million contain external links (URLs outside of Twitter). Many such URLs start with a *base URL*, such as youtube.com, followed by a part which specifies which subpage (e.g. which particular video) the link points to, as well as various API calls, etc. We extracted the base URL for each such link, resulting in around 100 thousand
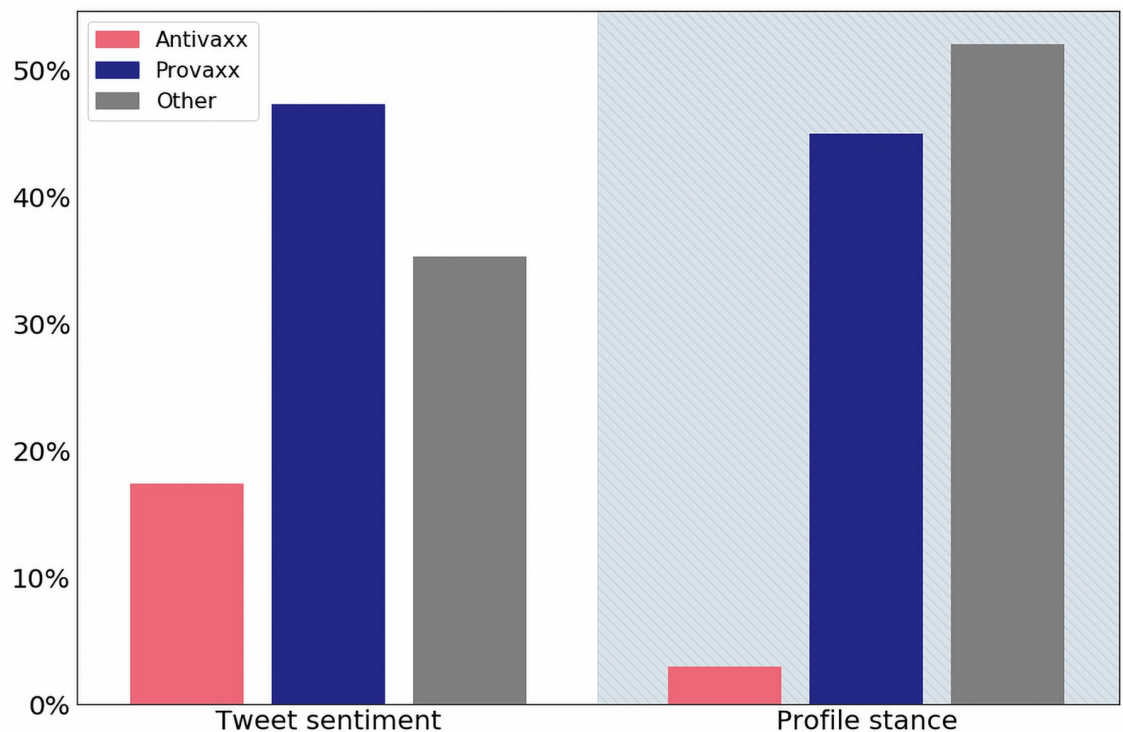
Fig 1. Distribution of tweet sentiment and profile stance. Tweets expressing anti-vaccine sentiment constitute an estimated 17% of vaccination-related tweets, where only about 3% of profiles stance are classified as antivaxx. Error bars are too small to depict visually, see S1 Appendix for uncertainty analyses.

https://doi.org/10.1371/journal.pone.0263746.g001
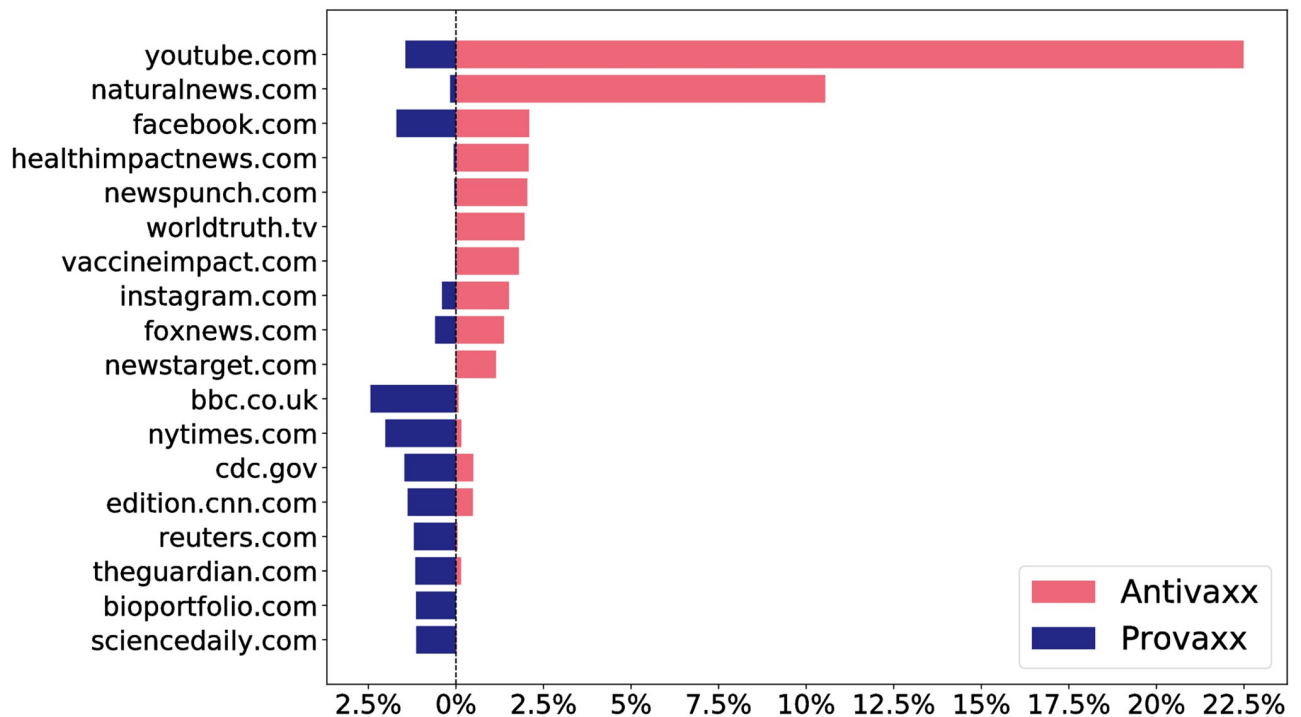
**Fig 2. The top 10 most linked to domains by strongly antivaxx and provaxx profiles.** Bar length shows percentage of the total number of links shared by profiles in the given category and hence do not sum to 100. For each domain, the red bars going right represent antivaxxers and blue bars going left provaxxers. Antivaxxers rely heavily on links to YouTube, and the page 'natural news', which promulgates pseudoscience and sells products related to health and nutrition. Provaxxers link to a wide array of news and science sites, which is why a lower overall percentage of their links are contained in the top 10. Error bars are too small to depict visually, see S1 Appendix for uncertainty analyses.

https://doi.org/10.1371/journal.pone.0263746.g002

distinct base URLs. Of those, we identified the 10 most frequently used by anti- and provax-xers, resulting in 18 URLs due to two domains (facebook.com and youtube.com) appearing among the ten most common URLs for both groups. The lists of top ten URLs contained 47% of links posted by antivaxx users, and 15% of links posted by provaxx users. Comparing the most frequently used base URLs with sentiment results reveals that profiles with different stances share highly dissimilar content, as shown in Fig 2. In the top ten URLs, profiles with a pro-vaccine stance typically share content from mainstream news sites, medical or technology/science sites, and various social media sites, whereas anti-vaccine profiles tend to share content from YouTube, social media sites, and a number of sites specializing in alternative health products, pseudoscience, and conspiracy theories. For details on the categorization of links, see the Materials and methods section. The absolute number of links posted to each such domain varies a lot over time for some domains, yet the relative frequency of posts for each stance is relatively constant; visualizations and statistics are provided in S3 Appendix.

Expanding upon the above, we assign to each of the popular base URLs one or more of the following labels: "news", "social", "science", "conspiracy", "pseudoscience", "commercial". By 'commercial' we here mean sites which sell products related to (alternative) health, and so have a direct financial interest in the vaccination discourse. Exact definitions of all labels are provided in the methods section. Fig 3 shows that fraction of links posted by profiles of different stances which belong to each such category. The results summarized in Fig 3 are generally robust to changes in the sentiment threshold for stance attribution, with one exception: When increasingly strict thresholds are applied, i.e. when we consider only users who share very
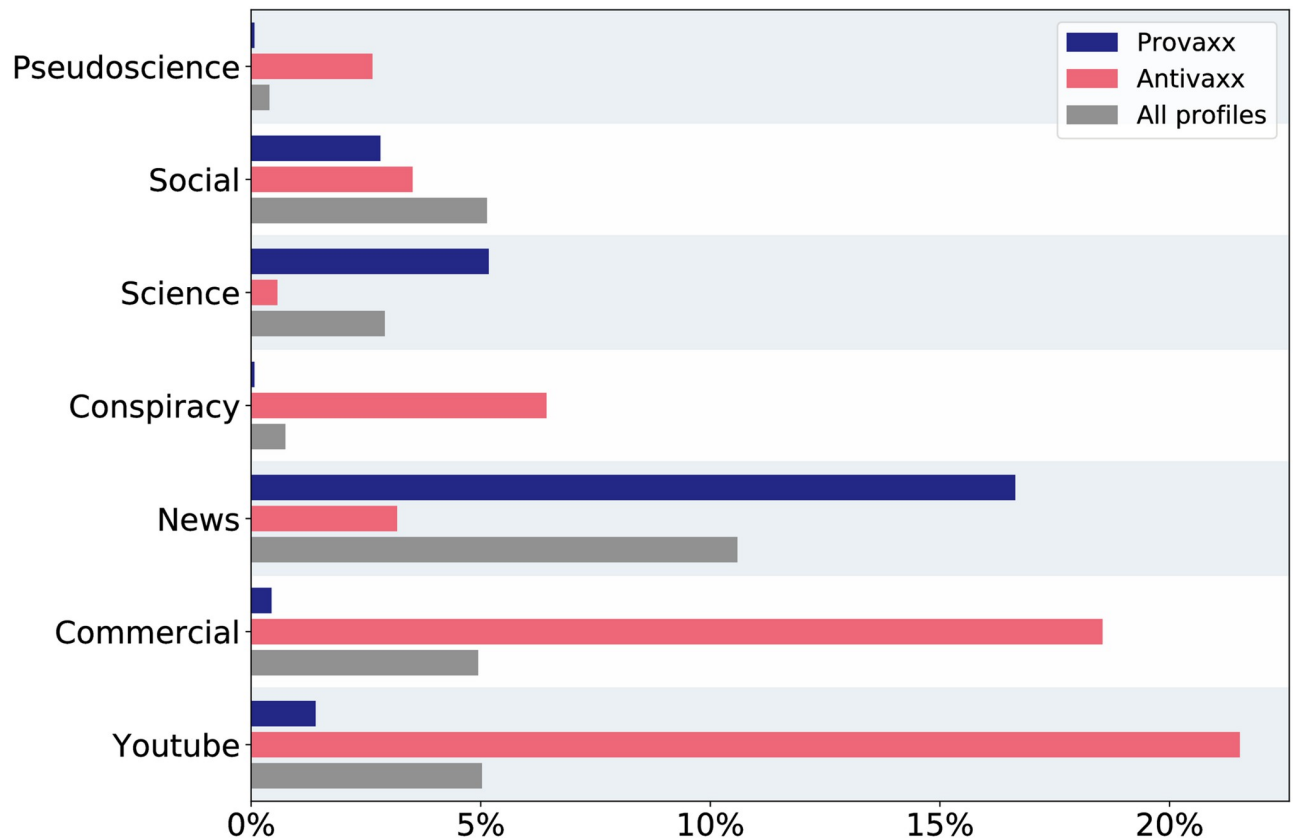
Fig 3. Frequency of various categories of links for profiles grouped by vaccination stance (provaxx, antivaxx, or neutral). Antivaxx profiles often post links to Youtube videos, and to sites that sell health related products and thus have a vested financial interest in the vaccine discourse. Error bars are too small to depict visually, see S1 Appendix for uncertainty analyses.

https://doi.org/10.1371/journal.pone.0263746.g003

strong vaccine sentiments, the "news" category becomes more frequently linked to among anti-vaccine profiles. This is due to strongly anti-vaccine profiles disproportionally posting URLs which point to Fox News. Details and figures regarding that analysis are presented in S1 Appendix.

The most frequently occurring link category for pro- and anti-vaxxers are news sites, and Youtube links, respectively. The second most frequent category among antivaxx profiles is commercial sites profiting from selling health related products. This finding is unexpected, as common reason for vaccine hesitancy is mistrust in medical research due to perceived financial conflicts of interest and industry ties to pharmaceutical companies [44]. Links to pseudoscience and conspiracy sites are also posted disproportionally by profiles with a strong anti-vaccination stance.

Polarization and epistemic echo chambers

Using Dataset 1, we construct a large network representing observed mutual interactions between profiles on Twitter. In this network, profiles are linked if there exists a reciprocal @-mention, or a reciprocal retweet, within a 3-month time window. We refer to the interaction network as the MMR (mutual mention/retweet) network for this reason. Additional details regarding the MMR network, as well as some analyses of the network structure and temporal stability, are provided in the materials and methods section. Links are constructed in

this fashion for consecutive time windows, where the number of such 3-month time windows in which two users have interacted can then be viewed as the weight of the link. In addition to thresholding on this weight, which is illustrated in Fig 5, the graph may be thresholded according to the number of vaccine-related tweets from each user, such that only nodes corresponding to users who posted at least a desired number of tweets with vaccination-related keywords are retained in the graph.

We initially consider a version of the MMR graph constructed using very strict criteria for node and link inclusion, then subsequently investigate the effects of easing those criteria. We first include, in each time window, only nodes that are assigned a pro- or anti-vaccine stance. Further, we only include links between nodes that interacted in several windows. The strictness of these criteria retains only nodes which consistently express strong vaccine-sentiments, in interact repeatedly nodes that do so as well. As a consequence of the strict criteria, the resulting graph contains only 4894 nodes, of which 3359 (69%) nodes form a giant connected component. The remaining connected components are loosely scattered, have fewer than 30 nodes in each, and contain only 5.6% anti-vaccine users. 395 nodes (11.76%) in the giant connected component represent antivaxx profiles. A representation of the graph using a force layout algorithm [45] is shown in Fig 4. The interplay between the stances of users and their neighborhoods, as well as user connectivity and activity, is visualized in S4 Appendix.

The graph is heavily stratified with regards to vaccination stance. The assortativity coefficient (Pearson correlation between stance and connectedness in the graph) is $r = 0.813$.

The analyses above, however, dependend on discretely partitioning users into two distinct categories. Considering instead user stance as a continuous variable—given by e.g. the average anti-/pro-vaccine sentiment expressed in their tweets—we obtain similar findings. Discarding users with fewer than 5 vaccination related tweets, we rebuilt the interaction graph while
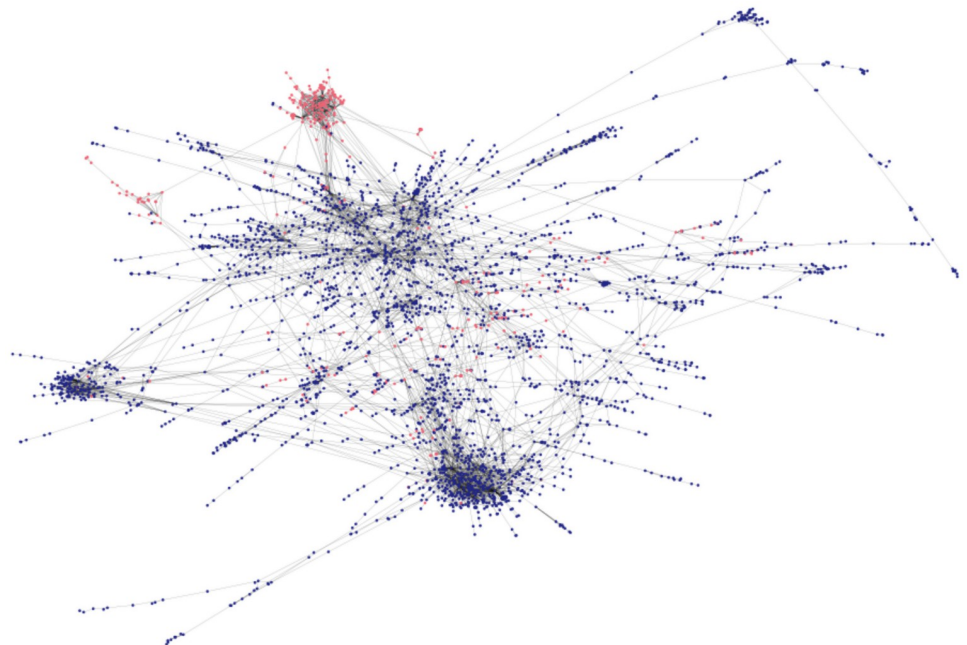
Fig 4. Representation of the repeated mutual interaction graph from 2013–2016. Profiles frequently interact with others who share their own stance, and antivaxx profiles are localized in relatively few, tightly nit clusters. Profiles with and anti- and provaccine stances are illustrated in red and blue, respectively. Only the giant conected component of the interaction graph is depicted.

https://doi.org/10.1371/journal.pone.0263746.g004

varying the minimum number of 3-month time windows in which users must have interacted before being connected in the graph. Results on the interplay of (continuous) stance and (repeated) connectivity are summarized in Fig 5.

Correlations between the mean vaccine sentiment expressed in neighboring users' tweets were roughly the same independently of how frequently the users interacted, as shown in Fig 5a, although the number of users in the interaction graph decreases quickly when using strict inclusion criteria (additional analyses in the methods section). Similarly, we observe an anti-correlation between pro- and anti-vaxx probabilities of neighbors, which seems to diminish somewhat when considering repeated interactions. However, this decrease appears to be driven by a few nodes which have many connections, yet do not frequently discuss vaccines, as shown in Fig 5b.
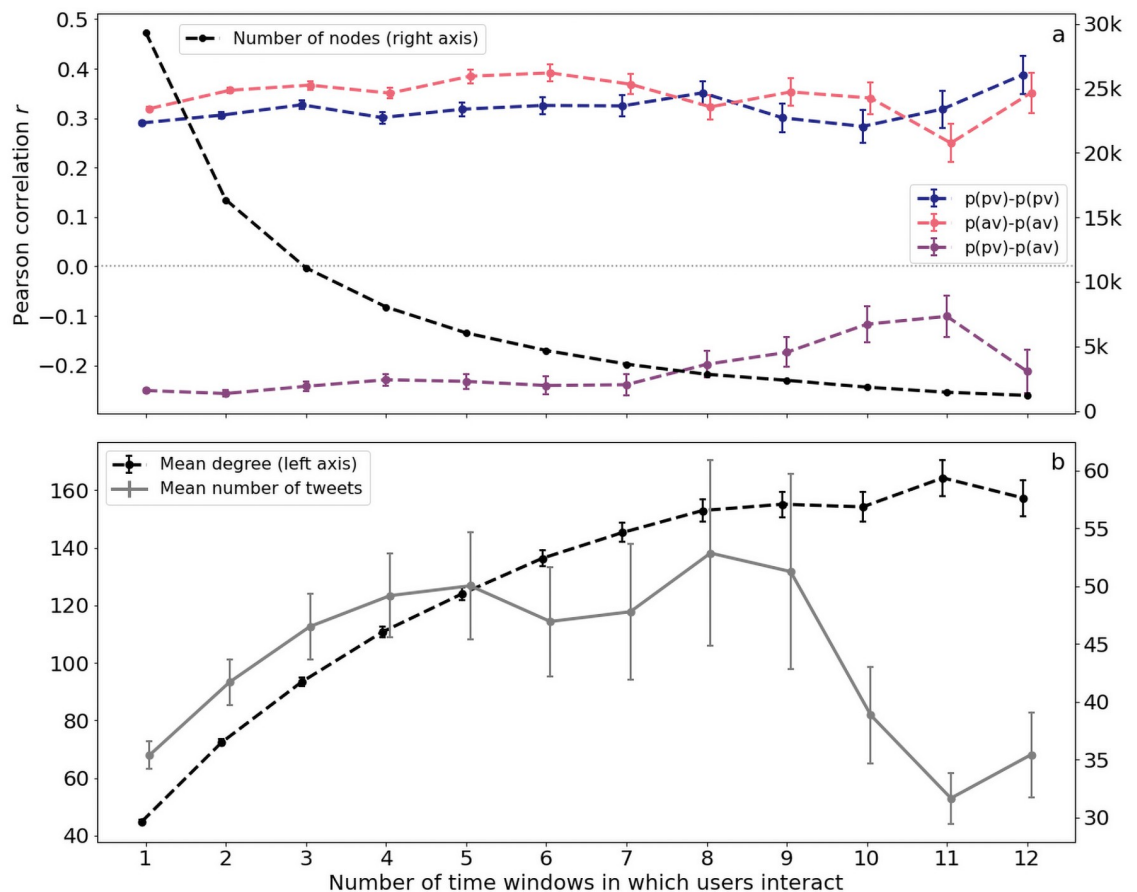
Fig 5. Interplay between average vaccination sentiment and user interactions. a: Users tend to disproportionally interact with users of similar stance, both in cases where users only interact during a single, and multiple, three-month time windows. Specifically, we compute for all users the average probabilities of that user's tweets expressing pro/anti-vaccine sentiment. Comparing these averages for all nodes and their neighbors, we find a positive correlation between the average pro- and antivaccine sentiments. Similarly, the average pro-vaccine sentiment of nodes exhibits a negative correlation with the anti-vaccine sentiments of their neighbors. The number of nodes in the interaction network decreases exponentially as the minimum number of time windows is increased. The negative correlation between pro- and antivaxx probabilities of neighbors tends slightly toward zero as the threshold for repeated interaction grows. **b**: As increasingly repeated interactions are considered, users in the interactions graph are increasingly well connected. However, the number of vaccination-related tweets posted by users decreases for interactions occurring very frequently, indicating that at this point, the graph likely includes users who are highly active on Twitter, yet do not discuss vaccination-related topics very often. Error bars on Pearson correlations represent one standard deviation of the Fisher-transformed variables z, i.e. the bounds on the error bar on a correlation r of n data points, is given by $\tanh(z \pm \sigma_z)$, where $z = \text{arctanh}(r)$ and $\sigma_z = 1/\sqrt{n-3}$.

The finding that users interact disproportionally with other users sharing their stance, aligns with previous findings that long time anti-vaccine users of social media tend to form tightly knit clusters which exhibit a high degree of in-group solidarity [46], and in which misinformation may thrive unquestioned [47]. To qualify the latter, we turn again to the URLs most frequently shared by users discussing vaccines, shown in Fig 2, we probe regions in the MMR network around individuals of various stances and assess whether the URLs shared in those regions differ more or less from a normal distribution depending on stance.

Considering only the approximately 32 thousand users with at least 5 vaccination-related tweets, we group users based on the mean probability of antivaxx (p_{av}) of their tweets. We computed the deciles of p_{av} for all tweets and grouped users based on which deciles their mean score fell between, i.e. one bin for mean p_{av} values below the first decile, one for values between the first and second deciles, and so forth. For each such group of users, we observed the regions surrounding them in the MMR network, and extracted the URLs shared by all users who were located in that region, and who had shared at least 5 URLs and posted at least 5 vaccine-related tweets. We then computed the frequency for each of the top URLs for the regions (locally), and observed the difference from the overall (global) frequency distribution. The frequency distributions may be interpreted as maximum likelihood estimates of the probability distributions over links shared in the regions around specific users, and globally. Therefore, we quantify the difference between such distributions using the Jensen-Shannon (JS) distance [48]—an information-theoretical measure of distance between probability distributions which take values in the range between zero (no overlap between distributions) and one (identical distributions).

Fig 6 shows the JS-distances between overall link frequencies, and links shared by users adjacent to users with a given mean p_{av}. The figure shows that Twitter profiles that engage in online vaccine discourse are not only disproportionately connected to other users who share their stance, but that users with stronger anti-vaccine stances are also exposed to increasingly atypical sources of information. This is indicative of 'epistemic echo chambers' in online vaccine discourse in the sense that users, depending on their stance, are exposed not only to a skewed distribution of stances from other users (i.e. network homophily), but also to information sources that are highly dissimilar to those typically partaking in the overall discussion. Although we do not attempt to explain how these echo chambers arise in the first place, we can point to some mechanisms described in the literature which are consistent with our results. First, it is a well-known result in sociology and network science that links tend to form between nodes that share similar attributes [49, 50]. Second, some studies indicate that people are highly selective in sharing information that aligns well with their convictions [51], which in term can cause polarization by opinion reinforcement [52], and by users cutting ties to avoid exposure to information causing cognitive dissonance [53].

Discussion

In summary, our findings paint a picture of the vaccine discourse on Twitter as highly polarized, where users who express similar sentiments regarding vaccinations are more likely to interact with one another, and tend to share contents from similar sources. Focusing on users whose vaccination stances are the positive and negative extremes of the spectrum, we observe relatively disjoint 'epistemic echo chambers' which imply that members of the two groups of users rarely interact, and in which users experience highly dissimilar 'information landscapes' depending on their stance. Finally, we find that strongly anti-vaccine users much more frequently share information from actors with a vested commercial interest in promoting medical misinformation.
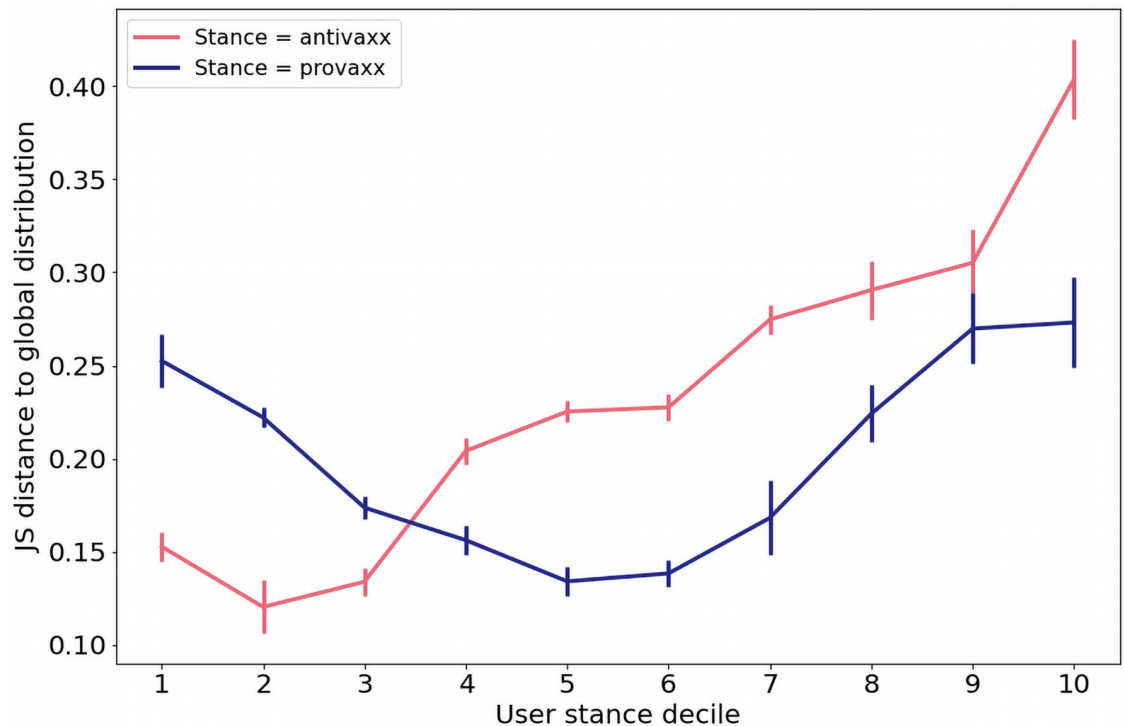
Fig 6. Profiles that express fringe vaccine sentiments are also exposed via their interaction networks to sources of information that are highly dissimilar to link frequencies in the overall discussion. Here we consider users who posted a minimum of 5 tweets containing vaccine-related keywords, and partition them into deciles based on their tweets' mean probability of expressing anti- and pro-vaccine sentiment. For each such decile and vaccine stance, the plot shows the Jensen-Shannon distance between the frequencies at which links from the domains shown in Fig 2 are shared in the vicinity of users in that decile, and in the interaction network overall. The error bars are computed using a bootstrap technique in which users in the target stance-decile combination where randomly sampled with replacement and the JS-distance to the overall distribution calculated. The error bars depict the standard deviations of each 1000 such samples.

One implication of these findings is that online (medical) misinformation may present an even greater problem than previously thought, because beliefs and behaviors in tightly knit, internally homogeneous communities are more resilient [37, 54], and provide fertile ground for fringe narratives [55, 56], while mainstream information is attenuated [57]. Furthermore, such polarization of communities may become self-perpetuating, because individuals avoid those not sharing their views [58], or because exposure to mainstream information might further entrench fringe viewpoints [59].

A further problem exacerbated by the structure of the debate is that, parents often base their vaccination decisions on their impression of what other parents do [60], so vaccine hesitant parents who encounter a strongly anti-vaccine community might get the impression that not vaccinating is the norm and opt not to. This risk is compounded by the fact that anti-vaccine communities are highly effective at reaching out to undecided individuals [61], which highlights the need to reach undecided individuals with accurate information to overcome vaccine hesitancy [62].

In summary, the characteristics of the online vaccine discourse may contribute to increasing vaccine hesitancy, possibly into the extreme of vaccine denial. A brief discussion of measures that have proven successful in decreasing hesitancy and increasing vaccine uptake therefore seems in order. One such measure is encouraging direct communication between hesitant individuals and healthcare professionals. Parents who interact with health care

professionals are significantly more likely to vaccinate their children [63, 64], whereas parents of underimmunized children are significantly more likely to obtain medical information online [23]. Another measure is implementing policies which incentivize vaccination or discourages rejection [65–67].

In terms of digital interventions, our findings highlight the need for measures based—not just on whether content is true or false—but on a more nuanced understanding of the interplay between vaccination attitudes, social network structure, and information sources, including actors with a vested interest in promoting false beliefs. With disinformation campaigns aiming to erode consensus [24, 68], fact-checking at the level of individual stories being shared online might need to be complemented by an understanding of the complex interplay between community structure and information content.

Future work based on the findings presented here could investigate e.g. the text content of the communication between users with highly similar and dissimilar stances regarding vaccination, as well as interactions between text topics and community structure.

Materials and methods

This section provides details of the data analyzed and the methods employed.

Twitter data

The work presented here relies on a large collection of data from Twitter. For clarity, we describe below two subsets of the data used in the analysis. The reason for this is that one part of the data comes from a large collection of data not specific to vaccination but well suited, due to its size, to analysis overall interactions on Twitter, whereas another was obtained by querying for vaccination-related keywords, and thus is better suited for analyses specific to vaccination. All data were collected through Twitter's public search API—no terms of service agreements were violated in collecting the data.

Dataset 1 consists of a large collection (approximately 60 billion) of tweets [69], collected as a random 10% in 2013–2016. These data were used to a general 'interaction network' in which nodes represent Twitter profiles, and connections between nodes represent cases where both profiles either mention or retweet each other. We refer to this as the Mutual Replies/Retweets (MMR) network.

Dataset 2 was constructed using Python to paginate backwards through the official search API for tweets containing various keywords pertaining to vaccination. A full list of the keywords queried for is: "unvaccinated", "unvaccined", "vaccinate", "vaccinated", "vaccinating", "vaccination", "vaccinations", "vaccinator", "vaccinators", "vaccine", "vaccined", "vaccinering", "vaccines", "vaccinology", "vaxx".

When a match occurred, the tweet was analyzed and stored in a database. The analysis involved evaluating the sentiment expressed by the tweet's contents, and following any external links it contained. In the following, we present details regarding the link analysis, the MMR network, and the sentiment classification process.

Ethics statement

The study, including the data collection process, has received written approval from the Institutional Review Board at the Technical University of Denmark (IRB number COMP-IRB-2021–09).
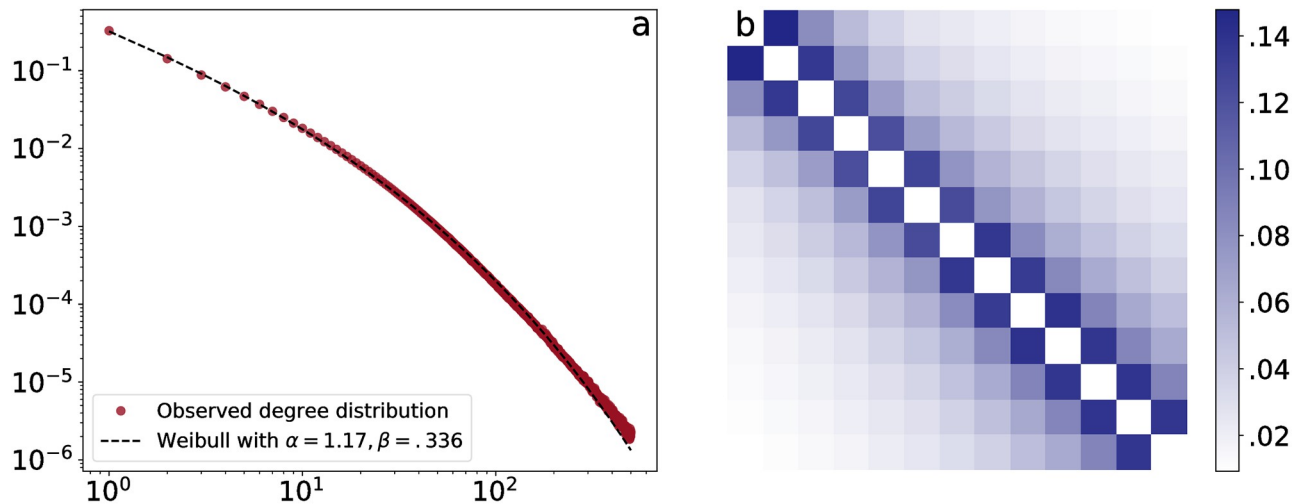
Fig 7. a: The degree distribution of the MMR graph, truncated at degree 500 to exclude automated profiles. The dashed line indicates the best fit for a stretched exponential (Weibull) function. **b**: The Jaccard similarity index of the sets of edges in the MMR graph for different 3-month periods. Each row and column correspond to a three-month time windows in the period from 2013–2016. The diagonal is therefor left out, as it represents the self-similarity of the interaction network in each time window, and so the Jaccard similarity is 1 by construction.

Network analysis

Using the sample of approximately 60 billion tweets (Dataset 1), we construct an interaction network, in which profiles are connected only if both profiles interact with the other, by retweeting content, or by replying, within a three month time window. Each such time window contained an average of 19,588,474 nodes and 47,115,188 edges. Combining the graph for all time windows resulted in a graph with a total of 89,577,277 nodes and 434,193,958 edges. The degree distribution—the probability distribution over k, the number of such connections to each profile—for the combined graph is showed in Fig 7a. The degree distribution closely matches a Weibull distribution,

$$p_k = \frac{\beta}{\alpha}\left(\frac{k}{\alpha}\right)^{\beta-1} e^{(k/\alpha)^\beta}, \tag{1}$$

with the parameter values specified in the figure legend.

To assess the stability of the interaction network over time, we constructed an MMR network for each three-month time window in the period 2013–2016. For every pair of such networks, we compared how similar the connections in the two networks were, using the Jaccard similarity index

$$H_{S_i,S_j} = \frac{|S_i \cap S_j|}{|S_i \cup S_j|}. \tag{2}$$

If the Jaccard similarity between the networks at two time points is 1, the network is completely unchanged, and if it is 0, no connection between any two profiles exists at both time points. Fig 7b shows the Jaccard similarities between the MMR networks for all of the three-month time windows. It shows that the similarity is relatively large at neighboring time points, with 14% of connections appearing at both time points, following which the self-similarity over time gradually reduces to almost zero over the period of 3 years.

Link analysis

Links contained in tweets are shortened by Twitter, and sometimes by external URL shorteners as well. In order to analyze external URLs contained in the vaccine-related tweets, we used python to crawl each URL, repeatedly follow redirects, end noting the domain of the final destination. For profiles that were categorized as strongly anti- or pro-vaccine, we recorded the ten most frequently used such domains. This resulted in a total of 18 domains, due to youtube and facebook occurring in both top tens.

   We manually assigned one or more categories to these 18 domains. The categories, along with classification criteria, are outlined below. For classifying pages as conspiracy or pseudo-science sites, we looked up the domains on the online service *media bias/fact check* (MBFC).

- **Commercial**—Pages that include an online store selling health related products.
  naturalnews.com, articles.mercola.com, go.thetruthaboutvaccines.com, greenmedinfo.com, healthimpactnews.com, healthnutnews.com, infowars.com, newstarget.com, vaccineimpact. com.

- **Conspiracy**—Classified using MBFC.
  awarenessact.com, newspunch.com, newstarget.com, worldtruth.tv, collective-evolution. com, inshapetoday.com, realfarmacy.com.

- **News**—Known news sites.
  bbc.co.uk, bbc.com, bioportfolio.com, cbc.ca, choice.npr.org, cnn.com, edition.cnn.com, forbes.com, foxnews.com, huffingtonpost.com, medicalnewstoday.com, nbcnews.com, nytimes.com, reuters.com, sciencedaily.com, statnews.com, theguardian.com, time.com, whitehouse.gov.

- **Pseudoscience**—Classified using MBFC.
  collective-evolution.com, inshapetoday.com, realfarmacy.com, seattleorganicrestaurants. com.

- **Science**—Sites promoting mainstream science/medical information.
  bioportfolio.com, cdc.gov, medicalnewstoday.com, sciencedaily.com, statnews.com, webmd. com.

- **Social Media**—Large social media platforms.
  facebook.com, instagram.com, reddit.com.

- **Youtube**—The popular video-sharing platform.
  youtube.com.

Classification of vaccine-sentiment in tweets

Using machine learning techniques in general often requires a large amount of 'ground truth' data on which ones model can be trained, and this is particularly true for deep learning models. Establishing a ground truth dataset often requires human work and is thus often costly. One technique to work around this issue is *transfer learning*, in which a model is first trained on a large 'source' dataset, in which the ground truth has already been established, and then applied to a 'target' dataset. In the case of deep neural networks, this approach typically consists of first training a complex model to the source data, then stripping off the layer of output neurons and using the output of the second last layer, often called a representation layer, in conjunction with another model to predict the desired target dataset. This sometimes increases
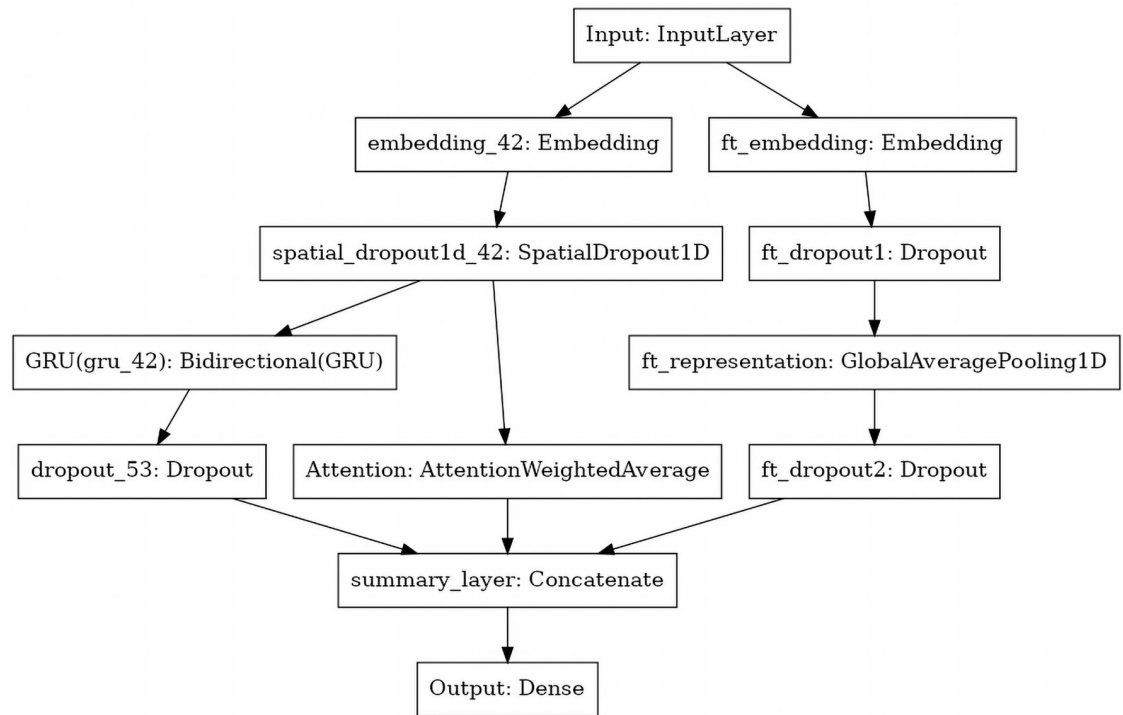
Fig 8. Representation of the final classifier. An initial input layer, in which strings are represented by a sequence of one-hot encoded words, is passed to a) a deep neural network similar to the DeepMoji classifier [70], and b) a fasttext classifier [71]. After being pre-trained to predict hashtags from the surrounding text (source dataset), the model is fine-tuned to instead predict vaccine sentiment from tweet text (target dataset).

https://doi.org/10.1371/journal.pone.0263746.g008

model performance, as it allows one to 'reuse' higher-order representations of the input data learned by the original classifier. We here describe the source and target datasets, along with the final model architecture, which is summarized in Fig 8.

The **target** data consisted of 10000 randomly selected tweets containing vaccine related keywords. We hired workers on Amazon's Mechanical Turk (MTurk) platform to classify tweets as being either for, or against human vaccination, or as undecideable or unrelated. To ensure high-quality ratings, we first manually rated 100 tweets, then hired a number of MTurk workers for a test assignment which clearly stated that top performers would receive offers for additional tasks. The payment was set to be very high compared to typical MTurk to provide incentive for good performance. We then identified top performers whose scores where most similar to our own, and launched the remaining tasks, allowing only the identified workers to participate. We hired workers such that each tweet would be rated by 3 distinct raters. We then kept only the tweets for which all 3 raters agreed on a label, which reduced the data set to 5358, the distribution of labels in which was 18.8% antivaxx, 45.67% provaxx, and 35.50% neutral/unrelated.

As the **source** dataset, we chose to train the classifier to predict a number of hashtags which we presumed to be related to the sentiment prediction task. From an initial qualitative analysis of the data, and from a brief review of the literature, we noted that

- Anti-vaccine narratives occasionally supposes underlying conspiracies, as represented by hashtags such as #cdctruth, or #cdcwhistleblower.

- Many tweets that that mention vaccine-related keywords are not concerned with vaccination of humans, but rather of pets. To help the classifier disambiguate, we included hashtags such as #dog and #cat.

- There is a relatively popular indie rock band called *The Vaccines*. To help disambiguate, we included hashtags like #music and #livemusic.

Based on the above observations, we opted to scrape for our source dataset a large number of tweets containing any of the following hashtags: #endautismnow, #antivax, #autism, #autismismedical, #cat, #cdctruth, #cdcwhistleblower, #dog, #ebola, #flu, #health, #hearthiswell, #hpv, #immunization, #livemusic, #measles, #medication, #music, #polio, #sb277, #science, #vaccination, #vaccine, #vaccines, #vaccinescauseautism, #vaccineswork, #vaxxed.

Using a large number of tweets ($\approx$ 10,670,000 in total) of tweets containing either of those hashtags, and trained a deep neural network classifier to predict the hashtags from text. These tweets were obtained in a similar fashion to dataset 2. We used a random upsampling approach to achieve a balanced dataset within each training sample when doing cross-validation [70].

The classifier consisted of an embedding layer, a spatial dropout, then a parallel sequence of a) a bi-directional GRU (gated recurrent unit) and a dropout layer, and b) a weighted attention average layer [70]. Those were then concatenated into a representation layer.

After fitting the hashtag model, we removed the output layer and 'froze' the remaining layers, to prohibit training of the weights contained in the original model. We then added a fasttext network [71] in parallel with the pretrained classifier. The rationale for this was that, while the initial classifier might have learned to recognize highly complex patterns in text, it might not do a good job of making simpler connections between input text and target probabilities. After fitting the fasttext part of the classifier, we used the chain-thaw approach of [70] to further improve performance.

On the three-class prediction task, the classifier attained a micro-averaged F1-score of 0.762. The score was computed by aggregating true and false positives/negatives over a 10-fold stratified cross-validation procedure [72]. For comparison with the literature, we also trained the classifier for binary prediction (i.e. predicting simply whether a text snippet was anti-vaxx or not). The accuracy on the binary case was 90.4±1.4% over a 10-fold stratified cross-validation evaluation, an increase over what to our knowledge is state of the art performance [46].

Looking qualitatively at the performance of the classifier, the tweets that were labeled with high confidence demonstrate some capability of the classifier to recognize relatively subtle indications of the correct label for the tweet, as shown in Table 1.

Table 1. Qualitative summary of classifier performance. The classifier correctly assigns a large probability of antivaxxness to text snippets the express conspiracist notions about vaccines being part of a global scam. Similarly, texts highlighting the positive qualities of vaccinations are assigned a high probability of being provaxx. In addition, text snippets concerning the band named The Vaccines are recognized as irrelevant. A text snippet expressing how much more expensive it is to kill, rather than vaccinate, badgers is also categorized as irrelevant with a high certainty, despite containing negative words like 'kill'.

Text	Class
"January is Cervical Health Awareness Month.Join the HPV vaccine campaign to prevent cervical cancer"	Provaxx
"Getting a flu shot will help prevent transmission to new babies who are too young to be vaccinated. #flukills"	Provaxx
"things i love about the vaccines: they change their set list every gig but norgaard is always the last song they play"	Neutral
"Hi ⟨USER⟩ £7.3m to kill 1771 Badgers—£4100 per Badger—Vaccination looks cheap now? £662 per Badger?"	Neutral
"Bill Gates Admits #Vaccines Are Used for Human #Depopulation"	Antivaxx
"Lead Developer Of HPV Vaccines Comes Clean, Warns Parents & Young Girls It's All A Giant Deadly Scam"	Antivaxx

Categorization of Twitter profiles from tweets

For each user, we considered tweets containing vaccination-related keywords (see description of Dataset 2 above). For each such tweet, we estimated the probability of the tweet expressing sentiment that is pro-vaccine, anti-vaccine, or neutral/unrelated, using the machine learning classification method described above. We then label profiles as anti/pro-vaxx if the classifier assigns more than 50% of the profile's tweets a probability of at least 50% of being anti/pro-vaxx. Note that this leaves the majority of profiles not assigned to either of the two categories, as illustrated in Fig 1. This strong criterion is intended to reduce the number of profiles falsely assigned into either category.

Note on uncertainties and robustness

Most of the figures presented here are produced using a very large number of data points. For this reason, some quantities, such as the tweet sentiment and user stance distributions presented in Fig 1, will have very small error bars that are difficult to meaningfully visualize. Meanwhile, the distributions turn out to be more sensitive to changes in the arbitrary threshold for labeling user stances from tweet sentiments, although this does not qualitatively alter the results. In such cases, we have opted to present the figures without error bars in the main paper, referring the reader to S1 Appendix for a more detailed overview of uncertainties, as well as analyses of robustness to the aforementioned sentiment threshold.

Supporting information

S1 Appendix. Robustness and uncertainties. More details on uncertainties and robustness to sentiment thresholds is provided in S1 Appendix.
(PDF)

S2 Appendix. Geographical analysis of tweets originating from the USA. A short analysis of tweet sentiment by American state, of potential relevancy to researchers interested in the interplay between state policy/regulations and Twitter discourse, is presented in S2 Appendix.
(PDF)

S3 Appendix. Temporal evolution of link frequencies. S3 Appendix presents illustrations of how the number of links posted to each external domain by changes over time.
(PDF)

S4 Appendix. Explorative analyses of user and neighborhood stances. S4 Appendix contains some explorative visualizations on the interplay between user stance strength, the strength of disagreement with the user neighborhood, and the neighborhood activity and number of neighbors, for the data underlying Fig 4.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Bjarke Mønsted, Sune Lehmann.

Data curation: Bjarke Mønsted, Sune Lehmann.

Formal analysis: Bjarke Mønsted.

Funding acquisition: Sune Lehmann.

Investigation: Bjarke Mønsted.

Methodology: Bjarke Mønsted.

Project administration: Bjarke Mønsted, Sune Lehmann.

Software: Bjarke Mønsted.

Supervision: Bjarke Mønsted, Sune Lehmann.

Validation: Bjarke Mønsted.

Visualization: Bjarke Mønsted.

Writing – original draft: Bjarke Mønsted.

Writing – review & editing: Bjarke Mønsted, Sune Lehmann.

References

1. MacDonald NE, et al. Vaccine hesitancy: Definition, scope and determinants. Vaccine. 2015; 33 (34):4161–4164. https://doi.org/10.1016/j.vaccine.2015.04.036 PMID: 25896383

2. WHO. Report of the SAGE working group on vaccine hesitancy; 2014.

3. Tomaszewski T, Morales A, Lourentzou I, Caskey R, Liu B, Schwartz A, et al. Identifying False Human Papillomavirus (HPV) Vaccine Information and Corresponding Risk Perceptions From Twitter: Advanced Predictive Models. Journal of medical Internet research. 2021; 23(9):e30451. https://doi.org/10.2196/30451 PMID: 34499043

4. Larson HJ, Jarrett C, Eckersberger E, Smith DM, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. Vaccine. 2014; 32(19):2150–2159. https://doi.org/10.1016/j.vaccine.2014.01.081 PMID: 24598724

5. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger JA. Vaccine hesitancy: an overview. Human vaccines & immunotherapeutics. 2013; 9(8):1763–1773. https://doi.org/10.4161/hv.24657 PMID: 23584253

6. Truong J, Bakshi S, Wasim A, Ahmad M, Majid U. What factors promote vaccine hesitancy or acceptance during pandemics? A systematic review and thematic analysis. Health promotion international. 2021;. https://doi.org/10.1093/heapro/daab105 PMID: 34244738

7. Crescitelli MD, Ghirotto L, Sisson H, Sarli L, Artioli G, Bassi M, et al. A meta-synthesis study of the key elements involved in childhood vaccine hesitancy. Public Health. 2020; 180:38–45. https://doi.org/10.1016/j.puhe.2019.10.027

8. Yaqub O, Castle-Clarke S, Sevdalis N, Chataway J. Attitudes to vaccination: a critical review. Social science & medicine. 2014; 112:1–11. https://doi.org/10.1016/j.socscimed.2014.04.018 PMID: 24788111

9. Jamison A, Broniatowski DA, Smith MC, Parikh KS, Malik A, Dredze M, et al. Adapting and extending a typology to identify vaccine misinformation on Twitter. American Journal of Public Health. 2020; 110 (S3):S331–S339. https://doi.org/10.2105/AJPH.2020.305940 PMID: 33001737

10. Dubé E, Vivion M, Sauvageau C, Gagneur A, Gagnon R, Guay M. "Nature does things well, why should we interfere?" Vaccine hesitancy among mothers. Qualitative Health Research. 2016; 26(3):411–425. https://doi.org/10.1177/1049732315573207 PMID: 25711847

11. Bonnevie E, Gallegos-Jeffrey A, Goldbarg J, Byrd B, Smyser J. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. Journal of communication in healthcare. 2021; 14 (1):12–19. https://doi.org/10.1080/17538068.2020.1858222

12. Thelwall M, Kousha K, Thelwall S. Covid-19 vaccine hesitancy on English-language Twitter. Profesional de la información (EPI). 2021; 30(2).

13. Phadke VK, Bednarczyk RA, Salmon DA, Omer SB. Association Between Vaccine Refusal and Vaccine-Preventable Diseases in the United States. JAMA. 2016; 315(11):1149. https://doi.org/10.1001/jama.2016.1353 PMID: 26978210

14. Salmon D, Haber M, Gangarosa E, Phillips L. Health consequences of religious and philosophical exemptions from immunization laws: individual and societal risk of measles. Jama. 1999;. https://doi.org/10.1001/jama.282.1.47 PMID: 10404911

15. Feikin D, Lezotte D, Hamman R, Salmon D. Individual and community risks of measles and pertussis associated with personal exemptions to immunization. Jama. 2000;. https://doi.org/10.1001/jama.284.24.3145 PMID: 11135778

16. Smith P, Chu S, Barker L. Children who have received no vaccines: who are they and where do they live? Pediatrics. 2004;. https://doi.org/10.1542/peds.114.1.187 PMID: 15231927

17. Atwell JE, Van Otterloo J, Zipprich J, Winter K, Harriman K, Salmon Da, et al. Nonmedical vaccine exemptions and pertussis in California, 2010. Pediatrics. 2013; 132(4):624–30. https://doi.org/10.1542/peds.2013-0878 PMID: 24082000

18. Omer S, Pan W, Halsey N, Stokley S, Moulton L. Nonmedical exemptions to school immunization requirements: secular trends and association of state policies with pertussis incidence. Jama. 2006;. https://doi.org/10.1001/jama.296.14.1757 PMID: 17032989

19. André F. Vaccinology: past achievements, present roadblocks and future promises. Vaccine. 2003;. PMID: 12531323

20. Gangarosa E, Galazka A, Wolfe C, Phillips L. Impact of anti-vaccine movements on pertussis control: the untold story. The Lancet. 1998;. https://doi.org/10.1016/S0140-6736(97)04334-1 PMID: 9652634

21. Dyda A, Shah Z, Surian D, Martin P, Coiera E, Dey A, et al. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. Human vaccines & immunotherapeutics. 2019; 15(7-8):1488–1495. https://doi.org/10.1080/21645515.2019.1596712 PMID: 30978147

22. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. Journal of health psychology. 2010; 15(3):446–455. https://doi.org/10.1177/1359105309353647 PMID: 20348365

23. Salmon DA, Moulton LH, Omer SB, DeHart MP, Stokley S, Halsey NA. Factors associated with refusal of childhood vaccines among parents of school-aged children: a case-control study. Archives of pediatrics & adolescent medicine. 2005; 159(5):470–476. https://doi.org/10.1001/archpedi.159.5.470 PMID: 15867122

24. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Global Health. 2020; 5(10): e004206. https://doi.org/10.1136/bmjgh-2020-004206 PMID: 33097547

25. Hall V, Banerjee E, Kenyon C, Strain A, Griffith J, Como-Sabetti K, et al. Measles Outbreak—Minnesota April–May 2017. MMWR Morbidity and Mortality Weekly Report. 2017; 66(27):713. https://doi.org/10.15585/mmwr.mm6627a1 PMID: 28704350

26. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. Vaccine. 2010;. https://doi.org/10.1016/j.vaccine.2009.12.022 PMID: 20045099

27. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. jama. 2007;. https://doi.org/10.1001/jama.298.21.2482 PMID: 18056901

28. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Human vaccines & immunotherapeutics. 2020; 16(11):2586–2593. https://doi.org/10.1080/21645515.2020.1780846 PMID: 32693678

29. DeStefano F, Bodenstab HM, Offit PA. Principal Controversies in Vaccine Safety in the United States. Clinical Infectious Diseases. 2019;. https://doi.org/10.1093/cid/ciz135 PMID: 30753348

30. Hviid A, Hansen JV, Frisch M, Melbye M. Measles, Mumps, Rubella Vaccination and Autism. Annals of Internal Medicine. 2019;. https://doi.org/10.7326/M18-2101

31. Motta M, Callaghan T, Sylvester S, Lunz-Trujillo K. Identifying the prevalence, correlates, and policy consequences of anti-vaccine social identity. Politics, Groups, and Identities. 2021; p. 1–15. https://doi.org/10.1080/21565503.2021.1932528

32. Kahan DM, Jenkins-Smith H, Braman D. Cultural cognition of scientific consensus. Journal of risk research. 2011; 14(2):147–174. https://doi.org/10.1080/13669877.2010.511246

33. Kahan DM. Ideology, motivated reasoning, and cognitive reflection: An experimental study. Judgment and Decision making. 2012; 8:407–24.

34. Kahan DM, Braman D, Cohen GL, Gastil J, Slovic P. Who fears the HPV vaccine, who doesn't, and why? An experimental study of the mechanisms of cultural cognition. Law and Human Behavior. 2010; 34(6):501–516. https://doi.org/10.1007/s10979-009-9201-0 PMID: 20076997

35. Featherstone JD, Barnett GA, Ruiz JB, Zhuang Y, Millam BJ. Exploring childhood anti-vaccine and pro-vaccine communities on twitter–a perspective from influential users. Online Social Networks and Media. 2020; 20:100105. https://doi.org/10.1016/j.osnem.2020.100105

36. Gunaratne K, Coomes EA, Haghbayan H. Temporal trends in anti-vaccine discourse on Twitter. Vaccine. 2019; 37(35):4867–4871. https://doi.org/10.1016/j.vaccine.2019.06.086 PMID: 31300292

37. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine. 2018; 36(25):3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040 PMID: 29773322

38. Nguyen CT. Echo chambers and epistemic bubbles. Episteme. 2020; 17(2):141–161. https://doi.org/10.1017/epi.2018.32

39. Cornwall W. Officials gird for a war on vaccine misinformation; 2020.

40. Schmid P, MacDonald NE, Habersaat K, Butler R. Commentary to: How to respond to vocal vaccine deniers in public. Vaccine. 2018; 36(2):196–8. https://doi.org/10.1016/j.vaccine.2016.09.065 PMID: 27745953

41. Burki T. Vaccine misinformation and social media. The Lancet Digital Health. 2019; 1(6):e258–e259. https://doi.org/10.1016/S2589-7500(19)30136-0

42. Ozawa S, Sripad P. How do you measure trust in the health system? A systematic review of the literature. Social science & medicine. 2013; 91:10–14. https://doi.org/10.1016/j.socscimed.2013.05.005

43. Myers LB, Goodwin R. Determinants of adults' intention to vaccinate against pandemic swine flu. BMC Public Health. 2011; 11(1):1–8. https://doi.org/10.1186/1471-2458-11-15 PMID: 21211000

44. Smith TC. Vaccine Rejection and Hesitancy: A Review and Call to Action. Open Forum Infectious Diseases. 2017; 4(3). https://doi.org/10.1093/ofid/ofx146 PMID: 28948177

45. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one. 2014; 9(6):e98679. https://doi.org/10.1371/journal.pone.0098679 PMID: 24914678

46. Mitra T, Counts S, Pennebaker J. Understanding Anti-Vaccination Attitudes in Social Media. ICWSM. 2016;.

47. Sunstein CR, Vermeule A. Conspiracy theories: Causes and cures. Journal of Political Philosophy. 2009; 17(2):202–227. https://doi.org/10.1111/j.1467-9760.2008.00325.x

48. Endres DM, Schindelin JE. A new metric for probability distributions. IEEE Transactions on Information theory. 2003; 49(7):1858–1860. https://doi.org/10.1109/TIT.2003.813506

49. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. Annual review of sociology. 2001; 27(1):415–444. https://doi.org/10.1146/annurev.soc.27.1.415

50. Newman ME. Mixing patterns in networks. Physical review E. 2003; 67(2):026126. https://doi.org/10.1103/PhysRevE.67.026126 PMID: 12636767

51. Giese H, Neth H, Moussaïd M, Betsch C, Gaissmaier W. The echo in flu-vaccination echo chambers: Selective attention trumps social influence. Vaccine. 2020; 38(8):2070–2076. https://doi.org/10.1016/j.vaccine.2019.11.038 PMID: 31864854

52. Baumann F, Lorenz-Spreen P, Sokolov IM, Starnini M. Modeling echo chambers and polarization dynamics in social networks. Physical Review Letters. 2020; 124(4):048301. https://doi.org/10.1103/PhysRevLett.124.048301 PMID: 32058741

53. Evans T, Fu F. Opinion formation on dynamic networks: identifying conditions for the emergence of partisan echo chambers. Royal Society open science. 2018; 5(10):181122. https://doi.org/10.1098/rsos.181122 PMID: 30473855

54. Mønsted B, Sapieżyński P, Ferrara E, Lehmann S. Evidence of complex contagion of information in social media: An experiment using Twitter bots. PLOS ONE. 2017; 12(9):e0184148. https://doi.org/10.1371/journal.pone.0184148 PMID: 28937984

55. Johnson NF, Zheng M, Vorobyeva Y, Gabriel A, Qi H, Velásquez N, et al. New online ecology of adversarial aggregates: ISIS and beyond. Science. 2016; 352(6292):1459–1463. https://doi.org/10.1126/science.aaf0675 PMID: 27313046

56. Johnson N, Leahy R, Restrepo NJ, Velasquez N, Zheng M, Manrique P, et al. Hidden resilience and adaptive dynamics of the global online hate ecology. Nature. 2019; 573(7773):261–265. https://doi.org/10.1038/s41586-019-1494-7 PMID: 31435010

57. Larson HJ. Blocking information on COVID-19 can fuel the spread of misinformation. Nature. 2020; 580 (7803):306. https://doi.org/10.1038/d41586-020-00920-w PMID: 32231320

58. Frimer JA, Skitka LJ, Motyl M. Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. Journal of Experimental Social Psychology. 2017; 72:1–12. https://doi.org/10.1016/j.jesp.2017.04.003

59. Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MF, et al. Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences. 2018; 115(37):9216–9221. https://doi.org/10.1073/pnas.1804840115 PMID: 30154168

60. Majid U, Ahmad M. The Factors That Promote Vaccine Hesitancy, Rejection, or Delay in Parents. Qualitative Health Research. 2020; p. 1049732320933863. https://doi.org/10.1177/1049732320933863 PMID: 32597313

61. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro-and anti-vaccination views. Nature. 2020; p. 1–4. PMID: 32499650

62. Perkins RB, Legler A, Jansen E, Bernstein J, Pierre-Joseph N, Eun TJ, et al. Improving HPV Vaccination Rates: A Stepped-Wedge Randomized Trial. Pediatrics. 2020; 146(1). https://doi.org/10.1542/peds.2019-2737 PMID: 32540986

63. Smith P, Kennedy A, Wooten K, Gust D. Association between health care providers' influence on parents who have concerns about vaccine safety and vaccination coverage. Pediatrics. 2006;. https://doi.org/10.1542/peds.2006-0923 PMID: 17079529

64. Omer SB, Salmon DA, Orenstein WA, Dehart MP, Halsey N. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. New England Journal of Medicine. 2009; 360(19):1981–1988. https://doi.org/10.1056/NEJMsa0806477 PMID: 19420367

65. Omer SB, Richards JL, Ward M, Bednarczyk RA. Vaccination policies and rates of exemption from immunization, 2005–2011. New England Journal of Medicine. 2012; 367(12):1170–1171. https://doi.org/10.1056/NEJMc1209037

66. Bradford WD, Mandich A. Some state vaccination laws contribute to greater exemption rates and disease outbreaks in the United States. Health Affairs. 2015; 34(8):1383–1390. https://doi.org/10.1377/hlthaff.2014.1428 PMID: 26240253

67. Vaz OM, Ellingson MK, Weiss P, Jenness SM, Bardají A, Bednarczyk RA, et al. Mandatory vaccination in Europe. Pediatrics. 2020; 145(2). https://doi.org/10.1542/peds.2019-0620 PMID: 31932361

68. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. American journal of public health. 2018; 108(10):1378–1384. https://doi.org/10.2105/AJPH.2018.304567 PMID: 30138075

69. Helps C, Leask J, Barclay L, Carter S. Understanding non-vaccinating parents' views to inform and improve clinical encounters: a qualitative study in an Australian community. BMJ Open. 2019; 9(5): e026299. https://doi.org/10.1136/bmjopen-2018-026299 PMID: 31142523

70. Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. arXiv preprint arXiv:170800524. 2017;.

71. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. arXiv preprint arXiv:160701759. 2016;.

72. Forman G, Scholz M. Apples-to-apples in cross-validation studies. ACM SIGKDD Explorations Newsletter. 2010; 12(1):49. https://doi.org/10.1145/1882471.1882479