Original Paper

Global Misinformation Spillovers in the Vaccination Debate Before and During the COVID-19 Pandemic: Multilingual Twitter Study

Jacopo Lenti[1,2], MSc; Yelena Mejova[3], PhD; Kyriaki Kalimeri[3], PhD; André Panisson[2], PhD; Daniela Paolotti[3], PhD; Michele Tizzani[3], PhD; Michele Starnini[2,4], PhD

[1]Department of Computer, Control, and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy

[2]CENTAI Institute S.p.A., Turin, Italy

[3]Institute for Scientific Interchange Foundation, Turin, Italy

[4]Departament de Fisica, Universitat Politecnica de Catalunya, Barcelona, Spain

Corresponding Author:
Michele Starnini, PhD
Departament de Fisica, Universitat Politecnica de Catalunya
Campus Nord B4
Barcelona, 08034
Spain
Phone: 34 934 01 62 00
Email: michele.starnini@upc.edu

Abstract

Background: Antivaccination views pervade online social media, fueling distrust in scientific expertise and increasing the number of vaccine-hesitant individuals. Although previous studies focused on specific countries, the COVID-19 pandemic has brought the vaccination discourse worldwide, underpinning the need to tackle low-credible information flows on a global scale to design effective countermeasures.

Objective: This study aimed to quantify cross-border misinformation flows among users exposed to antivaccination (no-vax) content and the effects of content moderation on vaccine-related misinformation.

Methods: We collected 316 million vaccine-related Twitter (Twitter, Inc) messages in 18 languages from October 2019 to March 2021. We geolocated users in 28 different countries and reconstructed a retweet network and cosharing network for each country. We identified communities of users exposed to no-vax content by detecting communities in the retweet network via hierarchical clustering and manual annotation. We collected a list of low-credibility domains and quantified the interactions and misinformation flows among no-vax communities of different countries.

Results: The findings showed that during the pandemic, no-vax communities became more central in the country-specific debates and their cross-border connections strengthened, revealing a global Twitter antivaccination network. US users are central in this network, whereas Russian users also became net exporters of misinformation during vaccination rollout. Interestingly, we found that Twitter's content moderation efforts, in particular the suspension of users following the January 6 US Capitol attack, had a worldwide impact in reducing the spread of misinformation about vaccines.

Conclusions: These findings may help public health institutions and social media platforms mitigate the spread of health-related, low-credibility information by revealing vulnerable web-based communities.

Introduction

Background

The COVID-19 pandemic has extended vaccination from the purview of parents and health-compromised individuals to the purview of the broader public. Restrictions around vaccination created an additional potential to impact one's personal freedom and the world economy, as well as one's health. However, vaccination hesitancy continues to limit the impact of this highly effective intervention [1]: hundreds of thousands of lives were

lost to COVID-19 that could have been prevented with vaccinations in the United States alone [2].

Vaccination hesitancy is a complex issue that has been associated with science denial [3], alternative health practices [4], and belief in conspiracy theories [5]. Among the many factors contributing to vaccine hesitancy is the spread of misinformation, especially on the web [6,7]. The impact of antivaccination content on online social media (OSM) may be compounded by the so-called echo chamber effect [8], in which users' beliefs are reinforced through interactions with like-minded peers [9-11]. Created by the interplay of (1) homophily between users' interactions and (2) polarization of the debate, echo chambers arise from a combination of the psychological tendencies of confirmation bias and selective exposure [12-14] together with algorithmic optimization for greater engagement at the cost of content diversity [15]. Importantly, echo chambers have also been found on OSM in the discussions around vaccination [16-19].

Thus far, scientific studies of the debate around vaccination on OSM have focused on specific countries [17,18,20,21] or English-speaking users [19]. Nevertheless, the COVID-19 pandemic has brought the vaccination discourse to a global scale [22], creating a deluge of international news around the development and deployment of COVID-19 vaccines, including low-quality content and misinformation [23]. The danger of this "infodemic" was acknowledged in mid-2020 by the United Nations and World Health Organization, which called for the member states to develop and implement the necessary action plans [1,24]. Thus, it is imperative to understand the flow of antivaccine—or no-vax—information not only nationally but also internationally to obtain a bird's-eye view on the topic and inform effective communication campaigns.

To address this need, in this work, we focused on the Twitter (Twitter, Inc) platform by leveraging 316 million tweets related to vaccines in 18 different languages from a pre–COVID-19 pandemic era to April 2021 to quantify misinformation flows among users in no-vax communities across national borders and identify which countries are central in the global vaccination debate. To this end, we first investigated (1) how polarized, in terms of echo chambers phenomenon, the vaccination debate is in different countries over time to identify users in no-vax communities and (2) how susceptible, in terms of circulation of information, these no-vax communities are to low-quality information. We proposed a flexible, language-neutral community detection approach and combined it with human-in-the-loop expert knowledge to track polarization and echo chambers in different countries and time periods. We show that communities in which no-vax content was shared (1) increased in number during the pandemic, (2) became less isolated in the national vaccination debate, and (3) displayed much stronger cross-border connections than the rest of the users. Alarmingly, users in these communities tend to heavily rely on low-credibility information sources and to spread it across national borders, resulting in international spillovers of misinformation through a global no-vax network.

Related Works

Vaccination deliberation on Twitter has been studied mainly in English and in the United States [25-27]. However, recently, the platform has gained attention from researchers also focusing on European countries. Before the pandemic, an analysis of the Dutch Twitter revealed an antagonistic relationship between an "anti-establishment" community and the community of journalists and writers, reinforcing the "arrogance of the elite" world view in the former [28]. On the Italian Twitter, the debate around vaccination revealed polarization in terms of retweets (RTs), where vaccine skeptics often mentioned vaccine advocates (generally in attacks), whereas the advocates seemed to ignore the skeptics altogether [17]. Outside Europe, a randomized study on Indonesian Twitter showed the importance of celebrity endorsement in message engagement and that the inclusion of the information source is associated with decreased propagation [29]. The COVID-19 pandemic has spurred increased attention to this topic. A recent examination of vaccine-critical actors on Francophone Twitter found that their place in discussions on vaccines has remained relatively constant during the pandemic compared with the mainstream media [20]. Furthermore, Crupi et al [18] studied the Italian Twitter during the rollout of the COVID-19 vaccinations, showing greater engagement across vaccine-supporting and hesitant communities in terms of mentions and similarity between the communities in the topics discussed.

Attempts to study the flows of vaccination discussion across borders have thus far been limited to dyadic relationships and English. A study of Canadian Twitter users found that most misinformation circulating on Twitter that was shared by Canadian accounts was retweeted from US-based accounts and that increased exposure to US-based information on Twitter is associated with an increased likelihood to post misinformation [30]. Beyond Twitter, Ng et al [22] examined news articles about COVID-19 from 20 countries, identifying the shift in narratives as the pandemic occurred. However, the data were limited to the English language and failed to capture the local language coverage. Unlike the previous studies, our study tracked the vaccination debate in the native languages of numerous countries to systematically study the flow of information (and potential misinformation) across national borders.

The most concerning aspect of the vaccination debates studied here is that misinformation may damage the confidence in the procedure. Controlled exposure studies have shown that web-based misinformation—especially misinformation that sounds scientific—negatively impacts vaccination intent in participants in the United States, United Kingdom [31], and New Zealand [32]. A panel study of US Twitter users found that the risk of average users occasionally sharing misinformation was alarmingly high, despite social bots' contribution to misinformation sharing being "surprisingly low" [33]. Although some efforts have been made toward using high-quality, manually annotated data sets for identifying misinformation [34], the quality of the cited URL domains is often used as a gauge of the quality of the tweet's content [35,36]. In this study, we used a similar approach by combining

lists of low-credibility domains from several languages and countries.

Beyond content analysis, an important aspect of information and misinformation spread is the network structure underlying such dynamic processes. Echo chambers in the Twitter debate around the impeachment of former Brazilian President Dilma Rousseff have been shown to alter the diffusion of information between the supporters and opponents of the impeachment [37]. A similar methodology has been used to compare different topics across social media [8], highlighting that Facebook (Meta Platforms, Inc) shows a higher segregation of news consumption than Reddit (Reddit Inc). Along the same research line, the Random Walk Controversy (RWC) score [38] quantifies how controversial the topics discussed over a certain social network are as the probability of an average user being exposed to information from their own side versus from the opposing side. Although several studies address the presence of echo chambers on social media and their effect on information diffusion, little to no efforts have been devoted to understanding the echo
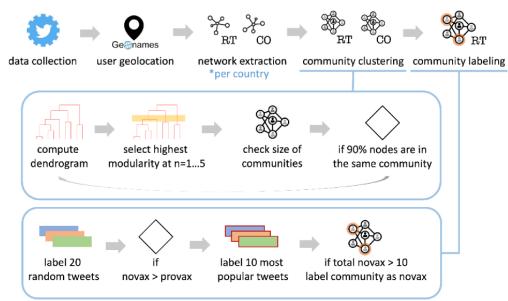
chamber effects within cross-border information spreading, which we examined in this study.

Methods

Overview

The methodology of the data processing pipeline is outlined in the flowchart in Figure 1. First, we used the Twitter Streaming application programming interface (API) to collect a multilingual data set, which we geolocated using the GeoNames database [39]. To identify potential misinformation, we found lists of low-credibility domains in different languages. For the selected countries, we built 2 networks, RT and cosharing (CO; identified by users sharing the same URLs), and applied clustering to find communities. We then manually labeled (in 2 stages) samples of tweets from these communities to identify communities in which users were likely to encounter no-vax content. Finally, we computed several measures to quantify network polarization and information CO, as well as the intensity of cross-national interactions among no-vax communities.

Figure 1. Flowchart of data processing, network extraction, community clustering, and community labeling. RT: retweet; CO: cosharing; novax: number of tweets labelled as discrediting vaccines per community; provax: number of tweets labelled as supporting vaccines per community.
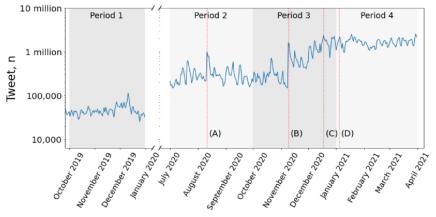

Data Set

We began by assembling a list of vaccine-related words translated into 18 different languages (vaccine, novax, measles, MMR, vaccinated, etc), obtaining a set of 459 keywords (see queries here [40]). An existing list from previous work [17] was expanded by iterative querying of Twitter and expanding the list until no new keywords could be found. Native speakers were then recruited to translate the words into other languages and were instructed to include different common grammatical variations or local relevant keywords. For each language, we query the Twitter Streaming API [41] for the tweets containing the keywords in that language (translated by volunteer native speakers) and keywords in English by applying a language filter. For analysis, we chose four 3-month periods: (1) pre–COVID-19 pandemic period, from October 1, 2019, to December 31, 2019;

(2) prevaccine period, from July 1, 2020, to September 30, 2020; (3) vaccine development period, from October 1, 2020, to December 31, 2020; and (4) vaccine rollout period, from January 1, 2021, to March 31, 2021. Figure 2 presents a summary of the daily volume of the data set. The volume increased by 2 orders of magnitude during the pandemic, from 6 million tweets in the 3-month pre–COVID-19 pandemic period to 39 million tweets in the prevaccine period to 91 million tweets in the vaccine development period to 178 million tweets in the vaccine rollout period. To check the completeness of our data, we ran an Historical API [42] in the pre–COVID-19 pandemic period with the same keywords. Owing to account suspension or post removal by the users themselves, a wide fraction of the tweets (72%) was not retrieved by this API, showing that such a data set cannot be retrieved by a retrospective search. Moreover, we took advantage of the passage of time to revisit the most notable

accounts (present in the networks described subsequently) using the Twitter Get User API call [43] to check on their status, specifically noting whether the accounts have been suspended by the platform or deleted by the users.

Figure 2. Volume of the vaccination debate on Twitter. Some external events with a substantial impact: (A) August 11, 2020: Sputnik V vaccine announced; (B) November 9, 2020: Pfizer-BioNTech vaccine announced; (C) December 18, 2020: Moderna vaccine announced; and (D) January 4, 2021: first AstraZeneca vaccine inoculation.

Geolocation

To capture country-specific dynamics of the social networks, we geolocated the users: we matched the location they provided in their description with the geographical database of locations from GeoNames. Manually verifying the matching accuracy, we filtered out >500 words often associated with nonlocations in this field. To further limit incorrect geolocations, we (1) removed the geolocation of users who changed their country locations during the observed period and (2) manually inspected users responsible for >50% of RTs between 2 pairs of countries in 1 period, assuming that a user who is heavily retweeted from another country is more likely to be wrongly geolocated. Under these conditions, we geolocated 48.7% of the users. This then allowed us to select countries for the study (as the focus was on the Western languages, we selected countries from Europe, North America, South America, and Oceania). To this end, we filtered countries with >2000 unique users in each period, obtaining 28 countries spanning 11 languages. Figure S1 in Multimedia Appendix 1 provides further details on the volume of tweets per language. The total number of different users geolocated in the chosen countries is 14.9 million, corresponding to 39.4% of the total number of users of the data set.

Low-Credibility Domains

Following the previous literature on misinformation tracking [44,45], we collected a list of low-credibility domains. As sources of low-credibility websites, we relied on Bufale (Italian) [46], Wikipedia (English) [47], Media Bias/Fact-Check (English) [48], Le Monde (French) [49], and dwrean (Greek) [50], obtaining a list of 1732 domains. The fact that we were unable to find lists for less-used languages is an important limitation of this work, which we discuss in the *Discussion* section.

Network Reconstruction

For each country, for each period, we built an RT network and a CO network. To limit the number of geolocation mismatches and filter users belonging to debates in other countries, we constrained the tweets considered for each country to the most common language in our data from that country among the languages spoken in the country (according to Wikipedia). The RT network is a directed weighted graph, where each node is a user, and the weight of the directed link ij is the number of times that user i retweeted user j. The CO network is an undirected weighted graph, where each node represents a user, and the weight of the undirected link ij is the number of unique URLs shared by both the users. In order to alleviate the computational cost of the network analyses, we filtered out the edges with weight equal to 1 (just 1 retweet) for the networks with more than 200.000 nodes. This filter affects the RT and CO networks in the United States (all periods), Brazil (PD, VD, and VR periods), Great Britain (VD and VR), and Spain and Mexico (VR period). This filter affects only the country-specific analyses of the RT networks, without influencing the later cross-country analysis and the findings about suspended accounts. When considering the constructed networks, we focused on the Giant Connected Component (GCC). On average, the GCC of the RT network contains 92% of its nodes, while the GCC of the CO network contains 76% of its nodes. In addition, we measured the Overlap Coefficient (OC) between the sets of users in the RT and CO networks. The OC is defined as the ratio of the size of the intersection of 2 sets, A and B, to the size of the smaller set, that is, $OC(A,B) = |A \cap B| / \min(|A|, |B|)$. During the vaccine rollout period, the OC between the sets of users in the RT and CO networks increases from 0.72 (pre–COVID-19 pandemic period) to 0.86, indicating that more people are sharing URLs. The total number of users in the reconstructed RT and CO networks is 2.7 million.

Hierarchical Clustering

Next, we applied a community detection algorithm to cluster the users of the RT and CO social networks. Because the goal was to find a small number of large groups of users, we adopted hierarchical clustering, instead of unsupervised algorithms (eg, Louvain), which finds the optimal partitioning with a very large number of often small communities. We used Paris, an agglomerative hierarchical clustering algorithm induced by the probability of sampling node pairs [51]. Next, we cut the

dendrogram to have a reasonable number of communities that are not too unbalanced following the steps listed here:

1. Build the dendrogram of the hierarchical clustering.

2. Compare the partitions obtained with cutoffs at heights 2, 3, 4, and 5 (ie, having this number of communities).

3. Pick the partition with the highest modularity.

4. If >90% of the nodes are in the same community, compare the partitions with cutoffs at the following 5 heights and repeat from step 3.

Using this procedure, we ensured that 90% of the users were partitioned into at least 2 communities but no >5 communities (although it is possible to have many small communities that comprise <10% of the nodes). Using this method, we found, on average, 6.7 communities in the RT networks and 5.0 communities in the CO networks, with a maximum of 20.

Labeling

To identify communities in which users were exposed to no-vax content, we labeled a sample of tweets shared in each community. First, we filtered out small communities by considering only those with >1% of the users of the network, resulting in 400 communities. Next, we randomly sampled 20 tweets from each community, resulting in a total of 8000 tweets. A total of 12 people were involved in labeling, all of whom had a background in vaccine debate and knowledge of the language used in the tweet to label. Furthermore, we translated all tweets into English using Google Translate to allow for cross-checking. Each person labeled between 600 and 1000 tweets, with an overlap of 20 tweets with other annotators. The tweets were labeled as "pro-vax," "no-vax," or "other." We labeled tweets as pro- or no-vax only if they were clearly supporting or discrediting vaccines, respectively. Therefore, more than half of the labels were "other," comprising nonrelevant posts, posts with unclear positions, discussions on other policies, and all generic pieces of news that did not express a stance. The task of distinguishing between pro vaccination and antivaccination stances proved to be fairly easy, with Cohen κ computed on an overlapping set at κ=0.84 (only 3% received different labels). By contrast, the task involving the "other" label proved to be more difficult, with κ=0.51 for the 3-class setting (disagreement of 26%), mainly because of the confusion between "other" and "pro-vax" labels (disagreement of 20%). However, note that we were only interested in distinguishing between the antivaccination stance and the rest.

To improve the quality of the labels, we then proceeded to a second round of annotation, focusing on the communities that have a majority of content with a no-vax stance. Specifically, we chose the communities with a majority of no-vax tweets and annotated the 10 most popular tweets in each (excluding the 50 most popular tweets in the whole network). The second labeling stage encompassed 82 communities, totaling in 820 tweets. At this stage, the Cohen κ for the 3 classes was 0.64. Finally, we defined a community as no-vax if the total number of "no-vax" labels in the rounds was >10, resulting in 58 communities. Because some networks had >1 antivax community, we had 52 networks with a no-vax community, that is, a community where

users were substantially exposed to no-vax content (Figure S2 in Multimedia Appendix 1).

Clustering Robustness

Next, we assessed the robustness of our approach to determine if our methodology influenced the results. To do so, we compared the communities previously identified in the RT networks with those obtained using 2 alternative partitioning algorithms: Louvain and Paris hierarchical clustering with a cutoff of 10. We chose the Louvain algorithm for its popularity in community detection problems in social network analysis and Paris with a cutoff of 10 for comparing the results obtained with a different parameter in the cutoff of the same dendrogram. We quantified the number of labeled tweets shared by users in the new clustering and categorized communities as "no-vax" using 2 different thresholds: the "majority threshold" and the "strict threshold." The former was applied when "no-vax" labels outnumbered "pro-vax" labels, while the latter was used when "no-vax" labels surpassed both "pro-vax" and "other" labels. This yielded 4 alternatives to our method: Louvain partitioning with majority threshold, Louvain partitioning with strict threshold, hierarchical clustering with majority threshold, and hierarchical clustering with strict threshold.

To evaluate the robustness of our methodology, we calculated the accuracy for each network as the proportion of users classified in the same group as the previous method (either "no-vax" or "not no-vax"). Our results demonstrate a high level of robustness, with average accuracies of 0.90 and 0.94 using Louvain partitioning with majority and strict thresholds (SD 0.15 and 0.10) and 0.92 and 0.95 using hierarchical clustering (SD 0.15 and 0.10), respectively. These findings support the consistency of the results presented in this paper, which are not overly dependent on the methodology used to detect and label communities.

RWC Score

Following previous literature [10,17,52], we used the RWC score to quantify the polarization between the communities labeled as no-vax and the rest of the network. Given an RT network, partitioned into 2 clusters X and Y, RWC is calculated as $P_{XX} P_{YY} - P_{XY} P_{YX}$, where $P_{XY}=P$ (a random walk ended in Y started in X). Intuitively, it represents the difference in probability for an average user in the network to be exposed to information from their own side versus that from the opposing side. Spanning (0, 1), an RWC close to 1 represents a polarized social network with 2 distinct groups that do not endorse each other's opinions, whereas an RWC close to 0 represents a noncontroversial topic where both opinions are equally likely to be received.

Normalized Mutual Information

We quantified the echo chamber effect by measuring the extent to which users from different RT communities shared the same sources of information (as quantified by the CO network), as a proxy for the information siloing in an echo chamber. To do this, we used normalized mutual information (NMI) [53] to gauge the similarity between the RT and CO communities obtained by hierarchical clustering, using the normalized_mutual_info_score module in the Python package

scikit-learn. Spanning (0, 1), an NMI of 1 means that the community structure is the same between the networks, whereas a low NMI indicates different community structures. Note that this metric did not use the opinion leaning determined in the labeling step and thus could be computed for any country/period network, regardless of whether it had a no-vax community.

Normalized RT Volume

To assess the extent to which one country retweets another, we computed a normalized retweeting volume for each pair of countries. To this aim, we started by the total number of RTs from country i to country j (which we have indicated as a_{ij}). Then, we divided it by the total number of RTs by users in country i (indicated as s_i^{out}) and the total number of RTs to users in country j (indicated as s_j^{in}), and we multiplied it by the total number of RTs by all countries (indicated by W). The normalized retweet volume n_{ij} is thus equal to the following:

$$n_{ij} = \frac{a_{ij}}{s_i^{out} s_j^{in}} W$$

Note that $W = \sum_i s_i^{out} = \sum_i s_i^{in}$, and $s_i^{out} s_j^{in}/W$ is the expected number of RTs from country i to country j in the random graph with the same node strengths. Hence, $n_{ij} > 1$ if i retweets j more than it would in a random baseline context. As the vast majority of RTs were within the same countries, $a_{ij} \gg s_i^{out} s_j^{in}/W$ if $i \neq j$, otherwise $a_{ij} \ll s_i^{out} s_j^{in}/W$. To focus on cross-border interactions, we consider $n_{ii} = 0$, for any i.

Cross-Border Interactions Between No-Vax Communities

We measured the strength of ties among the users in no-vax communities in different countries by comparing the number of RTs among these users with the number of RTs among the rest of the users in the same countries. In particular, we define V_i^K as the set of users in communities with stance K in a country i, where K can be A (antivax) or O (others). We define W_{ij}^K as the number of RTs from users in V_i^K to users in country j with the same stance K, V_j^K. Thus, one can measure the density, $\delta_{ij}^K = W_{ij}^K / V_i^K V_j^K$ the ratio of observed RTs; and the total possible pairs between sets V_i^K and V_j^K, that is, the probability that 2 random users in V_i^K and V_j^K, respectively, are connected. For each pair of countries, we analyzed if $\theta_{ij} = \delta_{ij}^A / \delta_{ij}^O > 1$, it means that the probability that 2 random users in no-vax communities in countries i and j are connected is higher than the probability that 2 random users in the rest of the network of the same countries are connected.

Ethical Considerations

Although the data were collected using the platform's own API, resulting only in posts that were posted publicly, it is possible that some users were unaware of the scope of their potential audience. Thus, we follow the platform's terms of service and share only the IDs of the tweets so that when the data are

recollected, deleted content will not be available (notably limiting the reproducibility of any Twitter-based study). Thus, the data that are shared do not contain any identifiable information about the poster or any other information except the numeric ID of the tweet, preserving the privacy of the user to the extent that they choose to keep their posts public on the platform. This will affect the reproducibility of the study, as some content may be deleted by the users over time. Furthermore, the multinational nature of the data captured wildly varying biases in the way people around the world are able to access the internet or Twitter specifically. Local barriers to access to the internet and local blocks of the platform itself shape the communities captured in this study. For instance, dissidents or those who wished to remain anonymous would likely not have shared their location information on their profile and would not have been captured as being a part of a country's discussion. The construction of retweet and cosharing networks also necessitates enough activity by the user to be included in the analysis, biasing our results to those who are more active in the conversation, especially in retweeting and sharing URLs. Moreover, the data may have captured vulnerable groups, including those who experienced or who were at risk of specific health conditions, those who had financial barriers to health care, and even those who were more susceptible to misinformation. Despite the negative connotation around "no-vax" communities, users found to propagate harmful information may first and foremost endanger themselves by following faulty advice. Thus, we would discourage the future researchers from publishing verbatim tweet text to preserve user privacy.

Finally, in this paper, we present tools that may be used to track and profile groups of Twitter users around a topic. These tools may then be used by both the platform and the government. However, such tools may also be used to target communities for harassment, doxing (providing private user information to harm or intimidate the person), and other abuses. On the one hand, it is the responsibility of the platforms and their communities to uphold the civil code of conduct and block the abusers. On the other hand, we call for the research community, as well as corporate and governmental actors, to use these tools ethically, with minimal harm to the participants.
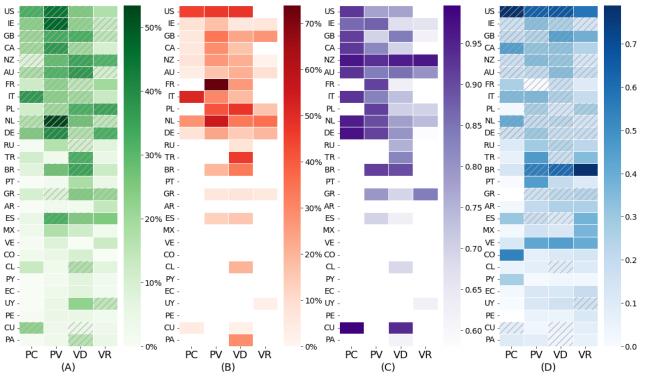
Results

Polarization of the Vaccination Debate

We began by examining different measures of polarization and no-vax activity in different countries over the 4 periods. Figure 3A shows that a high presence of no-vax tweets in a certain country and period is often associated with the presence of a community labeled as no-vax (dashed lines). This implies that no-vax content is generally clustered and not homogeneously distributed in the RT network, suggesting that the debate is polarized, as illustrated subsequently. Furthermore, we found that no-vax communities were generally present in the English-speaking countries (eg, compared with the Spanish-speaking ones). However, some of the relatively largest country-specific no-vax communities appeared in France, Italy, the Netherlands, Poland, and the United States (Figure 3B).

No-vax communities were particularly present in the prevaccine and vaccine development periods, where they also spanned a larger fraction of users compared with the other periods. Turning to potential echo chambers in these networks, we found that the RWC score was overall very high (Figure 3C), indicating that the vaccination debate was generally highly polarized. However, it decreased substantially over time, suggesting that the users in no-vax communities became less isolated in the vaccination discourse during the COVID-19 pandemic. Furthermore, we investigated whether the users in the no-vax communities were exposed to information sources that were different from those that the rest of the users were exposed to by considering NMI.

Despite the NMI being independent of the labeling of the communities, Figure 3D shows that, on average, the NMI of the networks with a no-vax community was higher than that of the others (0.27 vs 0.22, P<.05), indicating that the users in no-vax communities tended to have common information sources. Some countries, such as the United States and Brazil, showed an especially high NMI, indicating that the polarization in the RT network was reflected in the different content shared. The Spanish-speaking countries, conversely, were less polarized than the English-speaking countries (average 0.15 vs 0.33, P<.001).

Figure 3. Characterization of no-vax communities for each country and period considered via retweet (RT) and cosharing (CO) networks. (A) Proportion of tweets labeled as no-vax. (B) Proportion of users in no-vax communities with respect to the size of the RT networks (average in the 4 periods: 16.9% [SD 0.18], 30.9% [SD 0.18], 23.1% [SD 0.14], and 13.7% [SD 0.12]). (C) Random Walk Controversy between no-vax communities and the rest of the networks (average in the 4 periods: 0.94 [SD 0.04], 0.84 [SD 0.08], 0.76 [SD 0.10], and 0.73 [SD 0.12]). (D) Normalized mutual information between RT and CO communities. Countries with no-vax communities are marked with dashed lines. PC: pre–COVID-19 period; PV: prevaccine period; VD: vaccine development period; VR: vaccine rollout period. AR: Argentina; AU: Australia; BR: Brazil; CA: Canada; CL: Chile; CO: Colombia; CU: Cuba; DE: Germany; EC: Ecuador; ES: Spain; FR: France; GB: Great Britain; GR: Greece; IE: Ireland; IT: Italy; MX: Mexico; NL: Netherlands; NZ: New Zealand; PA: Panama; PE: Peru; PL: Poland; PT: Portugal; PY: Paraguay; RU: Russia; TR: Turkey; US: United States; UY: Uruguay; VE: Venezuela.

Characterizing the Users in No-Vax Communities

Considering the behavior of the users in no-vax communities, we found that they were more likely to retweet (Figure 4A) and share URLs (Figure 4B), especially URLs to YouTube (Figure 4C), than the other users. Furthermore, the URLs they posted were much more likely to have been from low-credibility domains (Figure 4D) compared with those posted in the rest of the networks. The difference is remarkable: 26% of the domains shared by no-vax communities came from lists of known low-credibility domains versus only 2.4% of those cited by the other users came from lists of known low-credibility domains (P<.001). The most common low-credibility websites among the no-vax communities were *Zero Hedge*, *LifeSiteNews*, *Daily Mail* (considered right-biased and questionably sourced), and

Children's Health Defense (conspiracy/pseudoscience). These findings extend the existing literature on English-language vaccination rhetoric to a multilingual, international scope by confirming the elevated social engagement in antivaccination communities [54] and provide additional evidence of the misleading nature of the popular COVID-19 pandemic–related YouTube videos [55].

Next, we investigated the effects of content moderation by Twitter on the vaccination debate. We found that the average proportion of suspended accounts in no-vax communities was much larger than that among the rest of the users for each country and period considered (average 13.3% vs 1.8%, P<.001; Figure 5A). The highest proportions of suspended accounts were found in the English-speaking countries, Germany, and the Netherlands, which also showed a larger presence of no-vax

content, than in the other countries. Furthermore, a large portion of suspensions came after the January 2021 US Capitol attack in Washington, DC [56] (Figure 5B). The proportion of suspended accounts from the United States increased from 38% before January 1 to 77% during the days around the Washington riots (January 1-12). Note that (1) 89% of the US users who were suspended belonged to the no-vax community in the vaccine development period; (2) the account realDonaldTrump (suspended on January 8) was one of the most popular accounts among the no-vax communities of the first 3 periods; and (3) in the last period, a no-vax community was not present in the US RT network, indicating that the suspension of US accounts following the Washington riots heavily impacted the vaccination debate on Twitter. These findings suggest that political leaning is often associated with strong stances taken in the vaccination debate (in line with previous literature [17,21]) and that actions taken in the political domain may greatly impact the quality of the public health discourse.

Figure 4. Behavior of users in no-vax communities versus those of other users. (A) Average retweets, (B) average URLs, (C) average YouTube URLs, and (D) proportion of low-credibility domains shared by users. Note that low-credibility domains were collected only in Italian, French, English, and Greek; therefore, the plots refer to countries speaking these languages. No-Vax: discrediting vaccines. Other: non-discrediting vaccines.
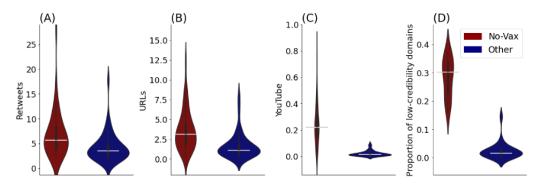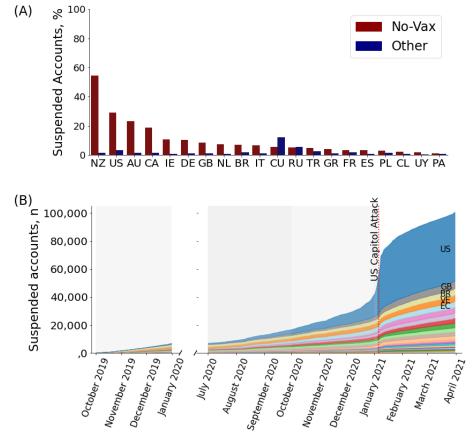

Figure 5. Suspended users per country in no-vax communities. (A) Average proportion of suspended accounts per country in the period in which no-vax community has been detected, computed separately for no-vax side and rest of users. (B) Number of suspended accounts as a function of the date they posted their last tweet, colored by country. No-Vax: discrediting vaccines. Other: non-discrediting vaccines. AR: Argentina; AU: Australia; BR: Brazil; CA: Canada; CL: Chile; CO: Colombia; CU: Cuba; DE: Germany; EC: Ecuador; ES: Spain; FR: France; GB: Great Britain; GR: Greece; IE: Ireland; IT: Italy; MX: Mexico; NL: Netherlands; NZ: New Zealand; PA: Panama; PE: Peru; PL: Poland; PT: Portugal; PY: Paraguay; RU: Russia; TR: Turkey; US: United States; UY: Uruguay; VE: Venezuela. No-Vax: plots refer to users in no-vax communities. Other: plots refer to users in other communities.

Cross-Border Information Spillover in the Global Vaccination Debate

Next, we quantified the information spillover across countries by considering the number of RTs from one country to another, normalized by the total number of RTs produced and received in the 2 countries (see Normalized RT Volume in the Methods section; Figure 6A). First, one can observe language homophily, indicated by the darker regions in the top left (English) and bottom right (Spanish) of the panels, as well as the pair Portugal-Brazil, in all periods. The darker patches corresponding to the interactions between Germany and the Netherlands and those between Germany and Turkey also reflect possible cultural or expat relationships. Second, the cross-border interaction matrices are not symmetrical: information generally flows in a preferred direction. For instance, the Spanish-speaking countries retweeted the English-speaking countries much more than the English-speaking countries retweeted the Spanish-speaking countries. Note that the United States is central in the global information flow, being a net exporter of information to the rest of the world when comparing inflows versus outflows of information for each country (Figure S3 in Multimedia Appendix 1). Interestingly, from the prevaccine period, Russia also became a net exporter, especially to South American countries (Figure S3 in Multimedia Appendix 1): some of the most used hashtags in the prevaccine and vax development periods are *#sputnikesesperanza* and *#sputnikparaelpueblo*.

In Figure 6B, we quantified the strength of the cross-border interactions among the users in no-vax communities compared with that among the rest of the users (see Cross-Border Interactions Between No-Vax Communities in the Methods section). We found that cross-border interactions among the users in no-vax communities were generally much stronger, sometimes by orders of magnitude, than the interactions among the rest of the users, creating a tightly knit global no-vax network. In particular, the users in the no-vax communities of the English-speaking countries, Germany, and the Netherlands were tightly connected in all periods. By contrast, the users in the no-vax communities from Cuba and Russia were isolated (adding to their unusual user suspension statistics). Again,
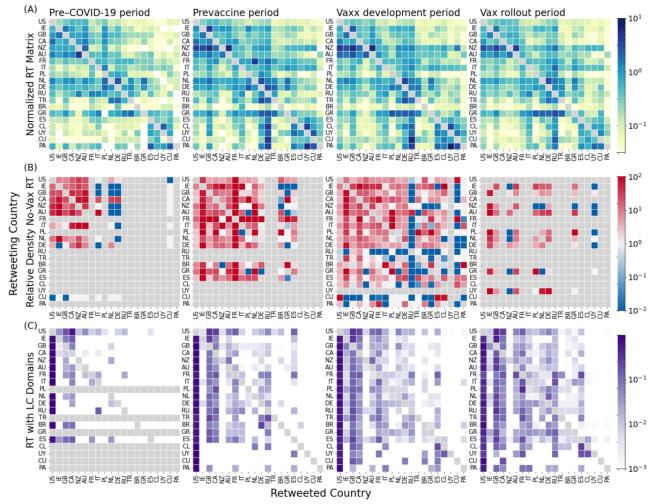
cross-border interactions can be asymmetrical: for instance, in the pre–COVID-19 pandemic period, the users in the no-vax communities in Germany and the Netherlands retweeted the users in the other countries, but not vice versa.

Finally, we focused on misinformation flows across countries by considering the fraction of low-credibility domains imported per country (Figure 6C), that is, the fraction of tweets pointing to low-credibility URLs, over the total number of RTs from one country to another. We stress that we considered flows of low-credibility information across borders spread by both humans and bots, without engaging in the difficult task of distinguishing them, as we were interested in quantifying how exposed a certain country A is to misinformation coming from country B. As in the previous case, the matrices show a clear asymmetry. The US users were responsible for exporting a large fraction of misinformation to the rest of the world: 68% of all the low-credibility URLs retweeted worldwide came from the United States (average over the 4 periods), a proportion much higher than the total volume of URLs (42%) retweeted from the United States.

Interestingly, the fraction of low-credibility URLs from the United States dropped from 74% in the vax development period to 55% in the vax rollout period. This large decrease can be directly ascribed to Twitter's moderation policy: 46% of cross-border RTs of US users linked to low-credibility websites in the vax development period came from accounts that were suspended following the US Capitol attack (Figure S5A in Multimedia Appendix 1). Note that Twitter's account purge substantially impacted the misinformation spread worldwide: the proportion of low-credibility domains in the URLs retweeted from the United States dropped from 14% to 7%. Finally, despite not having a list of low-credibility domains in Russian, Russia was central in exporting potential misinformation in the vax rollout period, especially to Latin American countries. In these countries, the proportion of low-credibility URLs coming from Russia increased from 1% in the vax development period to 18% in the vax rollout periods (Figure S5B in Multimedia Appendix 1).

Figure 6. Cross-border information flows in the global vaccination debate. (A) Normalized retweet (RT) matrix. Normalized number of retweets (excluding diagonal elements, colored in gray). (B) The ratio between probability of interactions between users in no-vax communities and rest of the networks from the same pair of countries (see the Methods section). Darker red (blue) elements of the matrices represent higher (lower) tendency of cross-border interactions between users in no-vax communities with respect to other users. Countries without no-vax communities are colored in gray. (C) Proportion of URLs that come from the retweeted country among the low-credibility (LC) domains imported from the retweeting country. Countries importing ≤10 low-credibility URLs are colored in gray. Diagonal elements are also gray, as our focus was on cross-country interactions. Element a$_{ij}$ of each matrix represents the information flow from country j to country i. In the plots, only the countries with at least 1 no-vax community in the 4 periods are represented. The extended version is Figure S4 in Multimedia Appendix 1.

Discussion

Principal Findings

The international, multilingual nature of the data we present here supports the ongoing efforts in monitoring the non-English debate around the topic of vaccination [57-59]. Using this information, we reveal the increasingly globalized nature of the vaccination debate as the COVID-19 vaccines were proposed, developed, and deployed. This increased globalization had a marked impact on vaccine-hesitant discourse: not only did the prominence of the no-vax communities increase within individual countries but their cross-border connections also strengthened around the world. We showed that the users in these communities are much more prone to sharing potential misinformation than other users, even across national borders.

Furthermore, the real-time nature of the data collection allowed us to capture Twitter's content moderation efforts, which proved to be uneven both across countries and time. The users blocked immediately following the January 6 Washington riots were

responsible for a substantial amount of misinformation spread—both within the United States and, crucially, internationally. Thus, we paint a picture of a "global no-vax Twitter network" that calls for the international collaboration of both public health and technology experts.

Limitations

Our study has several important limitations. First and foremost, it is well known that Twitter users are not a representative sample of the real population but are biased toward more educated, urban, younger, and male individuals [60]. Furthermore, Twitter use wildly differs among the countries under consideration, so cross-country comparisons should be taken with caution. With respect to this, the geolocation task also introduced some bias in the results due to the different fraction of missing accounts across countries. However, to the best of our knowledge, there are no means to collect real-time, representative data at this spatial and temporal scale. Note that we did not engage in bot detection, as this task is notoriously difficult [61], and, most importantly, misinformation can be

spread by complex interactions by bots and humans [44]. Moreover, our study was limited to 11 chosen languages and to the 4 languages for which low-credibility domains were collected, a limitation necessary to control the cultural heterogeneity of the data analyzed. However, the fact that low-credibility domains in other languages were not found (for instance, it was challenging to find a reputable list of low- or high-credibility domains in Russian) means that the potential misinformation flows presented here are a lower bound—one which should be expanded using additional resources. Furthermore, the content considered for this study is limited to the keywords and the query processing of the Twitter search engine. Note that we did not check for spelling errors, which may lead to underdetection of some tweets. However, we did perform a relevance check on a random selection of 300 tweets, resulting in 7% of pet-related tweets, 6% of nonrelevant tweets, and the remaining tweets relevant to vaccination. As we did not modify the keywords as events unfolded—notably when new vaccines were developed—in the aim of keeping a consistent methodology allowing for comparison over time, some content pertaining to time-specific keywords was missed (although we were still able to capture a large amount of discussion around these developments with existing keywords).

Future work should also be devoted to including countries from Africa and Asia, as well as to update and extend the list of low-credibility information sources to other languages. For the latter task, one could leverage the identification of no-vax communities—more susceptible to share low-credibility information—proposed in this study. Other possible limitations of this study include the method used to identify no-vax communities, hierarchical clustering of the RT network and labeling of the popular tweets in the resulting communities, which may have been sensitive to the thresholds adopted. However, note that this method was not aimed at detecting the stance of single users about vaccination but at identifying large clusters of users exposed to a certain kind of antivaccine narrative.

Broader Impact

Despite the platform's tweet flagging and removal policies around COVID-19 [62,63], it is the bout of account suspensions around the Washington riots that made the most impact on the national and international spread of vaccine-related misinformation, suggesting that political concerns elicit much stronger curbing of the freedom of speech than health concerns. It is possible that the effects of this event changed the social

media landscape itself, with platforms such as Truth Social appearing in the aftermath of the event. More documentation of the causal link between web-based misinformation and adverse health outcomes may provide a more solid ground for making censorship decisions for both the platforms and the politicians governing them. For instance, a randomized controlled trial in the United Kingdom and the United States showed that "relative to factual information, recent misinformation induced a decline in intent of 6.2 percentage points" [31].

Further, the Centers for Disease Control and Prevention and the Kaiser Family Foundation estimate that the lack of action early in the pandemic may have contributed to deaths of hundreds of thousands by June 2021 [2]. Furthermore, this study illustrates the impact of 1 social media platform's editorial policies on the international public health discourse, especially when the country involved is as culturally influential as the United States. Without examining in detail the content shared by the suspended accounts, we cannot be certain that the accounts indeed were sharing harmful content. Monitoring the censorship activities of major platforms (triggered by either internal policies or governments' requests) is important for assessing the users' freedom of speech. For instance, the Electronic Frontier Foundation has recently criticized social media platforms for blocking political dissidents who a decade ago used the same platforms to "push for political change and social justice" [64]. Fortunately, "de-platforming as censorship" is a topic of ongoing deliberation at the European Union's Internet Governance Forum involving civil society and government representatives [65].

An international perspective may also benefit the tracking of malicious actors, such as semiautomated or fully automated accounts, networks of colluding agents, and sources of poor-quality content. It has been shown that accounts identified as Russian trolls were more likely to tweet about vaccination before the pandemic [66]. During the pandemic, Russian trolls often posted misinformation concerning the personal dangers of vaccines, purported civil liberty violations, and vaccine conspiracies [67]. Since the beginning of the Ukraine war, it has been noted that antivaccine content has diminished dramatically, potentially because of the additional blocking of Twitter in Russia and refocusing of the conspiratorial attention on Ukraine [68]. Our findings suggest that changes in governance and censorship may encourage or discourage the flow of potential misinformation from states with known affinities.

Data Availability

All codes generated in the project have been deposited to GitHub [69]. All data collected in the project have been deposited to Zenodo [70].

Authors' Contributions

All authors contributed to the development of the main ideas of this study. JL analyzed the data and prepared the figures. MT contributed to the data collection. YM and MS led the interpretation of the results and preparation of the manuscript with contributions from all the authors. All the authors critically reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional figures for the data selection, annotation and analysis.
[DOCX File , 1124 KB-Multimedia Appendix 1]

References

1. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2023-04-27]
2. Amin J, Ortaliza J, Cox C, Michaud J, Kates J. COVID-19 mortality preventable by vaccines. Health System Tracker. 2022 Apr 21. URL: https://www.healthsystemtracker.org/brief/covid19-and-other-leading-causes-of-death-in-the-us/ [accessed 2023-04-27]
3. Browne M, Thomson P, Rockloff MJ, Pennycook G. Going against the herd: psychological and cultural factors underlying the 'Vaccination Confidence Gap'. PLoS One 2015 Sep 1;10(9):e0132562 [FREE Full text] [doi: 10.1371/journal.pone.0132562] [Medline: 26325522]
4. Kalimeri K, Beiró MG, Urbinati A, Bonanomi A, Rosina A, Cattuto C. Human values and attitudes towards vaccination in social media. In: Companion Proceedings of The 2019 World Wide Web Conference. 2019 May Presented at: WWW '19; May 13-17, 2019; San Francisco, CA, USA p. 248-254. [doi: 10.1145/3308560.3316489]
5. Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. PLoS One 2014 Feb 20;9(2):e89177 [FREE Full text] [doi: 10.1371/journal.pone.0089177] [Medline: 24586574]
6. Burki T. Vaccine misinformation and social media. The Lancet Digit Health 2019 Oct;1(6):e258-e259. [doi: 10.1016/s2589-7500(19)30136-0]
7. Carrieri V, Madio L, Principe F. Vaccine hesitancy and (fake) news: quasi-experimental evidence from Italy. Health Econ 2019 Nov;28(11):1377-1382 [FREE Full text] [doi: 10.1002/hec.3937] [Medline: 31429153]
8. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
9. Del Vicario M, Vivaldo G, Bessi A, Zollo F, Scala A, Caldarelli G, et al. Echo chambers: emotional contagion and group polarization on Facebook. Sci Rep 2016 Dec 01;6:37825 [FREE Full text] [doi: 10.1038/srep37825] [Medline: 27905402]
10. Garimella K, De Francisci Morales G, Gionis A, Mathioudakis M. Political discourse on social media: echo chambers, gatekeepers, and the price of bipartisanship. In: Proceedings of the 27th World Wide Web Conference. 2018 Presented at: WWW '18; April 23-27, 2018; Lyon, France. [doi: 10.1145/3178876.3186139]
11. Zollo F, Novak PK, Del Vicario M, Bessi A, Mozetič I, Scala A, et al. Emotional dynamics in the age of misinformation. PLoS One 2015 Sep 30;10(9):e0138740 [FREE Full text] [doi: 10.1371/journal.pone.0138740] [Medline: 26422473]
12. Bakshy E, Messing S, Adamic LA. Political science. Exposure to ideologically diverse news and opinion on Facebook. Science 2015 Jun 05;348(6239):1130-1132. [doi: 10.1126/science.aaa1160] [Medline: 25953820]
13. Cinelli M, Brugnoli E, Schmidt AL, Zollo F, Quattrociocchi W, Scala A. Selective exposure shapes the Facebook news diet. PLoS One 2020 Mar 13;15(3):e0229129 [FREE Full text] [doi: 10.1371/journal.pone.0229129] [Medline: 32168347]
14. Quattrociocchi W, Scala A, Sunstein CR. Echo chambers on Facebook. SSRN J 2016 Jun 15. [doi: 10.2139/ssrn.2795110]
15. Schmidt AL, Zollo F, Del Vicario M, Bessi A, Scala A, Caldarelli G, et al. Anatomy of news consumption on Facebook. Proc Natl Acad Sci U S A 2017 Mar 21;114(12):3035-3039 [FREE Full text] [doi: 10.1073/pnas.1617052114] [Medline: 28265082]
16. Bello-Orgaz G, Hernandez-Castro J, Camacho D. Detecting discussion communities on vaccination in Twitter. Future Gener Comput Syst 2017 Jan;66:125-136. [doi: 10.1016/j.future.2016.06.032]
17. Cossard A, De Francisci Morales G, Kalimeri K, Mejova Y, Paolotti D, Starnini M. Falling into the echo chamber: the Italian vaccination debate on Twitter. Proc Int AAAI Conf Web Soc Media 2020 May 26;14(1):130-140. [doi: 10.1609/icwsm.v14i1.7285]
18. Crupi G, Mejova Y, Tizzani M, Paolotti D, Panisson A. Echoes through time: evolution of the Italian COVID-19 vaccination debate. Proc Int AAAI Conf Web Soc Media 2022 May 31;16(1):102-113. [doi: 10.1609/icwsm.v16i1.19276]
19. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018 Jun 14;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]

XSL•FO
RenderX

20. Faccin M, Gargiulo F, Atlani-Duault L, Ward JK. Assessing the influence of French vaccine critics during the two first years of the COVID-19 pandemic. PLoS One 2022 Aug 4;17(8):e0271157 [FREE Full text] [doi: 10.1371/journal.pone.0271157] [Medline: 35925962]

21. Motta M, Stecula D, Farhart C. How right-leaning media coverage of COVID-19 facilitated the spread of misinformation in the early stages of the pandemic in the U.S. Can J Pol Sci 2020 May 01;53(2):335-342. [doi: 10.1017/s0008423920000396]

22. Ng R, Chow TY, Yang W. News media narratives of COVID-19 across 20 countries: early global convergence and later regional divergence. PLoS One 2021 Sep 1;16(9):e0256358 [FREE Full text] [doi: 10.1371/journal.pone.0256358] [Medline: 34469446]

23. Rovetta A, Bhagavathula AS. Global infodemiology of COVID-19: analysis of Google web searches and Instagram hashtags. J Med Internet Res 2020 Aug 25;22(8):e20673 [FREE Full text] [doi: 10.2196/20673] [Medline: 32748790]

24. Managing the COVID-19 infodemic: promoting healthy behaviours and mitigating the harm from misinformation and disinformation: joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC. World Health Organization. 2020 Sep 23. URL: https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation [accessed 2023-04-27]

25. Blankenship EB, Goff ME, Yin J, Tse ZT, Fu KW, Liang H, et al. Sentiment, contents, and retweets: a study of two vaccine-related twitter datasets. Perm J 2018;22:17-138 [FREE Full text] [doi: 10.7812/TPP/17-138] [Medline: 29911966]

26. Kennedy A, Lavail K, Nowak G, Basket M, Landry S. Confidence about vaccines in the United States: understanding parents' perceptions. Health Aff (Millwood) 2011 Jun;30(6):1151-1159. [doi: 10.1377/hlthaff.2011.0396] [Medline: 21653969]

27. Meadows CZ, Tang L, Liu W. Twitter message types, health beliefs, and vaccine attitudes during the 2015 measles outbreak in California. Am J Infect Control 2019 Nov;47(11):1314-1318. [doi: 10.1016/j.ajic.2019.05.007] [Medline: 31266661]

28. Lutkenhaus RO, Jansz J, Bouman MP. Mapping the dutch vaccination debate on Twitter: identifying communities, narratives, and interactions. Vaccine X 2019 Apr 11;1:100019 [FREE Full text] [doi: 10.1016/j.jvacx.2019.100019] [Medline: 31485580]

29. Alatas V, Chandrasekhar AG, Mobius M, Olken BA, Paladines C. When celebrities speak: a nationwide twitter experiment promoting vaccination in Indonesia. National Bureau of Economic Research. 2019 Feb. URL: https://www.nber.org/papers/w25589 [accessed 2023-04-27]

30. Bridgman A, Merkley E, Zhilin O, Loewen PJ, Owen T, Ruths D. Infodemic pathways: evaluating the role that traditional and social media play in cross-national information transfer. Front Polit Sci 2021 Mar 29;3:20. [doi: 10.3389/fpos.2021.648646]

31. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]

32. Thaker J, Subramanian A. Exposure to COVID-19 vaccine hesitancy is as impactful as vaccine misinformation in inducing a decline in vaccination intentions in new zealand: Results from pre-post between-groups randomized block experiment. Front Commun 2021 Aug 19;6:721982. [doi: 10.3389/fcomm.2021.721982]

33. Teng X, Lin Y, Chung W, Li A, Kovashka A. Characterizing user susceptibility to COVID-19 misinformation on Twitter. Proc Int AAAI Conf Web Soc Media 2022 May 31;16(1):1005-1016. [doi: 10.1609/icwsm.v16i1.19353]

34. Weinzierl M, Hopfer S, Harabagiu SM. Misinformation adoption or rejection in the era of COVID-19. Proc Int AAAI Conf Web Soc Media 2021 May 22;15(1):787-795. [doi: 10.1609/icwsm.v15i1.18103]

35. Muric G, Wu Y, Ferrara E. COVID-19 vaccine hesitancy on social media: building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642 [FREE Full text] [doi: 10.2196/30642] [Medline: 34653016]

36. Yang KC, Pierri F, Hui PM, Axelrod D, Torres-Lugo CT, Bryden J, et al. The COVID-19 infodemic: Twitter versus Facebook. Big Data Soc 2021 May 05;8(1):205395172110138. [doi: 10.1177/20539517211013861]

37. Cota W, Ferreira SC, Pastor-Satorras R, Starnini M. Quantifying echo chamber effects in information spreading over political communication networks. EPJ Data Sci 2019 Dec 09;8(1):35. [doi: 10.1140/epjds/s13688-019-0213-9]

38. Garimella K, De Francisci Morales G, Gionis A, Mathioudakis M. Quantifying controversy on social media. ACM Trans Comput Syst 2018;1(1):3. [doi: 10.1145/3140565]

39. GeoNames. URL: http://www.geonames.org/ [accessed 2023-04-27]

40. Vaccine multilingual queries. URL: https://tinyurl.com/344rbjcx [accessed 2023-04-27]

41. Stream Tweets in real-time. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/tutorials/stream-tweets-in-real-time [accessed 2023-04-27]

42. Choosing a historical API. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/tutorials/choosing-historical-api [accessed 2023-04-27]

43. Follow, search, and get users. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/follow-search-get-users/api-reference/get-users-show [accessed 2023-04-27]

44. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nat Hum Behav 2020 Dec;4(12):1285-1293. [doi: 10.1038/s41562-020-00994-6] [Medline: 33122812]

45.   Pierri F. The diffusion of mainstream and disinformation news on Twitter: the case of Italy and France. In: Companion Proceedings of the 2020 Web Conference. 2020 Presented at: WWW '20; April 20-24, 2020; Taipei, Taiwan p. 617-622. [doi: 10.1145/3366424.3385776]

46.   The Black List. Bufale. URL: https://www.bufale.net/the-black-list-la-lista-nera-del-web/ [accessed 2023-04-27]

47.   List of fake news websites. Wikipedia - The Free Encyclopedia. URL: https://en.wikipedia.org/wiki/List_of_fake_news_websites [accessed 2023-04-27]

48.   Media Bias / Fact Check. URL: https://mediabiasfactcheck.com/ [accessed 2023-04-27]

49.   Sénécat A. Les fausses informations circulent de moins en moins sur Facebook. Le Monde. 2018 Oct 17. URL: https://www.lemonde.fr/les-decodeurs/article/2018/10/17/les-fausses-informations-perdent-du-terrain-sur-facebook_5370461_4355770.html [accessed 2023-04-27]

50.   Oikonomidis K. List of Greek websites that have produced Fake News. dwrean.net. 2018 Oct 12. URL: https://www.dwrean.net/2018/12/fake-news-greek-sites.html [accessed 2023-04-27]

51.   Bonald T, Charpentier B, Galland A, Hollocou A. Hierarchical graph clustering using node pair sampling. arXiv Preprint posted online June 5, 2018. [doi: 10.48550/arXiv.1806.01664]

52.   Garimella K, De Francisci Morales G, Gionis A, Mathioudakis M. Quantifying controversy in social media. In: Proceedings of the 9th ACM International Conference on Web Search and Data Mining. 2016 Feb Presented at: WSDM '16; February 22-25, 2016; San Francisco, CA, USA p. 33-42. [doi: 10.1145/2835776.2835792]

53.   Danon L, Díaz-Guilera A, Duch J, Arenas A. Comparing community structure identification. J Stat Mech 2005;2005(9):P09008. [doi: 10.1088/1742-5468/2005/09/p09008]

54.   Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: a behavioral analysis. PLoS One 2021 Mar 3;16(3):e0247642. Retracted in: PLoS One. December 22, 2022. doi: 10.1371/journal.pone.02797964 [FREE Full text] [doi: 10.1371/journal.pone.0247642] [Medline: 33657152]

55.   Dutta A, Beriwal N, Van Breugel LM, Sachdeva S, Barman B, Saikia H, et al. YouTube as a source of medical and epidemiological information during COVID-19 pandemic: a cross-sectional study of content across six languages around the globe. Cureus 2020 Jun 15;12(6):e8622 [FREE Full text] [doi: 10.7759/cureus.8622] [Medline: 32685293]

56.   Twitter Safety. An update following the riots in Washington, DC. Twitter Blog. 2021 Jan 12. URL: https://blog.twitter.com/en_us/topics/company/2021/protecting--the-conversation-following-the-riots-in-washington-- [accessed 2023-04-27]

57.   Di Giovanni M, Pierri F, Torres-Lugo C, Brambilla M. VaccinEU: COVID-19 vaccine conversations on Twitter in French, German and Italian. Proc Int AAAI Conf Web Soc Media 2022 May 31;16(1):1236-1244. [doi: 10.1609/icwsm.v16i1.19374]

58.   Haouari F, Hasanain M, Suwaileh R, Elsayed T. ArCOV-19: the first Arabic COVID-19 Twitter dataset with propagation networks. In: Proceedings of the 6th Arabic Natural Language Processing Workshop. 2021 Presented at: WANLP '21; April 19, 2021; Kyiv, Ukraine p. 82-91 URL: https://aclanthology.org/2021.wanlp-1.9.pdf

59.   Pierri F, Tocchetti A, Corti L, Di Giovanni M, Pavanetto S, Brambilla M, et al. VaccinItaly: monitoring Italian conversations around vaccines on Twitter and Facebook. arXiv Preprint posted online January 11, 2021. [doi: 10.48550/arXiv.2101.03757]

60.   Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2023-04-27]

61.   Martini F, Samula P, Keller TR, Klinger U. Bot, or not? Comparing three methods for detecting social bots in five political discourses. Big Data Soc 2021 Jul;8(2):205395172110335. [doi: 10.1177/20539517211033566]

62.   COVID-19 Misinformation. Twitter Transparency. URL: https://transparency.twitter.com/en/reports/covid19.html#2021-jul-dec [accessed 2023-04-27]

63.   Crisis misinformation policy. Twitter Help Center. URL: https://help.twitter.com/en/rules-and-policies/crisis-misinformation [accessed 2023-04-27]

64.   Kelley J. A decade after the Arab Spring, platforms have turned their backs on critical voices in the Middle East and North Africa. Electronic Frontier Foundation. 2020 Dec 17. URL: https://www.eff.org/deeplinks/2020/12/decade-after-arab-spring-platforms-have-turned-their-backs-critical-voices-middle [accessed 2023-04-27]

65.   De-platforming as Censorship Means in the Digital Era. Internet Governance Forum. URL: https://www.intgovforum.org/en/content/igf-2022-ws-52-de-platforming-as-censorship-means-in-the-digital-era [accessed 2023-04-27]

66.   Broniatowski DA, Jamison AM, Qi A, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]

67.   Warner EL, Barbati JL, Duncan KL, Yan K, Rains SA. Vaccine misinformation types and properties in Russian troll tweets. Vaccine 2022 Feb 07;40(6):953-960. [doi: 10.1016/j.vaccine.2021.12.040] [Medline: 35034832]

68.   Schreiber M. 'Bot holiday': COVID disinformation down as social media pivot to Ukraine. The Guardian. 2022 Mar 4. URL: https://www.theguardian.com/media/2022/mar/04/bot-holiday-covid-misinformation-ukraine-social-media [accessed 2023-04-27]

69.   Lenti J. global-vax-misinfo. GitHub. URL: https://github.com/jaclenti/global-vax-misinfo [accessed 2023-05-09]

70.   Lenti J. Global misinformation spillovers in the online vaccination debate before and during COVID-19. Zenodo. 2023 Mar 10. URL: https://zenodo.org/record/7716817#.ZFoj--xBxpR [accessed 2023-05-09]

XSL•FO
RenderX

Abbreviations

API: application programming interface
CO: cosharing
NMI: normalized mutual information
OSM: online social media
RT: retweet
RWC: Random Walk Controversy

XSL·FO
RenderX