

Mis-tweeting communication: a Vaccine Hesitancy analysis among twitter users in Italy

Davide Gori^{1,*}, Francesco Durazzi^{2,*}, Marco Montalti¹, Zeno Di Valerio¹, Chiara Reno¹, Maria Pia Fantini¹, Daniel Remondini²

¹Department of Biomedical and Neuromotor Sciences, University of Bologna, 40138 Bologna, Italy; ²Department of Physics and Astronomy, University of Bologna, 40138 Bologna, Italy; *D.G. and F.D. contributed equally to the paper.

Abstract. *Background and aim:* A previously unseen body of scientific knowledge of varying quality has been produced during the ongoing COVID-19 pandemic. It has proven extremely difficult to navigate for experts and laymen alike, giving rise to the so-called “Infodemic”, a breeding ground for misinformation. This has a potential impact on vaccine hesitancy that must be considered in a situation where efficient vaccination campaigns are of the greatest importance. We aimed at describing the polarization and volumes of Italian language tweets in the months before and after the start of the vaccination campaign in Italy. *Methods:* Tweets were sampled in the October 2020-January 2021 period. The characteristics of the dataset were analyzed after manual annotation as Anti-Vax, Pro-Vax and Neutral, which allowed for the definition of a polarity score for each tweet. *Results:* Based on the annotated tweets, we could identify 29.6% of the 2,538 unique users as anti-Vax and 12.1% as pro-Vax, with a strong disagreement in annotation in 7.1% of the tweets. We observed a change in the proportion of retweets to anti-Vax and pro-Vax messages after the start of the vaccination campaign in Italy. Although the most shared tweets are those of opposite orientation, the most retweeted users are moderately polarized. *Conclusions:* The disagreement on the manual classification of tweets highlights a potential risk for misinterpretation of tweets among the general population. Our study reinforces the need to focus Public Health’s attention on the new social media with the aim of increasing vaccine confidence, especially in the context of the current pandemic. (www.actabiomedica.it)

Key words: Vaccine Hesitancy, Twitter, Social Media, Public Health, Vaccination, COVID-19, SARS-CoV-2, anti-vax

Introduction

The still ongoing COVID-19 pandemic has been accompanied by a continuous and overwhelming stream of information and scientific evidence, which has proven invaluable in pooling scientific knowledge and expertise across the globe. This remarkable collective effort has since shown some less benign consequences, well encompassed by the term “Infodemic” defined by the World Health Organization (WHO) as an overabundance of information, both online and offline (1). This overwhelming amount of data of varying quality and from different sources could endanger

the public health efforts to curb the spread and effects of the pandemic by fostering mistrust in health authorities. It could also prove harmful to individual health as it could lead to uncertainty regarding which information to trust and rely on and thus potentially leading to a higher risk of adopting harmful behaviours.

Said uncertainty and difficulty in acquiring reliable and scientifically sound information have the potential to exacerbate the phenomenon of Vaccine Hesitancy (VH) which has been identified as a major public health concern by WHO in 2019 (2). The WHO Scientific Advisory Group of Experts (SAGE) working group on VH defined it as a “delay in acceptance or

refusal of vaccination despite availability of vaccination services”: this definition not only encompasses utter opposition to vaccination but the whole spectrum of negative stances toward it, including reluctant acceptance (3). As the efforts to curb the harrowing effects of the pandemic on healthcare systems and societies heavily rely on efficient mass-vaccination campaigns, VH is now more than ever an issue that needs to be firmly addressed. In a recent concise systematic review, only a narrow majority of survey studies conducted among the general population stratified by country (29 out of 47 studies, 62%) showed an acceptance of COVID-19 vaccination $\geq 70\%$, (4). Unsatisfactory rates of COVID-19 vaccine acceptance have been reported in the Middle East, Russia, Africa, and several European countries, including Italy (5-9). In many countries, values under 60% have been registered: a figure below the estimated range of immune individuals (60-75%) needed to halt the transmission and spread of the virus (10, 11).

In this context, social media represent a relatively new factor at play; with a potential to accelerate the spread of (good and bad) information and to offer a mean of rapidly selecting what is relevant to public discourse. If properly managed, this instrument might prove invaluable in enabling the diffusion of scientifically validated and useful information, but if left unchecked it shows some detrimental aspects. In fact, it has been proven to be a fertile environment for misinformation, with an apparent advantage of unscientific views over official and scientifically validated information (12). This can be inferred by evidence showing how anti-vaccine tweets have a 4.13-fold chance of being retweeted if compared to neutral tweets (13); moreover, a majority (65%) of YouTube videos on the topic of vaccines have been shown to convey anti-vaccine messages (14). Information diffusion patterns on these platforms have also shown an echo chamber effect, which originates well-segregated communities and increases polarization (15). Additionally, previous evidence coming from several studies highlighted not only that VH is fostered by misinformation (16-19), but also actively weaponized for political purposes (20). Similar phenomena have been recently observed during the COVID-19 pandemic(21), when the scientific community initially played the role of boundary spanner but then lost its relevance and became more isolated within a context of increasing politicisation of the debate.

In addition to these social phenomena, social media are sensitive to targeted disinformation campaigns that actively spread malware and unsolicited content. The Twitter discourse around vaccines has been found to be influenced by Twitter bots and “trolls” (22), the former being artificial entities spreading antivaccine messages, and the latter users actively promoting discord and eroding public consensus. The overall picture regarding vaccines on Twitter during the last ten years before COVID-19 shows an increasing antivaccine user-base with a minimal amount of inter-communication between communities (23).

As a possible effect of the division about vaccines, previous literature observed that the struggle between pro- and anti-vaccinationists on various social media (i.e. Twitter, Facebook) leads to an increase in the number of undecided people that are more likely to cluster around anti-vaccination movements rather than official/institutional sources of information (24, 25). This might be one of the reasons behind the steeping increase of VH in the years following the introduction of social media, thus leading the WHO to label it as a major Public Health concern (2).

The aim of our work is to characterize a dataset of tweets in Italian language referred to both vaccines and vaccination practice in order to provide the scientific community with potentially useful information about Pro- and AntiVax users on Twitter. Such knowledge on prevalence, diffusion patterns, polarization, and other characteristics of tweets regarding vaccination could be helpful in understanding the machinery behind the successful online communication of scientific messages, and on a more practical side in planning more targeted and efficient campaigns on social media aimed at improving vaccine acceptance, especially considering their growing role in the public debate.

Materials and Methods

An Italian-based dataset of tweets related to vaccines and vaccination practice was collected from October 2020 to January 2021 through the continuous filter streaming endpoint of the Twitter API implemented in the Python’s package Tweepy. Tweets were downloaded if they contained at least one of the following words within their text: “vaccino”, “vaccini”,

“vaccinazione” and “vaccinazioni”, which are the Italian translations of “vaccine(s)” and “vaccination(s)”. Starting from a collection of 764K original tweets, we randomly sampled 7,004 tweets to be manually annotated. A characterization was performed by a panel of MD Public Health Residents and Physics students into different predefined classes:

- “AntiVax” (AV): expressing negative stances towards vaccines.
- “Neutral” (N): absence of expression of a clear judgement even if they are related to the topic.
- “ProVax” (PV): expressing positive stances towards the vaccine.
- “OffTopic” (OT): the vaccine-related word or topic is not dealt with in the right context (e.g., metaphors or analogies). Tweets considered as “OffTopic” for at least 1 over 3 annotators were removed.

In particular, the annotators were asked to consider as explicitly AntiVax or ProVax only tweets clearly containing an opinion of the author toward the acceptance of a vaccine. For this reason, messages expressing aversion to or support for pharmaceutical companies and distribution strategies were asked to be classified as Neutral.

To characterize the polarization between anti-vax and pro-vax tweets, we can define a polarity score for each tweet as

$$P_T = \frac{\sum_c w_c n_c}{n}$$

where the sum is over $C = \{\text{AntiVax}, \text{Neutral}, \text{ProVax}\}$ with $w_c = -1$ for AV, 0 for N and +1 for PV tweets, n_c is the number of annotators assigning the tweet to class c , and $n = 3$ is the total number of annotators. The polarity score spans from -1 (3/3 annotations as AV) to +1 (3/3 annotations as PV). Based on this score, we can assign a polarity class to each tweet: AV if $P_T \leq -2/3$, PV if $P_T \geq 2/3$ and N otherwise. This method allows to distinguish between annotations with “weak disagreement”, such as AntiVax-AntiVax-Neutral annotations by the three evaluators (producing a value of $P_T = -2/3$) and annotations with “strong disagreement”, such as AntiVax-AntiVax-ProVax annotations (producing a value of $P_T = -1/3$). Following the definition of polarity class written before, the tweets annotated with strong disagreement will be classified as Neutral, while the tweets annotated with weak disagreement will be classified in accordance with the majority annotation.

We also calculate the polarity score (P_u) for the Twitter users in the database as the average P_T scores of the tweets they wrote (in the range $[-1; 1]$). The user is then assigned to a user polarity class, that is AV if $P_u < -1/3$, PV if $P_u > 1/3$ and N otherwise.

Finally, following a previously conducted study in which the volume of tweets and retweets was considered (26), a stratified analysis for AV, PV, and N over time was conducted considering the time series of Tweets volumes. The visual representation of tweet volumes over time allows for considerations regarding volume peaks and potential events that might have caused them. The number of retweets to each one of the annotated tweets was interpreted as a proxy of the attention received by the tweets and by the polarity class associated to that tweet. In this sense, when we refer to AntiVax, Neutral or ProVax retweets, we refer to the labels of the manually annotated original tweets.

Results

A corpus of 7,004 tweets was manually annotated, of which 495 tweets (7.0%) were considered as OT by at least one annotator and consequently removed. The remaining 6,509 represented the final sample analysed.

In Figure. 1 the percentage of agreement and disagreement on tweets annotated with at least 2/3 annotations (based on a majority criterion) in the same class is shown.

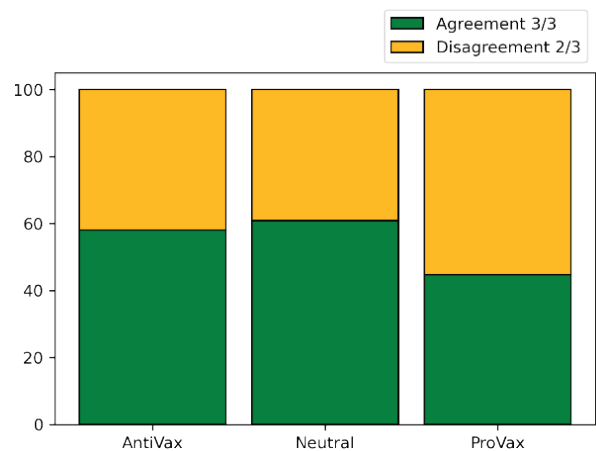


Figure 1. Percent of agreement and disagreement on tweets annotated with at least 2/3 annotations in the same class. The 100% agreement percentage is 58.0% for AV, 60.8% for Neutral and 44.7% for PV.

As we show in Figure 1, N and AV tweets share a higher percentage of 100% consensus between annotators (60.8% and 58.0% respectively), while the agreement on tweets to be considered as PV is lower (44.7%).

We then classified each tweet using the polarity classes estimated through the polarity measure P_T , which considers all tweets containing annotations of opposite polarity as Neutral, as detailed in Materials and methods. The class frequencies according to this method are thus 61.2% for N, 24.1 for AV and 14.7 for PV. Polarity scores and annotation frequencies within the dataset are listed in Table 1 and visualized in Figure 2. Tweets considered as AV by 3/3 annotators were 1,007 (15.5% of the dataset) while 563 (8.7%) by 2/3. On the other hand, 473 (7.3%) and 479 (7.4%) were labelled as PV by 3/3 and 2/3 annotators respectively.

Furthermore, 190 tweets received an annotation in each of the three categories (indicated as ANP in Table 1) and 272 tweets received partly opposite annotations (indicated as AAP or APP), leading to a strong disagreement in 7.1% of the total number of tweets. Even if these three groups of tweets, that we chose to classify as Neutral tweets, are the least represented categories in Table 2, their abundance is not negligible and show to some extent the risk of misinterpretation of the text content of the tweets. In Figure 2, AAP and APP annotations are represented by the two smallest spots in the AV-PV plane (light blue and light red respectively), while ANP annotations are represented

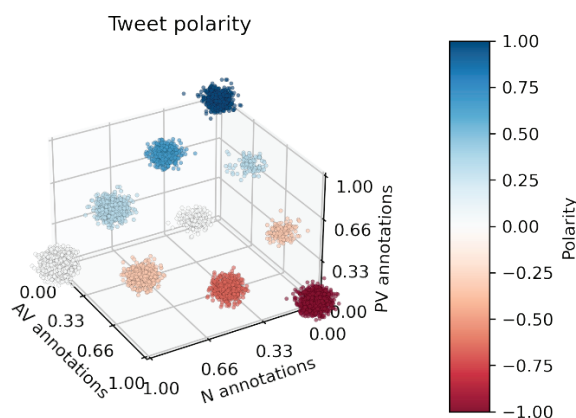


Figure 2. Scatter plot of agreement fraction for each single tweet. One point represents one tweet, with random noise added to help visualization. Tweets are coloured according to their polarity score.

by the smallest grey spot in the centre of the cartesian plot.

Twitter users

All the tweets analyzed were written by 2,538 unique users, meaning that on average each user wrote 2.6 original tweets. We grouped the users according to the polarity class defined above (shown in Figure 3a) and in Figure 3b we show how many tweets they wrote in each A N P category.

PV and AV users are mainly concentrated around the respective axes in Figure 3b, meaning that these

Table 1. Annotation frequencies and polarity scores.

| Short name | AntiVax (A) | Neutral (N) | ProVax (P) | Polarity | Polarity class | Tweets |
|------------|-------------|-------------|------------|----------|----------------|--------------|
| PPP | 0 | 0 | 3 | 1 | ProVax | 473 (7.3%) |
| NPP | 0 | 1 | 2 | 0.67 | ProVax | 479 (7.4%) |
| NNP | 0 | 2 | 1 | 0.33 | Neutral | 841 (12.9%) |
| NNN | 0 | 3 | 0 | 0 | Neutral | 2143 (32.9%) |
| APP | 1 | 0 | 2 | 0.33 | Neutral | 106 (1.6%) |
| ANP | 1 | 1 | 1 | 0 | Neutral | 190 (2.9%) |
| ANN | 1 | 2 | 0 | -0.33 | Neutral | 540 (8.3%) |
| AAP | 2 | 0 | 1 | -0.33 | Neutral | 166 (2.5%) |
| AAN | 2 | 1 | 0 | -0.67 | AntiVax | 563 (8.7%) |
| AAA | 3 | 0 | 0 | -1 | AntiVax | 1008 (15.5%) |

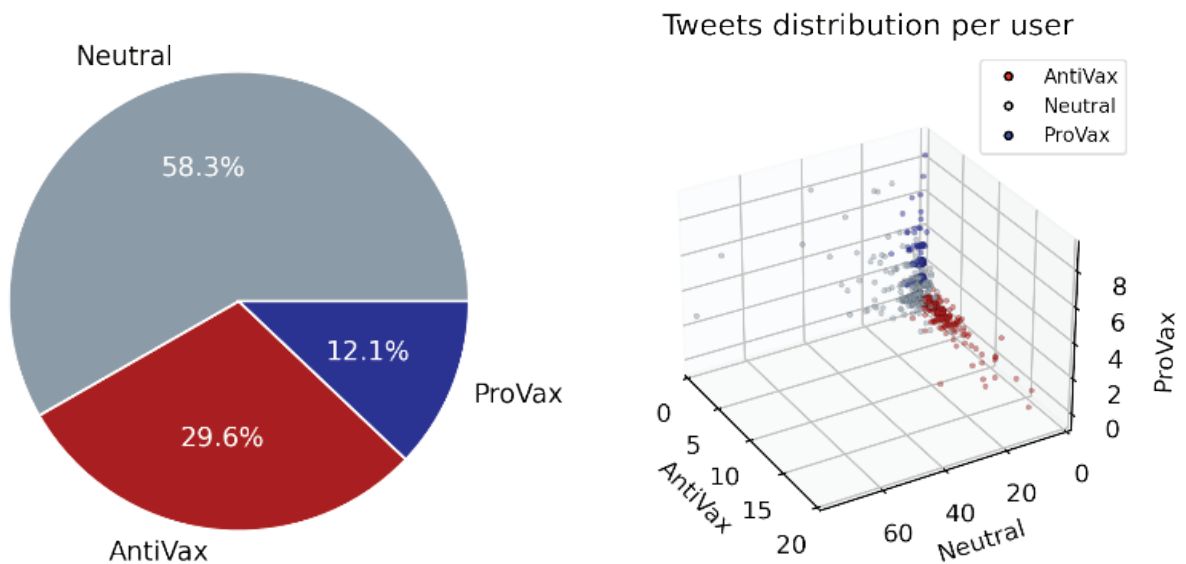


Figure 3. a) Percentage of users assigned to each polarity class. b) Scatter plot of tweets count in each polarity class for each single user. One point represents the number of tweets written by one single user, with a small random noise added to help visualization (i.e. remove overlap due to discrete values). Users are colored as the polarity class assigned to them.

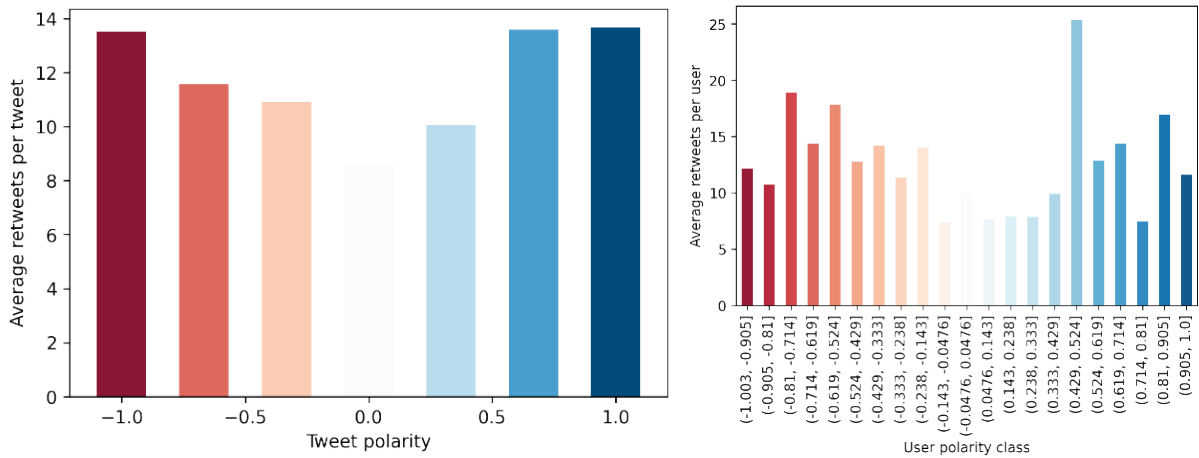
categories of users tend to write contents within the same polarity class. Only 4 PV users wrote 1 or 2 AV tweets, while 16 AV users wrote 1-3 PV tweets, but this difference is likely to be associated to the different total amount of AV and PV users and not necessarily to a different user behaviour. The dispersion of N users is higher, with users who wrote up to 8 AV or PV tweets and who are generally more represented in the ProVax-Neutral plane of Figure 3b.

In Figure 4a we show the average number of retweets done to the annotated tweets, stratified by polarity score, while in Figure 4b we show the average number of retweets received per user, stratified by user polarity score. On the one hand, tweets with extremal polarity values received more attention on average (Figure. 4a). On the other hand, the most retweeted users have an absolute value of polarization around intermediate values both for PV and AV: the peak is around +0.5 and -0.7 in Figure. 4b.

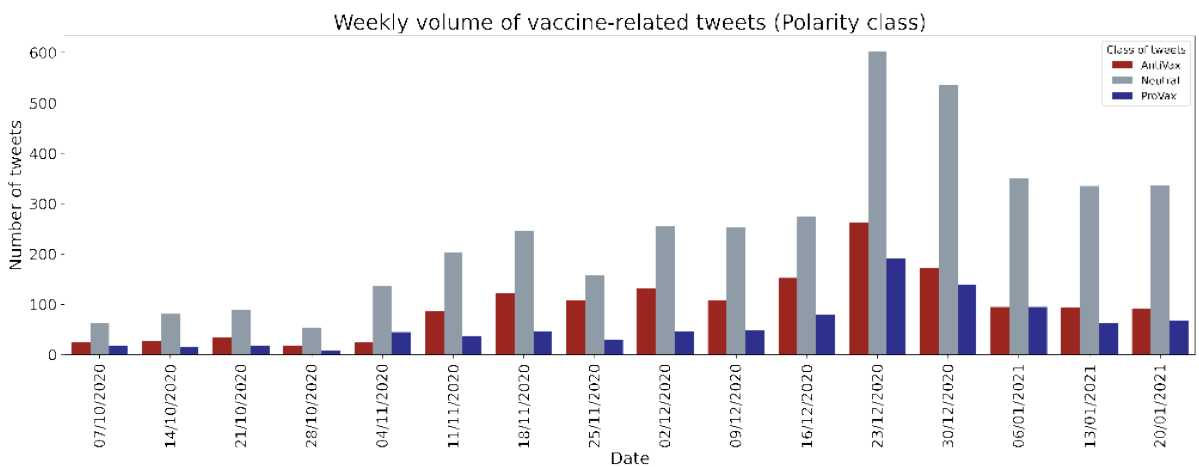
We also show the weekly time series of the volume of tweets between October 11, 2020 and January 31, 2021 (Figure. 4c) and their retweets (Figure. 4d) in the same time frame. Volumes are stratified for polarity classes (AV, PV and N). Peak volume was reached on

the week of December 23-30, 2020, but also the week of November 4-11 shows a strong relative volume increase. In relation to this, we note that on November 9th there was the Pfizer and BioNTech announce of a vaccine candidate against COVID-19 and on December 27th the vaccination campaign started in Italy, thus these changes in Tweet volume can be likely associated to these events with a large echo on the media in relation to COVID and vaccines.

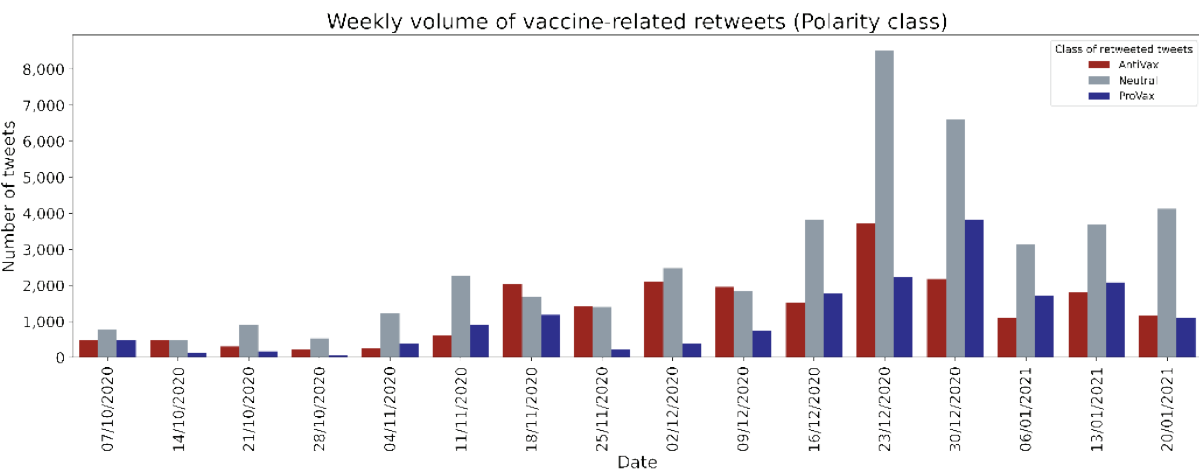
The volume of original tweets (Fig. 4c) does not show any particular trend reversal between the categories, with Neutral tweets being those written the most during each week, followed by AntiVax and finally ProVax tweets. On the other hand, the volume of retweets (Fig. 4d) shows that during the weeks starting in 18/11/2020 and 09/12/2020, the volume of AV retweets exceeded the volume of the Neutral ones, even though there were less AV original tweets than N ones. Furthermore, the fraction of Neutral tweets over time (Supplementary Fig. 1) follows a negative trend after the two main volume increases of 4 November and 23 December. In particular, the first decrease in the volume of retweets of Neutral content was characterized by an increase of AV retweets, while the



(a) (b)



(c)



(d)

Figure 4. Tweets and retweets volume. a) Average number of retweets to the annotated tweets, stratified by polarity score . b) Average number of retweets per user, stratified by user polarity score . c) Weekly volume of annotated tweets in each polarity class. d) Weekly volume of retweets done to the tweets in each polarity class.

second decrease was characterized by an increase of PV retweets.

Discussion

Our study analyzed the volumes of Pro- and Anti-Vax tweets in Italy in the period from October 2020 to January 2021. A panel of Public Health Residents and Physics MD students analyzed a total of 7,004 tweets categorizing them into “AntiVax”, “Neutral”, “ProVax” and “OffTopic”, recording a higher rate of disagreement for “ProVax” annotation. A greater disagreement in identifying a tweet as pro-vax is likely due to the inherent difficulty in recognizing a message actively supporting vaccine uptake as opposed to simply sharing the news. Since the majority of official and news sources support an implicit endorsement of the vaccine, it is sometimes difficult to distinguish forms of active support for the uptake. Our annotation identified users mainly interested in making neutral comments who nevertheless engage more in writing actively ProVax content than AntiVax content. This might be due to the fact that 1) Neutral users are more oriented towards ProVax opinions, or 2) potentially ProVax tweets could have been classified as Neutral to a major extent as compared to AntiVax tweets, which result more distinguishable.

Our results are similar to those found by a tweets sentiment analysis conducted between 2011 and 2019 (27), confirming the consistency of our sampling. In case of disagreement on a tweet with 2/3 annotations as PV or AV, the remaining annotation is more often N than at the opposite pole. However, 2.9% of the total number of tweets received an annotation within each category (ANP in Table 1) underlying that even within annotators we found some grade of disagreement. Remarkably, we observed a strong disagreement on nearly 500 tweets (7.1%), which received opposite classification by one of the annotators. This underscores a potential ambiguity associated to the messages conveyed through social media, which may lead to misinterpretation of the author's opinion or intention. In the perspective of devising even more sensitive and specific algorithms to perform sentiment analyses, this figure shows how much progress is still needed, as even highly educated annotators often struggle to

interpret the meaning of messages. On the other hand, the risk of misunderstanding is possibly enhanced by the lack of proper context in the task of manual annotation, where annotators were shown just the text of the tweets without information about the author, comments or replies as it usually happens scrolling down the Twitter feed.

We also found that tweets at the extremes of the opinion polarization are much more actively shared, with a prevalence of attention (number of retweets) for tweets with AV content versus PV content, similarly to other cases of dissemination of misleading information (28-30). Attempts at explaining this known online behaviour have included the observation that messages are deemed more convincing if they align with the reader's opinion (31): this might mean that individuals tend to share posts that are more clearly polarized and thus more recognizably aligned (or not) with their own opinion. This could be amplified even more by the segregation of groups of users with similar opinions and thus exposed to similar sources of information. Moreover, it has been shown that messages with a more emotionally charged content, or evoking powerful imagery, are more likely to be shared online (32). It could be that such characteristics could both make the detection of the stance expressed in a tweet more evident, resulting in a higher polarization score, and have a higher chance to be shared. Our result shows that the users are more willing to share extremely polarized content written by not extremely polarized users. One possible explanation is that an overly extreme user sharing always highly polarized content is seen as less trustworthy than a less biased user who writes moderate posts from time to time. Possibly, these moderately polarized users could assume the function of boundary spanners for a more balanced discussion, reducing the gap between the two extremes.

In a previous study we showed how the analysis of social media can be useful to characterize the perception about vaccines (26). The analysis of tweet volumes over time, stratified by their content, could help in identifying two changes in the time series. An increase in the total volume of tweets associated to vaccines occurred very close after the Pfizer's announcement of the new vaccine release on November 9, 2020. Another increase in volume was observed around December 27th, 2020, the so-called V-day corresponding to

the beginning of vaccination campaign in Italy. After both events, we reported a decrease on the fraction of Neutral retweets, suggesting that users tend to amplify more polarized content after important newsbreaks that may affect real life, probably trying to convince people of their own opinion. Interestingly, while after the November announcement, the volume of Anti-Vax retweets increased, the ratio turned in favour of the ProVax group after the beginning of the vaccination campaign. This last result might be related to the considerable increase of tweets posted by official channels (i.e. Ministry of Health, ISS) regarding the campaign kick-off, combined with a tendency by individual users to share pictures and reports of their vaccine administration. After 20/01/2021, the retweet volume becomes balanced between the two opposite polarizations. However, in the weeks following the end of the study, public attention globally and in Italy was drawn towards the reports of several cases of thrombotic events after the administration of the Vaxzevria vaccine. This culminated in a report by the Pharmacovigilance Risk Assessment Committee (PRAC) of the European Medicines Agency (EMA) linking this vaccine with some rare coagulation disorders (33). This might have shifted the opinions about vaccination uptake in a significant way. In this perspective, it would be of great importance to continue monitoring through validated algorithms that analyse not only the volumes but also the orientation of tweets to understand how sentiment shifts across different population strata.

The analysis of social media Big Data through the methods of Data Science, in particular exploiting new tools of extracting meaning from a large volume of information, has the potential to drive real change in Public Health practice, or at least to allow an insightful monitoring of topics of concern, such as vaccination hesitancy. For this reason, the creation of algorithms specifically designed for infoveillance is increasingly necessary in this field (34-36).

This could help to design targeted campaigns towards the segment of the population that falls in the “undecided” category due to confusing messages conveyance, given their susceptibility to the anti-vax rhetoric.

Conclusions

Social media play an extremely relevant role in selecting what becomes matter for public debate and in steering its direction, with observable real-life consequences. Due to only a partial understanding of the fine machinery behind the failure or success of content in the social media arena, its potential for good remains yet partially untapped. In order to favor the spread of scientifically validated and useful information with a positive impact on society, more knowledge is needed on what makes a message relevant among the cacophony of posts we are subjected to on a daily basis. In this study, by characterizing a dataset of almost 7,000 tweets in Italian language related to vaccines, we showed that it is not straightforward to understand the stance of the messages even for educated readers. Furthermore, we observed that the most retweeted users were only moderately polarized in the anti-vax vs pro-vax debate, even if the most shared tweets were extremely polarized. These results offer some insight into the social mechanisms that govern the amplification of messages in online social networks, particularly with regard to sensitive and debated topics such as vaccine uptake.

Acknowledgements: D.R. and F.D. were funded through the EU H2020 project No. 874735 “Versatile emerging infectious disease observatory - forecasting, nowcasting and tracking in a changing world (VEO)”.

Conflict of Interest: Each author declares that he or she has no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangement etc.) that might pose a conflict of interest in connection with the submitted article.

References

1. Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation. Available online: <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 on 25 May 2021).
2. Ten Health Issues WHO Will Tackle This Year. Available online: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> (accessed on 25 May 2021).

3. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine Hesitancy: Definition, Scope and Determinants. *Vaccine* 2015; 33: 4161–4.
4. Sallam M. COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates. *Vaccines (Basel)* 2021; 9.
5. Neumann-Böhme S, Varghese NE, Sabat I et al. Once We Have It, Will We Use It? A European Survey on Willingness to Be Vaccinated against COVID-19. *Eur J Health Econ* 2020. 21: 977–982.
6. Salali GD, Uysal MS. COVID-19 Vaccine Hesitancy Is Associated with Beliefs on the Origin of the Novel Coronavirus in the UK and Turkey. *Psychol Med* 2020. 1–3.
7. Lazarus JV, Ratzan SC, Palayew A et al. A Global Survey of Potential Acceptance of a COVID-19 Vaccine. *Nature Medicine* 2021. 27: 225–8.
8. Sallam M, Dababseh D, Eid H, Al-Mahzoum K et al. High Rates of COVID-19 Vaccine Hesitancy and Its Association with Conspiracy Beliefs: A Study in Jordan and Kuwait among Other Arab Countries. *Vaccines (Basel)* 2021. 9(1): 42.
9. Reno C, Maietti E, Fantini MP et al. Enhancing COVID-19 Vaccines Acceptance: Results from a Survey on Vaccine Hesitancy in Northern Italy. *Vaccines* 2021. 9(4): 378. <https://doi.org/10.3390/vaccines9040378>
10. Britton T, Ball F, Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* 2020;369(6505): 846–9.
11. Billah MA, Miah MM, Khan MN. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. *PLoS One* 2020. 15:1–17.
12. 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. *Hum Vaccin Immunother* 2020. 16: 2586–2593.
13. Blankenship EB, Goff ME, Yin J et al. Sentiment, contents, and retweets: a study of two vaccine-related twitter datasets. *Perm J*. 2018. 22:17–138.
14. Basch CH, Zybert P, Reeves R, Basch CE. What do popular YouTube TM videos say about vaccines? *Child Care Health Dev* 2017. 43(4):499.
15. Schmidt, A.L.; Zollo, F.; Scala, A.; Betsch, C.; Quattrocchi, W. Polarization of the Vaccination Debate on Facebook. *Vaccine* 2018. 36: 3606–3612.
16. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The Influence of Vaccine-Critical Websites on Perceiving Vaccination Risks. *J Health Psychol* 2010. 15: 446–455.
17. Nan X, Madden K. HPV Vaccine Information in the Blogosphere: How Positive and Negative Blogs Influence Vaccine-Related Risk Perceptions, Attitudes, and Behavioral Intentions. *Health Commun* 2012. 27: 829–836.
18. Ahmed N, Quinn SC, Hancock GR, Freimuth VS, Jamison A. Social Media Use and Influenza Vaccine Uptake among White and African American Adults. *Vaccine* 2018. 36: 7556–7561.
19. Bhattacharyya S, Vutha A, Bauch CT. The Impact of Rare but Severe Vaccine Adverse Events on Behaviour-Disease Dynamics: A Network Model. *Scientific Reports* 2019. 9: 7164.
20. UNICRI, United Nations Interregional Crime and Justice Research Institute. Available online: <http://www.unicri.it/index.php/News/COVID19-Disinformation-Malicious-Use-Social-Media-Terrorists-Organizedcrime> (accessed on 25 May 2021).
21. Durazzi F, Müller M, Salathé M, Remondini D, Clusters of science and health related Twitter users become more isolated during the COVID-19 pandemic. Preprint at arXiv: <https://arxiv.org/abs/2011.06845>
22. Broniatowski DA, Jamison AM, Qi SH et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health* 2018. 108: 1378–1384.
23. Gunaratne K, Coomes EA, Haghbayan H. Temporal trends in anti-vaccine discourse on Twitter. *Vaccine* 2019. 35: 4867–4871
24. Johnson NF, Velásquez N, Restrepo NJ et al. The Online Competition between Pro- and Anti-Vaccination Views. *Nature* 2020. 582: 230–233.
25. Germani F, Biller-Andorno N. The Anti-Vaccination Infodemic on Social Media: A Behavioral Analysis. *PLoS One* 2021.16(3): e0247642.
26. Gori D, Reno C, Remondini D, Durazzi F, Fantini MP. Are We Ready for the Arrival of the New COVID-19 Vaccinations? Great Promises and Unknown Challenges Still to Come. *Vaccines* 2021. 9: 173.
27. Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera J et al. Vaccine Hesitancy on Social Media: Sentiment Analysis from June 2011 to April 2019. *Vaccines* 2021. 9: 28.
28. Albalawi Y, Nikolov NS, Buckley J. Trustworthy Health-Related Tweets on Social Media in Saudi Arabia: Tweet Metadata Analysis. *Journal of Medical Internet Research* 2019. 21: e14731.
29. Pierri F, Piccardi C, Ceri S. Topology Comparison of Twitter Diffusion Networks Effectively Reveals Misleading Information. *Sci Rep* 2020. 10(1): 1372.
30. Shahi GK, Dirkson A, Majchrzak TA. An Exploratory Study of COVID-19 Misinformation on Twitter. *Online Soc Netw Media* 2021. 22: 100104.
31. Giese H, Neth H, Moussaïd M, Betsch C, Gaissmaier W. The echo in flu-vaccination echo chambers: selective attention trumps social influence. *Vaccine* 2019. 38(8): 2070–2076.
32. Broniatowski DA, Hilyard KM, Dredze M. Effective vaccine communication during the disneyland measles outbreak. *Vaccine* 2016. 34(28): 3225–28.
33. European Medicine Agency, Signal assessment report on embolic and thrombotic events (SMQ) with COVID-19 Vaccine (ChAdOx1-S [recombinant]) – COVID-19 Vaccine AstraZeneca (Other viral vaccines). Available online: https://www.ema.europa.eu/en/documents/prac-recommendation/signal-assessment-report-embolic-thrombotic-events-smq-covid-19-vaccine-chadox1-s-recombinant-covid_en.pdf (accessed on 25 May 2021)

34. Hulsen T, Jamuar SS, Moody AR et al. From Big Data to Precision Medicine. *Front. Med.* 2019. 6: 34.
35. Cossin S, Thiébaud R, Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics Public Health and Epidemiology Informatics: Recent Research Trends Moving toward Public Health Data Science. *Yearb Med Inform* 2020. 29: 231–234.
36. Eysenbach G. Infodemiology and Infoveillance: Tracking Online Health Information and Cyberbehavior for Public Health. *American Journal of Preventive Medicine* 2011. 40: S154–S158.

Correspondence:

Received: 7 September 2021

Accepted: 22 September 2021

Daniel Remondini, PhD

Department of Physics and Astronomy, University of Bologna;

Viale Berti Pichat 6/2, 40138 Bologna, Italy

Phone: +39 051 20 9 5127

E-mail: daniel.remondini@unibo.it

APPENDIX

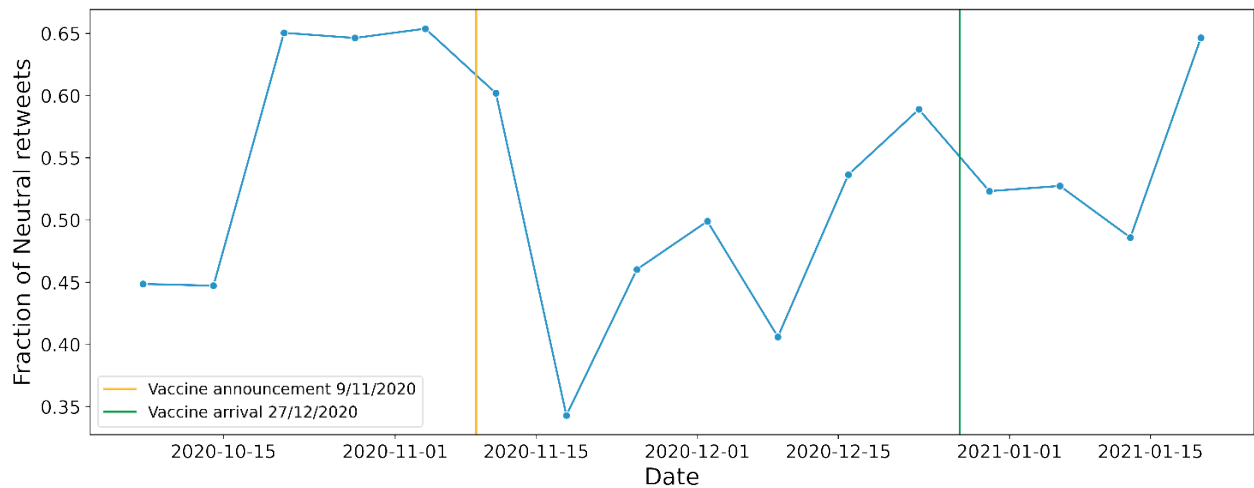
This appendix has been provided by the authors to give readers additional information about the list of investigators.

Supplementary Table S1

| | | |
|----------------------------|---|---------------------------------------|
| Dr. Alessandra Anastasia | Public Health Resident, University of Bologna | alessandra.anastasia@studio.unibo.it |
| Chiara Bernardo | MD student in Physics, University of Bologna | chiara.bernardo3@studio.unibo.it |
| Stefano Bianchi | MD student in Physics, University of Bologna | stefano.bianchi18@studio.unibo.it |
| Filippo Caleca | MD student in Physics, University of Bologna | filippo.caleca@studio.unibo.it |
| Dr. Angelo Capodici | Public Health Resident, University of Bologna | angelo.capodici@studio.unibo.it |
| Manuela Carriero | MD student in Physics, University of Bologna | manuela.carriero@studio.unibo.it |
| Giuseppe Cordella | MD student in Physics, University of Bologna | giuseppe.cordella@studio.unibo.it |
| Francesco Dalmonte | MD student in Physics, University of Bologna | francesco.dalmonte5@studio.unibo.it |
| Sara Dalmonte | MD student in Physics, University of Bologna | sara.dalmonte4@studio.unibo.it |
| Dr. Francesco De Dominicis | Public Health Resident, University of Bologna | francesco.de7@studio.unibo.it |
| Federico Dell'Anna | MD student in Physics, University of Bologna | federico.dellanna2@unibo.it |
| Matteo Fuligni | MD student in Physics, University of Bologna | matteo.fuligni@studio.unibo.it |
| Dr. Elisa Gabrielli | Public Health Resident, University of Bologna | elisa.gabrielli4@studio.unibo.it |
| Claudia Ghinato | MD student in Physics, University of Bologna | claudia.ghinato@studio.unibo.it |
| Dr. Giorgia Gribaudo | Public Health Resident, University of Bologna | giorgia.gribaudo@studio.unibo.it |
| Dr. Giusy La Fauci | Public Health Resident, University of Bologna | giusy.lafauci@studio.unibo.it |
| Stefano Lusoli | MD student in Physics, University of Bologna | stefano.lusoli@studio.unibo.it |
| Federico Magnani | MD student in Physics, University of Bologna | federico.magnani5@studio.unibo.it |
| Tommaso Marzi | MD student in Physics, University of Bologna | tommaso.marzi@studio.unibo.it |
| Dr. Alessandro Melis | Public Health Resident, University of Bologna | alessandro.melis6@studio.unibo.it |
| Pietro Micaletti | MD student in Physics, University of Bologna | pietro.micaletti@studio.unibo.it |
| Mattia Damia Paciarini | MD student in Physics, University of Bologna | mattia.damia@studio.unibo.it |
| Daniele Pignedoli | MD student in Physics, University of Bologna | daniele.pignedoli@studio.unibo.it |
| Alessandro Piovani | MD student in Physics, University of Bologna | alessandro.piovani2@unibo.it |
| Simone Puntin | MD student in Physics, University of Bologna | simone.puntin@studio.unibo.it |
| Domingo Ranieri | MD student in Physics, University of Bologna | domingo.ranieri@studio.unibo.it |
| Ettore Rocchi | MD student in Physics, University of Bologna | ettore.rocchi2@studio.unibo.it |
| Marco Saguatti | MD student in Physics, University of Bologna | marco.saguatti@studio.unibo.it |
| Dr. Maria Giulia Salomoni | Public Health Resident, University of Bologna | mariagiulia.salomoni@studio.unibo.it |
| Dr. Aurelia Salussolia | Public Health Resident, University of Bologna | aurelia.salussolia@studio.unibo.it |
| Dr. Francesca Scognamiglio | Public Health Resident, University of Bologna | francesc.scognamigli5@studio.unibo.it |
| Dr. Giorgia Soldà | Public Health Resident, University of Bologna | giorgia.solda@studio.unibo.it |
| Laura Soprani | MD student in Physics, University of Bologna | laura.soprani@studio.unibo.it |
| Lorenzo Squadrani | MD student in Physics, University of Bologna | lorenzo.squadrani@studio.unibo.it |
| Dr. Lara Storbini | Public Health Resident, University of Bologna | lara.storbini@studio.unibo.it |
| Dr. Tommaso Tarchiani | Public Health Resident, University of Bologna | tommaso.tarchiani@studio.unibo.it |
| Simone Tentori | MD student in Physics, University of Bologna | simone.tentori@studio.unibo.it |
| Daniele Trisciani | MD student in Physics, University of Bologna | daniele.trisciani@studio.unibo.it |
| Marco Uguccioni | MD student in Physics, University of Bologna | marco.uguccioni4@studio.unibo.it |

Supplementary Table S1 (continued)

| | | |
|---------------------|---|---------------------------------|
| Chiara Vece | MD student in Physics, University of Bologna | chiara.vece@studio.unibo.it |
| Adriano Zaghi | MD student in Physics, University of Bologna | adriano.zaghi@studio.unibo.it |
| Andrea Zanellati | MD student in Physics, University of Bologna | andrea.zanellati2@unibo.it |
| Dr. Anna Zannoner | Public Health Resident, University of Bologna | anna.zannoner2@studio.unibo.it |
| Dr. Giorgia Zanutto | Public Health Resident, University of Bologna | giorgia.zanutto@studio.unibo.it |
| Pietro Zito | MD student in Physics, University of Bologna | pietro.zito@studio.unibo.it |



Supplementary Figure S1. Fraction of the volume of Neutral retweets with respect to the total volume of retweets. We have drawn two vertical lines corresponding to the date of the Pfizer and Biontech announce of a vaccine candidate against COVID-19 and the vaccination campaign beginning in Italy.