Social-Cyber Maneuvers Analysis During the COVID-19 Vaccine Initial Rollout

Janice T. Blane[1[0000-0001-8719-3367]] Daniele Bellutta[1[0000-0002-6131-9846]] and Kathleen M. Carley[1[0000-0002-6356-0238]]

School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh PA 15213, USA

Abstract. Before and after the release of Pfizer's COVID-19 vaccine, the first COVID-19 vaccine to be approved and distributed in the western world, many users took to social media to discuss getting vaccinated or to persuade others not to vaccinate. The methods of persuasion or manipulation users employ on social media can be characterized under the BEND maneuver framework. In this study, we examine Twitter data from the time periods before, during, and after the rollout of the Pfizer vaccine and separate users into pro-vaccine and anti-vaccine communities. We then conduct a network analysis for these time periods and communities to find the important members and see how the different groups used BEND maneuvers to influence their target audiences and the network as a whole. Our analysis shows how each community attempts to *build* their own communities while simultaneously *narrowing* the opposing community. Pro-vaccine groups used *excite* and *explain* messages to encourage vaccination, while anti-vaccine groups relied on *dismaying* messages about side effects and death. Furthermore, *nuking* through platform policies showed to be effective in reducing the size of the anti-vaccine online community and the quantity of anti-vaccine messages.

Keywords: social cybersecurity · social network analysis · disinformation · BEND maneuvers· COVID-19 · vaccine

1 Introduction

COVID-19 claimed the lives of 2.6 million people in the first year since its discovery [11]. In an effort to reduce the cases and deaths resulting from the COVID-19 pandemic, governments and major health organizations have pushed for the development of COVID-19 vaccines and have promoted their quick distribution. Because of the rush to create the the COVID-19 vaccine, many feel the vaccines were inadequately tested and refuse the vaccine without long-term studies. Additionally, some believe conspiracies stating that the vaccines were created to microchip the population for some malicious purpose [14].

Social media has become a platform for COVID-19 vaccine discussion. Twitter is a popular platform on which government leaders, public health officials, and news organizations spread pertinent information. However, some malicious users spread disinformation by conducting influence campaigns that attempt to manipulate people's beliefs and ideas. To counter misinformation on Twitter during the initial administration of the vaccine, the social media giant expanded its policy for removing false and misleading tweets about COVID-19 vaccines, labeling potentially misleading COVID-19 vaccine information, and creating a five-strike system for suspending misleading accounts [16, 17].

This work focuses on the time period around the approval and initial administration of the Pfizer vaccine. We describe a methodology for determining pro-vaccine or anti-vaccine stances within tweets and identifying key players and communities within the social network. We further use bot detection and linguistic cues to analyze the content and significance of tweets, and we evaluate how various groups applied social cyber-maneuvers to persuade their targets.

2 Related Works

2.1 Pro-vaccine and Anti-vaccine Communities

Studies have been conducted on anti-vaccine and pro-vaccine communities to identify some of the methods used for spreading vaccine-related messages. Different communities have varying characteristics based on the nature and support for the two stances. Pro-vaccine messages tend to be supported my public health officials and governments seeking to reduce the spread of infectious diseases. Anti-vaccine communities are more niche and maintain a smaller following. In 2019, a study was conducted that looked at themes and influential actors within the anti-vaccine community [4]. Bonnevie et al. concluded that top tweeters rely on highly networked communities that emphasize specific narratives and select messages based on receptivity. This is different to the standard public official messages, which tended to repeat the same types of messages to the same communities limiting the extent of the message reach. In an analysis on Facebook vaccine group clusters, Johnson et al. observed that anti-vaccination clusters entangle more often with undecided clusters, where as pro-vaccination clusters tended to be more peripheral. Furthermore, Schmidt et al. examined how echo chambers reinforce the opinions of groups and how involvement within these groups could be an effective way of countering anti-vaccine beliefs [13].

2.2 Bots and Influence Campaigns

Bots have been used to spread disinformation, manipulate social discourse, and influence elections [10]. They have affected the course of online discussions with misinformation on public health topics such as e-cigarettes, diets, and medications [1]. Because of the influence of these bots on public opinion, several studies have been conducted on the use of bots for spreading vaccine information [18] [15]. Prior to the COVID-19 pandemic, Broniatowski et al. examined the extent to which bots spread anti-vaccine messages, showing the high rates of vaccine

Filter	Keywords
Coronavirus Tweets	coronaravirus, coronavirus, wuhan virus, wuhanvirus, 2019nCoV, NCoV, NCoV2019, covid-19, covid19, covid 19
Vaccine Tweets	vaccine, vax, mRNA, autoimmuneencephalitis, vaccination, getvaccinated, covidisjustacold, autism, covidshotcount, dose1, dose2, VAERS, GBS, believemothers, mybodymychoice, thisisourshot, killthevirus, proscience, immunization, gotmyshot, igottheshot, covidvaccinated, beatcovid19, moderna, astrazeneca, pfizer, johnson & johnson, j&j, johnson and johnson, jandj

Table 1. Keywords used to collect COVID-19 vaccine-related tweets.

content they spread and comparing it to the effects of Russian trolls, whose message primarily sought to increase discord online [5]. Dyer et al. determined that after Russian trolls, bots were the most prolific vaccine-related tweeters [9].

2.3 The BEND Framework

Social cybersecurity lies at the intersection between cyberspace and human interaction. It studies how humans can be influenced using tactful messaging and interlinking between people and content. Key players can conduct influence maneuvers to change beliefs and affect behavior [7]. The BEND Framework consists of 16 maneuvers for conducting online influence. They are divided into two types: narrative and network maneuvers. These types are further divided into positive and negative directions of influence. Narrative maneuvers focus on the information and content of messages. These maneuvers affect what is being discussed and how it is discussed. Network maneuvers focus on how the network and communities are shaped, affecting their structure and the positions of key actors.

3 Data

3.1 Data Collection

The data used in this work were a subset of COVID-19 tweets collected using the Twitter API and keywords related to COVID-19. The data set was then further filtered using the vaccine-related terms shown in Table 1. Furthermore, tweets from users not writing in English were removed from the data.

We divided the data into three one-week time periods surrounding the introduction of the Pfizer vaccine: 1-7 December 2020 (the week before the rollout), 8-10 December 2020 (during the week of the rollout in the United States and the United Kingdom), and 25-31 January 2021 (six weeks after the rollout). The three time periods contained 471,962, 694,200, and 662,776 users and 935,709, 1,511,344, and 1,368,035 tweets, respectively.

3.2 Data Preparation

Bot Detection Using the Tier-1 *Bothunter* model by Beskow and Carley [2, 3], we determined the probability that each user within the data set was a

	Pre-Labeled Hashtags		Users Labeled by Stance Detection	
Time Period	Pro-Vaccine	Anti-Vaccine	Pro-Vaccine	Anti-Vaccine
Pre-Rollout	219	227	216,156	36,609
During Rollout	267	306	195,334	47,566
After Rollout	305	138	430,278	19,519

Table 2. The number of hashtags labeled as pro-vaccine and anti-vaccine before running the stance detector, along with the number of users assigned as pro-vaccine and anti-vaccine after running the stance detector.

bot. *Bothunter* is a random forest regressor trained on labeled Twitter data sets developed from forensic analyses of events with extensively reported bot activity, such as the attack against the Atlantic Council Digital Forensic Research Lab in 2017. This machine learning model takes into account network-level features (such as number of followers and friends), user-level attributes (including screen name length and account age), and tweet-level features (such as timing and content). For this work, any score of 75% or greater was labeled as a bot to reduce the chance of false positives and ensure that the accounts classified as bots were truly bots (at the expense of missing some bots).

Linguistic Cues The NetMapper software [8] was used to extract linguistic cues from the tweet text, such as the frequency of positive and negative terms, types of pronouns, etc. These are metrics useful for identifying a tweet's sentiment and author's emotional state [7]. These features are used by the ORA software [8] to identify BEND maneuvers and actors participating in such maneuvers.

Stance Detection We used the stance detector [12] built into ORA to divide the data set into the pro-vaccine and anti-vaccine communities. This stance detector starts with a set of hashtags that are initially labeled as pro- or anti- with respect to an issue. Table 2 shows the number of hashtags we manually labeled as being pro- or anti-vaccine. These hashtags are used by the stance detector to label the stance of the Twitter accounts that used them. The algorithm then uses influence propagation through the social network to label the stance of users who did not use any of the pre-labeled hashtags. This propagation occurs by repeating two steps. First, stance is spread from users that have already been classified to other users and the hashtags with user bases largely made up of classified users. Then, stance is propagated from those other users to more users and from the labeled hashtags to the accounts that made extensive use of them. The algorithm also provides a confidence level for each stance classification. After running the stance detector on our data, pro-vaccine users had a mean confidence level of approximately 99% to 100% for each time period. However, the anti-vaccine users for the before, during, and after rollout time periods had mean confidence levels of 84%, 85%, and 67%, respectively. Table 2 shows the number of users classified by time period and stance. Neutral nodes were excluded in this study.

4 Methodology

We used ORA's BEND and Community Assessment feature, which seeks to detect BEND maneuvers [6] and provides details on topic-oriented communities, key actors, individual tweets, and the entire network. We ran this analysis on each of the three time periods, and then ran it again on each stance community within each time period. By examining an entire time period, we were able to observe the interactions of the users between those of opposing or neutral stance. Focusing on the individual stances then allowed us to do more fine-grained analysis.

4.1 Agent and Tweet Analysis

ORA's BEND report allows us to identify key agents within each community. The first type of key actor we identified was super friends, who are users that exhibit frequent two-way communication with others, such as reciprocal mentioning or retweeting. The next type we identified were super spreaders. These users generate content that is shared often, facilitating the diffusion of information across the network. We then identified influencers, which are users who tweet and mention others often. They are also positioned in central parts of conversations such as by using prominent hashtags or mentioning prominent users. Other interesting users are accounts that tweet the most, users who are mentioned the most, and hashtags tied to the most accounts.

The ORA software also uses NetMapper's linguistic cues to detect BEND maneuvers in tweets. We found which accounts made use of or were targeted by BEND maneuvers the most. We then examined the propagation of tweets and concepts across the network, analyzed the spheres of influence for key users, and explored the effects of bots on vaccine-related influence campaigns.

4.2 Topic-Oriented Communities

ORA identifies topic-oriented communities using the Leiden algorithm to cluster the user communication network connecting users to other users and the hashtag usage network connecting users to the hashtags they used. Given these topic oriented communities, we selected communities to evaluate based on their echo-chamberness, E/I index, number of users, and predominant hashtags.

5 Results and Discussion

5.1 Initial BEND Results

Table 3 shows the three most common maneuvers across the three time periods and by stance community. For the pro-vaccine community, *excite* messages were initially used to use joy and happiness to encourage others to get the vaccine because of the benefits such as ending the lockdown and restrictions from COVID-19 policies and being part of stopping the spread of the virus. *Back* is

Time Period	Pro-Vaccine	Anti-Vaccine	All Users
Pre-Rollout	narrow, excite, back	narrow, nuke, build	excite, narrow, build
During Rollout	enhance, bridge, narrow	narrow, build, bridge	excite, narrow, back
After Rollout	bridge, build, excite	enhance, distract, excite	bridge, distract, build

Table 3. The top three BEND maneuvers for each time period per user stance.

used to support those messaging pro-vaccine sentiment, and *narrow* is to discourage anti-vaccine or neutral users from not getting the vaccine. During the rollout, pro-vaccine messages continue to be encouraged as important topics of discussion and attempts are made to bridge groups by mentioning many individuals. After the rollout, *excite* becomes more focused on users feeling joyful or happy about getting the vaccine and then *building* communities of people who have the vaccine or should get the vaccine through multiple mentioning.

Anti-vaccine maneuvers were initially more negative. They aimed to *narrow* and *nuke* the pro-vaccine communities by discrediting the information they spread. Simultaneously, they sought to *build* their own anti-vaccine communities throughout all three time periods. After the rollout, *distract* was also used to deflect pro-vaccine counter-narratives by discussing different or irrelevant topics.

5.2 Key Actors

Before the rollout, the top super spreaders were primarily health organizations, vaccine manufacturers, and senior government leaders. These agents were mostly a pro-vaccine or neutral stance. The top influencers were primarily news media, vaccine manufacturers, and bots—both pro and anti-vaccine. Two world leaders were identified as major influencers with one being the top influencer of the time period. When examining the differences in top users for each of the stances, the pro-vaccine mirrored the larger group with the types of users for super spreaders. The anti-vaccine top influencers were not popularly known actors, and seven out of top ten super friends have been banned from Twitter.

During the rollout, the top pro-vaccine influencers continued to be health organizations and vaccine manufacturers. However, government leader influence declined as news organization began to support administering the vaccine. Anti-vaccine super spreaders and super friends mostly remained relatively unknown actors. However, anti-vaccine accounts continued to be suspended, including two of the top ten super friends and three of the top ten super spreaders. One prominent super spreader was classified as pro-vaccine before the rollout, classified as anti-vaccine during the rollout, and then suspended after the rollout.

After the rollout, pro-vaccine users proliferated while the number of anti-vaccine users decreased. This was likely the result of Twitter's updated policies.

5.3 Key Tweets

In addition to the maneuvers identified by the ORA software, we manually examined the key tweets by top users to find undetected maneuvers. While inspecting pro-vaccine tweets, proponents such as influential community leaders and celebrities would attempt to *engage* with the public to get their vaccine and *enhance* their messages on the benefits of being vaccinated. These included messages about returning to 'normal' and doing their part to prevent further deaths. Another unusually low metric for pro-vaccine tweets was *explain*. While observing the tweets from health organizations, many would provide information about the vaccines including safety and side effects to *explain* the benefits of getting the vaccine. Furthermore, a *bridging* tactic observed was the use of #antivaxxer or similar hashtags with pro-vaccine messages. These hashtags, formerly used for anti-vaccine messages, have been hijacked to either *explain* for those who are vaccine hesitant or to mock anti-vaccine groups.

For anti-vaccine tweets, the first thing observed was the lack of detected *dismay* messages, which are messages that evoke worry, sadness, or anger. Many of the tweets from anti-vaccine accounts used vaccine injuries and deaths to worry users about the effects of receiving the quickly developed COVID-19 vaccine, so this points out a weakness in the BEND maneuver detection. Another difficult maneuver to observe is the extreme effect of *nuke* maneuvers. While *nuke* maneuvers are actions that lead to a group being dismantled, the software is unable to examine the extreme cases where key leaders are suspended from Twitter and the effects that has on the network. For example, in the pre-rollout dataset, a prominent world leader was labeled as the top influencer, but because of the suspension, this person and their associated tweets no longer ranked in the after rollout results. Furthermore, the anti-vaccine community after the rollout was much smaller, likely due to Twitter enforcing its policies on disinformation.

5.4 Bots

The Bothunter results revealed that, as shown in Table 4, anti-vaccine agents had a higher percentage of bots than did the pro-vaccine agents. We analyzed the relationship between the probability of being a bot and the number of users who retweeted them before and after the rollout using in-degree centrality of the retweet networks. The correlation for before, during, and after periods were .022, .028, and .008, respectively. Though there was a low correlation between the two values, there were bots in all time periods who gained a large following of retweeters compared to human users.

Furthermore, we calculated the number of bots within the top 100 super spreaders and super friends. Bots ranged from 15% to 25% of super friends and 0% to 13% of super spreaders. The high percentage of super friends shows that users are interacting with the bots and engaging in two-way communications. The super spreaders show that bots are effectively diffusing tweets through the network. These bots have managed to gain the trust of users on Twitter, which make them susceptible to the information or disinformation that these bots can be posting. The summary of these key influencers are shown in Table 4.

	Pro-Vaccine			Anti-Vaccine		
Time Period	Prop. Bots	Super Friends	Super Spreaders	Prop. Bots	Super Friends	Super Spreaders
Pre-Rollout	22.7%	22	5	31.7%	36	13
During Rollout	21.9%	18	8	29.6%	35	11
After Rollout	23.4%	15	0	26.8%	25	7

Table 4. The proportion of pro-vaccine and anti-vaccine bots over time, along with the number of bots per top 100 key influencers.

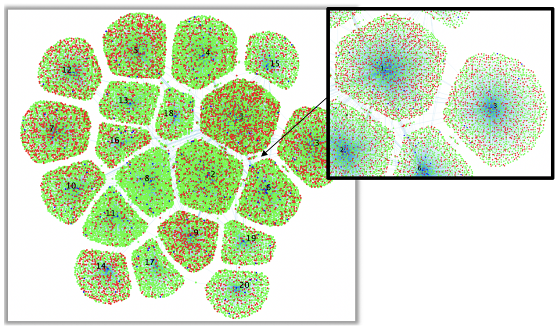

Fig. 1. Pro-vaccine topic oriented groups before rollout (*bridge maneuver*).

5.5 Topic-Oriented Groups

In Figure 1, we visualized the twenty largest topic-oriented communities for the pro-vaccine group prior to the vaccine rollout. The nodes are colored by bot identity (red is bot). Group 1 is the largest group with an echo chamber value of .03, and as shown in the inset in Figure 1, several agents are *bridging* the group with Group 6 with an echo chamber value of .06. Though in this case, the pro-vaccine communities primarily discuss pro-vaccine topics and the echo chamber values are relatively low, one implication is that bots and users that are able to infiltrate the conversations within the communities could be able to gain the trust of the members and exploit them.

Similarly, we looked at the twenty largest topic-oriented communities during the rollout, except for anti-vaccine users. When observing the topic-oriented communities by echo-chamberness, we identified high echo-chamber groups *bridging* to other communities, meaning they are mutually reinforcing their anti-vaccine bias and are transmitting that bias to other communities.

5.6 Limitations

A major limitation was the difficulty of separating the pro-vaccine and anti-vaccine communities. Many pro-vaccine maneuvers use neutral hashtags, and hashtag latching made it difficult to detect the stance of hashtags that were used to get the attention of members of the other community. In addition, anti-vaccine confidence levels for each time period were lower than the pro-vaccine confidence levels. A second limitation was ORA's detection of BEND maneuvers, which may have missed many of the more complex maneuvers. Finally, many of the collected tweets and users have either been deleted or suspended. This made it difficult to manually examine tweets with images as well as chains of replies.

6 Conclusion

In this paper, we analyzed how pro-vaccine and anti-vaccine communities attempted to persuade others of their stance on Pfizer's COVID-19 vaccine using strategies that fit the BEND maneuver framework. We found that one of the most effective maneuvers was Twitter's *nuking* of vaccine disinformation. The result was a decrease in offensive accounts and their associated anti-vaccine tweets.

Influential pro-vaccine actors tended to be prominent leaders such as government officials, vaccine manufacturers, public health officials, and as the vaccine began to be administering, secondary leaders, community organizations, news media, and celebrities began to act as key influencers within the network. Many of the maneuvers used were intended on *explaining* why people should vaccinate, getting *excited* about vaccination, and discouraging vaccine hesitant individuals from not getting the vaccine. Anti-vaccine users had a lesser presence online, particularly as key anti-vaccine proponents were banned from Twitter. These users were typically not prominent individuals in society. However, they expressed concerns with *dismaying* messages about vaccine injuries and vaccine death and made attempts to *narrow* the pro-vaccine communities as well as *build* their own. A higher percentage of bots existed within the anti-vaccine than in the pro-vaccine communities. Bots were also more prevalent among key influencers, which implies the potential for a greater reach in anti-vaccine influence.

Acknowledgements

The research for this paper was supported in part by the Office of Naval Research (ONR) under grant N00014182106, the Knight Foundation, the United States Army, and by the center for Informed Democracy and Social-cybersecurity (IDeaS). The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, the ONR, the United States Army, or the US Government.

References

1. Allem, J.P., Ferrara, E.: Could social bots pose a threat to public health? American journal of public health **108**(8), 1005 (2018)
2. Beskow, D., Carley, K.: Bot-hunter: A tiered approach to detecting characterizing automated activity on twitter. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (July 2018)

3. Beskow, D.M., Carley, K.M., Carley, K.M.: Its all in a name: detecting and labeling bots by their name. Computational and mathematical organization theory **25**(1), 24–35 (2019)
4. Bonnevie, E., Goldbarg, J., Gallegos-Jeffrey, A.K., Rosenberg, S.D., Wartella, E., Smyser, J.: Content themes and influential voices within vaccine opposition on twitter, 2019. American journal of public health (1971) **110**(Suppl 3), S326–S330 (2020)
5. Broniatowski, D.A., Jamison, A.M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S.C., Dredze, M.: Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. American journal of public health (1971) **108**(10), 1378–1384 (2018)
6. Carley, K.M.: Bend: a framework for social cybersecurity. Future Force **6**(2), 20–25 (2020)
7. Carley, K.M.: Social cybersecurity: an emerging science. Computational and Mathematical Organization Theory pp. 1–17 (2020)
8. Carley, L.R., Reminga, J., Carley, K.M.: Ora & NetMapper. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Springer (2018)
9. Dyer, O.: Vaccine safety: Russian bots and trolls stoked online debate, research finds. BMJ (Clinical research ed.) **362**, k3739–k3739 (2018)
10. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Commun. ACM **59**(7), 96–104 (Jun 2016)
11. John Hopkins University: John Hopkins University Coronavirus Resource Center (Feb 2019), `https://coronavirus.jhu.edu/map.html`, (Last accessed on 03-21-2021)
12. Kumar, S.: Social Media Analytics for Stance Mining: A Multi-Modal Approach with Weak Supervision. Ph.D. thesis, Carnegie Mellon University (2020), section 4.3.1
13. Schmidt, A.L., Zollo, F., Scala, A., Betsch, C., Quattrociocchi, W.: Polarization of the vaccination debate on facebook. Vaccine **36**(25), 3606–3612 (2018)
14. Shahsavari, S., Holur, P., Tangherlini, T.R., Roychowdhury, V.: Conspiracy in the time of corona: Automatic detection of covid-19 conspiracy theories in social media and the news (2020)
15. Subrahmanian, V., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., Menczer, F.: The darpa twitter bot challenge. Computer **49**(6), 38–46 (2016). https://doi.org/10.1109/MC.2016.183
16. Twitter: Covid-19: Our approach to misleading vaccine information (Dec 2020), `https://blog.twitter.com/en_us/topics/company/2020/covid19-vaccine.html`, (Last accessed on 05-16-2021)
17. Twitter: Updates to our work on covid-19 vaccine misinformation (Mar 2021), `https://blog.twitter.com/en_us/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation.html`, (Last accessed on 05-16-2021)
18. Yuan, X., Schuchard, R.J., Crooks, A.T.: Examining emergent communities and social bots within the polarized online vaccination debate in twitter. Social Media+ Society **5**(3), 2056305119865465 (2019)