Unraveling the use of disinformation hashtags by social bots during the Covid-19 pandemic: a networks analysis and community detection

Victor Suarez-Lledo, Esther Ortega-Martin, Jesus Carretero-Bravo, Begoña Ramos-Fiol, Javier Alvarez-Galvez

Submitted to: JMIR Infodemiology
on: June 16, 2023

Table of Contents

Unraveling the use of disinformation hashtags by social bots during the Covid-19 pandemic: a networks analysis and community detection

Victor Suarez-Lledo[1, 2*] MSc; Esther Ortega-Martin[1, 2*] MSc; Jesus Carretero-Bravo[1, 2*] MSc; Begoña Ramos-Fiol[1, 2*] MSc; Javier Alvarez-Galvez[1, 2*] PhD

[1]Department of Biomedicine, Biotechnology and Public Health University of Cadiz Cadiz ES
[2]Computational Social Science DataLab University Research Institute on Social Sciences University of Cadiz Jerez de la Frontera ES
[*]these authors contributed equally

Corresponding Author:
Victor Suarez-Lledo MSc
Department of Biomedicine, Biotechnology and Public Health
University of Cadiz
Av. Ana de Viya
Cadiz
ES

Abstract

Background: During the COVID-19 pandemic, social media platforms have been a venue for the exchange of messages, including those related to fake news. There are also accounts programmed to disseminate and amplify specific messages, which can affect individual decision-making and present new challenges for public health.

Objective: This paper aims to analyze how social bots use hashtags compared to human users on topics related to misinformation during the outbreak of the Covid-19 pandemic.

Methods: We selected tweets on specific topics related to infodemics such as vaccines, hydroxychloroquine, military, conspiracy, laboratory, Bill Gates, 5G, and UV. We built a network based on the co-occurrence of hashtags and classified the tweets based on their source. Using network analysis and community detection algorithms, we identified hashtags that tend to appear together in messages. For each topic, we extracted the most relevant subtopic communities, which are groups of interconnected hashtags.

Results: The distribution of bots and nonbots in each of these communities was uneven, with some sets of hashtags being more common among accounts classified as bots or nonbots. Hashtags related to the Trump and Qanon social movements were common among bots, and specific hashtags with anti-Asian sentiments were also identified. In the subcommunities most populated by bots in the case of vaccines, the group of hashtags including #billgates, #pandemic, and #china was among the most common.

Conclusions: The use of certain hashtags varies depending on the source, and some hashtags are used for different purposes. Understanding these patterns may help address the spread of health misinformation on social media networks.

(JMIR Preprints 16/06/2023:50021)
DOI: https://doi.org/10.2196/preprints.50021

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Unraveling the use of disinformation hashtags by social bots during the Covid-19 pandemic: a networks analysis and community detection

Authors: Victor Suarez-Lledo[1,2]; Esther Ortega-Martin[1,2]; Jesús Carretero-Bravo[1,2]; Begoña Ramos-Fiol[1,2]; Javier Alvarez-Galvez[1,2]

Affiliations:
1. Department of Biomedicine, Biotechnology and Public Health, University of Cadiz, Cádiz, Spain
2. Computational Social Science DataLab, University Research Institute on Social Sciences, University of Cadiz, Jerez de la Frontera, Spain

Corresponding Author:
Victor Suarez-Lledo,
Department of Biomedicine, Biotechnology and Public Health, University of Cadiz
Av Ana de Viya, Cádiz, 11009 Spain
Email: victor.sanz@uca.es

Background: During the COVID-19 pandemic, social media platforms have been a venue for the exchange of messages, including those related to fake news. There are also accounts programmed to disseminate and amplify specific messages, which can affect individual decision-making and present new challenges for public health.

Objectives: This paper aims to analyze how social bots use hashtags compared to human users on topics related to misinformation during the outbreak of the Covid-19 pandemic.

Methods: We selected tweets on specific topics related to infodemics such as vaccines, hydroxychloroquine, military, conspiracy, laboratory, Bill Gates, 5G, and UV. We built a network based on the co-occurrence of hashtags and classified the tweets based on their source. Using network analysis and community detection algorithms, we identified hashtags that tend to appear together in messages. For each topic, we extracted the most relevant subtopic communities, which are groups of interconnected hashtags.

Results: The distribution of bots and nonbots in each of these communities was uneven, with some sets of hashtags being more common among accounts classified as bots or nonbots. Hashtags related to the Trump and Qanon social movements were common among bots, and specific hashtags with anti-Asian sentiments were also identified. In the subcommunities most populated by bots in the case of vaccines, the group of hashtags including #billgates, #pandemic, and #china was among the most common.

Conclusions: The use of certain hashtags varies depending on the source, and some hashtags are used for different purposes. Understanding these patterns may help address the spread of health

misinformation on social media networks.

Keywords: social media, misinformation, COVID-19, bot, hashtags

Introduction

From the H1N1 pandemic in 2009 and the subsequent outbreak of the H7N9 virus, also known as bird flu, in 2013, Twitter has increasingly become a popular platform for sharing health information [1,2]. Using tweets, users can express their thoughts and opinions on many health topics. That is why specific interaction task have attracted the attention of researchers. This research can inform public policy by encouraging governments and healthcare professionals to allocate necessary resources, act, and plan accordingly [3,4]. These social media platforms have played a crucial role in providing information to the public during the COVID-19 pandemic. However, there were an increase in low-quality information, as well as the infodemic phenomenon. The infodemic, defined as an excess of information that makes it difficult for people to find reliable sources [5], can have harmful consequences [6].

The COVID-19 pandemic would trigger mandatory lockdowns, social distancing, quarantines, and SARS-COV-2 protective measures that would give rise to all sorts of opinions and behaviors [7]. During the COVID-19 pandemic, mandatory lockouts drastically altered people's daily routines (work, travel and leisure activities) to levels never before experienced by the population of the different countries affected by the new disease [8]. The situation of uncertainty in the face of an invisible threat would turn previously normal situations into situations of risk. Direct social interaction with people outside the nuclear family, attending a concert, meeting for dinner with friends and family, shaking hands with someone and even hugging would or kissing become exceptional situations during the most uncertain periods of the pandemic, situations that, as has been observed retrospectively, would have a significant impact on the mental health of the population [9]. Likewise, the health crisis would give rise to the infodemic that, through social media platforms, would open the door to fake news, misconceptions, hoaxes, and anecdotal evidence about the origin of the pandemic, the social agents to blame for the situation, and the possible measures to be taken at a time of maximum uncertainty [10] .

To understand how during the new context of health emergency misinformation spreads on these platforms, studies analyzed different elements, including the quality of information sources through URL analysis, identification of topics that generate misinformation, and analysis of online communities that spread misinformation, such as the anti-vaccine movement [11–14]. Others focused on the use of hashtags to describe the organization of the debate around the COVID related topics. Researchers examined the frequency of use and the topic analysis of hashtags and emphasized their main role in certain conversations [15,16]. By analyzing specific hashtags, studies also demonstrated how anti-vaccine communities, the proliferation of racist sentiments, or the spread of conspiracy theories are articulated on social media [17–19] . Some studies paid particular attention to how hashtags were used or combined in online conversations about Covid, using clustering techniques to describe the themes and also combining hashtags with semantic text analysis and natural language processing methods (NLP) to improve topic modeling [20–22]. Additionally, social network analysis (SNA) became useful to examine the co-occurrence of hashtags [23]. These studies demonstrate how the combination of different approach is useful to analyze online conversations more thoroughly.

Recently, the role of social bots has contributed to the spread of misinformation on social media platforms in various ways [24]. This issue garnered more attention as fake news and misinformation were significant factors during the Covid-19 pandemic. In this sense, some studies analyzed the role of bots regarding the spread of misinformation in general, while others have focused specifically on topics such as vaccines, conspiracy theories, hate speech, or reactions to other political actions [25–31]. However, a small amount of research compared the behavior of bots and humans [32,33].

To better understand the influence of bots on social media conversations, a previous study used topic modeling to segment the Twitter conversation and compare differences between accounts [34]. Nevertheless, the analysis did not focus on the usage of hashtags, which is the primary focus of this study. We aim to identify patterns and trends in hashtag usage to describe how bots and non-bots differ in their use of hashtags.

Only few studies analyzed how social media bots utilize hashtags compared to humans. Most studies in this field examine specific hashtags [17–19,35–37]. To fill this gap of knowledge we how social bots employ hashtags particularly in relation to certain infodemics topics (i.e., topics that generate or spread fake news, misinterpretations, or discriminatory phenomena). By analyzing how frequently hashtags co-occur, we aim to understand of how they appeared in the conversation and how they are combined. Besides we will consider the context in which hashtags are used. They can be used ironically or convey disagreement. Our goal is to address three key questions: 1) What are the most common hashtag co-occurrences? 2) What are the differences in hashtag usage between bots and nonbots? 3) Do bots and nonbots use certain hashtags in different ways?

Methods

Data Collection

Data collection for this study took place from March 16 to June 15, 2020, using the Twitter Streaming API. The hashtags covid_19, covid19, covid, and coronavirus were used to capture conversations about the first wave of COVID-19 and only English language tweets were selected.

Based on previous research, we created a list of topics that were commonly associated with fake news or misinformation. This list includes ozone, laboratory, 5G, conspiracy, Bill Gates, milk, military, and UV. Vaccines were also identified as a controversial topic in multiple studies, so we add them to the list [38–40].

Bot Classification

To identify whether accounts on Twitter were bots or not, we used Botometer (formerly known as BotOrNot) [41]. This publicly available application uses over a thousand criteria to determine how closely a Twitter account matches the typical characteristics of social bots.

To create a binary classification (bot or nonbot) and prioritize identifying true positives over true negatives, we set a threshold value of 0.8 [34]. Using this threshold, we classified approximately 14.8% of the accounts as bots, which is in line with the findings of other research that found bot levels to be between 9% and 15% of the total number of Twitter accounts [42].

Botometer also provides rankings for six main types of bots, including echo-chamber, fake follower, financial, self-declared, spammer, and others, in addition to the overall likelihood of being a bot. In this study, we focused on analyzing the behavior of social bot accounts, particularly those that were not identified as automated accounts. These types of accounts are often associated with press agencies, companies, newspapers, or journals, and their primary purpose is to automatically publish information about a specific topic. These accounts may indicate that they are automated, for example by including the word "bot" in their screen name or being identified as bots on Botwiki [41].

Therefore, we chose to exclude self-declared bots from our analysis due to their different characteristics compared to other social bots [41].

For the purposes of this study, we classified accounts as nonbots if their probability of being a bot was less than 0.8, self-declared bots if their probability of being a self-declared bot was greater than 0.8, and bots if their probability of being a bot was greater than 0.8 and their probability of being a self-declared bot was less than 0.8. We then filtered out self-declared bots and considered both bots and nonbots for analysis.

Network analysis

To identify patterns in the usage of hashtags, we applied network analysis. We constructed a network by analyzing the co-occurrence of hashtags in tweets and comparing the use of hashtags by bots and nonbots. In the network, hashtags were represented as nodes, and they were connected if they appeared in the same tweet. The weight of the connection between two hashtags was determined by the number of times they co-occurred.

We also calculated various metrics of connection, distribution, and segmentation of the hashtag network. We used the PageRank algorithm to identify the most important nodes in the network and the degree value, which represents the number of connections each hashtag has [43]. We also used the betweenness metric, which measures centrality [44]. In addition, we used Louvain's algorithm to detect the most important communities in the network. This algorithm maximizes a modularity score for each community, where the modularity measures the quality of the assignment of nodes to communities. This allowed us to identify hashtags that often co-occur together. We computed each metric separately considering whether the hashtags appear in Tweets posted by a bot or a nonbot. Figure 1 contains a flow diagram for the entire process.

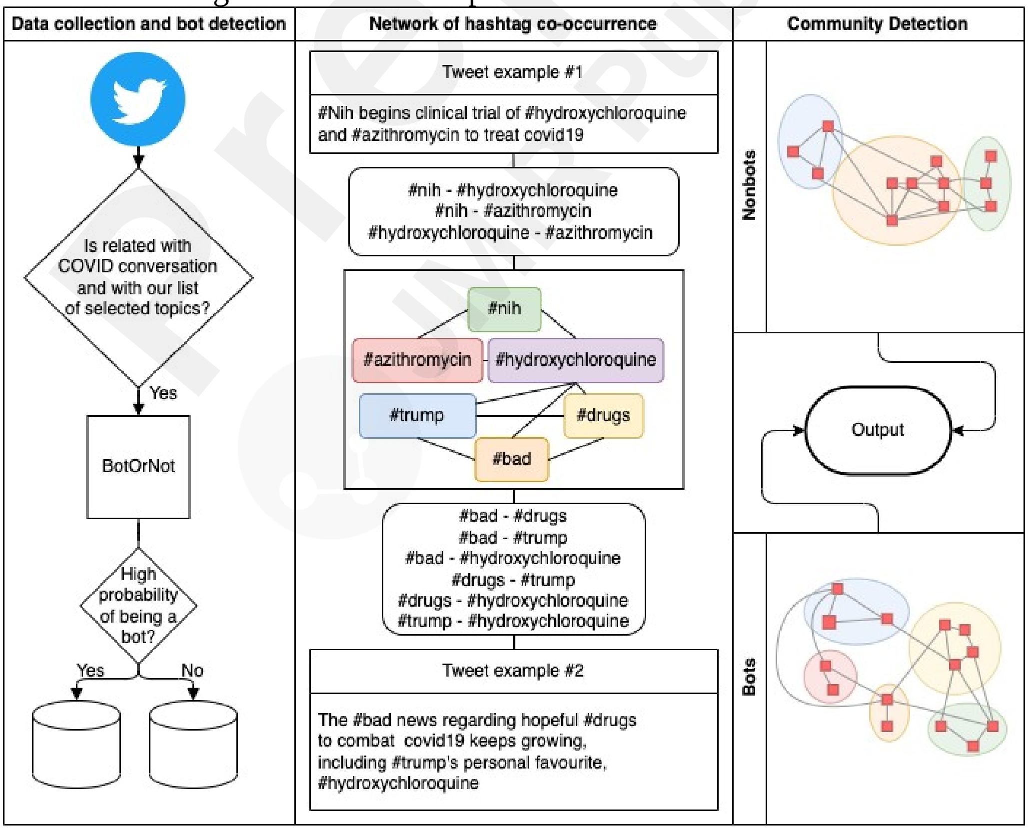

Figure 1 Flowchart from data collection to analysis

In the following section, we first present the results for the entire network. In the following sub-sections, one for each topic, we segment the overall network of hashtag co-occurrences by extracting tweets that specifically mention words related to each topic. For example, the network for vaccines will show the co-occurrences of all hashtags that appeared in tweets about vaccines.

Results

In total, we extracted around 107,173 tweets from March to July 2020 that related to the topics on our list. Most of these tweets were about vaccines (55.1%, 59,090/107,173), hydroxychloroquine (16.5%, 17,731/107,173), or military (11.5%, 12,548/107,173). Out of all the accounts analyzed, 85.2% were identified with a low likely of being bots, i.e. nonbots. Approximately 14.8% of the tweets were classified as likely being from bot accounts. As shown in Figure 2, the number of tweets related to vaccines was consistently higher throughout the period, except for two specific moments. The first of these coincides with a message from Trump recommending the use of hydroxychloroquine, an unproven drug. The second date also coincides with a message from Trump suggesting the injection of disinfectant to beat COVID-19.

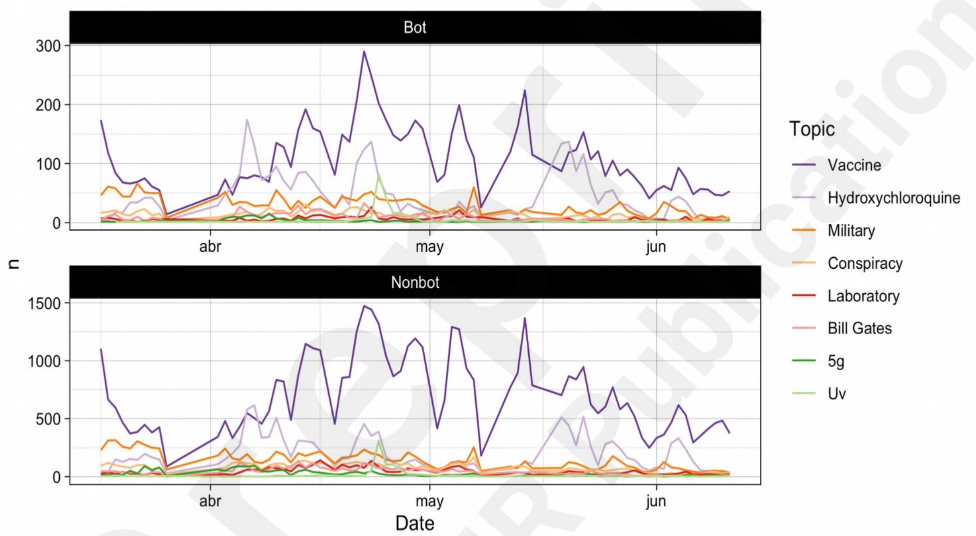

Figure 2 Bot and nonbot distribution by topic and date.

We created a graph of the full network of hashtags. For clarity, we selected a random sample from the entire collection of tweets and depicted it in Figure 3. We also applied color to the Louvain communities and highlighted some hashtags that represent the topics analyzed in the study. This process is like the one we will use for each topic in the list.

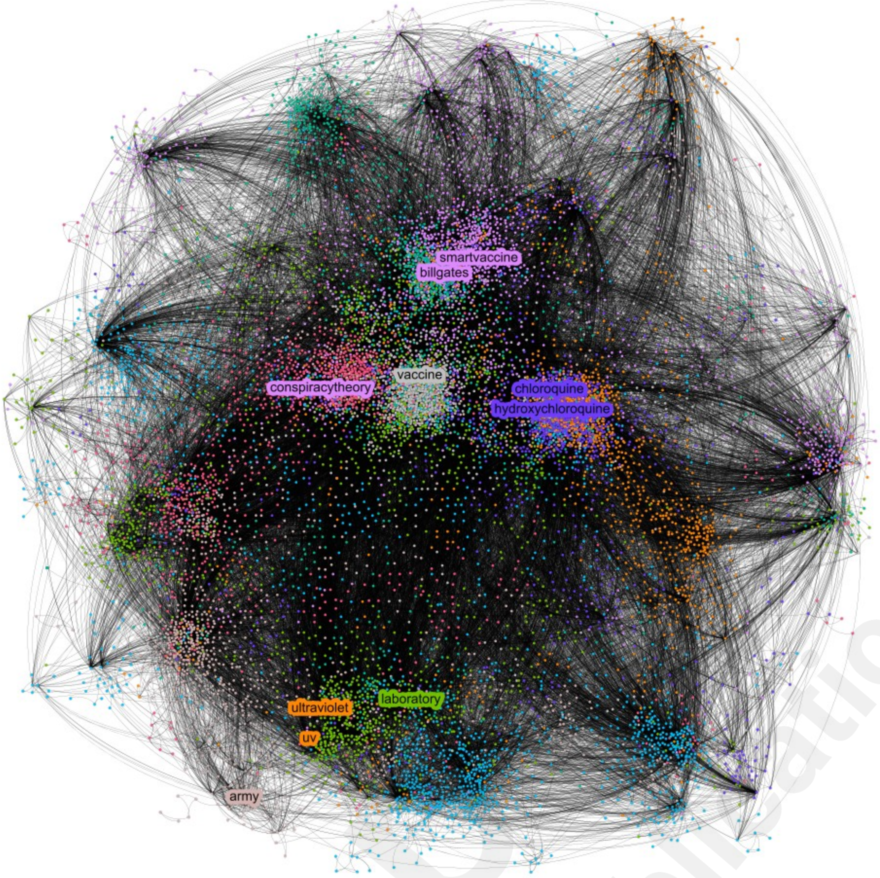

Figure 3 Hashtag network

In Table 1, we present statistics for the overall hashtags network to provide a broad overview. As mentioned earlier, we calculated the metrics separately for each type of account. There are some differences in the most used hashtags between the two groups. For example, hashtags such as #Trump, #China, and #BillGates appear in both groups. However, the hashtag #vaccineswork is one of the most used by nonbots, while the hashtag #lka (which is the country code for Sri Lanka) is more frequently used by bots.

Table 1 Most common co-occurrences by bot and nonbot.

Bots			NonBots		
Hashtags	Weight	Percentage	Hashtags	Weight	Percentage
#chloroquine - #hydroxychloroquine	537	15.52	#hydroxychloroquine - #trump	133	20
#hydroxychloroquine - #trump	490	14.17	#climatechange - #sustainability	106	15.94
#africaisnotalaboratory - #changeyourworld	437	12.63	#lka - #srilanka	86	12.93
#azithromycin - #hydroxychloroquine	345	9.97	#chloroquine - #hydroxychloroquine	84	12.63
#coronavirushoax - #prisonearth	280	8.09	#azithromycin - #hydroxychloroquine	72	10.83
#digitalvirus - #policestate	280	8.09	#kag - #maga	53	7.97
#digitalvirus - #prisonearth	280	8.09	#pandemic - #vaccine	35	5.26
#policestate - #prisonearth	280	8.09	#billgates - #vaccines	33	4.96
#coronaviruslockdown - #lockdownextension	267	7.72	#kag - #qanon	33	4.96
#changeyourworld - #coronacrisisuk	263	7.6	#china - #vaccine	30	4.51

There are also some similarities in the co-occurrence of hashtags between the two groups. For example, hashtags #hydroxychloroquine and #trump appear in the same tweets with higher frequency in both cases, at 14.2% and 20%, respectively. However, other hashtag pairs such as #kag - #maga, #billgates - #vaccines, or #kag - #qanon are common among bots. "KAG" stands for "Keep America Great," which was President Trump's campaign slogan in 2020, while "MAGA" stands for "Make America Great Again," which was his campaign slogan in 2016. Both slogans have been associated with American nationalism, and the hashtag #MAGA has sometimes been used by white supremacist groups and Trump supporters.

On the other hand, nonbots tend to use other hashtag pairs such as #coronavirushoax - #prisionearth, #digitalvirus - #policestate, and #digitalvirus - #prisionearth. These hashtags, especially "#prisionearth," were often used ironically to mock false rumors or exaggerations that were circulated online.

Vaccines

The most common co-occurrent hashtags used by nonbots regarding vaccines are #uk - #usa, #research - #science, #vaccineswork - #worldimmunizationweek. However, the most common hashtags in those tweets posted by bots are #trump - #votebluetosaveamerica, #healthcare - #ppe, or even #healthcare - #ventilators. In addition, these last mentioned are exclusive of bots. That is, they only co-occur in tweets from accounts classified as bots. Besides, it is worth mentioning that #billgates, along with #pandemic or #china, are the hashtags with the highest degree of connections, as seen in Table 2.

The algorithm extracted five different communities (Figure 4). We found significant differences in the hashtags that made up the Louvain Communities. The first community contains hashtags related to news (#breaking, #usnews, #breakingnews), countries (#canada, #france, #japan, #spain, #africa), and others related to fake news like #wuhanvirus, #ccpvirus, #bioweapon, #hiddenhand, #psychopaths, #chinaisassho, #madeinchina. This community is the most populated by bots, and the difference between bots and nonbots is the highest.

The second community contains hashtags related to famous people (#billgates, #anthonyfauci, #georgesoros). These people played a leading role by holding pro-vaccine positions like Bill Gates and Anthony Fauci. As in the previous case, we also found some hashtags related to fake news or conspiracy theories such as #billgatesisevil, #billgatesvaccine, #vaccinemafia, or #newworldorder. In this community, the quantity of nonbots is slightly higher than the number of bots.

On the other hand, the number of bots is also higher in the third community. In this case, the hashtags mention politics, such as #trump, #biden, and #borisjohnson. In addition, there were also some hashtags related to measures to curb the pandemic #stayhome, #socialdistancing, or #lockdown. Only a few infodemic-related hashtags were found #methanemouth, #pussygrabber, or #bananarepublic. The number of nonbots is higher in the other two communities. The fourth and fifth communities contain hashtags related to research and vaccines (#research, #health, #medicine) or diseases and public health campaigns (#vaccineswork, #measles, #endpolio, #healthforall), respectively. In particular, #vaccineswork is a hashtag used by health institutions such as the World Health Organization. Conversations on these hashtags were related to second waves and the importance of vaccines to fight against COVID-19.

Hydroxychloroquine

Hashtags related to Trump and the Republican movement were common in the case of hydroxychloroquine. These hashtags, such as #kag, #maga, #gop, #qanon, and #tcot were more common in bot tweets. Although #trump also appears in the case of nonbots, there were other hashtags related to news: #breaking-#breakingnew, #chinavirus-#wuhanvirus. Consequently, #trump has the highest degree of connection and the one with the highest betweenness. This hashtag, along with #chloroquine or #coronaviruspandemic, is the hashtag with the highest number of connections. There is a big difference between the first and the rest of the hashtags shown in Table 2. This difference indicates the leading role that #trump plays in the conversation about hydroxychloroquine.

We identified 8 different communities (Figure 4). Regarding the composition of the communities, it is worth mentioning the difference between the two most important ones. On the one hand, the first contains hashtags related to drugs, vaccines, or the pharmaceutical industry #azithromycin, #biotech, #chloroquine, #lupus, #malaria, #cdc, or #hcq. In the same line, in the fourth community, the predominance of nonbots is noticeable. This time the hashtags mention countries (#uk, #us, #coronavirusuk, #france, #italy, #germany), news (#worldnews, #usnews), TV series (#greysanatomy, #littlefireseverywhere) and supporting hashtags (#inthistogether).

On the other hand, in the second community, most of the hashtags are related to Trump or social movements related to him (#trump, #gop, #maga, #donaldtrump). Nonetheless, some are against him (#notaleader, #worstpresidentinhistory, #putinpuppet). In addition, the number of bots is higher than the number of nonbots. Contrary to what happens in the first one.

Military

In this case, hashtags are related to countries. For nonbots, those most mentioned are #china - #us, #italy - #russia, #lka - #srilanka. The latter is the most common among Bots, followed in fourth place by (#italy - #russia). Among the sets that do not mention countries, we find hashtags related to Trump (#gop - #trump, #kag - #maga, #kag - #qanon).

These hashtags have similarities to those of hydroxychloroquine. The bots' unique hashtags are related to the Trump movement or Republican movements (#gop, #kag, #qanon). In addition, #trump has the highest degree of connectivity and betweennes. This situation is also present in the communities (Figure 4). The first community detected contains hashtags related to Trump, and the second is related to military and veterans (#usmc, #veterans, or #usairforce). In both cases, these relationships take place in tweets posted by bots.

Conspiracy

In this group we found some hashtags related to (mis)information, and others related to countries. Regarding Bots, the most common hashtags are #fakenews - #technology, #conspiracytheories - #socialmedia, #donthecon - #trumplies. In line with this, for the nonbots the most common hashtags are #conspiracytheory - #woke. The hashtags used only by Bots are also related to racism: #racism - #sinophobia or against related to the economic system: #capitalismfails - #socialismworks.

Of the six most prominent communities (Figure 4), three of them have only nonbots. Topics in these communities are about minorities (#blackpeople, #lgbt, #amerikkka), about Trump (#maga, #bananarepublic, #qanon), and about the pandemic (#coronavirusoutbreak, #coronaviruspandemic, #pandemictech). Of the other three, in the first one, the number of nonbots is slightly higher than the number of bots. Some of the hashtags have to do with conspiracy theories (#conspiracytheory, #disinformation, #propaganda), others with media (#qanonnfoxnews, #propaganda, #fakenews), and others in a derogatory tone (#covidiot, #plandemic, #plandemicdocumentary). On the other hand, in the second and fifth communities, the number of Bots is higher. In this case, the most common hashtags are related to countries (#china, #us, #iran) and others specifically to Iran (#irancovidtruth, #iranregimechange) or against right-wing political parties (#rightwingignorance).

Laboratory

In this case, there are apparent differences in the geographical areas of the most used hashtags. On the one hand, nonbots mostly use #africaisnotalaboratory, while bots use #srilanka and #lka (country code for Sri Lanka). The hashtag #indiafightscorona is also common for bots. The hashtags #china - #wuhan are very common in both cases. This explains why #wuhan is the hashtag with the highest PageRank value and the highest degree of connection (Table 2). Followed by #laboratory, and in third place by #africaisnotalaboratory.

The differences between hashtags and the type of account that wrote the message were very clear in this case. On the one hand, in the first and fourth communities, the presence of bots is higher than nonbots (Figure 4). The first was focused on China with some examples such as #ccpvirus, #chinamustexplain, or #chinaliedpeopedied. The second in Southeast Asia, such as #armenia, #abudhabi, or #masdarcity.

Bill Gates

The data from the Bill Gates conversation are similar to those obtained in the case of hydroxychloroquine. Trump-related hashtags were very common (#kag, #maga, #qanon) in both bots and nonbots. The centrality and degree values are among the highest, as can be seen in Table 2. There were also new hashtags related to this type of political movement that only appears in this conversation, such as #crimesagainsthumanity, #gatesofhell, or #greatawakening. In addition, hashtags disparaging the figure of Bill Gates are also common, such as #saynotobillgates, or #billgatesisevil.

We identified 5 communities of hashtags (Figure 4). Among the three largest communities, the number of bots is higher than the number of nonbots in the second one. In this community, the most frequent hashtags are #trump, #depopulationagenda, #eugenetics, #repubicans, #auspol, #qanon, and #americafirst. The hashtags, as mentioned above, are related to Trump or against some figures who have publicly supported vaccines. Examples are #trump, #americafirst, or #faucifraud. These hashtags can also be found in the first community, where the percentage of both account types is similar. However, in this community, the number of bots is not higher than that of nonbots. In the third community, the number of nonbots is higher than bots. Most hashtags in this community mention covid (#coronaviruschallenge, #coronavirusbill, #coronaviruschina, #coronavirusnewyork) but other hashtags such as #hoaxvirus, #tedconnnect, #freedomovefear or #trumpisevil also appear.

5G

Regarding 5G, hashtags related to technology or news were the predominant ones in the case of nonbots, such as #techwar - #tradewar or #bbcaq - #itvnews. On the other hand, in the case of bots, the hashtags continue to mention geographical areas #america - #china, #america - #lka. There are other hashtags with higher intensity, for example, #chinesecoronavirus - #democratshateamerica, or #conspiracytheories - #technology. As can be seen in Table 2, the #china hashtag gets the highest PageRank value, followed by #pandemic and #wuhan. In addition, #china has 42 degrees of connectivity, doubling the value of the second, which is #pandemic with 27 connections. But above all, these values indicate the central place these hashtags take in the conversation. On the one hand, the high degree indicates they co-occur with many different hashtags. On the other hand, a high betweenness value indicates a central place in the network.

This time, the algorithm found five different communities of hashtags (Figure 4). The presence of bots is higher than nonbots in the first three. The first is related to #tech, #bigdata, #cibersecurity, and so on. The second one is focused on #conspiracytheories, #digitalskynet, and #misinformation. The third is focused on China, with hashtags such as #batflu, #chinesevirus, and #huaweithis. The last two communities, where the level of nonbots is higher, are formed by varied hashtags. The fourth community is formed by hashtags such as #kag or #maga. The fifth one contains hashtags mentioning rumors or disinformation #fakenews, #disinformation, #democrathoax. In this community, it is worth mentioning the appearance of hashtags related to #blacklivesmatter, such as #racism, #blacklivesmatteraustralia, or #policebrutality.

UV

In this case, the appearance of technology-related hashtags (#ai, #healthtech) is even more noticeable, especially in the case of bots (Table 2). On the other hand, the most common hashtags are #batflu - #quarantine in the case of nonbots. Concerning the six communities we found (Figure 4), in the first three, the number of nonbots is higher. The subject matter of these communities is related to politicians (#trump, #joebiden, #berniesanders), technology (#artificialintelligence, #bioinformatics, #machinelearning), or more specifically to technological innovation (#health, #innovation, #coronavirusnewyorkty, #smartcities).

Table 2 Most important hashtags by topic.

Vaccine				Hydroxychloroquine			
Hashtags	Degree	PageRank	Betweenness	Hashtags	Degree	PageRank	Betweenness
billgates	44	0.025	22,728	trump	54	0.074	10,106
pandemic	39	0.019	26,196	chloroquine	20	0.028	2,538
china	35	0.019	12,380	coronaviruspandemic	15	0.020	1,515
usa	30	0.013	7,375	kag	14	0.017	897
vaccineswork	28	0.019	8,833	maga	13	0.017	2,197
trump	28	0.015	15,704	coronavirusoutbreak	12	0.016	1,089
stayhome	22	0.011	4,583	india	12	0.017	855
uk	21	0.010	2,703	hcq	12	0.020	1,468
science	21	0.011	5,048	usa	12	0.015	2,095
france	19	0.008	2,064	gop	11	0.014	636
Military				**Conspiracy**			
Hashtags	Degree	PageRank	Betweenness	Hashtags	Degree	PageRank	Betweenness
trump	34	0.042	8,032	conspiracy	35	0.084	1,872
china	27	0.030	3,733	conspiracytheory	25	0.054	2,111
usa	22	0.026	5,561	conspiracytheories	16	0.037	686
italy	16	0.023	4,219	pandemic	16	0.033	878
us	16	0.019	1,667	china	15	0.032	785
iran	15	0.020	1,938	trump	12	0.030	732
russia	11	0.015	1,353	disinformation	10	0.022	77
maga	10	0.012	620	fakenews	10	0.023	321
wuhan	10	0.012	497	usa	10	0.024	778
breaking	9	0.012	2,372	us	9	0.020	213
Laboratory				**Bill Gates**			
Hashtags	Degree	PageRank	Betweenness	Hashtags	Degree	PageRank	Betweenness
wuhan	36	0.045	8,422	billgates	68	0.056	17,637
laboratory	26	0.033	11,660	qanon	29	0.023	4,043
africaisnotalaboratory	21	0.041	4,641	pandemic	27	0.024	7,341
china	20	0.023	3,470	maga	23	0.017	1,650
staysafe	11	0.017	7,566	vaccines	19	0.016	5,232
stayhome	10	0.013	9,242	stopbillgates	15	0.011	862
us	8	0.009	476	kag	13	0.009	104
pandemic	8	0.009	8,614	trump	13	0.011	1,049

coronaviruslockdown	7	0.011	1,676	microsoft	13	0.010	1,978
healthcare	7	0.009	1,331	usa	13	0.010	1,173
5G				UV			
Hashtags	Degree	PageRank	Betweenness	Hashtags	Degree	PageRank	Betweenness
china	42	0.020	31,413	ai	14	0.041	839
pandemic	27	0.012	25,136	trump	11	0.044	1,427
wuhan	19	0.009	13,463	health	8	0.025	491
iot	18	0.008	11,045	innovation	8	0.024	171
qanon	17	0.008	6,437	pandemic	8	0.029	428
bigdata	17	0.007	7,446	uvlight	8	0.028	1,617
technology	17	0.008	8,731	robots	7	0.023	754
ai	14	0.007	4,819	artificialintelligence	6	0.018	112
tech	14	0.006	4,455	lysol	5	0.018	122
fakenews	14	0.007	8,353	machinelearning	5	0.016	255

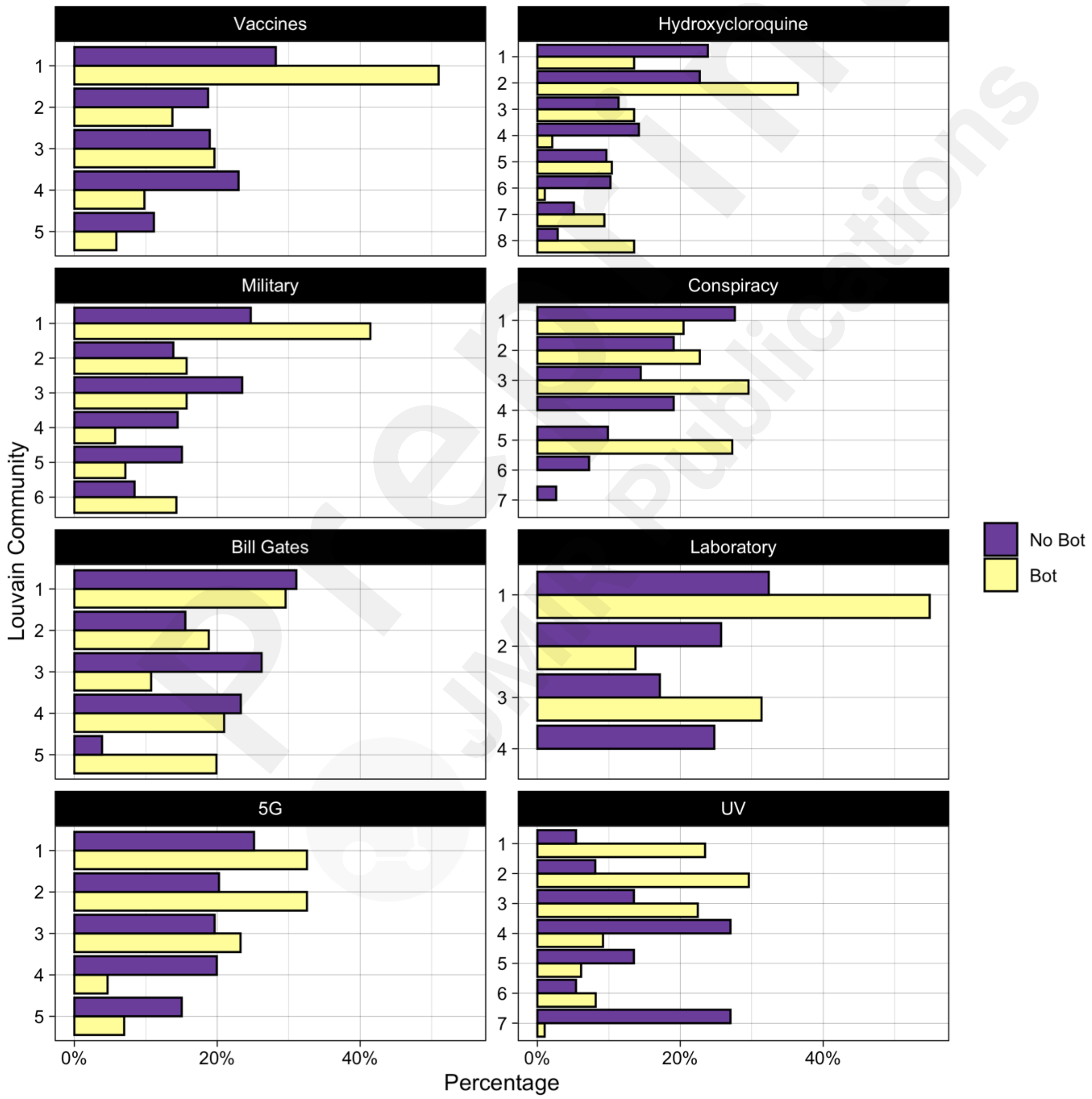

Figure 4 Bot distribution by topic.

Discussion

This study has examined the use of hashtags by social bots on Twitter during the early stages of the

COVID-19 pandemic. By analyzing the co-occurrence of hashtags, we were able to identify differences between accounts classified as bots and nonbots. We used Louvain communities to further classify these co-occurrences and found consistent differences in hashtag usage between the two groups. We employed social network analysis based on the co-occurrence of hashtags to take advantage of hashtags as key elements of online texts and understand how different users tag tweets.

The analysis of hashtags provided several key insights into attitudes towards COVID-19 and related behaviors. We consistently observed differences between bots and nonbots. In the case of bots, it was more common to find co-occurrences of hashtags related to political movements, particularly those on the right wing and related to Trump. This is consistent with findings in the literature showing a higher presence of conservatives in topics related to misinformation about COVID-19 [45].

In the conversation about vaccines, we observed that bots used hashtags related to fake news, such as #billgates and #china, more frequently. This analysis also identified specific uninformative hashtags (#ccpvirus, #chinesevirus) associated with anti-Asian sentiment [18]. Other hashtags expressed different opinions, such as criticism (#billgateisevil) or hate (#chinaliedpeopledied). It is worth noting that most of the tweets posted by non-bot users came from official accounts of institutions such as the World Health Organization, ministries of health or entities related to public health. These messages focused on reporting on the evolution of the pandemic, the number of deaths, infection rates and the health measures implemented, such as lockdowns and vaccination campaigns to contain the spread of the virus.

In our analysis of the conversation related to hydroxychloroquine, we identified two distinct communities of hashtags. One group was related to public health or medicine, while the other group was related to political movements and associated with Trump. Other studies have also found that Trump was involved in this conversation [46,47]. However, we also found that some of the hashtags in the conversation about hydroxychloroquine related to scientific facts. These differences suggest a highly polarized conversation with scientific arguments pitted against controversial political campaigns.

According to one of these studies [47], accounts with a higher impact on topics related to hydroxychloroquine disinformation were more likely to support President Trump. In addition, these types of content had a higher volume of tweets, longer duration in time, and greater echo. Our findings on the number of bots in these communities with politicized hashtags would partly explain the permanence over time and high echo values. Bots amplify these debates and increase the impact of the messages they disseminate [29,48,49]. However, our results also identify communities with anti-President Trump hashtags and higher numbers of bots. Liberals also engage in these conversations, although to a lesser extent than conservatives [45].

These findings are extensible to topics such as the military or Bill Gates, where the conversation has been highly politicized and permeated with fake news. According to the results obtained, Trump occupied a leading role in the Twitter conversation during the period analyzed. This fact has also been noted in other previous works. Trump publicly supported the use of hydroxychloroquine and other drugs to combat the advance of Covid with its corresponding impact on increased searches [50]. In addition, Bill Gates is often the protagonist in conspiracy theories [51].

Limitations and Strengths

There are several factors to consider when categorizing accounts as nonbot or bot. Botometer is backed by a large volume of research, but its effectiveness has been debated. It is important to remember that Botometer only provides a probability that an account is a bot, not a definitive classification. To get the most accurate results, it is recommended to compare probability distribution. However, in some cases it may be necessary to stablish a binary classification for research purposes. In such cases, previous research has shown that using a cutoff value and comparing the results is a successful strategy [52].

It is important to consider the language constraint of this study. Only selecting tweets written in English may limit the focus to actors and events from English-speaking countries. Additionally, no geographic limitations were placed on the collection of tweets, which allows for a larger volume of data but may also make it difficult to interpret results. It is also worth noting that the tweets analyzed in this study were from the early stages of the pandemic, and conversations and topics may have evolved over time.

Conclusion

Our analysis of hashtag usage on Twitter showed that there were differences in the patterns of use between bot and nonbot accounts. By grouping hashtags based on co-occurrence, we were able to identify distinct patterns in the usage of hashtags. On controversial or highly polarized issues, the hashtags used often pertained to the campaign or movement being promoted, with a significant portion related to Trump. In some cases, hashtags opposing these movements were also identified. On less polarized topics, hashtag usage was more diverse and included references to specific geographic locations or social groups. This analysis method can be useful in detecting hashtags that may be linked to fake news or misinformation, or in tracing the spread of such content on social media platforms.

References

1. Chew C, Eysenbach G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010/12/03. 2010;5(11):e14118. PMID:21124761
2. Vos SC, Buckner MM. Social Media Messages in an Emerging Health Crisis: Tweeting Bird Flu. J Health Commun 2016;21(3):301–308. doi: 10.1080/10810730.2015.1064495
3. Page BI, Shapiro RY. Effects of Public Opinion on Policy. American Political Science Review 1983;77(1). doi: 10.2307/1956018
4. Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China. Int J Environ Res Public Health 2020;17(8). doi: 10.3390/ijerph17082788
5. EPI-WIN: WHO Information Network for Epidemics. World Health Organization. Available from: https://www.who.int/teams/epi-win [accessed Sep 23, 2022]
6. Chou WYS, Gaysynsky A, Cappella JN. Where we go from here: Health misinformation on social media. Am J Public Health. 2020. doi: 10.2105/AJPH.2020.305905
7. Kyprianidou M, Konstantinou P, Alvarez-Galvez J, Ceccarelli L, Gruszczyńska E, Mierzejewska-Floreani D, Loumba N, Montagni I, Tavoschi L, Karekla M, Kassianos AP. Profiling Hesitancy to COVID-19 Vaccinations in Six European Countries: Behavioral, Attitudinal and Demographic Determinants. Behavioral Medicine 2023;49(4). doi: 10.1080/08964289.2022.2071198
8. Alvarez-Galvez J, Anastasiou A, Lamnisos D, Constantinou M, Nicolaou C, Papacostas S, Vasiliou VS, McHugh L, Lubenko J, Ruiz FJ, Paez-Blarrina M, Montesinos F, Valdivia-Salas S, Merwin RM, Karekla M, Gloster AT, Kassianos AP. The impact of government actions and risk perception on the promotion of self-protective behaviors during the COVID-19 pandemic. PLoS One 2023;18(4 April). doi: 10.1371/journal.pone.0284433
9. Gloster AT, Lamnisos D, Lubenko J, Presti G, Squatrito V, Constantinou M, Nicolaou C, Papacostas S, Aydın G, Chong YY, Chien WT, Cheng HY, Ruiz FJ, Garcia-Martin MB, Obando-Posada DP, Segura-Vargas MA, Vasiliou VS, McHugh L, Höfer S, Baban A, Neto DD, da Silva AN, Monestès

JL, Alvarez-Galvez J, Paez-Blarrina M, Montesinos F, Valdivia-Salas S, Ori D, Kleszcz B, Lappalainen R, Ivanović I, Gosar D, Dionne F, Merwin RM, Kassianos AP, Karekla M. Impact of COVID-19 pandemic on mental health: An international study. PLoS One 2020;15(12 December). doi: 10.1371/journal.pone.0244809

10. Balakrishnan V, Ng WZ, Soo MC, Han GJ, Lee CJ. Infodemic and fake news – A comprehensive overview of its global magnitude during the COVID-19 pandemic in 2021: A scoping review. International Journal of Disaster Risk Reduction. 2022. doi: 10.1016/j.ijdrr.2022.103144
11. Yang KC, Pierri F, Hui PM, Axelrod D, Torres-Lugo C, Bryden J, Menczer F. The COVID-19 Infodemic: Twitter versus Facebook. Big Data Soc 2021;8(1). doi: 10.1177/20539517211013861
12. Singh L, Bode L, Budak C, Kawintiranon K, Padden C, Vraga E. Understanding high- and low-quality URL Sharing on COVID-19 Twitter streams. J Comput Soc Sci 2020;3(2). doi: 10.1007/s42001-020-00093-6
13. Rovetta A, Bhagavathula AS. Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags. J Med Internet Res 2020;22(8):e20673 https://www.jmir.org/2020/8/e20673 Journal of Medical Internet Research; 2020 Aug 25;22(8):e20673. PMID:32748790
14. Blane JT, Bellutta D, Carley KM. Social-Cyber Maneuvers During the COVID-19 Vaccine Initial Rollout: Content Analysis of Tweets. J Med Internet Res 2022;24(3). doi: 10.2196/34040
15. Petersen K, Gerken JM. #Covid-19: An exploratory investigation of hashtag usage on Twitter. Health Policy (New York) 2021;125(4). doi: 10.1016/j.healthpol.2021.01.001
16. Xi W, Xu W, Zhang X, Ayalon L. A Thematic Analysis of Weibo Topics (Chinese Twitter Hashtags) Regarding Older Adults during the COVID-19 Outbreak. Journals of Gerontology - Series B Psychological Sciences and Social Sciences 2021;76(7). doi: 10.1093/geronb/gbaa148
17. Khadafi R, Nurmandi A, Qodir Z, Misran. Hashtag as a new weapon to resist the COVID-19 vaccination policy: a qualitative study of the anti-vaccine movement in Brazil, USA, and Indonesia. Hum Vaccin Immunother 2022;18(1). doi: 10.1080/21645515.2022.2042135
18. Hswen Y, Xu X, Hing A, Hawkins JB, Brownstein JS, Gee GC. Association of “#covid19” Versus “#chinesevirus” With Anti-Asian Sentiments on Twitter: March 9-23, 2020. Am J Public Health NLM (Medline); 2021 May 1;111(5):956–964. PMID:33734838
19. Gruzd A, Mai P. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter: https://doi.org/101177/2053951720938405 SAGE PublicationsSage UK: London, England; 2020 Jul 20;7(2). doi: 10.1177/2053951720938405
20. Tahamtan I, Potnis D, Mohammadi E, Miller LE, Singh V. Framing of and attention to covid-19 on twitter: Thematic analysis of hashtags. J Med Internet Res 2021;23(9). doi: 10.2196/30800
21. Vicient C, Moreno A. Unsupervised topic discovery in micro-blogging networks. Expert Syst Appl 2015;42(17–18). doi: 10.1016/j.eswa.2015.04.014
22. Alkhodair SA, Fung BCM, Rahman O, Hung PCK. Improving interpretations of topic modeling in microblogs. J Assoc Inf Sci Technol 2018;69(4). doi: 10.1002/asi.23980
23. Eriksson Krutrök M, Lindgren S. Continued Contexts of Terror: Analyzing Temporal Patterns of Hashtag Co-Occurrence as Discursive Articulations. Social Media and Society 2018;4(4). doi: 10.1177/2056305118813649
24. Caldarelli G, de Nicola R, del Vigna F, Petrocchi M, Saracco F. The role of bot squads in the political propaganda on Twitter. Communications Physics 2020 3:1 Nature Publishing Group; 2020 May 11;3(1):1–15. doi: 10.1038/s42005-020-0340-4
25. Ruiz-Núñez C, Segado-Fernández S, Jiménez-Gómez B, Hidalgo PJJ, Magdalena CSR, Pollo M del CÁ, Santillán-Garcia A, Herrera-Peco I. Bots’ Activity on COVID-19 Pro and Anti-Vaccination Networks: Analysis of Spanish-Written Messages on Twitter. Vaccines (Basel) Vaccines (Basel); 2022 Aug 1;10(8). PMID:36016126
26. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. International Journal of Infectious Diseases 2021;108.

doi: 10.1016/j.ijid.2021.05.059
27. Scannell D, Desens L, Guadagno M, Tra Y, Acker E, Sheridan K, Rosner M, Mathieu J, Fulk M. COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques, Sentiment and Mis/Disinformation. J Health Commun 2021;26(7). doi: 10.1080/10810730.2021.1955050
28. Ahmed W, Seguí FL, Vidal-Alaball J, Katz MS. COVID-19 and the "Film Your Hospital" conspiracy theory: Social network analysis of Twitter data. J Med Internet Res 2020;22(10). doi: 10.2196/22374
29. Uyheng J, Carley KM. Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. J Comput Soc Sci 2020;3(2):445–468. PMID:33102925
30. Uyheng J, Bellutta D, Carley KM. Bots Amplify and Redirect Hate Speech in Online Discourse About Racism During the COVID-19 Pandemic. Soc Media Soc 2022;8(3):20563051221104748. doi: 10.1177/20563051221104749
31. Green M, Musi E, Rowe F, Charles D, Pollock FD, Kypridemos C, Morse A, Rossini P, Tulloch J, Davies A, Dearden E, Maheswaran H, Singleton A, Vivancos R, Sheard S. Identifying how COVID-19-related misinformation reacts to the announcement of the UK national lockdown: An interrupted time-series study. Big Data Soc 2021;8(1). doi: 10.1177/20539517211013869
32. Chang H-CH, Ferrara E. Comparative analysis of social bots and humans during the COVID-19 pandemic. J Comput Soc Sci 2022;5(2):1409–1425. doi: 10.1007/s42001-022-00173-9
33. Zhang M, Chen Z, Qi X, Liu J. Could Social Bots' Sentiment Engagement Shape Humans' Sentiment on COVID-19 Vaccine Discussion on Twitter? Sustainability 2022;14(9). doi: 10.3390/su14095566
34. Suarez-Lledo Victor, and Alvarez-Galvez J, Suarez-Lledo V, Alvarez-Galvez J. Assessing the Role of Social Bots During the COVID-19 Pandemic: Infodemic, Disagreement, and Criticism. Journal of Medical Internet Research; 2022 Aug 25;24(8):e36085. PMID:35839385
35. Graham T, Bruns A, Angus D, Hurcombe E, Hames S. #IStandWithDan versus #DictatorDan: the polarised dynamics of Twitter discussions about Victoria's COVID-19 restrictions. Media International Australia 2021;179(1). doi: 10.1177/1329878X20981780
36. Ruffer N, Knitza J, Krusche M. #Covid4Rheum: an analytical twitter study in the time of the COVID-19 pandemic. Rheumatol Int 2020;40(12). doi: 10.1007/s00296-020-04710-5
37. Liu AKC, Ophir Y, Tsai S-A, Walter D, Himelboim I. Hashtag activism in a politicized pandemic: Framing the campaign to include Taiwan in the World Health Organization's efforts to combat COVID-19. New Media Soc 0(0):14614448221099172. doi: 10.1177/14614448221099173
38. Sharevski F, Alsaadi R, Jachim P, Pieroni E. Misinformation warnings: Twitter's soft moderation effects on COVID-19 vaccine belief echoes. Comput Secur 2022;114. doi: 10.1016/j.cose.2021.102577
39. Calac AJ, Haupt MR, Li Z, Mackey T. Spread of COVID-19 Vaccine Misinformation in the Ninth Inning: Retrospective Observational Infodemic Study. JMIR Infodemiology 2022;2(1). doi: 10.2196/33587
40. Herrera-Peco I, Jiménez-Gómez B, Romero Magdalena CS, Deudero JJ, García-Puente M, Benítez De Gracia E, Ruiz Núñez C. Antivaccine movement and covid-19 negationism: A content analysis of spanish-written messages on twitter. Vaccines (Basel) 2021;9(6). doi: 10.3390/vaccines9060656
41. Sayyadiharikandeh M, Varol O, Yang KC, Flammini A, Menczer F. Detection of Novel Social Bots by Ensembles of Specialized Classifiers. International Conference on Information and Knowledge Management, Proceedings 2020. doi: 10.1145/3340531.3412698
42. Varol O, Ferrara E, Davis C, Menczer F, Flammini A. Online Human-Bot Interactions: Detection, Estimation, and Characterization. Proceedings of the International AAAI Conference on Web and Social Media 2017 May 3;11(1):280–289. Available from: https://ojs.aaai.org/index.php/ICWSM/article/view/14871 [accessed Oct 20, 2021]
43. Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. World Wide Web Internet And Web Information Systems 1998;54(1999–66). doi: 10.1.1.31.1768
44. Freeman LC. A Set of Measures of Centrality Based on Betweenness. Sociometry 1977;40(1). doi:

10.2307/3033543
45. Havey NF. Partisan public health: how does political ideology influence support for COVID-19 related misinformation? J Comput Soc Sci 2020;3(2). doi: 10.1007/s42001-020-00089-2
46. Blevins JL, Edgerton E, Jason DP, Lee JJ. Shouting Into the Wind: Medical Science versus “B.S.” in the Twitter Maelstrom of Politics and Misinformation About Hydroxychloroquine. Social Media and Society 2021;7(2). doi: 10.1177/20563051211024977
47. Haupt MR, Li J, Mackey TK. Identifying and characterizing scientific authority-related misinformation discourse about hydroxychloroquine on twitter using unsupervised machine learning. Big Data Soc SAGE Publications Ltd; 2021 May 6;8(1). doi: 10.1177/20539517211013843/ASSET/IMAGES/LARGE/10.1177_20539517211013843-FIG2.JPEG
48. Xu W, Sasahara K. Characterizing the roles of bots on Twitter during the COVID-19 infodemic. J Comput Soc Sci 2021; doi: 10.1007/s42001-021-00139-3
49. Stella M, Ferrara E, de Domenico M. Bots increase exposure to negative and inflammatory content in online social systems. Proc Natl Acad Sci U S A 2018;115(49):12435–12440. PMID:30459270
50. Niburski K, Niburski O. Impact of trump’s promotion of unproven COVID-19 treatments and subsequent internet trends: Observational study. J Med Internet Res 2020;22(11). doi: 10.2196/20044
51. Shahsavari S, Holur P, Wang T, Tangherlini TR, Roychowdhury V. Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. J Comput Soc Sci 2020;3(2):279–317. doi: 10.1007/s42001-020-00086-5
52. Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F. The spread of low-credibility content by social bots. Nat Commun 2018;9(1):4787. doi: 10.1038/s41467-018-06930-7

Supplementary Files

Multimedia Appendixes

List of common hashtags.
URL: http://asset.jmir.pub/assets/4926dade899904311508cb1812a4ceaa.docx