Article

The Conversation around COVID-19 on Twitter—Sentiment Analysis and Topic Modelling to Analyse Tweets Published in English during the First Wave of the Pandemic

Javier J. Amores * , **David Blanco-Herrero** and **Carlos Arcila-Calderón**

Department of Sociology and Communication, University of Salamanca, 37008 Salamanca, Spain;
david.blanco.herrero@usal.es (D.B.-H.); carcila@usal.es (C.A.-C.)
* Correspondence: javieramores@usal.es

Abstract: The COVID-19 pandemic disrupted societies all over the world. In an interconnected and digital global society, social media was the platform not only to convey information and recommendations but also to discuss the pandemic and its consequences. Focusing on the phase of stabilization during the first wave of the pandemic in Western countries, this work analyses the conversation around it through tweets in English. For that purpose, the authors have studied who the most active and influential accounts were, identified the most frequent words in the sample, conducted topic modelling, and researched the predominant sentiments. It was observed that the conversation followed two main lines: a more political and controversial one, which can be exemplified by the relevant presence of former US President Donald Trump, and a more informational one, mostly concerning recommendations to fight the virus, represented by the World Health Organization. In general, sentiments were predominantly neutral due to the abundance of information.

Keywords: COVID-19; coronavirus; pandemic; Twitter; topic modelling; sentiment analysis; Donald Trump; World Health Organization

1. Introduction

The SARS-CoV-2 pandemic has been the biggest properly *global* event since the popularization of social media. No other episode in this century has had the social and economic impact, duration, or media coverage of this pandemic, declared as such on 11 March 2020 by the World Health Organization (WHO 2020b). In a globally connected world, in which news travels fast and in which citizens can express their fears and opinions in an almost immediate and unfiltered way on their social media, these platforms have become a very powerful tool for understanding how societies perceived and lived not only through the pandemic but also the measures that were implemented to control it, including strict lockdowns and prohibitions of non-essential activities.

Social media platforms such as Twitter play a key role in communication scenarios, not only as new scenarios for sharing feelings but also as relevant sources of information, both for citizens and for journalists (Hermida 2010; López-Meri 2015). That is why text mining Twitter's unstructured data is becoming a very powerful tool for multiple research fields, including epidemiology. In fact, previous studies have observed the potential of Twitter conversations for the forecasting of flu incidence (Paul et al. 2014) or for dengue surveillance (Gomide et al. 2011). In the field of communication, the most frequent application of text mining on Twitter in relation to epidemics has been the study of public awareness, sentiments, and the conversation around them. This was visible during the 2014 Ebola (Lazard et al. 2015; Guidry et al. 2017) and 2015 Zika outbreaks (Pruss et al. 2019; Miller et al. 2017; Fu et al. 2016).

Given the deadlier and broader condition of the COVID-19 pandemic and the more generalized use of social media as a public agora, many studies of Twitter conversations around the new coronavirus have been published. Some have focused on the outbreak and the initial stages (Prabhakar Kaila and Prasad 2020; Jahanbin and Rahmanian 2020),

as well as on the *infodemic*[1] and the spread of misinformation surrounding the pandemic (Bridgman et al. 2020; Singh et al. 2020). Other common topics have been the use of Twitter during the pandemic by mass media (Yu et al. 2020) or as a tool for political communication (Yunez 2020). Later studies have also studied vaccination attitudes (Yurtsever et al. 2023).

Most of these works have taken national approaches, which our study seeks to overcome by collecting content in English, the most spoken language on the platform (Vicinitas 2018), independently from countries. Following the line of previous works, such as the ones by Mutanga and Abayomi (2022), Wicke and Bolognesi (2020), and Xue et al. (2020), our study offers an exploratory analysis with the capacity to detect transnational topics and characters.

Thus, together with this global dimension, the novelty of the present work is its focus on the months in which the first international wave, which affected most Western countries in March and April 2020, was starting to stabilize and the cases began to decline. So far, studies have focused on the Chinese outbreak—December 2019 to January 2020—or on the spread of the disease into most countries—February and March 2020—whereas others have tried to cover longer periods of time including the year 2021. However, our research focuses on a specific sample of 70 days between April and June 2020. This is relevant because it is the moment in which public conversation was broadening, introducing topics such as politics or economics besides health and the pandemic, which practically monopolized the conversation during the first weeks after the declaration of the pandemic (Yu et al. 2020; Yunez 2020). Thus, our research provides a broader understanding of the discussion by focussing on a determinant yet under-researched period.

Research Questions

All that being said, the main goal of this article is to understand the discourse around the COVID-19 pandemic published in English on Twitter once the first wave started to stabilize in Western countries. Following the agenda-setting theory (McCombs and Shaw 1972), which has been proven to also apply to social media such as Twitter (Lee and Xu 2018), it is relevant to identify which individuals or institutions are leading a conversation, as they will be in a stronger position to determine the topics in the agenda that will be discussed. That is why the first step will be to study who the most active public actors were, as well as the most influential, based on the public metrics of retweets and mentions.

1. Which public accounts were most active and influential on Twitter's discourse in English around COVID-19 during the stabilization and decline of the first wave of the pandemic in Western countries?

In connection with this agenda approach, one of the most common attempts among studies analysing Twitter conversations around the COVID-19 pandemic has been the detection of topics. This is of great relevance because the topics addressed by citizens and their conversations can influence their posterior behaviours and their attitudes towards health issues, such as vaccines, or the level of trust in the measures implemented by governments to control the spread of the virus (Lim et al. 2020; Gozgor 2022).

The use of topic modelling has been common for studying the conversation around the pandemic on social media. Prabhakar Kaila and Prasad (2020) focused on the first stages of the pandemic and observed that Twitter is considered "one of the most preferred media for information spread during pandemics" (p. 133) and that misinformation did not play a great role at that moment. Yu, Lu, and Muñoz-Justicia (Yu et al. 2020) studied the frames used by Spanish news media on Twitter before, during, and after the lockdown. Mutanga and Abayomi (2022) analysed a variety of topics and the surge in fake news and conspiracy theories around the pandemic and the lockdown in South Africa. Alamoodi et al. (2022), Gourisaria et al. (2022), and Mathayomchan et al. (2022) offer very complete approaches, combining topic modelling with sentiment analysis in Malaysia, India, and Southeast Asian countries, respectively. These studies offer a basis for our work, but their national or regional dimension is replaced by a completely international approach.

Before addressing the latent topics, it is also interesting, as a preliminary observation, to identify the most frequent terms, as they can offer complementary hints for the topic modelling interpretation. Thus, we wonder about the following:

2. What were the most frequent words used in tweets around COVID-19 published in English during the stabilization and decline of the first wave of the pandemic in Western countries?

3. What topics were used in Twitter's discourse in English around the COVID-19 pandemic during the stabilization and decline of the first wave of the pandemic in Western countries?

Finally, in addition to agenda theory, framing theory (Goffman 1974) should also be considered, as it is relevant to determine the attributes used to discuss and frame a topic. This theory justifies the study of sentiments, as they provide the most relevant frame to analyse the pandemic. Previous works that have focused on the predominant sentiments on Twitter about the pandemic are the aforementioned ones by Alamoodi et al. (2022), Gourisaria et al. (2022), and Mathayomchan et al. (2022), as well as one by Lwin et al. (2020), which paid attention to the first stages of the pandemic. With the goal of studying the period in which the conversation around the pandemic was going beyond purely health issues, as well as to offer a more international understanding of the phenomenon, the following research question is posed:

4. Which is the predominant sentiment in tweets around COVID-19 published in English during the stabilization and decline of the first wave of the pandemic in Western countries?

2. Methods

A completely computational strategy was carried out, both for the search, download, and collection of the sample, as well as for its analysis. This methodological design allowed for the exploration and identification of the sentiments and topics underlying all international English-language tweets collected in the defined period. Each of the techniques developed in this work is detailed below.

2.1. Data Collection

The selected sample covers tweets in English including the terms "coronavirus", "covid-19", "covid_19", "covid2019", or "covid19" posted on Twitter from 13 April to 22 June. The selection of these dates seeks to cover the *stabilization* phase of the first wave of the pandemic in Western countries. More specifically, on 13 April, the UK reached its peak of daily new confirmed COVID-19 deaths (7-day rolling average) for the first wave according to figures from Johns Hopkins University data[2], and it was on 22 June when the Health and Social Care Secretary of the United Kingdom, Matt Hancock, announced that, the next day, the Prime Minister would set out the next steps to ease the national lockdown[3]. Although the selection of these specific dates is based on the UK's context, other European countries, as well as the United States, were following a similar trend: the peak of daily deaths for the first wave in Europe took place on 10 April, while the one in the USA took place on 24 April, and in both contexts, the curve was reaching its lowest levels at the end of June.

To download the sample, the Python programming language was used to access Twitter data through its REST API, since, at the time, it was not possible to access the academic API 2. This means that it was not possible to access the total number of tweets posted on those dates, so we sampled the tweet history, downloading between 5000 and 10,000 tweets per day within 10 days after their publication. For this, a strategy already designed by the authors in previous works was executed (e.g., Arcila-Calderón et al. 2017; Arcila-Calderón et al. 2022). Specifically, a language filter was used to ensure that all the collected tweets were written in English. All retweets and replies were also filtered out, thus collecting only original tweets. In total, 436,296 tweets in English published during the indicated period were downloaded, along with their aggregated data. This initial sample was used to extract exploratory data about the users and their public metrics, such as the total number of posts, retweets, and mentions in order to answer RQ1. After this, the downloaded dataset was cleaned, rejecting all duplicate or repeated messages, as well as those that did not contain

textual information—tweets with only emojis, links, or empty content. The final sample used for the topic modelling included a total of 180,509 tweets (after removing 255,787).

2.2. Word Frequency Distribution

After having collected and cleaned the tweet dataset, the first step was to apply basic techniques of natural language processing (NLP) to obtain the frequency distribution of words. NLP is a branch of computer science that is combined with applied linguistics and seeks to convert a text into a set of structured data that describe its meaning and the topics it transmits (Collobert et al. 2011). For this study, different Python libraries were used, such as *Numpy*, which adds greater support for vectors and matrices, as well as the *Natural Language Toolkit* (NLTK), which defines an infrastructure that allows for the development of NLP scripts.

Word frequency was used as a preliminary step for the identification of underlying topics after a filtering process. Knowing the most frequent words in the sample offers valuable exploratory information, useful for a better interpretation of the results of the topic modelling. The first step to correctly perform NLP techniques was the identification of *tokens*—the basic units—typically words or short sentences into which a text can be deconstructed for further analysis. For this process of tokenization, the aforementioned NLTK library was used. The next step was the removal of stop-words, which are those very frequent and common words that do not provide relevant information, such as articles or prepositions. In this phase, after running several tests, we removed the words and combinations of words that made up the search terms, as well as some directly related to COVID-19, since, as expected, these were too frequent and could bias the results. In specific, the removed terms were *coronavirus*, *covid*, *covid 19*, *covid19*, *covid_19*, *covid-19*, and *covid'19*. Punctuation marks and weblinks were also removed to avoid the repetition of terms and obtain homogeneous and coherent findings. Finally, we were able to look at the most repeated terms and their distribution and decide how many topics it was convenient to obtain. Once these adjustments were made and the stop-words removed, word clouds were generated with the most frequent words in order to better visualize the most and least frequent terms in the analysed messages.

2.3. Topic Modelling

Given the promising results and the growing trend of topic modelling use among text mining techniques (Bogović et al. 2021; Wright et al. 2022), particularly for Twitter (Karami et al. 2020), this method was employed for the detection of topics around which public discourse on Twitter related to the COVID-19 pandemic was built. In this case, the authors also followed a strategy developed in previous works to identify the main underlying topics in a dataset (e.g., Latorre and Amores 2021). Specifically, the Latent Dirichlet Allocation (LDA) algorithm was used, the most common for the identification of topics in a set of documents (Ramage et al. 2009; Grimmer and Stewart 2017). With this technique, topics are detected by automatically identifying patterns in the presence of groups of concurrent words in the documents (Jacobi et al. 2016). In this case, in addition to NLTK, the following Python libraries were used: *Pandas*, used for data analysis; *Gensim*, used for the topic modelling; and *pyLDAvis*, used for displaying inferred topics on maps. After importing all the requested libraries and modules and selecting the sample to model, the next step, once again, was to convert all text to lowercase and remove the punctuation marks, double spaces, and stop-words—a total of 864—to achieve a higher level of coherence in the identified topics. After this cleaning process, internal coherence values were extracted, which allowed us to decide the total number of topics that should be inferred. Similarly, the *pyLDAvis* library allowed us to print interactive display maps to visually explore the results of the modelling, which also helped to more reliably select the number of latent topics to detect. With all this, it was decided to model a total of 6 topics, as it was the most coherent number according to the visualizations and the number that presented the highest internal coherence (0.387). Finally, a manual validation was carried out exploring the tweets in which the different topics were most predominantly present.

2.4. Sentiment Analysis

The last stage was the identification of the latent sentiments in the sample. We used SentiStrength, an open-source tool developed by Thelwall et al. (2011) that allows for automatic sentiment analysis from lexicon dictionaries. Specifically, this validated software rates the relevance and presence of negative words (from −1 to −5) and positive words (from +1 to +5) for each text. The sum of these two values indicates the general emotions of the tweet in terms of language (language sentiment). To report global results regarding latent sentiments, the total mean of the coefficients obtained was extracted, as well as the percentages of all tweets with positive (from +5 to +1), negative (from −1 to −5), and neutral (0) sentiments, the last ones usually being purely informative texts.

3. Results

3.1. Most Active and Influential Twitter Users in the Conversation around COVID-19 during the Decline of the First Wave of the Pandemic

The public metrics data extracted from the original dataset with all 436,296 downloaded tweets answered RQ1. Specifically, the number of tweets posted during the period allowed us to identify the most active users talking about COVID-19 in English. The number of retweets and mentions was also observed to identify the most influential users. The number of followers and followed users by each of those accounts, the date of creation, and the declared country were used to determine their nature and whether they were public figures, common users, or potential bots and trolls.

It analyzes worth mentioning that among the most active users, one stands out, with a total of 727 posted tweets, but that account is no longer active, so it might have been removed or blocked by Twitter. Among the 10 most active accounts, 3 of them were deleted by the time of the analysis. It should also be noted that a total of three accounts include the word "bot" in their usernames. One of those accounts is no longer active, and another has only four followed users and was created in January 2020, just at the beginning of the COVID-19 pandemic. In addition, among the most active users, none seems to stand out as a public figure, with the possible exception of *HO_Wrestling*, an alleged news account about wrestling, as well as two accounts from *MyNation* Foundation members, an alleged "non-profit association of Self Help Support for Dowry Law victims". The total number of tweets posted by these most active accounts, together with their public metrics, are shown in Table 1.

Table 1. Data of the 10 most active users in the conversation about COVID-19 during the studied period.

User	Creation Date	Declared Country	Followers	Following	Total Tweets Posted	Tweets Posted in the Dataset
@naattuvarthakal	Deleted account	Deleted account	Deleted account	Deleted account	Deleted account	727
@Arikring	July 2019	Israel	88 K	62.8 K	62.8 K	146
@SmartUSAPat1	January 2017	USA	846	633	4427	141
@Clairebotai	Deleted account	Deleted account	Deleted account	Deleted account	Deleted account	117
@Sumanebot	June 2015	Sri Lanka	618	304	864.7 K	111
@Ourfuturebot	January 2020	Italy	2706	4	231.3 K	106
@Sweposten	Deleted account	Deleted account	Deleted account	Deleted account	Deleted account	103
@HO_Wrestling	July 2012	UK	2474	1272	8652	97
@MynationSos	July 2019	India	748	525	23.6 K	93
@mynation_pune	November 2019	-	542	277	12.5 K	91

Source: Own elaboration.

Among the most influential accounts in the dataset, profiles of international public actors can be recognized. The first and most influential account is that of Donald Trump,

which was later blocked and deleted by Twitter. In total, the posts about COVID-19 published by the former US President accumulated 15,667 retweets and mentions during the analysed period, more than 2.5 times as much as the second most influential account. Other relevant political figures or organizations from the USA present among the 10 most influential accounts are the Speaker of the House of Representatives of the US, Nancy Pelosi, and Dr Dena Grayson, as well as the Lincoln Project, a political committee formed in 2019 by several prominent Republicans and former Republicans with the objective of preventing the re-election of Donald Trump (Young 2020).

The second account in terms of influence is that of the World Health Organization, with 5820 retweets and mentions, something that is unsurprising during a health crisis. Other well-known personalities among the 10 most influential accounts were the writer Stephen King and the filmmaker Ava Duvernay. Another relevant account in the list is that of the Nigeria Centre for Disease Control. The number of retweets and mentions that the tweets posted by these accounts had in the original dataset, together with the public metrics of these users, can be seen in Table 2.

Table 2. Data of the 10 most influential users in the conversation about COVID-19 during the studied period.

| User	Creation Date	Declared Country	Followers	Following	Total Tweets Posted	Retweets and Mentions in the Dataset
@Realdonaldtrump	Blocked account	Blocked account	Blocked account	Blocked account	Blocked account	15.667
@WHO (World Health Organization)	April 2008	Switzerland	9.6 M	1738	63.6 K	5820
@Demetriachavon	November 2016	USA	2217	527	2786	3191
@NCDCgov (Nigeria Centre for Disease Control)	March 2016	Nigeria	1.1 M	382	13.4 K	1850
@ava (Ava Duvernay)	June 2018	-	2.7 M	15.8 K	52.1 K	1790
@Totallyjesss	February 2015	-	918	646	24.6 K	1671
@ProjectLincoln	December 2019	USA	2.7 M	795	14.6 K	1550
@TeamPelosi (Nancy Pelosi)	April 2014	USA	811.7 K	8242	9661	1387
@DrDenaGrayson	August 2013	USA	326.1 K	437	78.4 K	1233
@StephenKing	December 2013	-	6.5 M	132	6332	1159

Source: Own elaboration.

3.2. Most Frequent Words in the Conversation about COVID-19 during the Decline of the First Wave of the Pandemic

The application of NLP techniques allowed for the extraction of the most frequent words used in the collected tweets once the dataset had been cleaned and after deleting stop-words. The 50 most frequent terms in tweets about COVID-19 in English at the end of the first wave of the pandemic were

('covid19', 113312), ('coronavirus', 51346), ('covid', 33614), ('people', 17844), ('pandemic', 15223), ('cases', 15183), ('new', 14993), ('health', 10513), ('covid_19', 9433), ('help', 9177), ('deaths', 8588), ('today', 8584), ('lockdown', 8485), ('trump', 7995), ('time', 7921), ('need', 7308), ('support', 6591), ('home', 6585), ('world', 6531), ('work', 6210), ('day', 6187), ('crisis', 5802), ('virus', 5754), ('care', 5642), ('government', 5098), ('positive', 5046), ('testing', 4996), ('know', 4931), ('state', 4838), ('realdonaldtrump', 4739), ('spread', 4685), ('response', 4682), ('patients', 4680), ('workers', 4400), ('2020', 4397), ('stay', 4379), ('news', 4334), ('going', 4254), ('public', 4100), ('social', 4026), ('china', 3949), ('uk', 3948), ('total', 3906), ('fight', 3885), ('safe', 3857), ('death', 3840), ('country', 3738), ('good', 3660), ('week', 3506), ('test', 3370)

The presence of the 30 most predo"inan' words can be better visualized in Figure 1. The terms used to refer to the COVID-19 disease stand out, something to be expected since they were precisely those that were used as keywords in the search and download of the sample.
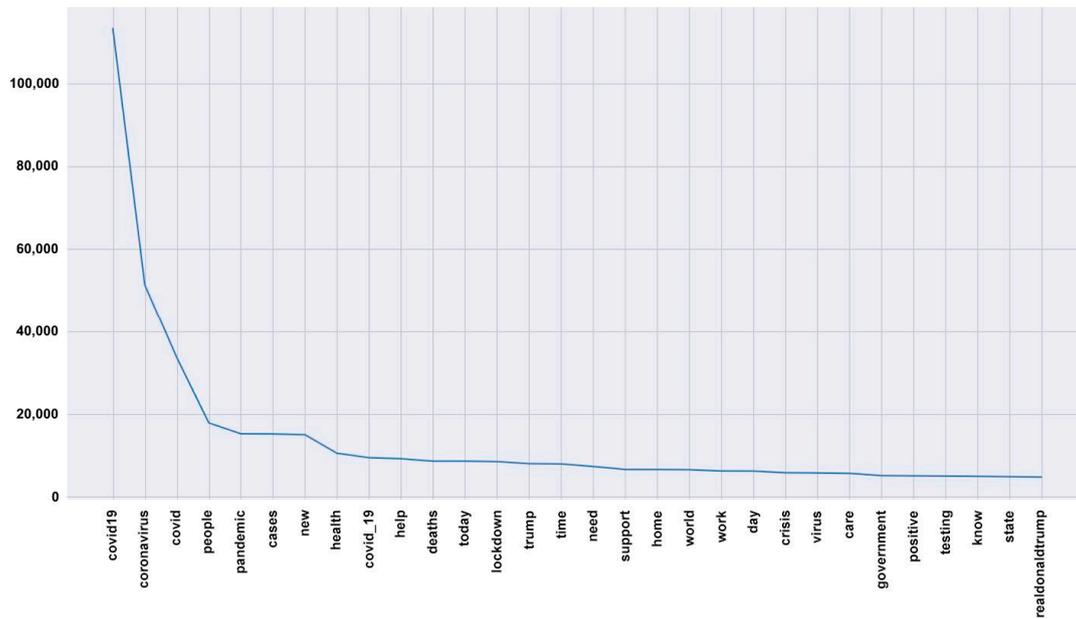
Figure 1. Thirty most frequent words in tweets about the pandemic in English during the studied period.

For this reason, we decided to include all used search terms among the stop-words, as well as all terms and hashtags used to refer to COVID-19, such as *coronavirus, covid, covid19, covid_19,* or similar variations. Consequently, the different frequencies of appearances of the rest of the terms can be better identified. The following 50 were the most frequent terms; Figure 2 shows the frequency of appearance of the first 30 terms.
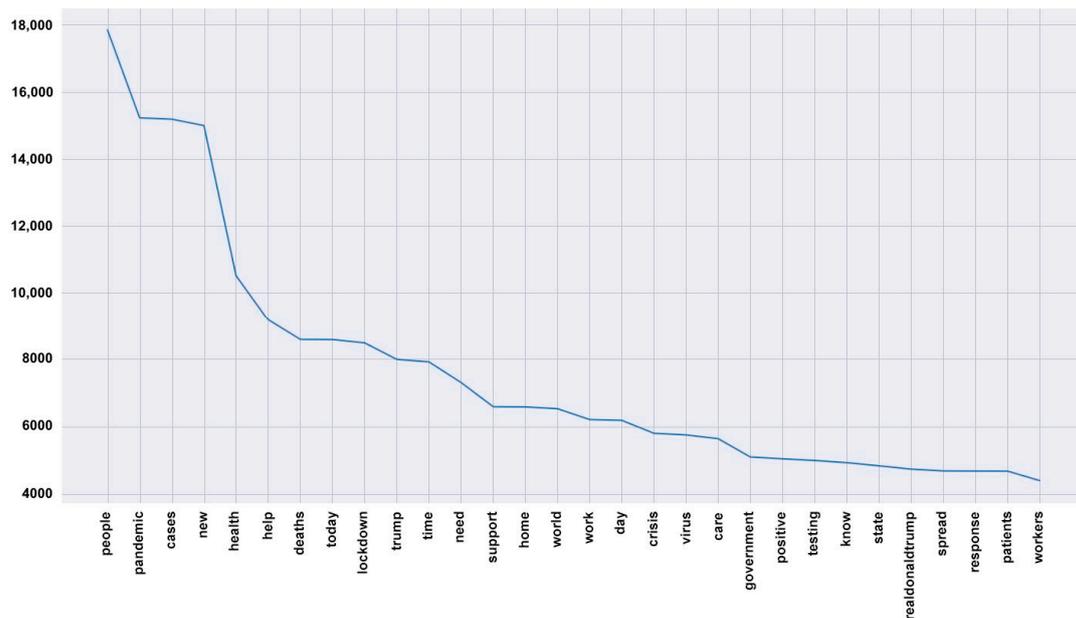
Figure 2. Thirty most frequent words in tweets about the pandemic in English during the studied period without terms referring to COVID-19.

('people', 17844), ('pandemic', 15223), ('cases', 15183), ('new', 14993), ('health', 10513), ('help', 9177), ('deaths', 8588), ('today', 8584), ('lockdown', 8485), ('trump', 7995), ('time', 7921), ('need', 7308), ('support', 6591), ('home', 6585), ('world', 6531), ('work', 6210), ('day', 6187), ('crisis', 5802), ('virus', 5754), ('care', 5642), ('government', 5098), ('positive', 5046), ('testing', 4996), ('know', 4931), ('state', 4838), ('realdonaldtrump', 4739), ('spread', 4685), ('response', 4682), ('patients', 4680), ('workers', 4400), ('2020', 4397), ('stay', 4379), ('news', 4334), ('going', 4254), ('public', 4100), ('social', 4026), ('china', 3949), ('uk', 3948), ('total', 3906), ('fight', 3885), ('safe', 3857), ('death', 3840), ('country', 3738), ('good', 3660), ('week', 3506), ('test', 3370), ('community', 3368), ('working', 3364), ('right', 3354), ('risk', 3343)

Answering RQ2, as shown in Figure 2, the most frequent words in the analysed tweets tend to be related to health and the measures taken to fight the pandemic: we can find words such as *people*, *pandemic*, *cases*, *health*, *help*, *deaths*, *lockdown*, *need*, *support*, *world*, *testing*, *patients*, *stay*, *public*, *crisis*, *safe*, *death*, or *risk*. All these terms refer to the health crisis and how to combat it and may imply that most of the messages report information about the evolution of the pandemic, as well as the measures and policies carried out.

Secondly, a large number of the most frequent words relate to politics and state issues, such as *trump*, *realdonaldtrump*, *government*, *state*, *china*, *uk*, *country*, or *right*. It can be seen that the surname and handle of the Twitter account of the former US President, Donald Trump, appear on this list. This shows what a relevant presence Donald Trump had in the conversation around COVID-19 at the end of the first wave of the pandemic, at least in the English-speaking context. Among these most frequent words, we also find the name of two countries, China and the UK, which also shows the prominence of both in these conversations about COVID in English, something to be partly expected considering that China was the country in which the virus originated and that tweets in English are being analysed, so the relevance of the UK, especially as it was the first of the English-speaking countries to suffer the most serious consequences of the pandemic, is understandable. Finally, we find a series of words with a more positive and encouraging tone, which seems to refer to the importance of the public and social union, as well as the need to fight and work together to overcome the pandemic. Some of these words are *work*, *workers*, *positive*, *social*, *fight*, *public*, *good*, and *community*. Figure 3 shows two word clouds with each of the referred samples, one in which the terms related to COVID-19 are maintained and the other without these terms.
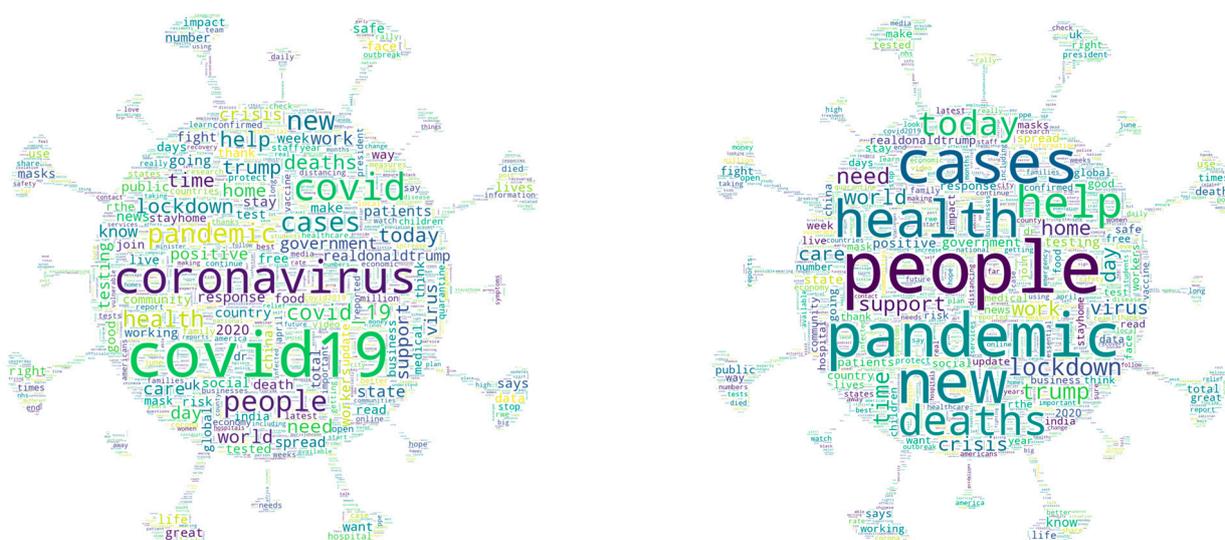
Figure 3. Word clouds with the most frequent terms in tweets about the pandemic in English during the studied period, with and without words referring to COVID-19.

3.3. Predominant Topics in the Conversation about COVID-19 during the Decline of the First Wave of the Pandemic

After obtaining the frequency distribution for the analysed tweets, topic modelling was conducted to automatically detect the main underlying topics in the conversation about COVID-19 that took place on Twitter during the analysed period, thus answering RQ3. The level of coherence was measured—the further from zero, the better—to determine an adequate number of topics, comparing several models with ten words for each topic, and we finally decided that the adequate number of topics was six. After removing the stop-words, the topics were detected and validated by exploring examples of tweets for each one:

Topic 1. Information about the pandemic (Figure 4). This topic focuses on cases, infections, incidence rates, deaths, lethality, and virulence. It may be health information offered by international politicians and institutions to control and combat the health and economic crisis or journalistic information on statistics and surveys; there are also advice and recommendations to fight the virus, and caution is requested. The most representative words are the following:

('0.028"cases" + 0.021*"new" + 0.017*"positive" + 0.015*"pandemic" + 0.010*"deaths" + 0.010*"health" + 0.009*"today" + 0.008*"people" + 0.008*"virus" + 0.008*"mask" + 0.007*"masks" + 0.007*"campaign" + 0.006*"day" + 0.006*"time" + 0.006*"total" + 0.005*"testing" + 0.005*"staffers" + 0.005*"think" + 0.005*"help" + 0.004*"death"')*

Here are some examples of tweets on this topic:

- "Released today: a free information book explaining the #coronavirus to children, illustrated by Gruffalo illustrator #AxelScheffler"
- "Today @UNDP has an even greater role to play in shaping responses to #COVID19, I told Administrator @AchimSteiner in our discussion this evening on how best Maldives & @UNDP can partner to control the virus. Also thanked him for his leadership in highlighting challenges #SIDS face https://t.co/qLJlQXamJo"
- "USAID donated two ambulances to the Rizgary Hospital today to support #Erbil Health Directorate's response to #COVID19. The U.S. continues to provide key resources to help save lives, build health institutions and reduce delays in communities receiving critical medical attention. https://t.co/pAPXCqRuhg"
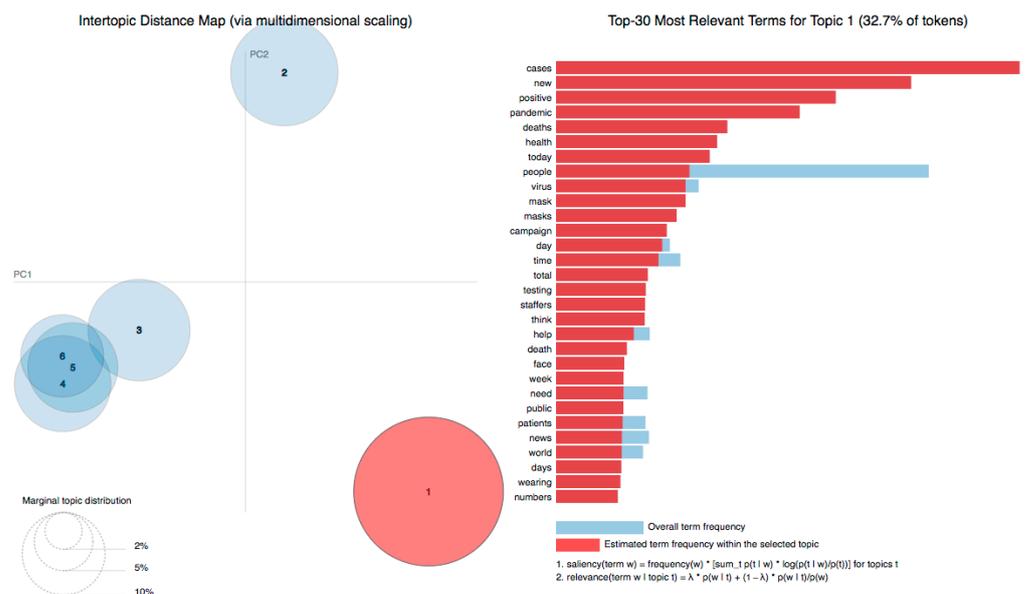
Figure 4. Interactive map of topic 1.

Topic 2. Information on the health and political crisis, specifically in the US, the country that dominates the discourse, with the figure of Trump (*RealDonaldTrump*) as the main

protagonist (Figure 5). These are not only messages launched by government and institutions about the pandemic but also responses to those messages and discussions with a more politics-related approach than a purely health-related one. This topic also includes the responses of citizens to the management of the pandemic; many of the messages are direct criticism of the Trump government and its handling of the pandemic. The main words are

('0.043*"trump" + 0.027*"rally" + 0.026*"people" + 0.015*"realdonaldtrump" + 0.012*"going" + 0.009*"state" + 0.007*"home" + 0.007*"make" + 0.006*"work" + 0.006*"want" + 0.006*"social" + 0.006*"covidiots" + 0.006*"reported" + 0.005*"tulsatrumprally" + 0.005*"lives" + 0.005*"crowd" + 0.005*"stay" + 0.005*"states" + 0.005*"know" + 0.005*"president"')

Some examples of tweets on topic 2 are the following:

- "Trump just threw a mega tantrum, cutting all funding to the World Health Organisation —in the middle of the #Covid19 pandemic! Now this massive public call to save the WHO is going viral! https://t.co/BQCyOkQ74w"
- "Trump delayed action on #Covid_19 so his buddies could sell off certain stocks. See, some of us are mistaken about who he is there to represent. Spoiler alert! It is not the 99%"
- "@realDonaldTrump Bill Gates, what a benevolent and kind person. Thanks for not feeding the starving masses or storing some PPE for the world. Oh thanks for your $100m donation for vaccines you will profit from. I don't give a shit if they jail me but know this #youcanshoveyourvaccine #covid19"
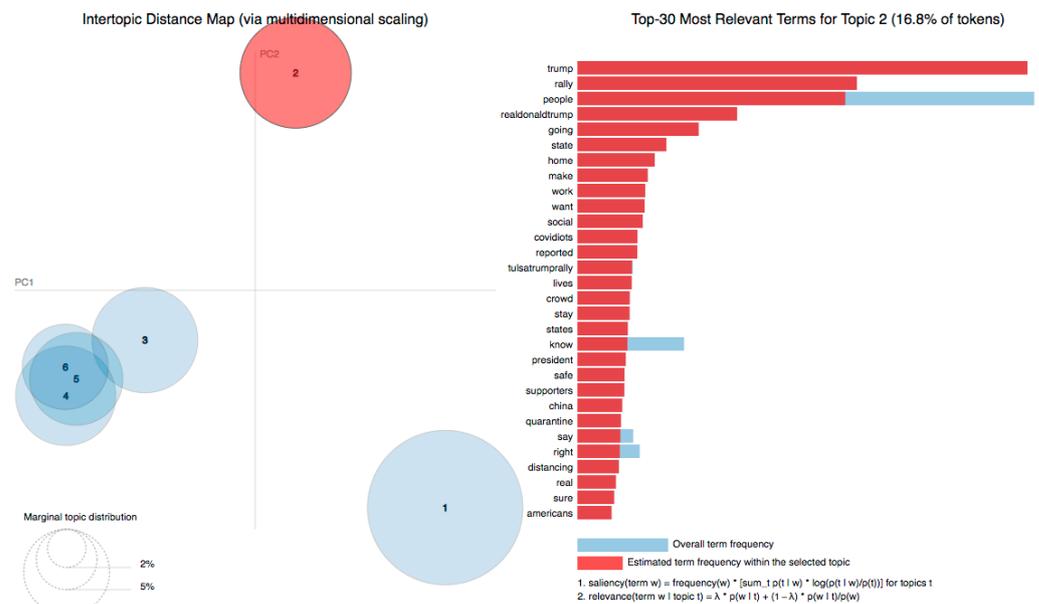
Figure 5. Interactive map of topic 2.

Topic 3. Negative and alarm messages about the risks of not respecting measures during the pandemic and the dangers posed by irresponsible citizens (Figure 6). This topic includes calls to prevention, warnings, and requests for responsible behaviours and community care. There seem to also be protests from the population due to the mismanagement of the pandemic and the (lack of) measures taken. The most representative terms in this topic are

(0.011*"risk" + 0.010*"florida" + 0.007*"died" + 0.007*"spread" + 0.007*"staff" + 0.006*"event" + 0.006*"know" + 0.006*"away" + 0.005*"members" + 0.005*"hope" + 0.005*"really" + 0.005*"family" + 0.005*"daily" + 0.005*"months" + 0.005*"infected" + 0.005*"community" + 0.005*"protests" + 0.004*"care" + 0.004*"increase" + 0.004*"symptoms"')

Some tweets that exemplify this topic are

- "Digitalgurucool request you to follow our PM Narendra Modi advice and stay safe at your home"
- "#covid_19 reminds us of our mortality. We all have to depart this physical body one day"
- "Lockdown has been extended till the 3rd of May, so let's stay untied and fight against COVID 19, stay, stay safe #lockdown #extended #letsfight #against #covid19 #gocorona #stayhome #staysafe #COVID2019 #FightAgainstCOVID19"
- "Don't listen to idiotic and denialist speeches and stay home"
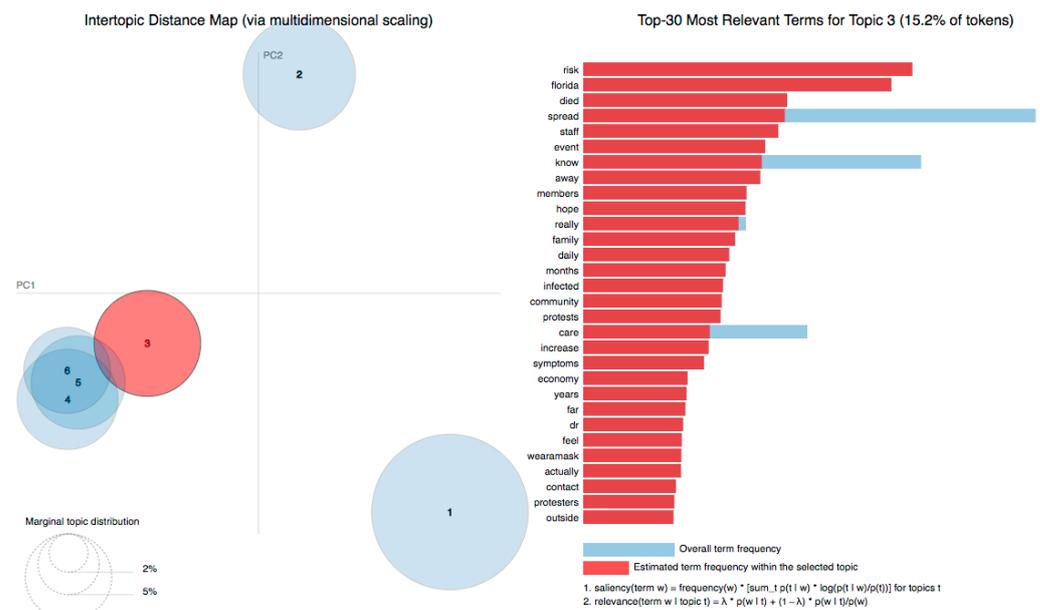
Figure 6. Interactive map of topic 3.

Topic 4. Objective information on health issues, comparative data, figures, and statistics (Figure 7). This is mostly information offered by the news media and public institutions. The main words of this topic are

('0.019"tested" + 0.018*"test" + 0.013*june" + 0.010*"good" + 0.007*"outbreak" + 0.007*"maybe" + 0.006*"data" + 0.005*"change" + 0.005*"big" + 0.004*"remember" + 0.004*"bad" + 0.004*"including" + 0.004*"long" + 0.004*"rise" + 0.004*"open" + 0.004*"check" + 0.004*"place" + 0.004*"arena" + 0.004*"half" + 0.003*"factors"')*

The following are some examples of tweets on this topic:

- "BREAKING NEWS: There were 2100 more deaths linked to #coronavirus in England and Wales by 3 April than reported by the government, according to the Office for National Statistics"
- "Britain's death toll from #coronavirus may be 15% higher than official numbers according to new Gov't figures. The Office for National Statistics says 15% is the additional figure for deaths in nursing & residential homes in England & Wales. The official toll up to y'day: 11,329"
- "Disparities in #COVID19 #testing rates are troubling. Delays in testing increase risk of a surge in silent spread & severe COVID19 cases. This epidemic is exacerbating large health #disparities across U.S. states"
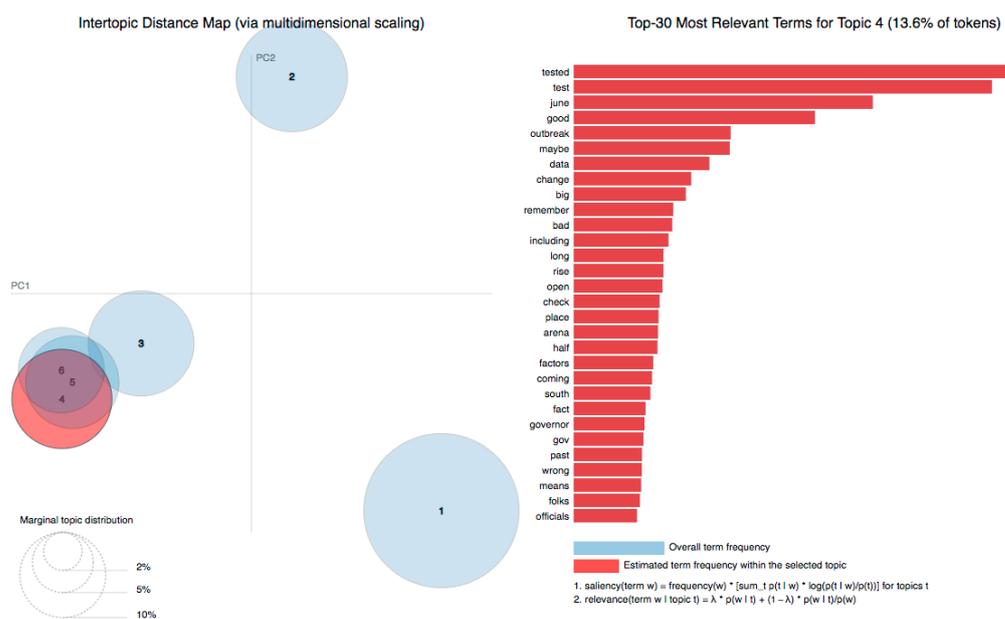
Figure 7. Interactive map of topic 4.

Topic 5. Mostly negative information on the evolution of the pandemic and the negative symptoms of the disease (Figure 8). These are messages that not only try to inform in a detailed and professional way but also raise awareness in society to be cautious and act carefully. Recommendations are offered based on research articles or reports, as well as rigorous and objective scientific data. The most representative words of this topic are

('0.011"confirmed" + 0.009*"weeks" + 0.008*"die" + 0.008*"study" + 0.007*"man" + 0.006*nigeria" + 0.005*"blame" + 0.005*"thousands" + 0.005*"love" + 0.005*"plan" + 0.004*"refugees" + 0.004*"infection" + 0.004*"research" + 0.004*"despite" + 0.004*"prevent" + 0.004*"trying" + 0.004*"case" + 0.004*"football" + 0.004*"worldrefugeeday" + 0.003*american"')*

Some examples of this topic are

- "The following information is relevant to assess the situation of #COVID-19 in Sindh as of 14 April at 8 AM: Total Tests 14,503, Positive Cases 1518 (today 66), Recovered Cases 427, Deaths 35"
- "The NIH is looking for blood samples from 10,000 healthy US adults for a research study to determine how many people without a confirmed history of #COVID19 infection have produced antibodies to the virus. #coronavirus"
- "Study Finds That Cloth Masks Can Increase Healthcare Workers Risk of Infection"
- "22% say they already can't afford essential items or housing costs, or think they are certain/very likely to during the crisis. @policyatkings surveyed the UK public on life under #Covid_19 lockdown. @RishiSunak https://t.co/LxOLfoAyh7"

Topic 6. Complementary issues besides the pandemic (Figure 9). In this case, health information is not the main issue but rather information or content that is only partially connected to the pandemic, including political, social, or current events, sometimes with sensationalist approaches. The main terms of this topic are

('0.013"wear" + 0.009*"media" + 0.009*"live" + 0.008*"black" + 0.007*"sick" + 0.007*"disease" + 0.006*"lot" + 0.006*"medical" + 0.006*"high" + 0.006*"watch" + 0.006*"god" + 0.006*"seen" + 0.006*"children" + 0.005*"second" + 0.005*"person" + 0.005*"close" + 0.005*"small" + 0.005*"patients" + 0.004*"happening" + 0.004*"wait"')*

Below are some tweets on this topic:

- "Exercise hour in the local park on a fine spring morninanalyze.#SpringTime #Spring #COVID19 #exercisewalk #fatheranddaughter #blossomwatch #Blossom #blossoms https://t.co/xSddbhevYO"
- "With many employees now working remotely as a result of #COVID19, organisations are starting to turn their attention to the challenge of managing a virtual workforce in the longer term. @PwC_UK shares tips on moving from crisis response to normality: https://t.co/re84rktq39 https://t.co/kK9JnpcAU6"
- "#DemiRose's boobs unleashed in riskiest bikini yet as she talks #coronavirus fears https://t.co/WTD5s8Hy0f"
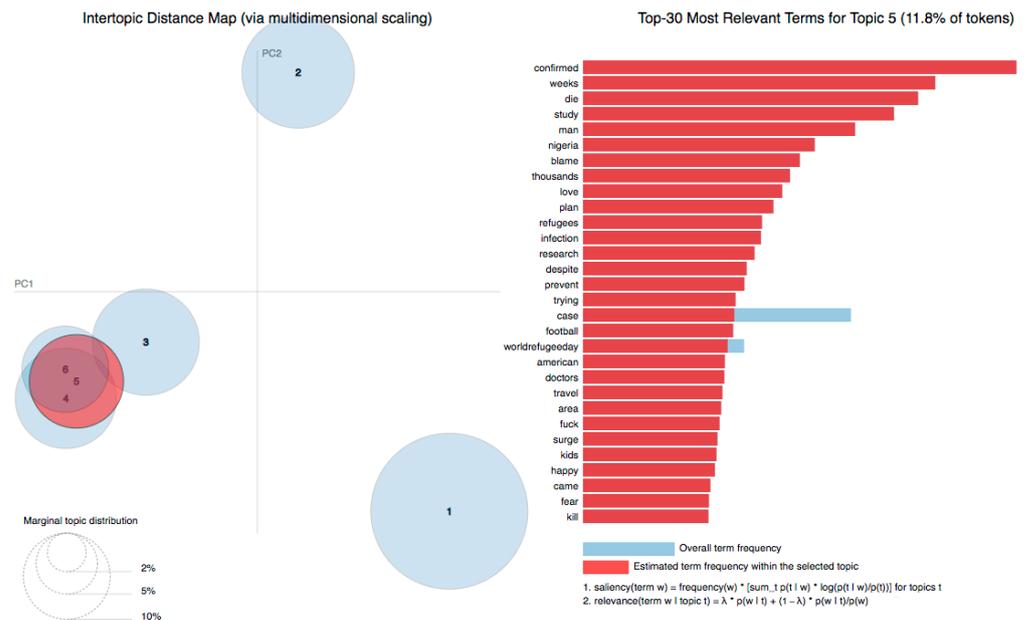
Figure 8. Interactive map of topic 5.
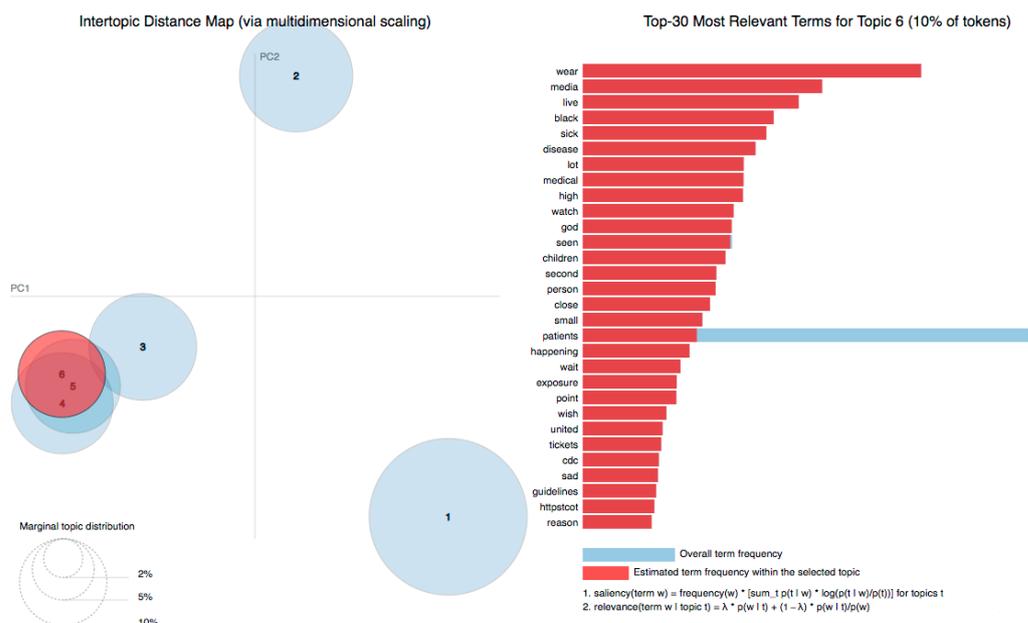
Figure 9. Interactive map of topic 6.

3.4. Sentiments in Tweets about COVID-19 during the Decline of the First Wave of the Pandemic

Finally, using SentiStrength, we conducted, first, a sentiment analysis with the total sample and, second, a longitudinal analysis, dividing the original sample into the 10 weeks of data collection. Considering the 180,509 clean tweets, a total of 41,160 messages had positive feelings (22.80% of the total), compared with 61,204 tweets with negative feelings (33.90%) and 78,146 completely neutral ones (43.29%). This is possibly explained by a large amount of merely informational tweets about the health crisis and everything that surrounds it. The mean of positive sentiments in the entire sample was 1.511, while the mean of negative sentiments was −1.744, which provided an overall mean result of −0.233, that is, a slightly negative trend, although close to neutrality.

Longitudinally, no large changes were observed; the mean sentiment was always negative and ranged between −0.160 during the most positive moment in the sixth analysed week—from 18 to 24 May—and −0.289 in the most negative moment during the following week—from 25 to 31 May. Figure 10 shows the evolution of the average sentiment throughout the period.
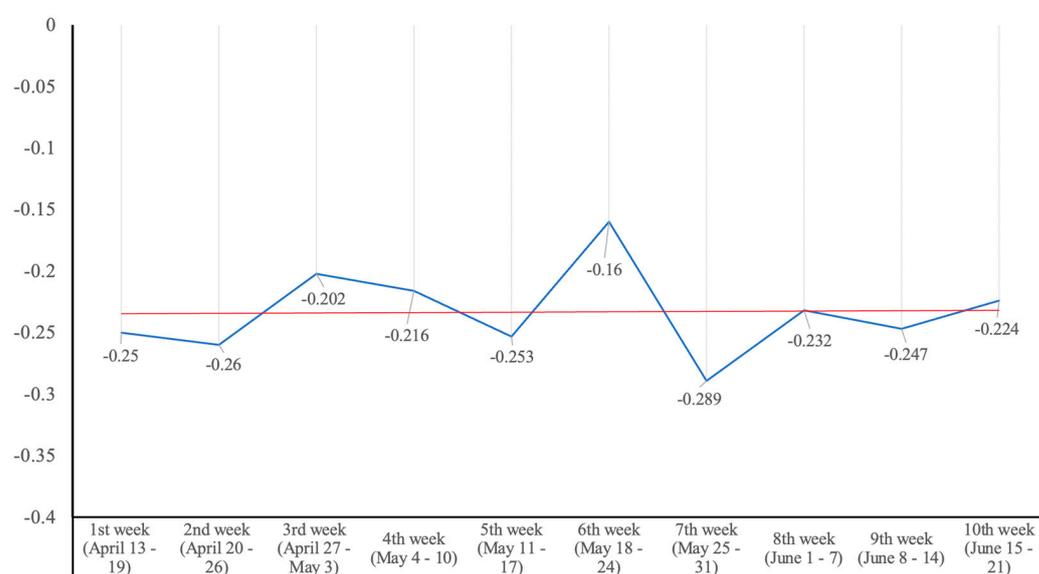
Figure 10. Evolution of mean sentiment in tweets about COVID-19 in English during the first wave of the pandemic.

Answering RQ4, it can be confirmed that the predominant sentiment in tweets about COVID-19 in English published at the end of the first wave of the pandemic was generally neutral but with a trend towards negative feelings. This can be associated with the observations made in our study of the most frequent words, as many of them had to do with the health crisis and how to combat it.

4. Discussion and Conclusions

This paper analysed the public conversation around COVID-19 that took place on Twitter during the weeks in which the first wave of the pandemic was receding in Western countries and the conversation was broadening beyond purely health topics. A large set of tweets related to the disease and published in English from 13 April to 22 June was downloaded, allowing us to access the conversation that took place mainly in the USA and in the UK, two of the countries with the most registered cases and deaths related to the virus during that first wave. Apart from the manual exploration of the collected public data extracted using Twitter's API, analyses based on computational techniques, such as word frequency distribution, topic modelling, and sentiment analysis, were used to produce more valuable information.

Among the main findings, we can highlight the relevance of Donald Trump as a key actor during this period, not only as one of the most repeated terms and as a central figure in the most important topics but also because his account was one of the most active and influential. In this sense, it should be noted that some months later, and due to the hostile behaviour of the former president spreading false or doubtful information and encouraging violent behaviours, especially in relation to the presidential elections that he lost in November 2020, the platform decided to block and delete his profile in January 2021. This helps us understand how a polarizing figure, who was also accused of not implementing adequate control and prevention measures against the virus, could lead the conversation around the pandemic. In this context, it is also important to keep in mind that 3 of the 10 most active accounts no longer exist and that 3 included the word "bot" in their profile names, which indicates that they might have been bots, although that cannot be confirmed. At any rate, this shows that conversations on Twitter might have been led by instability, lack of reliability, and confrontation.

On the other hand, Twitter was also a space for what we can consider a more useful conversation, given that the second most influential account in the sample was that of the World Health Organization, the main institution in charge of informing people about the pandemic and the one offering the most important recommendations at an international level. In fact, information and recommendations seem to have been essential during this uncertain period, in which the Twitter conversation also looked for data and ways to face the pandemic, something that the sentiment analysis, the word frequency, and the topic modelling confirm.

The word frequency analysis revealed that the most frequent words in the sample refer mainly to health-related topics, such as the evolution of the pandemic or the measures taken to fight it, once again indicating that most of the tweets could be informative. Part of the conversation also focuses on US politics, showing that the discussion, originally strictly focused on health elements (Yu et al. 2020; Yunez 2020), was broadening. Nonetheless, it might be surprising that no economic issues—strongly affected by the measures taken to control the virus—seem to be present.

The topic modelling of the sample of tweets confirms the prevalence of informative messages. Of the six main identified topics, three of them refer to general or specific information messages about the coronavirus, its effects, damage, evolution, and ways to control and combat it. Some of the topics also share a strong component of confrontation, criticizing measures or the behaviour of other people, as well as referring to protests or politically charged messages.

Following the postulates of the agenda-setting theory, it can be observed that the main topics strongly relate to observations about the most relevant figures—polarization around US politics and its handling of the pandemic and general recommendations to fight the virus, mainly coming from the WHO.

Finally, regarding the latent sentiment of the tweets, it was found that these sentiments are predominantly neutral, possibly even informative, although there is a slight tendency towards negativity, something that is understandable during a hard moment in which citizens were suffering the effects of the pandemic. Furthermore, although the number of cases were declining during the studied weeks, no relevant trends were observed in the evolution of the sentiment of the conversation.

As a general conclusion, it can be observed that the conversation around COVID-19 during the weeks in which the first wave of the pandemic was receding in most Western countries had two perspectives: first, a rather informative one, with neutral sentiments, led by public institutions or media with information, data, or recommendations about the pandemic; second, a more polarized and political one, with confrontation and complaints due to the mismanagement of institutions or irresponsible citizen behaviours. The most paradigmatic accounts for each type of conversation are the ones of the WHO and Donald Trump, respectively.

One relevant aspect to highlight is the possibility that bots participated in these conversations and transmitted information, which suggests that some of that information might have been fake or manipulated, leading to misinformation or polarization; this matches previous studies that have focused on misinformation or other information disorders during the pandemic (Bridgman et al. 2020; Singh et al. 2020), and it points out the important role that the *infodemic* might have played during this phenomenon.

Another important observation is the predominance of the USA in the public debate. The selected tweets had no country identification, but the conversation clearly focuses on the USA, despite the presence of or allusions to countries such as China, the UK, or Nigeria. Besides its large population and the great penetration of Twitter in the US, its worldwide influence, the discussed management of the pandemic by the Trump administration, the upcoming presidential elections, or the strong impact of the pandemic in this country can help explain this presence.

Finally, it is important to highlight the limitations of this work. Although the paper is extensive and a large sample has been collected and analysed from different perspectives, there are still both temporal and methodological limitations. On the one hand, not all the tweets published on the selected dates have been used; a random sample according to what the Twitter REST API allowed was used, so it would be advisable to use API 2 to access all the messages in future studies. It would also be advisable for future works to analyse a longer period, including the consecutive waves of the pandemic, and compare them with the specific moment studied here, studying this public debate around the coronavirus longitudinally. This will provide a better understanding of how international public opinion has evolved and how this has affected the institutional decisions and different measures taken to fight the pandemic, as well as how these have impacted conversations on Twitter. Of course, in future studies, it would be necessary to collect data from other social media sites, as well as include conversations in more languages, such as Spanish and Italian, spoken in the two Western countries where the virus first arrived, which were two of the countries hardest hit by the pandemic during the first wave. Similarly, given that sentiment analysis and topic modelling by themselves are not entirely adequate in identifying and analysing predominant frames in discourses and the formation and establishment of agendas, it would be also convenient to carry out analyses using other methods, such as network analysis or qualitative techniques, trying to identify ghost accounts that could be participating in these public debates, as well as delving into the different topics and discourses spread through these platforms and their possible effects on society.

Notes

[1] The World Health Organization has repeatedly demanded efforts to counter the 'infodemic' (WHO 2020a).

[2] COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Available at https://github.com/CSSEGISandData/COVID-19 (accessed on 19 March 2023).

[3] The transcription of the press conference can be found at https://www.gov.uk/government/speeches/health-and-social-care-secretarys-statement-on-coronavirus-covid-19-22-june-2020 (accessed on 19 March 2023).

References

Alamoodi, A. H., Mohammed Rashad Baker, O. S. Albahri, B. B. Zaidan, A. A. Zaidan, Wing-Kwong Wong, Salem Garfan, A. S. Albahri, Miguel A. Alonso, Ali Najm Jasim, and et al. 2022. Public Sentiment Analysis and Topic Modeling Regarding COVID-19's Three Waves of Total Lockdown: A Case Study on Movement Control Order in Malaysia. *KSII Transactions on Internet & Information Systems* 16: 2169–90. [CrossRef]

Arcila-Calderón, Carlos, Félix Ortega-Mohedano, Javier J. Amores, and Sofía Trullenque. 2017. Supervised sentiment analysis of political messages in Spanish: Real-time classification of tweets based on machine learning. *Profesional de la Información* 26: 973–82. [CrossRef]

Arcila-Calderón, Carlos, Patricia Sánchez-Holgado, Cristina Quintana-Moreno, Javier J. Amores, and David Blanco-Herrero. 2022. Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation. *Comunicar: Revista científica Iberoamericana de Comunicación y Educación* 30: 21–35. [CrossRef]

Bogović, Petar Kristijan, Ana Meštrović, Slobodan Beliga, and Sanda Martinčić-Ipšić. 2021. Topic Modelling of Croatian News During COVID-19 Pandemic. Paper presented at the 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, May 24–28; pp. 1044–51. [CrossRef]

Bridgman, Aengus, Eric Merkley, Peter John Loewen, Taylor Owen, Derek Ruths, Lisa Teichmann, and Oleg Zhilin. 2020. The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. *Harvard Kennedy School Misinformation Review* 1. [CrossRef]

Collobert, Ronan, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research* 12: 2493–537.

Fu, King-Wa, Hai Liang, Nitin Saroha, Zion Tsz Ho Tse, Patrick Ip, and Isaac Chung-Hai Fung. 2016. How people react to Zika virus outbreaks on Twitter? A computational content analysis. *American Journal of Infection Control* 44: 1700–2. [CrossRef]

Goffman, Erving. 1974. *Frame Analysis: An Essay on the Organization of Experience*. Harvard: Harvard University Press.

Gomide, Janaína, Adriano Veloso, Wagner Meira Jr., Virgílio Almeida, Fabrício Benevenuto, Fernanda Ferraz, and Mauro Teixeira. 2011. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. Paper presented at the 3rd International Web Science Conference, Koblenz, Germany, June 15–17; New York: ACM, pp. 1–8. [CrossRef]

Gourisaria, Mahendra Kumar, Satish Chandra, Himansu Das, Sudhansu Sheckhar Patra, Manoj Sahni, Ernesto Leon-Castro, Vijander Singh, and Sandeep Kumar. 2022. Semantic Analysis and Topic Modelling of Web-Scrapped COVID-19 Tweet Corpora through Data Mining Methodologies. *Healthcare* 10: 881. [CrossRef]

Gozgor, Giray. 2022. Global Evidence on the Determinants of Public Trust in Governments during the COVID-19. *Applied Research in Quality of Life* 17: 559–78. [CrossRef] [PubMed]

Grimmer, Justin, and Brandon M. Stewart. 2017. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21: 267–97. [CrossRef]

Guidry, Jeanine P. D., Yan Jin, Caroline A. Orr, Marcus Messner, and Shana Meganck. 2017. Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. *Public Relations Review* 43: 477–86. [CrossRef]

Hermida, Alfred. 2010. Twittering the news: The emergence of ambient journalism. *Journalism Practice* 4: 297–308. [CrossRef]

Jacobi, Carina, Wouter Van Atteveldt, and Kasper Welbers. 2016. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism* 4: 89–106. [CrossRef]

Jahanbin, Kia, and Vahid Rahmanian. 2020. Using Twitter and web news mining to predict COVID-19 outbreak. *Asian Pacific Journal of Tropical Medicine* 13: 378–80. [CrossRef]

Karami, Amir, Morgan Lundy, Frank Webb, and Yogesh K. Dwivedi. 2020. Twitter and Research: A Systematic Literature Review Through Text Mining. *IEEE Access* 8: 67698–717. [CrossRef]

Latorre, Juan Pablo, and Javier J. Amores. 2021. Topic modelling of racist and xenophobic YouTube comments. Analyzing hate speech against migrants and refugees spread through YouTube in Spanish. Paper presented at the Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21), Barcelona, Spain, October 26–29; pp. 456–60. [CrossRef]

Lazard, Allison J., Emily Scheinfeld, Jay M. Bernhardt, Gary B. Wilcox, and Melissa Suran. 2015. Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *American Journal of Infection Control* 43: 1109–11. [CrossRef]

Lee, Jayeon, and Weiai Xu. 2018. The more attacks, the more retweets: Trump's and Clinton's agenda setting on Twitter. *Public Relations Review* 44: 201–13. [CrossRef]

Lim, Vanessa W., Rachel L. Lim, Yi Roe Tan, Alexius S. E. Soh, Mei Xuan Tan, Norhudah Bte Othman, Sue Borame Dickens, Tun-Linn Thein, May O. Lwin, Rick Twee-Hee Ong, and et al. 2020. Government trust, perceptions of COVID-19 and behaviour change: Cohort surveys, Singapore. *Bulletin of the World Health Organization* 99: 92–101. [CrossRef]

López-Meri, Amparo. 2015. El impacto de Twitter en el periodismo; un estado de la cuestión. *Revista de la Asociación Española de Investigación de la Comunicación* 2: 34–41. [CrossRef]

Lwin, May Oo, Jiahui Lu, Anita Sheldenkar, Peter Johannes Schulz, Wonsun Shin, Raj Gupta, and Yinping Yang. 2020. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. *JMIR Public Health and Surveillance* 6: e19447. [CrossRef] [PubMed]

Mathayomchan, Boonyanit, Viriya Taecharungroj, and Walanchalee Wattanacharoensil. 2022. Evolution of COVID-19 tweets about Southeast Asian Countries: Topic modelling and sentiment analyses. *Place Branding and Public Diplomacy*, 1–18. [CrossRef]

McCombs, Maxwell E., and Donald L. Shaw. 1972. The agenda-setting function of mass media. *Public Opinion Quarterly* 36: 176–87. [CrossRef]

Miller, Michelle, Tanvi Banerjee, Roopteja Muppalla, William Romine, and Amit Sheth. 2017. What Are People Tweeting About Zika? An Exploratory Study Concerning Its Symptoms, Treatment, Transmission, and Prevention. *JMIR Public Health and Surveillance* 3: e38. [CrossRef] [PubMed]

Mutanga, Murimo Bethel, and Abdultaofeek Abayomi. 2022. Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. *African Journal of Science, Technology, Innovation and Development* 14: 163–72. [CrossRef]

Paul, Michael J., Mark Dredze, and David Broniatowski. 2014. Twitter Improves Influenza Forecasting. *PLoS Currents Outbreaks* 6. [CrossRef] [PubMed]

Prabhakar Kaila, Rajesh, and Krishna Prasad. 2020. Informational Flow on Twitter—Corona Virus Outbreak—Topic Modelling Approach. *International Journal of Advanced Research in Engineering and Technology* 11: 128–34.

Pruss, Dasha, Yoshinari Fujinuma, Ashlynn R. Daughton, Michael J. Paul, Brad Arnot, Danielle Albers Szafir, and Jordan Boyd-Graber. 2019. Zika discourse in the Americas: A multilingual topic analysis of Twitter. *PLoS ONE* 14: e0216922. [CrossRef]

Ramage, Daniel, David Hall, Ramesh Nallapati, and Christopher D. Manning. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. Paper presented at the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, August 6–7; pp. 248–56.

Singh, Lisa, Shweta Bansal, Leticia Bode, Ceren Budak, Guangqing Chi, Kornraphop Kawintiranon, Colton Padden, Rebecca Vanarsdall, Emily Vraga, and Yanchen Wang. 2020. A first look at COVID-19 information and misinformation sharing on Twitter. *arXiv* arXiv:2003.13907.

Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2011. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 62: 419. [CrossRef]

Vicinitas. 2018. 2018 Research on 100 Million Tweets: What It Means for Your Social Media Strategy for Twitter. Available online: https://www.vicinitas.io/blog/twitter-social-media-strategy-2018-research-100-million-tweets (accessed on 19 March 2021).

WHO (World Health Organization). 2020a. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19—8 April 2020. Available online: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--8-april-2020 (accessed on 19 March 2023).

WHO (World Health Organization). 2020b. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19—11 March 2020. Available online: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (accessed on 19 March 2023).

Wicke, Philipp, and Marianna M. Bolognesi. 2020. Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. *PLoS ONE* 15: e0240010. [CrossRef] [PubMed]

Wright, Liam, Elise Paul, Andrew Steptoe, and Daisy Fancourt. 2022. Facilitators and barriers to compliance with COVID-19 guidelines: A structural topic modelling analysis of free-text data from 17,500 UK adults. *BMC Public Health* 22: 34. [CrossRef]

Xue, Jia, Junxiang Chen, Chen Chen, Chengda Zheng, Sijia Li, and Tingshao Zhu. 2020. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLoS ONE* 15: e0239441. [CrossRef] [PubMed]

Young, Dannagal G. 2020. The Lincoln Project and the Conservative Aesthetic. *Society* 57: 562–68. [CrossRef]

Yu, Jingyuan, Yanqin Lu, and Juan Muñoz-Justicia. 2020. Analizing Spanish News Frames on Twitter during COVID-19—A Network Study of El País and El Mundo. *International Journal of Environmetal Research and Public Health* 17: 5414. [CrossRef]

Yunez, Julián. 2020. Twitter presidencial y el falso dilema entre salud y economía. *La Revista de ACOP* 56: 34–35. Available online: https://compolitica.com/wp-content/uploads/2021/01/N56_Eta2_La_revista_de_ACOP_Enero2021_E.pdf (accessed on 19 March 2023).

Yurtsever, Muhammet Mücahit Enes, Muhammad Shiraz, Ekin Ekinci, and Süleyman Eken. 2023. Comparing COVID-19 vaccine passports attitudes across countries by analysing Reddit comments. *Journal of Information Science*, 01655515221148356. [CrossRef]