Journal of Computational Social Science (2022) 5:687–729
https://doi.org/10.1007/s42001-021-00150-8

RESEARCH ARTICLE

A high-resolution temporal and geospatial content analysis of Twitter posts related to the COVID-19 pandemic

Charalampos Ntompras[1] · **George Drosatos**[2] · **Eleni Kaldoudi**[1,3]

Received: 8 February 2021 / Accepted: 6 October 2021 / Published online: 20 October 2021

Abstract

The COVID-19 pandemic has deeply impacted all aspects of social, professional, and financial life, with concerns and responses being readily published in online social media worldwide. This study employs probabilistic text mining techniques for a large-scale, high-resolution, temporal, and geospatial content analysis of Twitter related discussions. Analysis considered 20,230,833 English language original COVID-19-related tweets with global origin retrieved between January 25, 2020 and April 30, 2020. Fine grain topic analysis identified 91 meaningful topics. Most of the topics showed a temporal evolution with local maxima, underlining the short-lived character of discussions in Twitter. When compared to real-world events, temporal popularity curves showed a good correlation with and quick response to real-world triggers. Geospatial analysis of topics showed that approximately 30% of original English language tweets were contributed by USA-based users, while overall more than 60% of the English language tweets were contributed by users from countries with an official language other than English. High-resolution temporal and geospatial analysis of Twitter content shows potential for political, economic, and social monitoring on a global and national level.

Keywords Twitter · COVID-19 · Social media analysis · Topic modeling · Latent Dirichlet Allocation · Geospatial analysis

✉ George Drosatos
gdrosato@athenarc.gr

Charalampos Ntompras
cntompra@med.duth.gr

Eleni Kaldoudi
kaldoudi@med.duth.gr

1 School of Medicine, Democritus University of Thrace, Alexandroupoli, Greece

2 Institute for Language and Speech Processing, Athena Research Center, Xanthi, Greece

3 European Alliance for Medical and Biological Engineering and Science, Brussels, Belgium

Introduction

Social media, and especially microblogging, provide the means to explore how information, perceptions, and feelings diffuse during health crises and leverage this evidence to better manage evolving health crises, or even discover and apply a variety of health, social, and economic indicators and predictors. Twitter, a multilingual microblogging platform with more than 300 million users worldwide [1] was the most often studied social media platform for assessment of public's interest during past infectious disease outbreaks [2, 3], primarily to conduct content analysis, surveillance, and engagement [4].

During the current COVID-19 pandemic, Twitter has been explored in many different ways for related analyses. Some approaches exploited the geolocation metadata available in Twitter to derive a social mobility index [5] or predict risk of pandemic for a region [6]. Other studies investigated the content of limited number of tweets published by specific official health organizations [7]; of specific social media movements (e.g., the Free Open Access Medical Education twitter movement #FOAMed) [8]; and of specific user groups, for example, people with arthritis [9], electronic cigarette users [10], and people interested in plastic and esthetic surgery [11].

Several studies have employed text mining and machine learning techniques to assess large volumes of tweets for topic and sentiment analysis. Examples include studies that use conventional text mining techniques, e.g., descriptive text mining to identify primary topics of discussion in a set of 7301 global tweets [12], non-negative matrix factorization on COVID-19-related hate tweets in Arabic language [13], and a study that used transfer learning classifiers to analyze 60 million tweets originating from the USA to identify psychological effects of the pandemic [14].

Given the immense volume of Twitter data, most large-scale studies employed autonomous machine learning techniques, and especially the Latent Dirichlet Allocation (LDA) [15], probably the most popular statistical model for unsupervised discovering of abstract topics in large corpora. Notable examples are presented below.

A study (195,201 tweets with global origin) identified 12 topics and compared topics and sentiments in tweets posted by humans versus topics and sentiments in discussions posted by social bots [16]. Another study [17] used a daily sampling of the most popular tweets about COVID-19 during the early phase of the pandemic (December 2019–March 2020) to create a representative tweet dataset (to filter out outliers and low activity); sentiment analysis on 6 topics revealed a negative outlook toward COVID-19. Similarly, LDA was used to identify topics using daily COVID-19-related tweet corpuses originated in the USA (86,581,237 original English language tweets with USA origin) [18] and worldwide on 4,196,020 original English language tweets with global origin [19] and on another occasion on 1,963,285 original English language tweets with global origin [20], with a rather limited topical granularity of only 8, 13, and 11 topics, respectively. These studies focused mainly on sentiment analysis and reported negative effects on the overall USA population sentiment, while highlighting fear and anger for the global tweet studies. These results were confirmed by another study (126,049 original English language tweets with global origin) [21] which identified 10 salient topics at the early days of the pandemic outbreak and reported on fear as well as surprise. Another study on 13,937,906 original English language tweets posted by individuals only (organizations excluded) [22] identified 26 topics and confirmed negative sentiments for the topics of spread and growth of cases, symptoms, racism, source of the outbreak, and political impact; in contrast, the study reported a reversal of sentiments toward positive for topics related to prevention, impact on the economy and markets, government response, impact on the health care industry, and treatment and recovery. Mixed sentiments were also reported by another study on 167,073 original English language tweets with global origin, based on 12 topics; only 2 of them (deaths caused by COVID-19 and increased racism) exhibited a mean negative sentiment [23].

The above-mentioned studies used a rather limited granularity for topics (ranging from 6 to 26 topics, mean value 12.3 ± 6.5) with primary objective to assess sentiments across topics. Temporal differences in Twitter discussions before and after the official declaration of COVID-19 as a pandemic were explored on a set of randomly sampled tweets representing the 5% of the relevant Twitter activity for a 2-week period around the declaration of the pandemic (940,837 original English language tweets with global origin) [24] and results based on 9 topics indicate that concerns of the public vary as the pandemic progresses.

The objective of our study is to perform a high-resolution, temporal, and spatial analysis of discussion topics in a large-scale COVID-19-related Twitter data set. Primary research questions include:

1. What are salient topics of discussion and concern in English language tweets in fine granularity during the first period of the pandemic?
2. What is the temporal evolution of these topics and how temporal evolution correlates to real-world events?
3. What is the geospatial variance of discussed topics?

Secondary research questions include:

4. Can global English language tweets analysis identify local (country-specific) topics of discussion?
5. Can Twitter topics analysis lead to indications or metrics of economic or societal value?

Materials and methods

Twitter data collection

The open source Social Feed Manager (SFM) [25] tool was used to COVID-19-related tweets harvest from January 25 to April 30, 2020. Data collection complied with Twitter's Terms of Service and Developers Agreement and Policy and was performed using the Twitter filter streaming application programming interface (API) and the following search keywords:

> *"coronavirus, #coronavirus, sars virus, #SARSvirus, #SARS2020, #SARS2, sars-cov, sars cov, SarsCov, #SarsCov, severe acute respiratory coronavirus, severe acute respiratory syndrome, #WuhanCoronavirus, #WuhanSARS, Wuhan Coronavirus, Wuhan SARS, 2019-nCoV, 2019 nCoV, #2019nCoV, 2019nCoV, COVID-19, #COVID19, COVID19"*

This process retrieves invariably original tweets, retweets, quotes, and reply tweets generated globally and in any available language. For the purposes of this paper, further processing was employed to create a subset containing only original tweets written in English language (based on the values of the respective metadata fields of the originally retrieved dataset).

The duration of data collection was chosen based on landmarks provided by WHO COVID-19 timeline [26]. Specifically, start of collection was set on February 25, 2020, when the WHO Regional Director for Europe issued a public statement outlining the importance of being ready at the local and national levels for detecting cases, testing samples and clinical management, thus indicating the potential for a worldwide pandemic. Data were collected for 15 weeks till April 30, 2020, when the Director-General convened the WHO International Health Regulations Emergency Committee on COVID-19 convened for a third time to assess the pandemic and declare that the outbreak continued to constitute a public health emergency of international concern.

Geographical origin of tweets

The country of origin for each tweet was identified as the location of the user generating the tweet. In the retrieved dataset, this value is a free text metadata field that normally contains the name of a city and a country. However, this metadata field is often empty, or semi-completed, containing only city or country name, written in different languages and spellings and quite often containing typos.

Identification of country of origin was performed by basic natural language processing of the User Location metadata field. Free text was processed to identify potential place names, which were semantically annotated using the freely available GeoNames (www.geonames.org) geographical database, which contains over 11 million place names spanning all countries and written in various languages. Final geographical origin was identified via a heuristic scoring approach that considered the identification of city names with population bigger than 500 people and identification of country names to compute an overall probability for the country of origin for each twitter user contributing to the dataset. The pseudo-code of the heuristic scoring algorithm is included in "Appendix A".

Topic modeling

Topic modeling algorithms are probabilistic methods that automatically identify topics from a large and unstructured collection of documents. In this work, we used the LDA algorithm [15, 27] and specifically the scalable implementation in the MALLET toolkit (v2.0.8) [28], as proposed in [29]. To avoid any noise in the topic modeling from the free text of tweets, we applied the following cleaning process: (1) removed all URLs and usernames mentioned (i.e., @usernames); (2) removed all punctuation, special symbols and non-Latin characters; (3) converted all lowercase words to their lemmas by applying the Krovetz stemming procedure [30]; (4) excluded all stop words using the list in the Text Categorization Project [31]; (5) to avoid bias in topics, we excluded the keywords used in the search query (i.e., *"coronaviru, sar, sarsviru, sars2020, sars2, sars-cov, sarscov, wuhancoronaviru, wuhansar, 2019-ncov, ncov, 2019ncov, COVID-19"*); and (6) excluded tweets with less than two words.

When applying LDA analysis, the appropriate number of topics is a user-specified parameter without a subjectively appropriate answer. A common approach is to perform a number of LDA experiments for different values for topics and then choose the most suitable number of topics based either on a topic coherence metric [29, 32] or manual inspection for the most meaningful set of topics. Given our objective to perform high-resolution topic analysis (i.e., large number of topics) in a large-scale data set, the immense processing load required for coherence analysis proved impractical. Thus, to determine the appropriate number of topics, we performed the LDA experiment for 50, 100, and 150 topics at 10,000 iterations each to identify the most suitable number of topics based on manual topic inspection. The artificially generated topics, which consist of a weighted list of words, were screened manually by the authors, were labeled by a short descriptive title, and organized in conceptual categories.

Topics popularity evolution over time

Temporal evolution of topics popularity was calculated on weekly basis following the approach proposed in [29]. First, the weight of each topic for each tweet was calculated as the percentage of the tweet words belonging to a topic. The popularity of the topic was defined as the weekly topic contribution estimate $P(t, y)$ of the topic (t) for each week (y), and it was calculated as the average weight of this topic for all tweets published that week of the year D_y

$$P(t,y) = \frac{1}{\left|D_y\right|} \sum_{d \in D_y} \frac{\left|\{w \in d : \text{topic}(w) = t\}\right|}{|d|},$$

where t represents a topic and w is a word in tweet d of the tweets' collection D_y for week y. Accordingly, the overall popularity of a topic was defined as the overall topic contribution estimate, calculated as the average weight of this topic for all tweets included in the corpus.

Topics geographical distribution

Geospatial distribution of a topic was calculated following an approach similar to the approach used for popularity evolution over time. The weight of each topic was calculated as the percentage of the tweet words belonging to a topic. The popularity of the topic for each country of origin was defined as the topic contribution estimate $P(t, c)$ of the topic (t) for each country (c), and it was calculated as the average weight of this topic for all tweets D_c published by user accounts of the particular country (c)

$$P(t,c) = \frac{1}{|D_c|} \sum_{d \in D_c} \frac{|\{w \in d : \text{topic}(w) = t\}|}{|d|},$$

where t represents a topic and w is a word in tweet d of the tweets' collection D_c of the tweets by users of a particular country c. Accordingly, the overall popularity of a topic was defined as the overall topic contribution estimate, calculated as the average weight of this topic for all tweets included in the corpus of geographically annotated tweets.

Results

Twitter dataset

From January 25 to April 30, 2020, we retrieved 316,988,440 COVID-19-related tweets, produced by 33,488,183 unique Twitter accounts. Preprocessing to account for original tweets and English language resulted in a subset of 20,614,490 original tweets written in English and produced by 4,834,467 unique accounts. Figure 1 shows the distribution of English tweets per day, while the profile of the retrieved dataset is summarized in Table 1.

For the purposes of this paper, we considered only original tweets written in English language (20,614,490). Further preprocessing excluded 383,657 tweets with less than two words (1.86% of all retrieved original English tweets). The final corpus

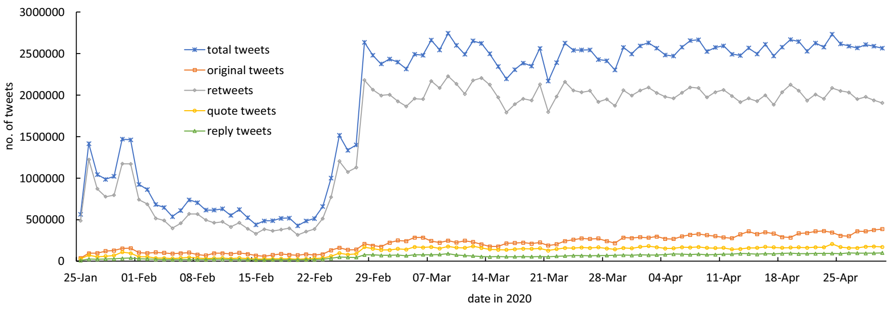


Fig. 1 Number of COVID-19-related English tweets per day from January 25 to April 30, 2020

Table 1 The profile of the COVID-19-related tweet dataset retrieved for this study, from January 25 to April 30, 2020

Twitter data	# In all languages	# In English language
Tweets		
Original tweets	36,064,666	20,614,490
Retweets	247,882,547	147,969,146
Reply tweets	10,076,553	5,668,059
Quote tweets	22,964,667	11,678,287
Total tweets	316,988,440	185,929,982
Twitter accounts		
Accounts of original tweets	8,407,815	4,834,467
Accounts of retweets	26,839,558	16,724,425
Accounts of reply tweets	3,368,506	1,835,153
Accounts of quote tweets	7,361,248	3,815,053
Total twitter accounts	33,488,183	20,483,486

consisted of 20,230,833 tweets, corresponding to a total of 197,728,410 words and a vocabulary of 2,139,369 unique words.

Topic modeling

LDA parametrization experiments resulted in a value of 100 topics as a suitable initialization parameter. Screening of the topics analysis results identified 91 meaningful topics which were organized into nine categories as follows:

1. *Life during the pandemic*: 18 topics corresponding to how twitter users went through the pandemic. Examples include expression of sentiments (e.g., anger and fear, and fear of dying). Also includes 'USA protests', 'USA primary elections (Wisconsin)', and others related to art in quarantine (e.g., 'movies and video games in quarantine' and 'musical bands and groups').
2. *Pandemic management*: 16 topics related to pandemic issues and how to manage them (e.g., 'relief bills in USA', 'US White House task force', 'donations and relief funds', etc.).
3. *Medical*: 12 topics discussing concepts related to medical issues (e.g., 'medical equipment and supplies', 'world health emergency declaration', etc.) as well as for the 'vaccine development' and the 'virus origin'.
4. *Outbreak*: 13 topics addressing different cases of outbreak (e.g., cases in CPAC 2020 conference, China and Wuhan outbreak, Diamond Princess cruise ship outbreak, Arabic countries outbreak etc.).
5. *Lockdown*: 9 topics pertaining to specific lockdowns worldwide (e.g., 'Nigeria lockdown', 'India lockdown', 'European countries and Japan lockdown', etc.), but also, in events that were canceled or postponed (e.g., football, basketball, etc.).

6. *Economy*: 10 topics that discuss the impact of COVID-19 on the economy. Examples include topics such as ’lockdown and economy restart’, ’cancellation fees and refunds’, and ’bitcoin and cryptocurrencies’. This category also includes topics related to the impact on economy (e.g., ‘impact on supply chain due to China lockdown’ and ‘impact on business and companies’).
7. *Cases and deaths*: 5 topics discussing about the number of cases and deaths caused by COVID-19, such as ’live data maps’,’ confirmed deaths and recoveries’, and ’death toll rising (China and Italy)’.
8. *News and Fake News*: 5 topics related to the ‘5G conspiracy theory’, the ‘misinformation spread in social media’, and the fact that ‘US President claims disinfectants can sure’.
9. *Preventive measures*: 3 topics addressing the ‘facemasks’, the ‘social distancing’, and the ‘hand washing’.

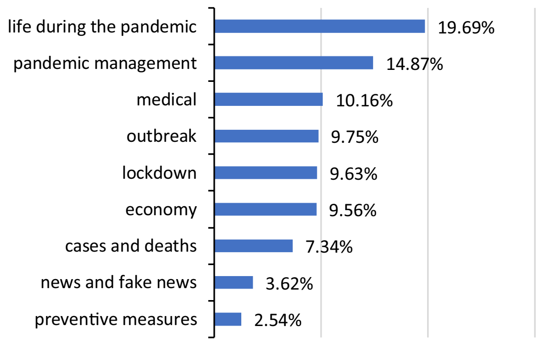
The relative popularity of each topic category (in descending order) is shown in Fig. 2. The entire list of topics is presented in“ Appendix B” along with the percentage of the overall topic contribution used to calculate the topic popularity rank and the top 15 significant words defining each topic.

The ten most popular topics for the over the entire time span of the study are presented as word clouds in Fig. 3. The first and third most popular topics are ‘expression of extreme sentiment (anger & fear)’ (3.32%) and ‘expression of extreme sentiment (fear of dying)’ (2.69%). Second most popular is the topic ‘USA President response’ (2.82%) and in the tenth position is the topic ‘information and guidelines updates’ (1.46%). There are 2 topics related to *cases and deaths* (‘death toll rising (China and Italy)’ and ‘number of cases and deaths’), another 2 topics from the *life during the pandemic* category (‘reading and writing in quarantine’ and ‘quarantine time eating and activities’) and the topics ‘fears for impact on stock market’ and ‘ban of flights to/from China’ from the *economy* and *lockdown* category, respectively.

Topics popularity evolution

Linear regression analysis showed a significant linear fit ($R^2 > 80\%$ and p value ≤ 0.05) for 15 topics; 13 topics showed a positive trend and 2 topics showed

Fig. 2 Relative overall popularity of the topics categories



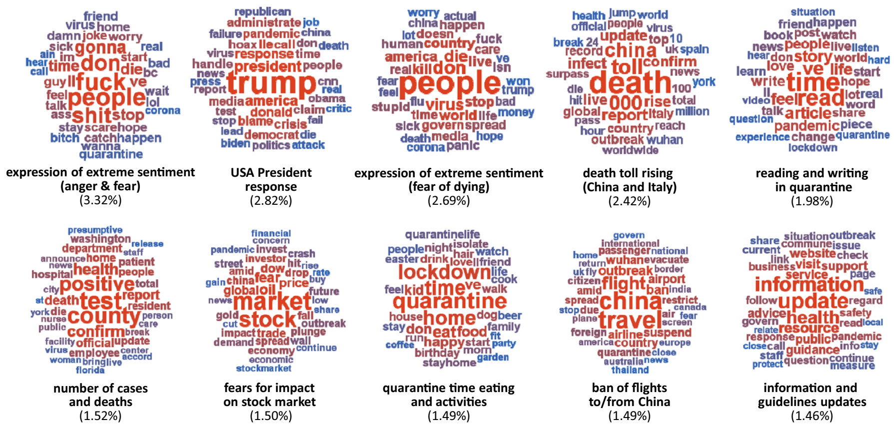


Fig. 3 The top eight popular topics for the entire time span

negative trend. The top three topics with the higher positive trend are: (a) 'impact on business and companies', (b) 'support and donations', and (c) 'mental health during the pandemic'. The two topics with the negative trend are: (a) 'disease spreading' and (b) 'live data maps'. Topics with significant linear fit are appropriately marked on Table 2 in "Appendix B".

The following diagrams present plots of topics' popularity per week for all topics for the 15-week span covered by this study, organized per category.

Figure 4 shows topics related to *Life during the Pandemic* category. The most popular topics 'expression of extreme sentiment (anger & fear)' and 'expression of extreme sentiment (fear of dying)' show an initial decreasing trend, followed by a clear peak on 11th week (Fig. 4a). Additionally, the topic 'stay home safe' peaks at 13th week (Fig. 4b). Most of the least popular topics (Fig. 4c) show brief peaks; exemptions are the increasing trend of the topic related to 'art in quarantine' and the decreasing trend of the topic "musical bands and groups".

Weekly variation of popularity for topics in the *Pandemic Management* category is shown in Fig. 5. Most popular topics (Fig. 5a) of this category are related to specific events or announcements and show peaks in specific weeks. One the other hand, topics related to various pandemic management strategies and instruments (Fig. 5b,c) show a linear fit with increasing trend.

Weekly variation of popularity for topics in the *Medical* category are shown in Fig. 6. The topic on 'world health emergency declaration' shows repeated consecutive peaks of a decreasing height for the duration of the study (Fig. 6a). A number of topics related to vaccines, treatment, and medical tools and supplies (Fig. 6b) present a variability in popularity for the time span of the study, with an overall increasing trend.

Figure 7 shows topics related to *Outbreak* category. The most popular topic is about 'China and Wuhan outbreak' with a strong prevalence up to the 8th week, while the topic 'Diamond Princess cruise ship outbreak' predominates during 6th to 8th week (Fig. 7a). Among the less popular topics (Fig. 7b), the ones related to

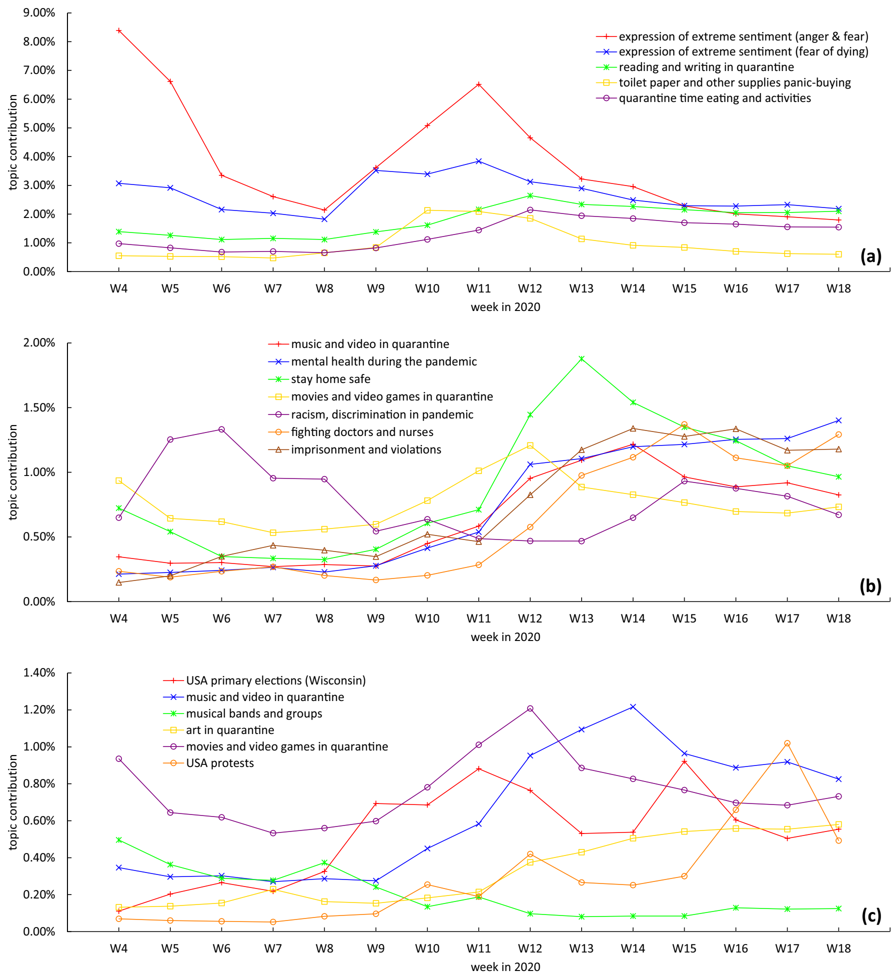


Fig. 4 Weekly variation of popularity for topics related to Life during the Pandemic category: **a** Most popular topics, with a peak of more than 2% for at least 1 week; **b** and **c** less popular topics

specific events, e.g., the topic 'cases in CPAC 2020 conference' and 'UK Prime Minister infection' show a single clear peak on 11th week and 15th week, respectively. County or region-specific outbreaks follow trend lines with recurrent peaks.

The topics related to *Lockdown* category are shown in Fig. 8. The topic 'ban of flights to/from China' shows a high peak on the 5th week, followed by a quick decline. Topics related to schools lockdown and events cancelation appear to be of equivalent popularity (Fig. 8a) and overall of more importance than the country or region-specific lockdown shown in Fig. 8b.

Topics related to *Economy* category are shown in Fig. 9. Topics about 'impact on supply chain (due to China lockdown)' and 'fears for impact in stock market'

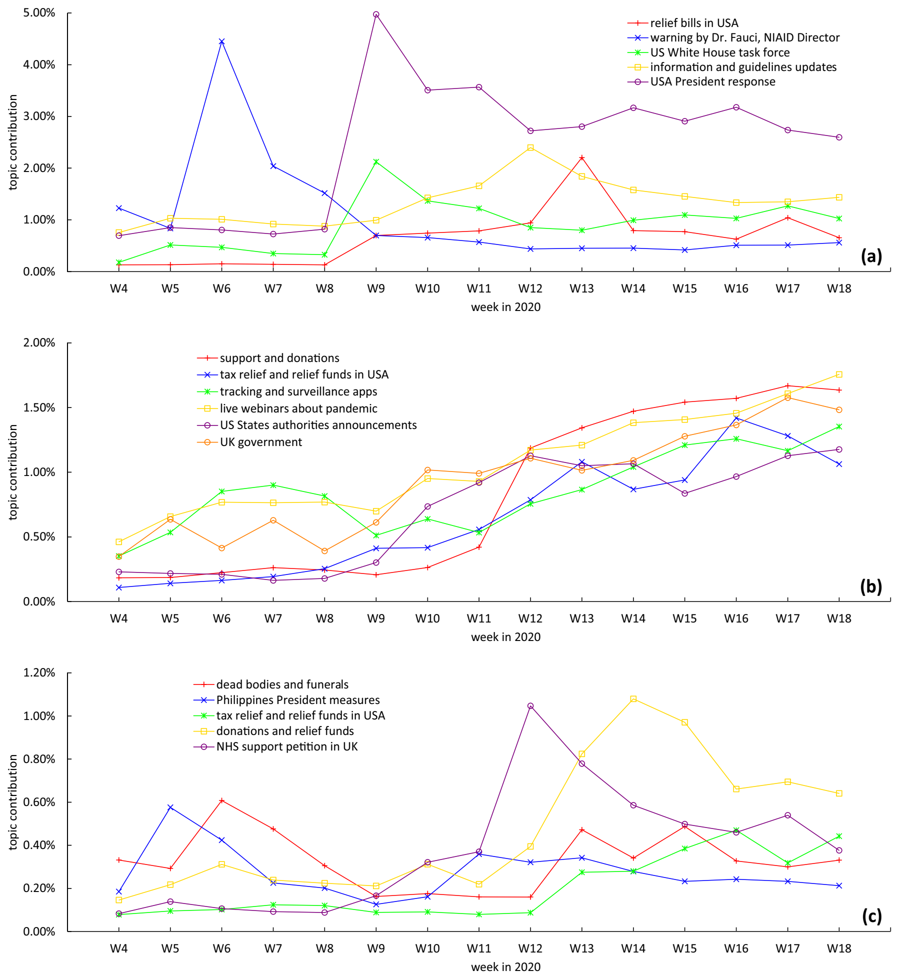


Fig. 5 Weekly variation of popularity for topics related to Pandemic Management category. **a** Most popular topics, with a peak of more than 2% for at least 1 week; **b** and **c** less popular topics

peak on 8th and 9th week, respectively (Fig. 9a). The topic 'lockdown and economy restart', albeit of a low popularity, shows a clearly increasing trend (Fig. 9b).

Figure 10 shows topics of the *Cases and Deaths* category. The topic 'number of cases and deaths' shows a peak on 7$^{\text{th}}$ week and the topic 'death toll rising (China and Italy)' shows a peak on 10$^{\text{th}}$ week (Fig. 10a). Only the topic on 'confirmed deaths and recoveries' shows an increasing trend (Fig. 10b).

Topics related on *News and Fake News* category are shown in Fig. 11. The topics on 'misinformation spread in social media' and 'news channels updates' are the most popular of this category show a slowly decreasing trend, while the topic on 'US President claims disinfectants can cure' shows a sharp peak on 17$^{\text{th}}$ week.

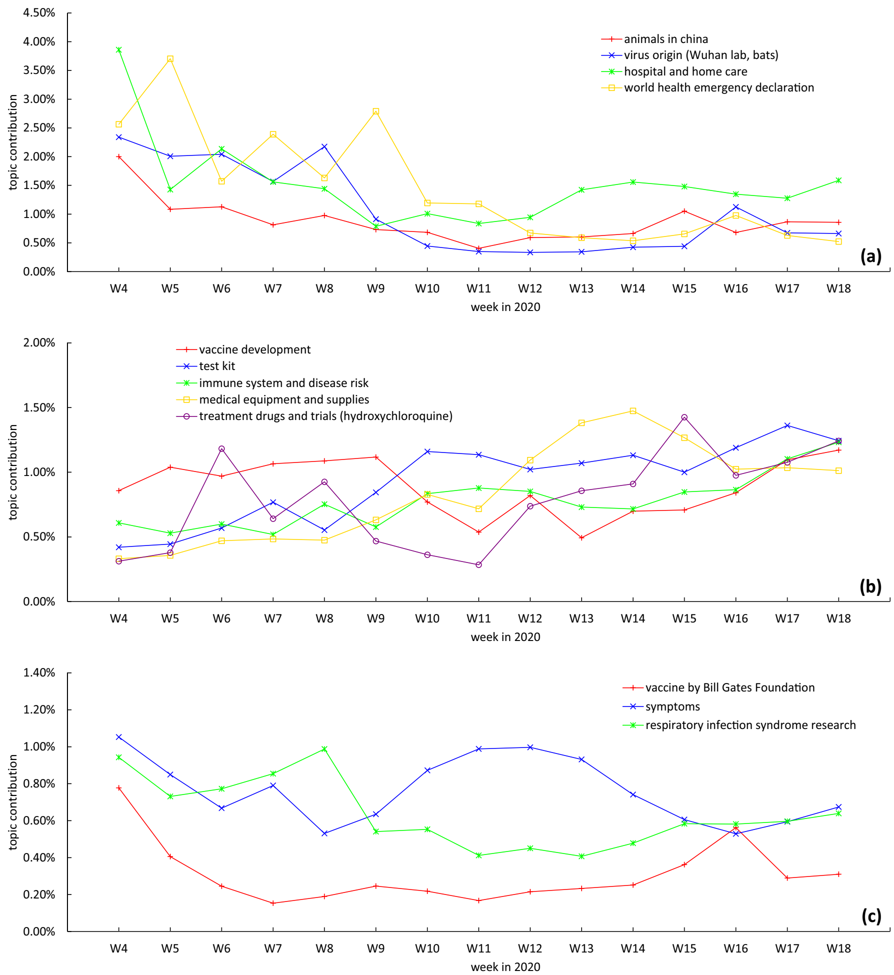


Fig. 6 Weekly variation of popularity for topics related to Medical category: **a** most popular topics, with a peak of more than 2% for at least 1 week; **b** and **c** less popular topics

Figure 12 shows topics related to *Preventive Measures* category. The only topic that reaches even for a week the 2% of the topic contribution is 'hand washing' showing a peak on 10th week, while the 'facemasks' and the 'social distancing' topics remain of comparatively low popularity.

Geographical distribution

Processing for geographic origin showed that tweets in the final corpus originated from 248 countries, while country of origin could not be determined for 6,822,526 tweets (33.7% of the tweets included in the final corpus). Figure 13 shows the tweet contribution of the 20 most contributing countries.

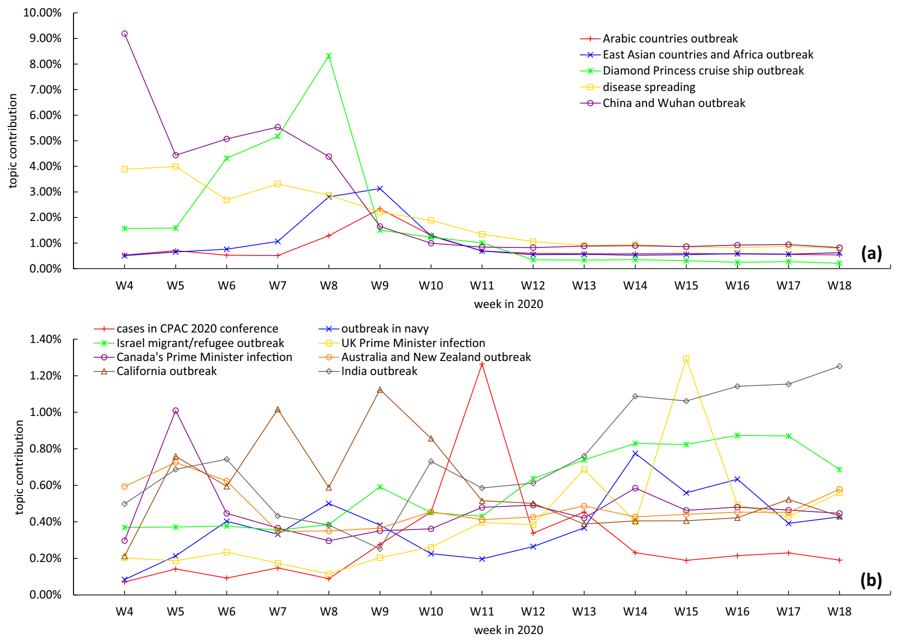


Fig. 7 Weekly variation of popularity for topics related to Outbreak category. **a** Most popular topics, with a peak of more than 2% for at least 1 week; and **b** less popular topics

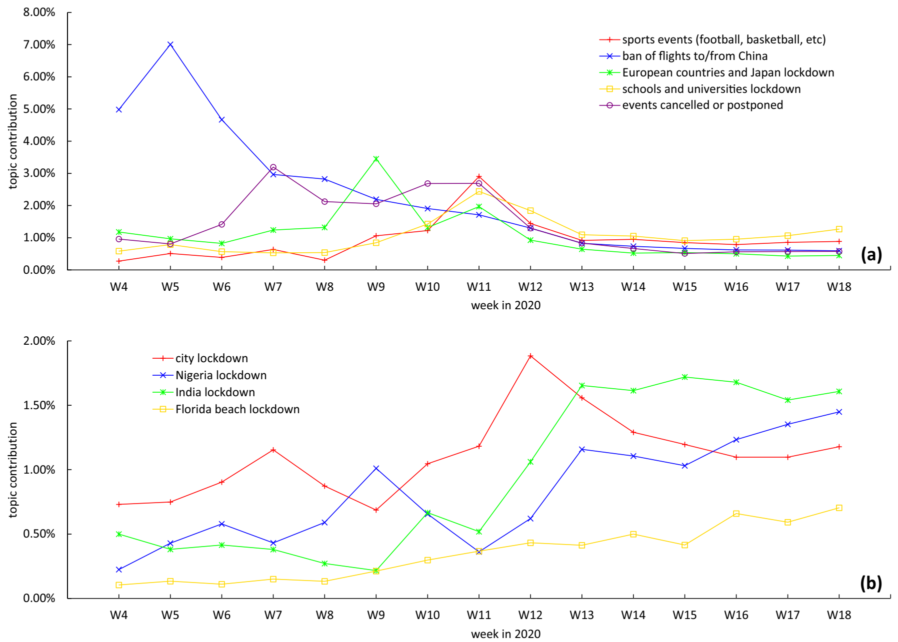


Fig. 8 Weekly variation of popularity for topics related to Lockdown category: **a** most popular topics, with a peak of more than 2% for at least 1 week; and **b** less popular topics

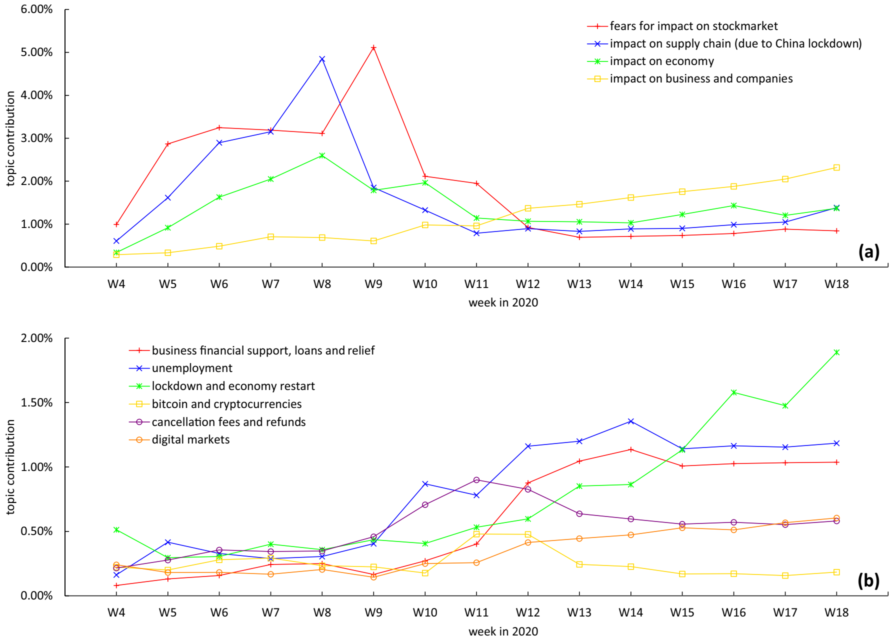


Fig. 9 Weekly variation of popularity for topics related to Economy category: **a** most popular topics, with a peak of more than 2% for at least 1 week, and **b** less popular topics

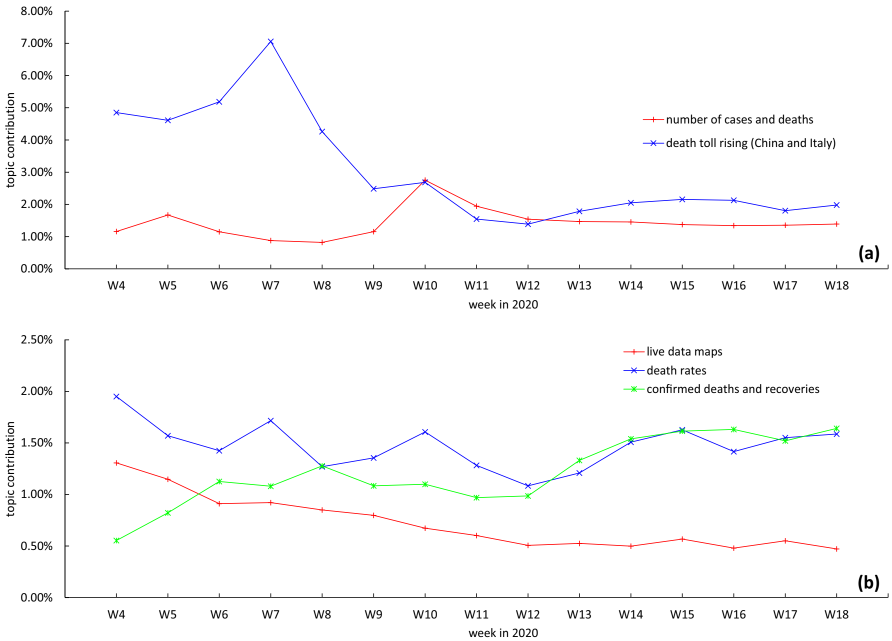


Fig. 10 Weekly variation of popularity for topics related to Cases and Deaths category: **a** most popular topics, with a peak of more than 2% for at least 1 week, and **b** less popular topics

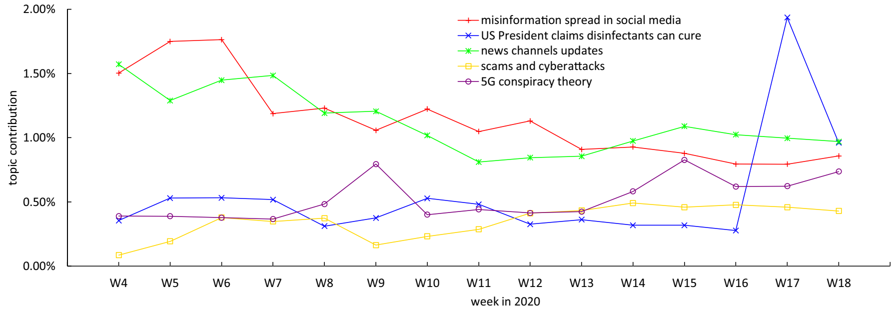


Fig. 11 Weekly variation of popularity for topics related to News and Fake News category

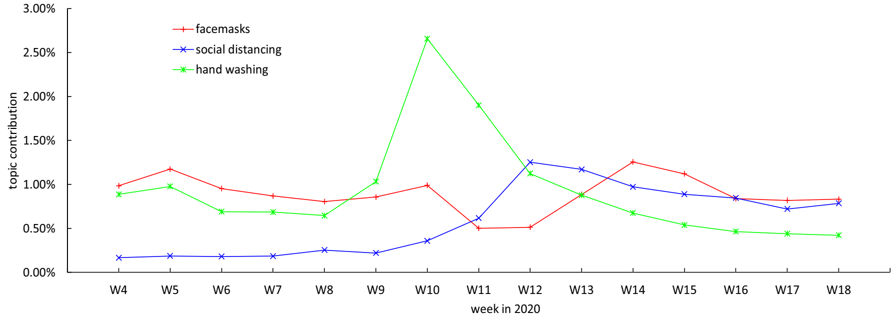


Fig. 12 Weekly variation of popularity for topics related to Preventive Measures category

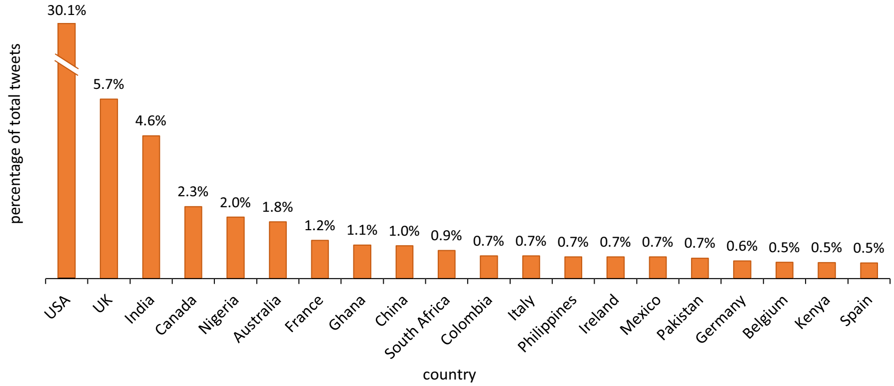


Fig. 13 Geographical origin of the tweets in the final corpus for the top 20 contributing countries

The 10 top most popular topics for the four countries with the highest number of tweets in the corpus are shown in Fig. 14. The most popular topic in each of these countries is different and, as expected, of regional interest. Topics related to expression of extreme sentiments, either anger and fear or of fear of dying, appear in the

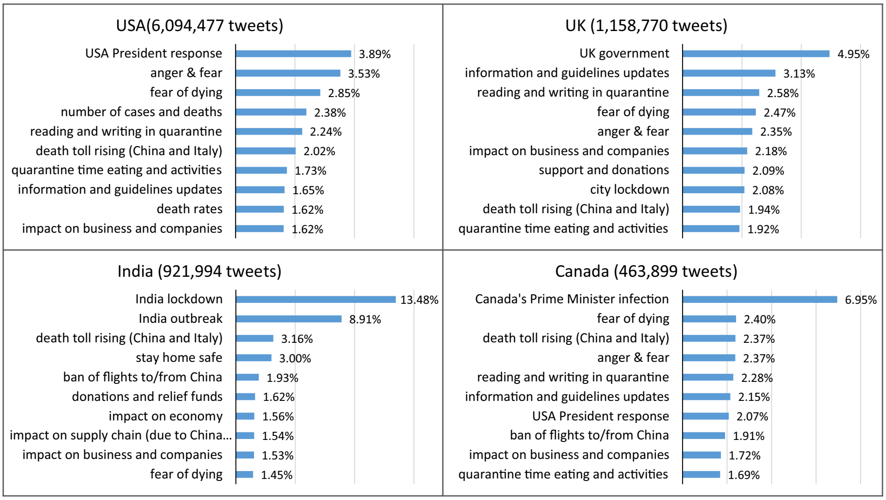


Fig. 14 Top 10 most popular tweets for the four countries with the highest number of tweets in the corpus

top 10 most popular list in all these countries. However, considering countries with more than 10,000 tweets, Greece was that the country with the highest expression of topic 'expression of extreme sentiment (anger & fear)' (11.3% of overall tweet corpus of this country), while Hungary was the country with the highest expression of topic 'expression of extreme sentiment (fear of dying)' (38.9% of overall tweet corpus for this country).

Synthesis of results: Correlation with real-world events

Real-world events of the period of the study were used as points in the weekly popularity timeline to identify correlations and provide evidence of the efficiency and effectiveness of the popularity evolution analysis and geographical distribution followed in the study.

Weekly variation of popularity for topics in the *Lockdown* category is shown in Fig. 15, together with related real-world events. On 27 February, Prime Minister Shinzo Abe requested that all Japanese elementary, junior high, and high schools close until early April to help curb the virus [33]. This coincides with the 9th week peak of the topic 'European and Japan. The topic 'Nigeria lockdown' shows also a first peak on 9th week; at the same time, the first confirmed case in Nigeria was announced (27 February, [34]). The same topics peak again during the 13th week and then follows an increasing trend; this is corroborated by the fact that on March 23th Ebonyi State government in southeastern Nigeria, banned all public gatherings [35], while at the same time, Kwara State and Lagos State announced the indefinite closure of their schools [36, 37]. On 24 March, the Government of India under Prime Minister ordered a nationwide lockdown [38], and this is depicted on the

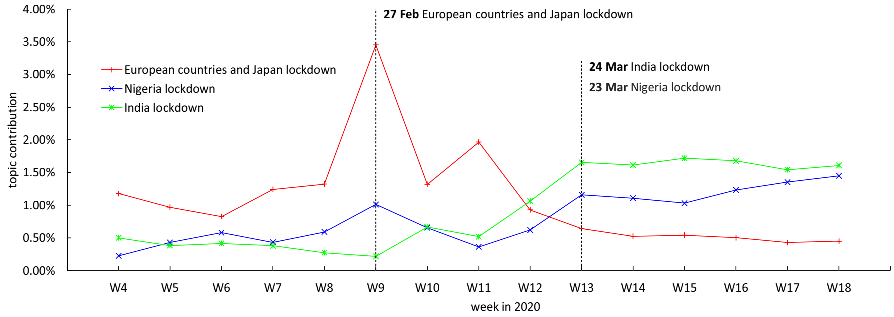


Fig. 15 Representative events related to lockdown

respective topic on 'India lockdown' which sharply increases popularity during the 2 previous weeks.

Figure 16 shows USA representative topics and events. Clearly, the topic 'warning by Dr. Fauci' peaks on the week of Dr. Fauci's announcement at a business show on CNBC that "good public health has limited outbreak in the US" [39] and is briefly discussed for the following 2 weeks. The topics 'US White House task force' and 'USA President response' peak sharply on the 9th week, as the President Trump, Vice President Pence, and members of the coronavirus task force gave a

Fig. 16 Representative USA-related topics and events. **a** Most popular topics, with a peak of more than 2% for at least 1 week; and **b** less popular topics

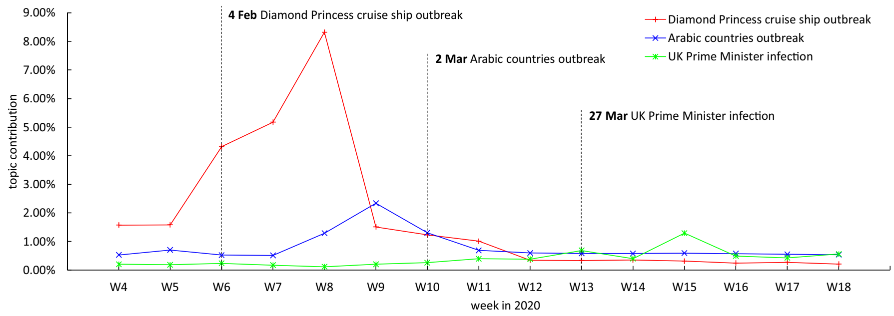


Fig. 17 Representative events related to disease outbreak

press conference on February 26th [40]. The topic 'tax relief and relief funds in USA' shows a peak on 13th week, coinciding with the U.S. President signing the Coronavirus Aid, Relief, and Economic Security Act on March 27th [41]. The topic on 'USA primary elections (Wisconsin)' shows a low but continuous popularity from the 8th till the 16th week and peaks during the week of the elections [42]. Finally, the topic 'USA protests' is spawned by the USA protest on 15 April [43], while the topic 'US president claims disinfectants can cure' is briefly popular during the week of the respective announcement [44].

Figure 17 shows representative events of *Outbreak* category. The topic on 'Diamond Princess cruise ship outbreak' shows an increase in popularity during the 6th week, and peaks during the 8th week when passengers were evacuated after a 2 week period of continual announcements of new positively tested passengers [45]. On 2 March, Saudi Arabia confirms its first case, a Saudi national returning from Iran via Bahrain [46] and this seems to spawn a 1 week earlier peak for the topic on the neighboring 'Arabic countries outbreak'. On 27 March, Britain's Prime Minister Boris Johnson tested positive [47], subsequently admitted in the hospital and discharged on 12 April [48] which coincides with the popularity of the 'UK Prime Minister infection' peaking on 13th and 15th week.

The time evolution of the popularity of the topic 'toilet paper and other supplies panic-buying' is shown in Fig. 18. The 5 countries where this particular topic was most popular were Australia (2% of overall tweet corpus of this country), Hong Kong (1.7%), New Zealand (1.5%), Ukraine (1.4%), and United Kingdom (1.3%) (considering countries with more than 10,000 tweets), so related events or announcements from these countries are also depicted in Fig. 18. The popularity of the topic starts increasing during 8th week, when armed robbers stole hundreds of toilet trolls in Hong Kong [49]. The increase continues next week, as Australian toilet paper company announced that it had completely sold out of stock on 26 February [50], and a supermarket rush evolves along the first confirmed case in New Zealand [51]. The popularity of this topic peaks during 10th week, when on March 6th, the British Health Secretary has urged the public to stop panic-buying [52]; a slow decline thereafter is followed with the Ukrainian announcement of quotas for shoppers [53].

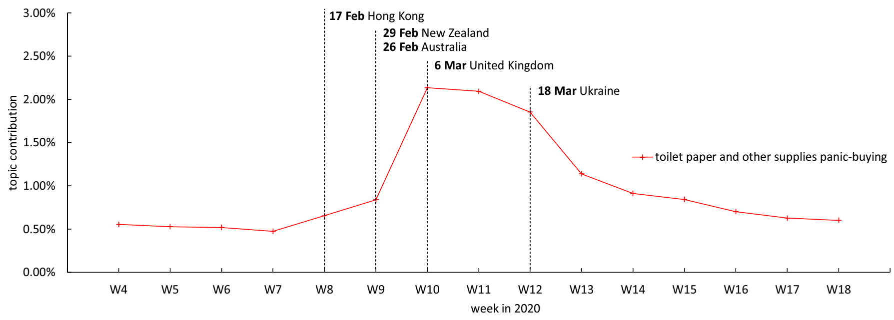


Fig. 18 The topic 'toilet paper and other supplies panic-buying' and related events in the five countries where this topic was most popular

Discussion

Principal findings

Our study performed a high-resolution topic analysis, including an investigation of temporal and geospatial variance of topics, of public discussions related to COVID-19 in a set of 20,230,833 original English language postings on Twitter, corresponding to a period from January 25 to April 30, 2020. Fine grain topic analysis identified 91 meaningful topics which give detailed insights on public concerns and discussions on Twitter during the period of the study. Calculation of the relative popularity of each topic showed that topics related to life during the pandemic were most popular, followed by discussions on pandemic management (at a social and political level and in terms of scientific response) and by information diffusion on local outbreaks and lockdown measures. Several topics related to concerns about economy were also identified.

Regression analysis on the temporal evolution of topic popularity revealed only 15 topics (16%) with a good linear fit ($R^2 > 80\%$ and p value ≤ 0.05). The majority of the topics showed a temporal evolution with local maxima (in most cases a single peak) of a duration spanning (in most cases) 1 week, underlining the short-lived character of discussions in Twitter, which reinforces the previous findings on the rather swift temporal evolution of discussions on Twitter [24]. When compared to real-world events, temporal popularity curves show a good correlation with and quick response to real-world triggers.

Geospatial analysis of topics showed that approximately 30% of original English language tweets were contributed by USA-based users, while overall more than 60% of the English language tweets were contributed by users from countries with an official language other than English. Topic popularity analysis per country for the countries with significant contribution revealed that the most popular topics were rather similar for each country, including topics on expression of negative sentiments, on aspects of life during the quarantine and information on cases, deaths, and guidelines. However, the ranking of most popular topics varied among countries; in most cases, the first most popular topic for each country was of local interest, related either to political issues, local outbreak, or local lockdown.

Comparison with prior work

A recently published study produced an LDA analysis of the daily USA tweets for approximately the same period as our study, namely from the 4th till the 15th week of 2020 [18]. They identified eight salient topic groups, in order of popularity: (1) China, (2) USA, (3) deaths, (4) lockdown, (5) USA president, (6) home, (7) pandemic, and (8) social distancing. Although the topical granularity reported is limited, we did identify similar topics within the 10 most popular topics in the USA as presented in Fig. 14. This verifies the ability of the LDA approach to identify similar topics of discussion in Twitter either applied on a daily tweet corpus or on a corpus of larger duration as in our study.

A similar LDA analysis of global English language tweets and a coarse topic resolution of 11 topics for the period week #4 to week #10 [20] was not possible to identify a topic related to symptoms or to treatments. Our findings reveal a topic on 'treatment drugs and trials' which shows an increasing popularity after week #11, outside the time period of the aforementioned study. Additionally, our high-resolution analysis revealed a topic on 'symptoms' which is prevalent from week #9 to week #10; this is consistent with other Twitter analysis findings which indicate discussion on COVID-19 symptoms in Twitter well before the first official announcement of the 3 major symptoms by the Centers for Disease Control and Prevention (CDC) on March 30 followed by a second announcement with more detailed list of symptoms on April 19, 2020 [54, 55].

Another study identified 26 topics for the period week #1 to week #19 in global tweets only by individuals [22], which suggests that topics related to economy and markets are most popular and correspond to about 20.51%. Our study confirms a 9.56% popularity of topics related to economy and financial issues, which brings this category lower in the relative popularity list (Fig. 2).

Another investigation looked into the content and credibility of tweets by Centers for Disease Control and Prevention (CDC) and the American College of Surgeons (ACS) for 5 months starting on February 1, 2020 [7]. These 2 accounts during this time mainly published national guidelines and personal protective measures for practitioners. The study concluded that the 2 accounts showed a greater public response with a higher mean of retweets and likes, which peaked in March for CDC and in April for ACS. Indeed, our study identified a relevant topic 'information and guideline updates' with a popularity which, as shown in Fig. 5, starts increasing in the beginning of March, peaks mid-March and continues in a rather steady popularity (of around 1.5%) during April.

Our findings are consistent with a previous study [19] that analyzed global tweets between weeks #11 and #16. In particular, we confirm all the topics identified, although our analysis did not identify topics related to the Chinese Communist Party. Our findings agree that handwashing is no more popular after the 12th week (with a peak during 10th week), and as new measures are becoming more popular, namely social distancing (peak on 12th week) and facemasks (peak on 14th week). Contrary to what has been suggested by this previous study, our analysis for this time interval revealed a number of topics (with high or increasing popularity or even popularity peaks), clearly related to events in countries other than USA (2 topics on India and 1 about Nigeria). We also identified 3 additional distinct UK-related topics (instead of 1) with different temporal evolution, and also 4 additional USA-related topics, again with distinct temporal evolution. Also, our analysis also reveals additional popular topics on (1) fear and anger (peak on w11); (2) lockdown and economy restart (significant increase from 14th week onwards); (3) three distinct topics on lifestyle during quarantine; (4) two topics related to donations and support funds; (5) two topics related to medical equipment and specific clinical trials; and (6) a topic on surveillance applications. This indicates that topic modeling of twitter datasets should use larger numbers of topics; this implies that topic coherence analysis to identify the number of topics should consider experiments with a large span of values and consider higher local minimization points.

Another similar study [12] using a different analysis approach (combining text mining and manual analysis) of a limited USA tweet set acquired during the 20th and 21st weeks of 2020 (just after the period of our study) identified six major topics: (1) surveillance, (2) prevention measures, (3) treatment and testing information, (4) symptoms and transmission, 5) fear, and 6) financial loss. These topics again are very similar to the ones we have identified as top most popular topics for USA tweets. An exception is the most popular topic on 'USA President response' which is not identified in the aforementioned study, which is of no surprise as we found this topic to peak during the 9th week, showing a steady decline thereafter (Fig. 5).

A similar LDA analysis of 203,191 original Chinese language microblogs posted in Sina Weido, a Chinese Twitter equivalent [56], identified 17 salient topics for a period from December 1, 2019 to July 31, 2020 (including the observation period of our study). The geospatial analysis performed on Chinese originating tweets (a total of 212,373 tweets) included in our study confirms 14 similar topics, ranked within the top 20 most popular for China; our study could identify the following topics: (1) joint prevention and control; (2) epidemics in neighboring countries; and (3) fueling and saluting anti-epidemic action.

A recent analysis [57] showed a substantial increase in purchases and searches for previously unpurchased and unsearched therapies by the general public following the backing of US president Donald J Trump. These increases correlated with his discussions in press conferences and personal social media posts advocating for hydroxychloroquine and chloroquine cures, followed the initial press conference on March 19, 2020. Our study confirms this, by the evolution of the topic 'treatment drugs and trials (hydroxychloroquine)' which after a deep low during week #11, shows a statistically significant increase of the popularity from week #12 (16th to 22nd March 2020) which peaks on week #15.

Finally, a study on economic uncertainty during the pandemic using several different indicators shows that implied volatility rose rapidly from late February, peaked in mid-March and fell back by late March [58]. This coincides with our finding on the time evolution of 2 related topics "fears for impact on stockmarket" and "impact on economy" (Fig. 9), which follow the same longitudinal behavior. The same study also shows that broader measures of economic uncertainty reporting on subjective business uncertainty peaked later; again, this agrees with the evolution of the topic ‘impact on business and companies’ and ‘lockdown and economy restart’ which during the period of our study are both statistically significantly increasing.

Strengths and limitations of this study

A major strength of the approach used in this study to explore Twitter content with temporal and geospatial analysis is the statistical unsupervised nature, which makes it ideal for the big data sets produced in Twitter, which in principle cannot be fully and homogeneously retrieved via the provided API. Additionally, our approach is language independent and can easily produce temporal and geospatial analysis.

A principal limitation pertaining to Twitter content analysis in our study arises from the original set of tweets which was generated using a query based on a specific search terms identified in the initial phases of the pandemic. As the pandemic evolves, more hashtags and keywords are emerging, which should be also considered for inclusion in the query. The findings of the study are clearly limited to the duration of collected data; more experiments are required to cover later phases of the pandemic and compare to findings of the onset explored here. Another limitation might arise from the fact that, as customary, only original tweets were included; calculating topic popularity should ideally include a weighting based on the impression of original tweets as expressed by retweets and likes. Additionally, tweets included in the study did not filter out postings by social bots which might skew findings on public concerns; a recent twitter analysis of COVID-19-related tweets during approximately the same period estimated that only 9.27% of related tweets can be attributed to social bots [16]. Finally, the findings of our study should be viewed under the assumption of Twitter users demographics; based on Twitter reports, users cover all age range (with most users in the range of 18–49 years old), with a clear bias toward male sex (70.4%) [1].

Questions answered and future work

Results of this study indicate that a high-resolution (i.e., large number of topics) topic analysis of Twitter datasets can reveal valuable information with a fine granularity on public discussions and concerns. Temporal evolution of topics popularity shows that most discussions on Twitter are short-lived and often closely driven by real-world events. Topics in English language Twitter are quite similar among users originating from different countries; however, the relative popularity of each topic within a country may vary; topics of local interest are most often the most popular in each country. Based on preliminary comparison with other studies, our findings indicate that exploring global, English language Twitter datasets might be adequate to identify salient themes of discussion for country-specific audiences. More experiments are necessary to compare identified topics using the same Twitter dataset for global and local analysis. Finally, based on our findings, Twitter discussions related to fears or expectations regarding economy and the markets seem to correlate with conventional economic indicators. Again, further analysis is needed to specify topic modeling methodologies that could lead to qualitative indices of economic uncertainty and its recovery.

Appendix A

Algorithm 1: Geographical origin of tweets.

```
Function getCountryOfUserLocation (ul: Text)
    h_score ← new hash table that contains the current score per country
    token[ ] ← split ul based on punctuation

    for i ← token.length − 1 to 0 do
        pn ← get the placename with the highest population for token[i]
        if pn.isNotNull then
            score_pn ← if pn.isCountry then set 1000 · (i + 1) else set 10 · (i + 1)
            if h_score.contain(pn.countryName) then
                h_score.add(pn.countryName, score_pn)
            else
                h_score.setInitial(pn.countryName, score_pn)
            end
        else
            token2[ ] ← split ul based on whitespace character
            for j ← token2.length − 1 to 0 do
                if token2[j].isNotStopWord then
                    pn2 ← get the placename with the highest population for token2[j]
                    if pn2.isNotNull then
                        score_pn2 ← if pn2.isCountry then set 100 · (j + 1) else set 1 · (j + 1)
                        if h_score.contain(pn2.countryName) then
                            h_score.add(pn2.countryName, score_pn2)
                        else
                            h_score.setInitial(pn2.countryName, score_pn2)
                        end
                    end
                end
            end
        end
    end

    return country with the highest score in h_score
end
```

Appendix B

See Table 2.

Table 2 List of topics organized in nine categories, showing the rank of the topics, the respective overall popularity metric, the regression analysis results, and the top 15 words of each topic

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Category: Life during the pandemic					
Expression of extreme sentiment (anger and fear)	1	3.32	– 0.002901	41.07	Fuck (0.019) People (0.017) Don (0.016) Shit (0.015) Gonna (0.010) Die (0.007) Time (0.007) ll (0.007) ve (0.007) Stop (0.007) Guy (0.006) Start (0.006) Joke (0.006) Feel (0.006) Im (0.006)
Expression of extreme sentiment (fear of dying)	3	2.69	– 0.000270	3.97	People (0.051) Don (0.020) Virus (0.013) Die (0.013) World (0.009) Kill (0.009) Country (0.008) America (0.008) Stop (0.008) Live (0.007) Govern (0.006) Time (0.006) Doesn (0.005) Care (0.005) Media (0.005)
Reading and writing in quarantine	5	1.98	0.000885	59.10	Time (0.024) Read (0.019) ve (0.013) Story (0.009) Article (0.009) Life (0.009) Feel (0.009) People (0.008) Pandemic (0.008) Love (0.006) Write (0.006) World (0.006) Hope (0.006) Don (0.006) Share (0.005)
Quarantine time eating and activities	8	1.49	0.000860	57.34	Quarantine (0.012) Time (0.010) Home (0.010) Lockdown (0.010) Eat (0.006) ve (0.006) Food (0.005) Kid (0.004) Don (0.004) Happy (0.004) Love (0.004) Drink (0.004) House (0.004) Dog (0.004) Walk (0.004)
Stay home safe	28	1.06	0.000740	43.90	Stay (0.112) Home (0.073) Safe (0.052) Stayhome (0.020) Staysafe (0.014) Healthy (0.012) Spread (0.011) Time (0.010) Family (0.010) People (0.010) Stayathome (0.009) Don (0.009) Live (0.008) Lockdown (0.008) Corona (0.008)
Toilet paper and other supplies panic-buying	31	1.04	0.000157	1.48	Toilet (0.026) Paper (0.026) Store (0.025) Buy (0.024) Shop (0.021) Food (0.019) Grocery (0.018) People (0.016) Panic (0.014) Hand (0.010) Stock (0.009) Roll (0.009) Supply (0.008) Supermarket (0.008) Mask (0.008)
Imprisonment and violations[a]	41	0.93	0.000933	85.25	Prison (0.032) Police (0.022) Release (0.022) Jail (0.015) Court (0.014) Law (0.012) Lockdown (0.012) Arrest (0.012) Inmate (0.010) Pandemic (0.009) Amid (0.008) Spread (0.008) People (0.007) Rule (0.006) violate (0.006)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Religion and prayers	42	0.93	0.000461	61.54	God (0.038) Church (0.020) Pray (0.017) Prayer (0.015) World (0.012) Jesus (0.012) Easter (0.010) Allah (0.009) Pastor (0.009) Pandemic (0.009) Bless (0.009) Christian (0.008) Time (0.008) Lord (0.008) People (0.008)
Mental health during the pandemic[a]	43	0.93	0.001022	89.04	Health (0.031) Mental (0.023) Pandemic (0.018) Children (0.015) Anxiety (0.014) Time (0.014) Care (0.013) Tip (0.012) Stress (0.011) Support (0.011) People (0.010) Parent (0.010) Lockdown (0.010) Home (0.010) Family (0.009)
Fighting doctors and nurses	54	0.80	0.000932	78.74	Fight (0.026) Nurse (0.023) Doctor (0.020) NHS (0.018) Frontline (0.016) Health (0.015) Staff (0.014) Care (0.014) Pandemic (0.014) Heroes (0.013) Healthcare (0.013) Line (0.012) Support (0.012) Live (0.010) Medical (0.010)
Movies and video games in quarantine	55	0.78	0.000069	2.81	Movie (0.023) Watch (0.021) Game (0.019) Film (0.016) Due (0.012) Pandemic (0.010) Time (0.010) Release (0.010) Video (0.009) Stream (0.009) Netflix (0.009) Delay (0.009) TV (0.009) Play (0.008) outbreak (0.007)
Music and video in quarantine	56	0.78	0.000633	67.26	Music (0.024) Song (0.017) Listen (0.013) Video (0.010) Quarantine (0.009) Love (0.008) Podcast (0.007) Live (0.007) Artist (0.006) Concert (0.006) Socialdistanc (0.006) Dance (0.006) Watch (0.005) Play (0.005) Singe (0.005)
Racism, discrimination in pandemic	61	0.74	– 0.000216	12.82	China (0.025) People (0.024) Asia (0.021) Black (0.020) Rac (0.018) Racist (0.018) America (0.015) Spread (0.012) Africa (0.011) Virus (0.011) Commune (0.011) Muslim (0.011) India (0.010) Fear (0.009) Attack (0.008)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Family separation in quarantine	65	0.67	0.000553	61.68	Die (0.046) Family (0.040) Friend (0.038) Italy (0.027) Time (0.018) Share (0.016) Mother (0.014) Colleague (0.013) Support (0.013) Baby (0.012) Stand (0.012) Con (0.011) Woman (0.010) Cari (0.009) Voi (0.009)
USA primary elections (Wisconsin)	69	0.61	0.000350	39.08	Vote (0.046) Elect (0.038) Biden (0.026) Trump (0.025) Bernie (0.016) Joe (0.014) Sand (0.013) Primary (0.011) Pandemic (0.010) Wisconsin (0.010) 2020 (0.010) Mail (0.010) Poll (0.009) November (0.009) Campaign (0.009)
Art in quarantine[a]	82	0.40	0.000382	87.81	Art (0.021) Lockdown (0.012) Photo (0.010) Image (0.010) Artist (0.008) Quarantine (0.008) Paint (0.007) Stayhome (0.007) Pandemic (0.007) Love (0.007) Time (0.006) Covid (0.006) Create (0.006) Photography (0.006) Beautiful (0.005)
USA protests	84	0.39	0.000491	65.28	Protest (0.058) Curve (0.041) Flatten (0.035) Michigan (0.026) Trump (0.026) Lockdown (0.017) Governor (0.016) Washington (0.014) Whitmer (0.012) Post (0.011) Stay-at-Home (0.011) Gov (0.010) Call (0.010) Rally (0.009) Spread (0.009)
Musical bands and groups	91	0.15	– 0.000244	71.99	BTS (0.023) Follow (0.019) Kpop (0.019) RT (0.018) NSFW (0.016) Fancam (0.015) GC (0.013) 18 (0.013) Daddy (0.012) Blackpink (0.012) Gain (0.011) Au (0.011) Trick (0.011) IFB (0.010) View (0.010)
Category: Pandemic management					
USA President response	2	2.82	0.001683	32.74	Trump (0.101) President (0.018) America (0.017) Response (0.013) Donald (0.010) Lie (0.010) Blame (0.007) Call (0.007) Pandemic (0.007) Crisis (0.007) Democrat (0.006) Administrate (0.006) Media (0.006) Hoax (0.006) time (0.006)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Information and guidelines updates	10	1.46	0.000542	31.52	Update (0.024) Information (0.023) Health (0.017) Resource (0.012) Service (0.010) Visit (0.010) Public (0.009) Website (0.009) Guidance (0.009) Support (0.009) Advice (0.008) Safety (0.008) Business (0.008) Pandemic (0.008) Response (0.007)
Live webinars about pandemic[a]	19	1.23	0.000852	95.12	Join (0.020) Dr (0.018) Live (0.015) Discuss (0.013) Webinar (0.012) Question (0.011) Pandemic (0.010) Talk (0.009) Watch (0.009) Impact (0.008) April (0.007) Answer (0.007) Listen (0.007) Register (0.007) Podcast (0.007)
UK government[a]	24	1.11	0.000855	89.62	UK (0.040) Govern (0.023) NHS (0.017) Johnson (0.009) Bori (0.009) Britain (0.008) News (0.008) Brexit (0.008) PPE (0.007) Tory (0.007) Crisis (0.006) Test (0.006) Minister (0.006) Scotland (0.006) Lockdown (0.006)
Support and donations[a]	25	1.09	0.001350	85.69	Support (0.036) Commune (0.024) Donate (0.021) Food (0.021) Fund (0.015) Local (0.014) Pandemic (0.013) Crisis (0.011) People (0.011) Family (0.010) Business (0.008) Response (0.008) Volunteer (0.008) Relief (0.008) Vulnerable (0.007)
US White House task force	30	1.05	0.000539	22.99	Trump (0.067) House (0.047) White (0.044) Force (0.033) Task (0.030) President (0.025) Press (0.023) Pence (0.020) Live (0.015) Response (0.015) Conference (0.012) Mike (0.011) Hold (0.010) Watch (0.010) News (0.007)
Tracking and surveillance apps	39	0.95	0.000556	66.36	App (0.021) Data (0.019) Google (0.019) Track (0.018) Ai (0.016) Technology (0.016) Tech (0.015) Fight (0.014) Trace (0.013) Contact (0.013) Apple (0.013) Launch (0.009) Develop (0.008) Privacy (0.007) Pandemic (0.007)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
US States authorities announcements	48	0.86	0.000825	79.53	Gov (0.036) Governor (0.033) York (0.026) Cuomo (0.023) Update (0.015) Mayor (0.013) Live (0.013) City (0.012) California (0.010) Ny (0.010) Business (0.009) Georgia (0.009) Announce (0.009) Close (0.009) Nyc (0.008)
Tax relief and relief funds in USA[a]	52	0.84	0.000941	89.43	Money (0.023) Stimulus (0.023) Pay (0.023) Check (0.021) Trump (0.019) Tax (0.018) Relief (0.013) Fund (0.013) America (0.013) Billion (0.010) Govern (0.010) Bailout (0.009) Business (0.008) Aid (0.008) Cut (0.008)
Relief bills in USA	53	0.83	0.000686	32.67	Bill (0.037) Senate (0.027) House (0.027) Trump (0.025) Relief (0.021) Package (0.019) Pelosi (0.017) Stimulus (0.016) Democrat (0.014) Pass (0.014) Congress (0.014) Fund (0.014) Vote (0.013) Billion (0.012) Aid (0.011)
Warning by Dr. Fauci, NIAID Director	63	0.73	– 0.001311	30.75	Flu (0.063) China (0.053) Dr (0.042) Virus (0.036) Doctor (0.029) Trump (0.023) Fauci (0.022) Detail (0.019) Warn (0.017) Die (0.015) Outbreak (0.013) Wuhan (0.011) Li (0.009) Disease (0.009) Anthony (0.009)
Donations and relief funds	71	0.59	0.000531	59.34	Donate (0.064) Fight (0.035) Fund (0.034) Million (0.026) Relief (0.025) Support (0.018) Pandemic (0.014) Effort (0.013) 000 (0.012) Raise (0.010) Rs (0.010) Join (0.009) Pledge (0.009) World (0.009) Ecoin (0.009)
NHS support petition in UK	77	0.47	0.000404	40.32	Sign (0.116) Petition (0.101) Govern (0.022) Uk (0.019) Crisis (0.016) Call (0.016) Nhs (0.015) Support (0.014) Protect (0.014) Pay (0.013) People (0.013) Pandemic (0.012) Frontline (0.011) Rent (0.009) Income (0.009)
Dead bodies and funerals	86	0.32	– 0.000022	0.51	Body (0.039) Victim (0.017) Funeral (0.017) Dead (0.016) Home (0.016) Yemen (0.015) Death (0.015) York (0.015) China (0.015) City (0.013) Wuhan (0.012) Mass (0.010) Kill (0.010) Hospital (0.009) Outbreak (0.008)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Philippines President measures	88	0.27	– 0.000069	7.11	Philippines (0.026) Duterte (0.016) City (0.013) Ph (0.011) Govern (0.011) Na (0.010) Test (0.010) President (0.009) Doh (0.009) Sa (0.009) Confirm (0.009) Health (0.008) Quarantine (0.008) Manila (0.008) Disease (0.007)
App downloads for outbreak management	89	0.25	0.000272	71.24	App (0.119) Download (0.087) Time (0.029) Epoch (0.026) Spread (0.026) Fight (0.025) Android (0.024) Symptom (0.023) Update (0.023) Share (0.023) Slow (0.023) Feel (0.022) Risk (0.021) Coverage (0.021) Identify (0.020)
Category: Medical					
Hospital and home care	17	1.32	– 0.000696	17.95	Hospital (0.080) Patient (0.063) Doctor (0.025) Nurse (0.025) Home (0.019) Treat (0.018) Die (0.018) Care (0.017) Ventilate (0.010) Wuhan (0.010) Bed (0.010) York (0.009) China (0.009) Staff (0.009) Medical (0.008)
World health emergency declaration	26	1.09	– 0.001817	67.15	Health (0.086) World (0.045) Emergency (0.034) Pandemic (0.030) Declare (0.029) Outbreak (0.027) Public (0.026) Global (0.026) Organize (0.026) Trump (0.025) Warn (0.022) Official (0.017) Spread (0.016) China (0.016) CDC (0.013)
Test kit[a]	27	1.06	0.000620	81.26	Test (0.194) Kit (0.031) Positive (0.016) Antibody (0.014) Result (0.014) Lab (0.013) Cdc (0.009) Negate (0.009) People (0.009) Rapid (0.008) 000 (0.007) Fda (0.007) Develop (0.006) Sample (0.005) Diagnostic (0.005)
Medical equipment and supplies	35	0.99	0.000696	68.19	Health (0.032) Medical (0.027) Protect (0.026) Care (0.025) Equip (0.017) Hospital (0.017) Healthcare (0.017) Supply (0.016) Ppe (0.012) Pandemic (0.012) Test (0.012) Patient (0.011) Ventilate (0.011) Fight (0.010) Doctor (0.010)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Treatment drugs and trials (hydroxychloroquine)	47	0.87	0.000497	36.75	Drug (0.061) Treat (0.061) Patient (0.038) Hydroxychloroquine (0.022) Trial (0.017) Doctor (0.015) Trump (0.014) Cure (0.013) Effect (0.012) Plasma (0.011) Study (0.011) Remdesivir (0.011) Chloroquine (0.011) Clinic (0.008) China (0.008)
Vaccine development	49	0.85	– 0.000068	2.01	Vaccine (0.114) Develop (0.030) Trial (0.021) Scientist (0.020) Test (0.016) Research (0.013) Treat (0.013) Human (0.010) Race (0.009) Drug (0.009) Company (0.009) World (0.008) Researcher (0.008) News (0.007) Cure (0.007)
Immune system and disease risk	51	0.84	0.000381	70.08	Immune (0.042) People (0.023) System (0.021) Risk (0.021) Patient (0.017) Health (0.015) Disease (0.014) Condition (0.011) Infect (0.011) Virus (0.011) Lung (0.010) Heart (0.009) Die (0.009) Smoke (0.008) Doctor (0.008)
Animals in China	57	0.77	– 0.000424	26.04	Animal (0.037) China (0.028) Dog (0.018) Cat (0.017) Pet (0.016) Human (0.016) Pollution (0.015) Air (0.014) Market (0.013) Wildlife (0.012) Eat (0.012) Trade (0.009) Positive (0.009) Test (0.009) Stop (0.009)
Virus origin (Wuhan lab, bats)	59	0.76	– 0.001261	55.40	China (0.072) Wuhan (0.041) Lab (0.038) Virus (0.027) Scientist (0.015) Origin (0.014) Bat (0.013) Weapon (0.012) Originate (0.011) Market (0.010) Research (0.010) Bioweapon (0.010) Outbreak (0.010) Biology (0.010) Create (0.009)
Symptoms	62	0.74	– 0.000152	14.94	Symptom (0.058) Test (0.030) Cough (0.019) Fever (0.017) People (0.015) Doctor (0.013) Patient (0.012) Cold (0.010) Infect (0.010) Feel (0.010) Flu (0.010) Sick (0.009) Don (0.009) Call (0.008) ve (0.008)
Respiratory infection syndrome research	72	0.57	– 0.000245	35.25	Virus (0.018) Sars-Cov-2 (0.016) Infect (0.015) Patient (0.014) Disease (0.012) Respiratory (0.010) Study (0.010) Cell (0.008) 2019 (0.008) Research (0.008) Human (0.007) Severe (0.007) Clinic (0.006) Syndrome (0.006) Acute (0.006)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Vaccine by Bill Gates Foundation	87	0.30	− 0.000050	1.78	Bill (0.087) Gate (0.079) Vaccine (0.023) Million (0.021) People (0.019) Economy (0.014) Urge (0.014) Debt (0.013) Pandemic (0.013) Student (0.013) Foundation (0.013) Package (0.012) 45 (0.012) Stimulate (0.012) Cancelstudentdebt (0.010)
Category:Outbreak					
China and Wuhan outbreak	11	1.45	− 0.004684	66.77	China (0.148) Wuhan (0.037) Outbreak (0.020) Virus (0.014) World (0.012) Xi (0.010) Spread (0.008) Coronavirusoutbreak (0.008) Report (0.008) Govern (0.008) Beije (0.007) People (0.007) Communist (0.007) City (0.006) Epidemic (0.006)
Disease spreading[a]	14	1.38	− 0.002456	86.90	Spread (0.053) Virus (0.035) Disease (0.027) Infect (0.024) China (0.023) People (0.013) Transmission (0.013) Symptom (0.012) Study (0.010) Expert (0.010) Outbreak (0.010) Prevent (0.009) Control (0.009) Health (0.009) CDC (0.008)
Diamond Princess cruise ship outbreak	33	1.02	− 0.003013	32.91	Cruise (0.076) Ship (0.073) Hong (0.035) Kong (0.035) Quarantine (0.032) Japan (0.032) Passenger (0.027) Test (0.023) Princess (0.022) Positive (0.016) Diamond (0.015) America (0.012) People (0.012) Infect (0.012) China (0.012)
India outbreak	46	0.88	0.000557	63.15	Positive (0.045) Test (0.037) India (0.027) Patient (0.019) Report (0.018) Total (0.018) Health (0.012) Delhi (0.012) Death (0.011) Hospital (0.010) People (0.008) Update (0.008) Kerala (0.008) News (0.007) District (0.007)
East Asian countries and Africa outbreak	50	0.85	− 0.000550	8.73	South (0.098) Korea (0.091) Africa (0.046) North (0.018) China (0.016) Report (0.015) Country (0.014) Confirm (0.014) Infect (0.011) Test (0.011) Outbreak (0.011) Italy (0.011) Spread (0.009) Iran (0.007) Lockdown (0.006)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Arabic countries outbreak	60	0.75	− 0.000237	4.51	Iran (0.061) Health (0.022) Minister (0.014) Saudi (0.012) Ministry (0.011) Country (0.011) Confirm (0.010) Report (0.009) Arabia (0.008) Infect (0.008) UAE (0.008) Outbreak (0.008) Spread (0.007) Official (0.007) Test (0.007)
Israel migrant/refugee outbreak	66	0.67	0.000406	78.68	Israel (0.022) Border (0.020) Spread (0.012) Migrant (0.011) Refugee (0.011) Pandemic (0.010) Camp (0.010) Amid (0.008) Palestinian (0.008) Fear (0.008) Mexico (0.007) Syria (0.007) People (0.007) Risk (0.007) Country (0.007)
California outbreak	74	0.55	− 0.000199	12.45	San (0.045) County (0.038) California (0.032) Los (0.026) Angeles (0.025) Francisco (0.019) Confirm (0.015) City (0.014) Diego (0.011) Test (0.011) Vega (0.011) Official (0.011) Santa (0.010) Homeless (0.010) Patient (0.010)
UK Prime Minister infection	75	0.49	0.000416	39.82	Johnson (0.072) Bori (0.065) Pakistan (0.031) Minister (0.030) Uk (0.025) Prime (0.022) Test (0.020) Pm (0.020) Hospital (0.016) Care (0.015) Positive (0.014) News (0.014) Britain (0.012) Intensive (0.012) Queen (0.011)
Canada’s Prime Minister infection	76	0.48	− 0.000019	0.25	Canada (0.072) Ontario (0.021) Health (0.019) Trudeau (0.016) News (0.016) Toronto (0.014) Cdnpoli (0.014) Test (0.011) Update (0.010) CBC (0.009) Alberta (0.009) Confirm (0.009) Positive (0.008) Minister (0.008) Province (0.008)
Australia and New Zealand outbreak	78	0.47	− 0.000063	6.77	Australia (0.068) Auspol (0.019) News (0.018) Virus (0.016) Minister (0.011) Live (0.011) Govern (0.010) Zealand (0.010) 7news (0.010) Pandemic (0.009) Morrison (0.009) Outbreak (0.008) Break (0.008) Health (0.008) NSW (0.008)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Outbreak in navy	80	0.43	0.000227	31.58	Navy (0.027) Military (0.024) Russia (0.019) Carry (0.017) Captain (0.017) Sailor (0.012) Ship (0.012) Aircraft (0.011) Outbreak (0.011) Army (0.011) Force (0.010) Test (0.010) Roosevelt (0.009) Command (0.009) Fire (0.008)
Cases in CPAC 2020 conference	85	0.33	0.000100	2.33	Test (0.050) Positive (0.036) Brazil (0.022) Germany (0.019) Trump (0.019) Cpac (0.014) Bolsonaro (0.013) Paul (0.012) President (0.012) Matt (0.012) Ted (0.011) Quarantine (0.011) Merkel (0.011) Cruz (0.010) Attendee (0.009)
Category:Lockdown					
Ban of flights to/from China	9	1.49	– 0.003859	78.75	China (0.055) Travel (0.055) Flight (0.035) Outbreak (0.018) Ban (0.017) Airline (0.015) Wuhan (0.014) Airport (0.013) Country (0.012) Suspend (0.011) Amid (0.010) Spread (0.010) Restrict (0.010) Passenger (0.009) Due (0.009)
Events canceled or postponed	20	1.21	– 0.000975	22.62	Due (0.044) Cancel (0.035) Postpone (0.030) Event (0.029) 2020 (0.023) Canceled (0.021) Concern (0.017) Outbreak (0.015) Olympic (0.013) Fear (0.012) Conference (0.011) World (0.010) Amid (0.010) Tokyo (0.010) Festival (0.009)
India lockdown	21	1.19	0.001203	77.12	India (0.031) Lockdown (0.023) Pm (0.015) Modi (0.014) Fight (0.013) Govt (0.011) Indian (0.009) People (0.008) Minister (0.008) Sir (0.007) Indiafightscorona (0.007) Delhi (0.005) App (0.005) Govern (0.005) Corona (0.005)
City lockdown	22	1.18	0.000364	26.37	Close (0.020) Drive (0.015) City (0.014) Lockdown (0.011) People (0.010) Restaurant (0.009) Park (0.009) Street (0.009) Car (0.008) Due (0.007) Service (0.007) Home (0.007) London (0.007) Public (0.006) Road (0.006)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Schools and universities lockdown	23	1.17	0.000441	14.05	School (0.072) Student (0.048) Close (0.022) University (0.020) Teach (0.018) Class (0.017) Online (0.016) College (0.016) Educate (0.015) Learn (0.013) Due (0.013) Parent (0.009) Home (0.009) Kid (0.009) Closure (0.008)
Sports events (football, basketball, etc.)	29	1.05	0.000397	7.76	Play (0.032) League (0.020) Sport (0.019) Football (0.018) Game (0.016) Season (0.015) Due (0.014) Nba (0.011) Fan (0.010) Pandemic (0.009) Club (0.008) Team (0.008) Amid (0.008) Test (0.008) Premier (0.007)
Nigeria lockdown	37	0.96	0.000771	76.60	Nigeria (0.066) Lago (0.022) Test (0.013) Lockdown (0.013) Confirm (0.012) Record (0.010) Buhari (0.009) Ghana (0.009) Govern (0.008) Break (0.008) Patient (0.008) Ncdc (0.008) President (0.007) Kano (0.007) Positive (0.007)
European countries and Japan lockdown	40	0.94	– 0.000756	18.58	Italy (0.072) France (0.023) Europe (0.018) Japan (0.016) Lockdown (0.016) Country (0.014) Spain (0.012) Close (0.012) People (0.012) Quarantine (0.012) Spread (0.011) Germany (0.011) Outbreak (0.011) Confirm (0.010) Northern (0.009)
Florida beach lockdown[a]	79	0.44	0.000442	92.84	Florida (0.052) Beach (0.028) Close (0.017) Hot (0.014) South (0.013) Plant (0.013) Governor (0.012) Spot (0.012) Park (0.011) Desanti (0.011) Miami (0.010) Disney (0.009) Pandemic (0.008) Dakota (0.007) Gov (0.007)
Category: Economy					
Fears for impact on stock market	7	1.50	– 0.001868	38.12	Market (0.050) Stock (0.041) Oil (0.019) Fear (0.018) Price (0.015) Global (0.011) Dow (0.010) Trade (0.009) China (0.009) Investor (0.009) Drop (0.008) Amid (0.008) Fall (0.007) Spread (0.007) Impact (0.007)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Impact on business and companies[a]	13	1.45	0.001448	97.06	Business (0.027) Pandemic (0.015) Crisis (0.015) Impact (0.013) Time (0.010) Company (0.010) Market (0.009) Change (0.009) Manage (0.008) Industry (0.007) Learn (0.007) Challenge (0.007) Read (0.006) Plan (0.006) Remote (0.006)
Impact on economy	15	1.36	– 0.000087	0.50	Economy (0.032) Economic (0.024) Impact (0.019) Global (0.017) China (0.013) Bank (0.013) Crisis (0.011) Pandemic (0.010) World (0.010) Outbreak (0.010) Hit (0.009) Rate (0.009) Country (0.008) Cut (0.008) Billion (0.008)
Impact on supply chain (due to China lockdown)	18	1.30	– 0.001102	17.69	China (0.036) Supply (0.025) Outbreak (0.016) Impact (0.014) Chain (0.014) Product (0.014) Due (0.013) Apple (0.011) Sale (0.011) Industry (0.010) Global (0.009) Hit (0.009) Food (0.009) Business (0.009) Company (0.008)
Unemployment[a]	36	0.98	0.000864	81.62	Job (0.033) Employee (0.022) Pay (0.021) Unemployment (0.014) Due (0.013) Pandemic (0.013) Leave (0.011) Million (0.011) Sick (0.011) Claim (0.010) Business (0.010) Company (0.010) Paid (0.010) Amazon (0.009) 000 (0.009)
Lockdown and economy restart	38	0.96	0.001020	78.79	Lockdown (0.063) Restrict (0.023) Ease (0.019) Lift (0.015) Country (0.014) Plan (0.013) Reopen (0.012) Economy (0.012) Measure (0.011) Sweden (0.011) UK (0.010) Reopene (0.009) Trump (0.008) Govern (0.008) Extend (0.008)
Business financial support, loans and relief[*]	58	0.77	0.000876	84.02	Business (0.037) Loan (0.022) Pay (0.016) Relief (0.014) Rent (0.012) Tax (0.012) Support (0.010) Financial (0.010) Pandemic (0.009) Impact (0.009) Govern (0.009) Bank (0.008) Program (0.008) Fund (0.007) Due (0.007)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Cancelation fees and refunds	70	0.59	0.000262	35.50	Due (0.028) Refund (0.020) Flight (0.016) Travel (0.016) Book (0.013) Customer (0.012) Cancel (0.011) Pay (0.011) Ticket (0.010) Canceled (0.009) Fee (0.008) Service (0.008) Change (0.008) Call (0.008) Insurance (0.007)
Digital markets[a]	81	0.41	0.000340	84.84	Free (0.025) Design (0.016) Business (0.015) Market (0.015) Book (0.014) Online (0.011) Check (0.008) Read (0.007) Brand (0.007) Website (0.007) Google (0.006) Search (0.006) Offer (0.006) Seo (0.006) Create (0.006)
Bitcoin and cryptocurrencies	90	0.24	– 0.000044	3.86	Bitcoin (0.052) crypto (0.037) Tom (0.031) Curfew (0.030) People (0.028) Hank (0.027) Cryptocurrence (0.025) Blockchain (0.022) Janta (0.020) Pledge (0.020) BTC (0.018) Fight (0.018) Rita (0.013) Wilson (0.013) News (0.013)
Category: Cases and deaths					
Death toll rising (China and Italy)	4	2.42	– 0.002898	58.54	Death (0.102) Toll (0.051) 000 (0.046) China (0.044) Report (0.026) Rise (0.021) Confirm (0.020) Update (0.020) Infect (0.016) Live (0.015) Italy (0.013) Country (0.011) Global (0.010) Outbreak (0.010) Record (0.009)
Number of cases and deaths	6	1.52	0.000141	1.85	Test (0.053) County (0.046) Positive (0.040) Confirm (0.027) Health (0.026) Report (0.017) Death (0.016) Official (0.015) Home (0.010) People (0.010) Department (0.010) Resident (0.009) Update (0.009) Die (0.009) Patient (0.009)
Death rates	12	1.45	– 0.000113	5.31	Death (0.050) Rate (0.034) 000 (0.032) People (0.030) Infect (0.021) Flu (0.018) Die (0.017) Million (0.016) Populate (0.010) Kill (0.009) Mortal (0.009) Data (0.009) America (0.009) Report (0.008) Country (0.007)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Confirmed deaths and recoveries	16	1.34	0.000630	74.42	Death (0.071) Total (0.053) Confirm (0.036) Recovere (0.023) Report (0.021) Update (0.018) Country (0.009) Bring (0.008) 10 (0.007) 24 (0.006) 19 (0.006) 2020 (0.006) Infect (0.006) 11 (0.006) Recovery (0.006)
Live data maps[a]	68	0.61	– 0.000527	82.24	Map (0.040) Time (0.033) Update (0.032) Track (0.028) Live (0.028) Data (0.025) York (0.024) John (0.023) Spread (0.020) Nyt (0.020) Hopkins (0.019) News (0.019) Article (0.016) Reuter (0.015) World (0.015)
Category: News and fake news					
News channels updates	32	1.03	– 0.000400	55.53	News (0.227) Bbc (0.063) Fox (0.027) World (0.019) Story (0.014) Update (0.014) China (0.014) Live (0.013) Uk (0.012) Outbreak (0.012) Break (0.011) Pandemic (0.011) ABC (0.010) Top (0.010) Report (0.009)
Misinformation spread in social media	34	1.02	– 0.000630	78.74	Media (0.033) News (0.030) Spread (0.020) Social (0.018) Information (0.017) Facebook (0.016) Twitter (0.016) Fake (0.015) Misinformation (0.015) Post (0.015) Video (0.014) Share (0.014) People (0.008) False (0.008) Report (0.008)
US President claims disinfectants can cure	67	0.61	0.000371	15.52	Trump (0.043) Disinfectant (0.038) Cure (0.030) Kill (0.023) Inject (0.023) Treat (0.017) Bleach (0.016) Light (0.014) Drink (0.013) Virus (0.012) Clean (0.012) Suggest (0.011) Lysol (0.010) People (0.009) President (0.008)
5G conspiracy theory	73	0.57	0.000231	41.22	5G (0.028) Conspiracy (0.025) Theory (0.017) Trump (0.013) Maga (0.012) Qanon (0.011) News (0.010) Media (0.008) Tucker (0.008) wwg1wga (0.007) Fox (0.007) Truth (0.006) Video (0.006) Watch (0.006) Pandemic (0.006)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Scams and cyberattacks	83	0.39	0.000215	57.21	Scam (0.043) Secure (0.019) Cybersecur (0.018) Hack (0.016) Email (0.014) Warn (0.013) Pandemic (0.012) Phish (0.012) Cyber (0.011) Attack (0.010) Malware (0.010) Fraud (0.009) Target (0.008) Zoom (0.007) Protect (0.007)
Category: Preventive measures					
Hand washing	44	0.92	– 0.000355	6.92	Hand (0.102) Wash (0.062) Sanitize (0.021) Spread (0.016) Touch (0.016) People (0.015) Don (0.014) Water (0.013) Clean (0.012) Avoid (0.012) Stay (0.012) Soap (0.012) Mask (0.011) Cough (0.010) Virus (0.009)
Facemasks	45	0.89	– 0.000065	1.99	Mask (0.145) Wear (0.054) Protect (0.027) N95 (0.013) Medical (0.013) Facemask (0.011) Glove (0.010) Surgical (0.008) People (0.008) ppe (0.007) Buy (0.007) China (0.006) Cover (0.006) Public (0.006) Supply (0.006)
Social distancing	64	0.73	0.000666	58.73	Social (0.112) Distanc (0.089) People (0.022) Spread (0.020) Stay (0.018) Home (0.017) Distance (0.013) Measure (0.013) Practice (0.012) Follow (0.009) Rule (0.009) Socialdistanc (0.008) Guidelines (0.008) Gather (0.007) Maintain (0.007)
Non-meaningful topics					
Topic 42[a]	–	2.16	0.001503	96.25	Pandemic (0.021) Crisis (0.018) World (0.014) Response (0.012) Global (0.011) Health (0.011) Change (0.010) Country (0.010) Lead (0.010) Govern (0.008) Time (0.008) Economic (0.007) Climate (0.007) Politics (0.006) Public (0.006)
Topic 88	–	1.19	0.000483	26.75	Covid (0.173) 19 (0.141) Corona (0.030) Coronavirusoutbreak (0.029) Coronavirusupdate (0.025) Covid2019 (0.022) Coronaviruspandemic (0.015) Virus (0.015) Stayhome (0.010) Coronaoutbreak (0.010) Lockdown (0.009) Coronavirususa (0.008) Pandemic (0.006) Stayathome (0.006) Quarantine (0.005)

Table 2 (continued)

Topic label	Topic popularity		Trend analysis		Top 15 words (weight in topic)
	Rank	Contrib. (%)	Reg. Coeff	R^2 (%)	
Topic 44	–	0.70	0.000228	66.41	2020 (0.090) April (0.059) March (0.046) Update (0.035) 20 (0.017) 00 (0.015) pm (0.014) 30 (0.013) Feb (0.012) 10 (0.012) 11 (0.011) February (0.010) 12 (0.010) apr (0.010) mar (0.010)
Topic 78	–	0.60	– 0.000613	45.49	War (0.028) World (0.022) 2020 (0.020) Pandemic (0.011) Predict (0.010) Plague (0.010) Kobe (0.009) America (0.009) Kill (0.008) Fire (0.007) Black (0.007) Happen (0.007) Outbreak (0.007) Australia (0.006) die (0.006)
Topic 53	–	0.49	0.000426	51.11	Test (0.051) Positive (0.044) Die (0.027) Kenya (0.026) Prince (0.017) People (0.013) Charles (0.012) Patient (0.010) Confirm (0.010) Condition (0.008) Star (0.008) Critic (0.007) Complicate (0.007) John (0.007) Moment (0.007)
Topic 31	–	0.29	– 0.000054	14.41	DE (0.026) EL (0.018) EN (0.018) LE (0.016) LA (0.012) Confine (0.012) Covid (0.010) Minute (0.009) 19 (0.009) ES (0.006) Spain (0.005) PASS (0.005) Mexico (0.005) del (0.005) Restezchezvous (0.005)
Topic 1	–	0.29	0.000126	6.36	Retweet (0.013) Follow (0.011) dm (0.011) Play (0.009) Send (0.009) Cashapp (0.008) Comment (0.008) Person (0.008) People (0.008) Sex (0.008) Win (0.008) Time (0.008) Pay (0.007) Link (0.007) Money (0.007)
Topic 75	–	0.23	0.000264	21.26	Deliver (0.124) Support (0.100) Sign (0.063) ll (0.063) Official (0.062) Copy (0.062) usp (0.038) Senate (0.029) Representative (0.017) Act (0.017) Stock (0.013) Capolitic (0.008) Capol (0.008) Dump (0.006) Sen (0.006)
Topic 57	–	0.07	0.000098	47.94	Covid (0.093) 19 (0.033) David (0.031) Jorge (0.031) Prior (0.030) Pillow (0.030) Trump (0.030) CDTV (0.029) Mike (0.028) Manu (0.028) Suho (0.028) Foraprior (0.028) Babu (0.028) Lindell (0.028) Bizbizeyeteriz (0.027)

[a]Topics with a good linear regression fit ($R^2 > 80\%$ and p value ≤ 0.05)

Acknowledgements This work has been conducted in part within the EU H2020 project PandeVITA (Pandemic Virus Trace Application for the Effective Knowledge Transfer Between Science and Society Inside the Quadruple Helix Collaboration) with Grant no. 101006316.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Iqbal, M. (2020). Twitter revenue and usage statistics (2020). *Business of Apps*. Retrieved January 5, 2021, from https://www.businessofapps.com/data/twitter-statistics/
2. Fung, I.C.-H., Duke, C. H., Finch, K. C., Snook, K. R., Tseng, P.-L., Hernandez, A. C., & Tse, Z. T. H. (2016). Ebola virus disease and social media: a systematic review. *American Journal of Infection Control, 44*(12), 1660–1671. https://doi.org/10.1016/j.ajic.2016.05.011
3. Tang, L., Bie, B., Park, S.-E., & Zhi, D. (2018). Social media and outbreaks of emerging infectious diseases: a systematic review of literature. *American Journal of Infection Control, 46*(9), 962–972. https://doi.org/10.1016/j.ajic.2018.02.010
4. Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: a systematic review. *American Journal of Public Health, 107*(1), e1–e8. https://doi.org/10.2105/AJPH.2016.303512
5. Xu, P., Dredze, M., & Broniatowski, D. A. (2020). The twitter social mobility index: measuring social distancing practices with geolocated tweets. *Journal of Medical Internet Research, 22*(12), e21499. https://doi.org/10.2196/21499
6. Li, Z., Li, X., Porter, D., Zhang, J., Jiang, Y., Olatosi, B., & Weissman, S. (2020). Monitoring the spatial spread of COVID-19 and effectiveness of control measures through human movement data: proposal for a predictive model using big data analytics. *JMIR Research Protocols, 9*(12), e24432. https://doi.org/10.2196/24432
7. Gill, S., Kinslow, K., McKenney, M., & Elkbuli, A. (2020). Twitter and the credibility of disseminated medical information during the COVID-19 pandemic. *The American Surgeon*. https://doi.org/10.1177/0003134820973391
8. Rashid, M. A., Yip, S. W. L., Gill, D., & Arfeen, Z. (2020). Sharing is caring: An Analysis of #FOAMed twitter posts during the COVID-19 pandemic. *Postgraduate Medical Journal*. https://doi.org/10.1136/postgradmedj-2020-139267
9. Berkovic, D., Ackerman, I. N., Briggs, A. M., & Ayton, D. (2020). Tweets by People with arthritis during the COVID-19 Pandemic: Content and sentiment analysis. *Journal of Medical Internet Research, 22*(12), e24550. https://doi.org/10.2196/24550
10. Gao, Y., Xie, Z., & Li, D. (2021). Electronic cigarette users' perspective on the COVID-19 pandemic: Observational study using twitter data. *JMIR Public Health and Surveillance, 7*(1), e24859. https://doi.org/10.2196/24859
11. Chen, Z., Chen, Z.-Y., Liu, W.-H., Sultan, M. T., & Li, G.-S. (2020). Challenges for plastic surgeons and estheticians during COVID-19: A twitter study. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. https://doi.org/10.1016/j.bjps.2020.12.010
12. Osakwe, Z. T., Ikhapoh, I., Arora, B. K., & Bubu, O. M. (2020). Identifying public concerns and reactions during the COVID-19 pandemic on twitter: A text-mining analysis. *Public Health Nursing*. https://doi.org/10.1111/phn.12843
13. Alshalan, R., Al-Khalifa, H., Alsaeed, D., Al-Baity, H., & Alshalan, S. (2020). Detection of hate speech in COVID-19-related tweets in the arab region: deep learning and topic modeling approach. *Journal of Medical Internet Research, 22*(12), e22609. https://doi.org/10.2196/22609
14. Saha, K., Torous, J., Caine, E. D., & De Choudhury, M. (2020). Psychosocial effects of the COVID-19 Pandemic: Large-scale quasi-experimental study on social media. *Journal of Medical Internet Research, 22*(11), e22600. https://doi.org/10.2196/22600
15. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

16. Shi, W., Liu, D., Yang, J., Zhang, J., Wen, S., & Su, J. (2020). Social Bots' sentiment engagement in health emergencies: A topic-based analysis of the covid-19 pandemic discussions on twitter. *International Journal of Environmental Research and Public Health, 17*(22), 8701. https://doi.org/10.3390/ijerph17228701
17. Boon-Itt, S., & Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on twitter: Sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance, 6*(4), e21978. https://doi.org/10.2196/21978
18. Valdez, D., ten Thij, M., Bathina, K., Rutter, L. A., & Bollen, J. (2020). Social media insights into US Mental health during the COVID-19 pandemic: Longitudinal analysis of twitter data. *Journal of Medical Internet Research, 22*(12), e21418. https://doi.org/10.2196/21418
19. Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. *Journal of Medical Internet Research, 22*(11), e20550. https://doi.org/10.2196/20550
20. Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet allocation for topic modeling on twitter. *PLoS ONE, 15*(9), e0239441. https://doi.org/10.1371/journal.pone.0239441
21. Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. (2020). An "Infodemic": Leveraging high-volume twitter data to understand early public sentiment for the Coronavirus disease 2019 outbreak. *Open Forum Infectious Diseases, 7*(7), ofaa258. https://doi.org/10.1093/ofid/ofaa258
22. Chandrasekaran, R., Mehta, V., Valkunde, T., & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal Infoveillance study. *Journal of Medical Internet Research, 22*(10), e22624. https://doi.org/10.2196/22624
23. Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 Pandemic: Infoveillance study. *Journal of Medical Internet Research, 22*(4), e19016. https://doi.org/10.2196/19016
24. Zheng, H., Goh, D.H.-L., Lee, C. S., Lee, E. W. J., & Theng, Y. L. (2020). Uncovering temporal differences in COVID-19 tweets. *Proceedings of the Association for Information Science and Technology*. https://doi.org/10.1002/pra2.233
25. Littman, J., Chudnov, D., Kerchner, D., Peterson, C., Tan, Y., Trent, R., & Wrubel, L. (2018). API-based social media collecting as a form of web archiving. *International Journal on Digital Libraries, 19*(1), 21–38. https://doi.org/10.1007/s00799-016-0201-7
26. WHO. (2021). Timeline: WHO's COVID-19 Response. *World Health Organization*. Retrieved January 5, 2021, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline
27. Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. https://doi.org/10.1145/2133806.2133826
28. McCallum, A. K. (2002). MALLET: A Machine learning for language toolkit. Retrieved August 9, 2020, from http://mallet.cs.umass.edu
29. Drosatos, G., & Kaldoudi, E. (2020). A probabilistic semantic analysis of ehealth scientific literature. *Journal of Telemedicine and Telecare, 26*(7–8), 414–432. https://doi.org/10.1177/1357633X19846252
30. Krovetz, R. (1993). Viewing morphology as an inference process. *Proceedings Annual International ACM SIGIR Conference on Research And Development In Information Retrieval*. https://doi.org/10.1145/160688.160718
31. Lambda. (2011). Lists of Stopwords. *Text Categorization Project*. Retrieved January 5, 2021, from https://github.com/lam8da/text-categorization
32. Korenčić, D., Ristov, S., & Šnajder, J. (2018). Document-based topic coherence measures for news media text. *Expert Systems with Applications, 114*, 357–373. https://doi.org/10.1016/j.eswa.2018.07.063
33. Chappell, B. (2020). Coronavirus updates: Italy reports 650 cases and 17 deaths. *NPR*. Retrieved January 5, 2021, from https://www.npr.org/sections/goatsandsoda/2020/02/27/809984502/coronavirus-update-japan-is-closing-all-its-schools-for-a-month?t=1605471829289
34. Ehanire, O. (2020). First case of corona virus disease confirmed in Nigeria. *Nigeria Centre for Disease Control*. Retrieved January 5, 2021, from https://ncdc.gov.ng/news/227/first-case-of-corona-virus-disease-confirmed-in-nigeria

35. Anioke, O. (2020). COVID-19: Ebonyi Bans Burials, weddings for one month. *The Nation*. Retrieved January 5, 2021, from https://thenationonlineng.net/covid-19-ebonyi-bans-burials-weddings-for-one-month/
36. Akinyemi, D. (2020). Coronavirus: Kwara shuts down schools indefinitely from monday. *Vanguard Media Limited*. Retrieved January 5, 2021, from https://www.vanguardngr.com/2020/03/coronavirus-kwara-shuts-down-schools-indefinitely-from-monday/
37. Ojerinde, D. (2020). Coronavirus: Lagos announces closure of schools. *The Punch Newspaper*. Retrieved January 5, 2021, from https://punchng.com/breaking-coronavirus-lagos-announces-closure-of-schools/
38. Gettleman, J., & Schultz, K. (2020). Modi orders 3-week total lockdown for all 1.3 billion Indians. *The New York Times*. Retrieved January 5, 2021, from https://www.nytimes.com/2020/03/24/world/asia/india-coronavirus-lockdown.html
39. Bursztynsky, J. (2020). NIH doctor says 25% of coronavirus cases in China are 'very serious,' requiring 'intensive care.' *CNBC*. Retrieved January 5, 2021, from https://www.cnbc.com/2020/02/03/nih-dr-anthony-fauci-25percent-of-china-coronavirus-cases-very-serious.html
40. Brady, J. S. (2020). Remarks by president trump, vice president pence, and members of the Coronavirus Task Force in Press Conference. *The White House*. Retrieved January 5, 2021, from https://www.whitehouse.gov/briefings-statements/remarks-president-trump-vice-president-pence-members-coronavirus-task-force-press-conference/
41. Trump, D. J. (2020). Statement by the president. *The White House*. Retrieved January 5, 2021, from https://www.whitehouse.gov/briefings-statements/statement-by-the-president-38/
42. Wikipedia. (2020). 2020 Wisconsin democratic presidential primary. *Wikipedia*. Retrieved January 5, 2021, from https://en.wikipedia.org/wiki/2020_Wisconsin_Democratic_presidential_primary
43. Wikipedia. (2020). 2020 United States anti-lockdown protests. *Wikipedia*. Retrieved January 5, 2020, from https://en.wikipedia.org/wiki/2020_United_States_anti-lockdown_protests
44. Noor, P. (2020). 'Please don't Inject bleach': Trump's wild coronavirus claims prompt disbelief. *The Guardian*. Retrieved January 5, 2021, from https://www.theguardian.com/us-news/2020/apr/24/trump-disinfectant-bleach-coronavirus-claims-reaction
45. Wikipedia. (2021). COVID-19 pandemic on diamond princess. *Wikipedia*. Retrieved January 5, 2021, from https://en.wikipedia.org/wiki/COVID-19_pandemic_on_Diamond_Princess
46. Arab News. (2020). Saudi Arabia announces first case of coronavirus. *Arab News*. Retrieved January 5, 2021, from https://www.arabnews.com/node/1635781/saudi-arabia
47. Austin, H. (2020). U.K. Prime Minister Boris Johnson tests positive for coronavirus. *NBC News*. Retrieved January 5, 2021, from https://www.nbcnews.com/news/world/u-k-prime-minister-boris-johnson-tests-positive-coronavirus-n1170196
48. Harding, L., Mason, R., Sabbagh, D., Busby, M., Campbell, D., & Bowcott, O. (2020). Boris Johnson and coronavirus: The inside story of his illness. *The Guardian*. Retrieved January 5, 2021, from https://www.theguardian.com/world/2020/apr/17/boris-johnson-and-coronavirus-inside-story-illness
49. BBC. (2020). Coronavirus: Armed robbers steal hundreds of toilet rolls in Hong Kong. *BBC News*. Retrieved January 5, 2021, from https://www.bbc.com/news/world-asia-china-51527043
50. Guardian. (2020). Coronavirus: Woolworths moves to ration toilet paper as panic buying empties shelves. *The Guardian*. Retrieved January 5, 2021, from https://www.theguardian.com/world/2020/mar/03/toilet-paper-is-trending-in-australia-heres-why
51. RNZ. (2020). Crowds rush to some supermarkets as Covid-19 enters NZ. *Radio New Zealand*. Retrieved January 5, 2021, from https://www.rnz.co.nz/news/national/410658/crowds-rush-to-some-supermarkets-as-covid-19-enters-nz
52. Merrifield, R. (2020). Question Time: Health secretary pleads with brits to stop panic buying over coronavirus. *Mirror*. Retrieved January 5, 2021, from https://www.mirror.co.uk/news/politics/question-time-health-secretary-pleads-21642304
53. Shulzhenko, D. (2020). Coronavirus causes panic buying in Ukraine, but few shortages. *KyivPost*. Retrieved January 5, 2021, from https://www.kyivpost.com/ukraine-politics/coronavirus-causes-panic-buying-in-ukraine-but-few-shortages.html
54. Guo, J., Radloff, C. L., Wawrzynski, S. E., & Cloyes, K. G. (2020). Mining twitter to explore the emergence of COVID-19 symptoms. *Public Health Nursing, 37*(6), 934–940. https://doi.org/10.1111/phn.12809

55. CDC. (2020). Symptoms of coronavirus. *Centers for Disease Control and Prevention*. Retrieved January 5, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html
56. Wang, J., Zhou, Y., Zhang, W., Evans, R., & Zhu, C. (2020). Concerns expressed by chinese social media users during the COVID-19 pandemic: Content analysis of sina weibo microblogging data. *Journal of Medical Internet Research, 22*(11), e22152. https://doi.org/10.2196/22152
57. Niburski, K., & Niburski, O. (2020). Impact of trump's promotion of unproven COVID-19 treatments on social media and subsequent internet trends: Observational Study. *Journal of Medical Internet Research, 22*(11), e20044. https://doi.org/10.2196/20044
58. Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., & Thwaites, G. (2020). Economic uncertainty before and during the COVID-19 pandemic. *Journal of Public Economics, 191*, 104274. https://doi.org/10.1016/j.jpubeco.2020.104274