JCOM

Politics, economy and society in the coverage of COVID-19 by elite newspapers in US, UK, China and Brazil: a text mining approach

Luiz Felipe Fernandes Neves and Luisa Massarani

Abstract   We analyzed 95,970 stories on COVID-19 published in 2020 by newspapers in US, UK, China and Brazil — countries marked by controversial management of the crisis. Through a text mining approach, we identified main topics, subjects, actors and the level of attention. The coverage was politicized in *The New York Times* and *Folha de S. Paulo*; focused on health aspects in *The Guardian*; and emphasized the economic situation in *China Daily*. In this sense, the pandemic has motivated a deeper approach to the multiple dimensions of science and health, pointing to a broader perspective of science communication.

Introduction

The media usually follows closely — and with a certain apprehension — the emergence of new diseases around the world [Ihekweazu, 2016; da Silva Medeiros & Massarani, 2010; Roche & Muskavitch, 2003; Tian & Stewart, 2005; Van den Bulck & Custers, 2009]. With the SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), it took only a few weeks for a "pneumonia of unknown cause" [WHO, 2020a], whose first cases were detected in Wuhan, China, to receive international media attention for its potential for large-scale spread. Negative headlines accompanied the escalation of the disease [Aslam, Awan, Syed, Kashif & Parveen, 2020; Sacerdote, Sehgal & Cook, 2020], then called COVID-19 (Coronavirus Disease 2019), and expressions such as *fear*, *afraid* and *killer virus* became part of the news [Wahl-Jorgensen, 2020]. The production of journalistic content also sought to meet the growing demand for information, consumed by an audience that was confined due to the strict social isolation measures adopted in several countries [Casero-Ripollés, 2020; Liu et al., 2020; Fitera, Abuín-Vences & Sánchez, 2021].

The current study is inserted in this context of intense media coverage and massive production of news provoked by the pandemic, aiming to address the following research questions:

RQ 1: What topics, subjects and actors were present in the press coverage related to COVID-19 in countries of different regions of the world throughout 2020, and to what extent were they addressed over the months?

RQ 2: What can the press attention devoted to these topics, subjects and actors throughout 2020 reveal about the multiple dimensions of science and health and their approaches from the perspective of science communication?

Our objective was to analyze, comparatively and longitudinally, the full coverage of important media outlets between January and December 2020, in order to understand the different trajectories during this period. We selected elite newspapers from strategic countries from an international or regional geopolitical perspective, which were initially or strongly affected by COVID-19, and which were marked by controversial government managements of the health crisis. They are *The New York Times* (United States), *The Guardian* (United Kingdom), *China Daily* (China) and *Folha de S. Paulo* (Brazil). To account for this large quantity of data, we used an exploratory text mining approach [Péladeau, 2021] to analyze 95,970 news stories with mentions of COVID-19 published in 2020 by these newspapers.

In this sense, this study contributes to filling a gap in research about the press coverage of COVID-19, especially in two aspects: firstly, with a big data methodological approach, we overcame some limitations inevitably resulting from a sample selection, mainly because it is a massive and diverse database; secondly, we included outlets from different geographic, political and economic perspectives, inserted in different media systems (corporate and state). In this broader perspective, we were able to verify how the media in different contexts, with distinct news values, reflected the social, political and economic dimensions of a public health emergency.

When science and journalism meet

Studies in different countries indicate that science, to a greater or lesser extent, has always been present in the media, either in the pages of traditional newspapers and magazines, or publications specifically focused on scientific themes aimed at a wider audience [Dunwoody, 2008; de Castro Moreira & Massarani, 2002; Patairiya, 2004]. Throughout the 20th century, the constitution of this area of journalism became more evident with the first "science correspondents" [Becker, 2011], the increase in the number of reporters dedicated and specialized in scientific topics — and consequently, of the stories on this subject –, and the gathering of these professionals in specific associations [Bauer & Gregory, 2007; Dunwoody, 2008].

Bauer and Gregory [2007] state that journalistic activity consists of investigating scientific issues through reliable sources, developing expertise in the subject and critical response. This notion is based on the assumption that the media is the classic locus of public opinion and that, in terms of the professional ethos, the journalist claims to be a trustee of public interests. More recently, rapid technological and economic changes, which offer new ways of engaging with sources and audiences, have expanded the science journalist's area of expertise

from the traditional roles of reporter, conduit, watchdog and agenda-setter to those of curator, convener, public intellectual and civic educator [Fahy & Nisbet, 2011].

In the systematization proposed by Bauer and Gregory [2007] regarding the changes in the ways of communicating science, the late 1980s and early 1990s are marked by the "medicalization" of science news, in the wake of innovative research in biomedical and biotechnology areas. Some of the characteristics of this media coverage are the traditional news pegs, focusing more on the discoveries than on the processes of science, and the privileging of values such as opportunity, conflict and novelty, in order to attract the audience's attention [Dunwoody, 2008].

The outbreaks and epidemics of the early of the 21st century, especially of airborne diseases, such as Severe Acute Respiratory Syndrome (SARS), Avian Influenza (H5N1), Influenza A-H1N1, Ebola and Middle East Respiratory Syndrome (MERS), have highlighted other functions of science journalism and the relationship between media and science communication. Research shows that, in addition of being a source of science and health information, the media can contribute not only to the public understanding of a health crisis [Van den Bulck & Custers, 2009; Vasterman & Ruigrok, 2013; Wallis & Nerlich, 2005], but also act on individual emotion and resilience [Giri & Maurya, 2021] and even on the public's adoption — or not — of certain preventive measures [Bonneux & Van Damme, 2006; W. Chen & Stoecker, 2020].

These and other aspects could be studied in the context of COVID-19. A survey in Australia in the early stages of the outbreak showed that people who said they were worried about the disease reported consuming more news than before [Park, Fisher, Lee & McGuinness, 2020]. The same reality was found in the United States, where people who had generally stayed away from information reconnected with the news [Casero-Ripollés, 2020]. In Germany, a study identified that the frequency, duration and diversity of media exposure were positively associated with more symptoms of depression and anxiety [Bendau et al., 2020]. In the United Kingdom, more than half of survey respondents said the media helped them understand the pandemic and how to combat it. However, 27% considered that the media exaggerated the crisis, and a third of them believed that the situation in the country had been made worse by the way the media reported it [Nielsen, Kalogeropoulos & Fletcher, 2020].

Regarding journalistic content, the trend has been to analyze the coverage of specific topics related to COVID-19, such as the vaccine. A study observed that newspapers in the US, UK and Brazil characterized the quest to develop a vaccine as a race between nations and blocs historically marked by economic, political and ideological disputes, such as the United States, Europe, China and Russia [Massarani & Neves, 2021]. In Nigeria, vaccine development received moderate media attention, with concise stories and the predominance of official sources [Asogwa, 2021]. A number of other studies have focused on media coverage of the pandemic on specific topics, such as immigration [Shomron, 2021], the economy [Atri, Kouki & imen Gallali, 2021], tourism [H. Chen, Huang & Li, 2020], childhood [Katz & Cohen, 2021], old age [Morgan, Wiles, Williams & Gott, 2021] and health workers [Lynch, Evans, Ice & Costa, 2021]. The present study extends this approach by analyzing the full coverage of the COVID-19 pandemic, without restricting it to a specific theme, over a considerably longer period (most studies

focus on the initial months of the pandemic, while we analyzed a full year), and in a cross-national perspective.

| **Material and methods** | *Selection of newspapers and countries* |

Guided by the objective of undertaking a longitudinal and comparative analysis of news coverage of the COVID-19 pandemic, we selected countries that, in addition to being important nations from a regional and international geopolitical perspective, were significantly impacted by the disease and that were marked, to a greater or lesser extents, by controversial management of the crisis by their government leaders. Thus, we intended to observe if and how the media articulated the coverage of a health emergency to social, political and economic issues and how the news reflected the dynamics of the pandemic in different nations. In this regard, the United States, the United Kingdom and Brazil stood out for the controversies involving, respectively, then President Donald Trump, Prime Minister Boris Johnson and President Jair Bolsonaro. At different times, these politicians actively resisted measures supported by scientific evidence, having been criticized by members of scientific communities in their countries and even by their own advisors [The Lancet, 2020; Yamey & Gonsalves, 2020; Pollock, Roderick, Cheng & Pankhania, 2020].

Although China managed to control the crisis with strict monitoring and social isolation, the initial conduct of the outbreak by the Chinese President Xi Jinping's government was also criticized for a lack of transparency and a possible delay in adopting containment measures before the coronavirus spread throughout the country [Ang, 2020]. Moreover, the inclusion of China in the current study would add an important comparative possibility with a non-Western context, based on the reality of the first country to record cases of people infected by the new coronavirus.

Regarding the choice of media outlets, we selected to analyze "print media" newspapers.[1] Despite being part of a communication environment in growing dispute, newspapers still maintain their status as a source of credible information, especially those historically consolidated and with national and international prestige. Printed newspapers (and their online versions) make up the so-called traditional or legacy media [Casero-Ripollés, 2020], and the most important titles in the news market are considered by the scholarly literature as prestigious or elite newspapers. These are newspapers with a large circulation, national reach, with resources for more specialized coverage, and that enjoy a reputation that increases the probability of more balanced news [Carpenter, 2007; Lacy, Fico & Simon, 1991]. Furthermore, elite newspapers are known to influence the agenda of other media outlets [Carpenter, 2007; K. Wu, 2021], which means that, by analyzing them, we can get a good idea of news coverage in general. Among the most prestigious international titles, we favored those that made all their content available online, since the extraction technique was applied from the URLs. In this way, we also

---

[1]In the context of a hybrid media system [Chadwick, 2013], the previously well-defined divisions between the platforms used by the different media outlets are increasingly converging. In this sense, when we use the expression "printed media", we refer to newspapers that originated as printed publications, but we considered for the composition of the corpus all the content published by them on the internet, which goes far beyond the stories of the printed version.

guaranteed the collection of all textual stories, as close as possible to the full coverage. Thus, we reach the newspapers *The New York Times* (United States), *The Guardian* (United Kingdom), *China Daily* (China) and *Folha de S. Paulo* (Brazil).

Collection and extraction

We started by collecting the URLs available on the websites of the four newspapers with content related to the health crisis, published from January 1 to December 31, 2020. Thus, it would be possible to encompass the emergence of the new disease, the global spread and the beginning of vaccination. For *The New York Times*, *China Daily* and *Folha de S. Paulo*, we used the websites' search engines, as they allow us to define specific dates and return all publications. The search was performed with the keywords *coronavirus* and *covid*, and we used the Octoparse software to extract the URLs. For *The Guardian*, this option was not possible, as the search tool on its website does not have the same functionality. On the other hand, the British newspaper created a section with all publications related to COVID-19, from which we collected the URLs with the Parsehub software. At the end of this first stage, we obtained a total of 116,044 URLs related to some type of content mentioning the new coronavirus or the disease caused by it. We used this raw database for the first part of the study, focusing on quantitative analysis.

To extract the texts related to the URLs, we used packages developed for the Python programming language — Beautifulsoup for *The New York Times* and Newsplease for *The Guardian*, *China Daily* and *Folha de S. Paulo*. From there, we started the database refinement step. First, we reduced the corpus to textual content only, as this is the modality supported by the analysis software. Therefore, URLs with exclusively audiovisual material, such as photography, video, podcast and infographic, were excluded. Second, we sought to reduce the duplication of content as much as possible, even content published with different URLs, in different sections of the websites, or even in summarized versions in the form of a newsletter. Based on this protocol, we stabilized the corpus for the second part of the study, focusing on a qualitative-quantitative analysis, of 95,970 journalistic publications, including news stories, reports, articles, columns and editorials, distributed as follows: 22,962 from *The New York Times*, 21,497 from *The Guardian*, 26,669 from *China Daily* and 24,842 from *Folha de S. Paulo* (the full dataset containing all the collected texts is available via the hyperlink in the Data Availability section).

Analytical procedures

Given the large amount of data and the objective of performing a longitudinal analysis during a year of intense textual production, we opted for an exploratory text mining approach [Péladeau, 2021]. This gave us the advantage of working as closely as possible with the full coverage. We used the WordStat software, developed by Provalis Research, which has been used in studies of large textual databases [Al-Rawi, Al-Musalli & Fakida, 2021; Al-Rawi, Grepin et al., 2021; Al-Rawi, Kane & Bizimana, 2021; Aweisi et al., 2021; Al-Zaman & Khan, 2021; Cogburn, 2019; Cogburn, Thomas & Lai, 2020; Massarani & Neves, 2021]. The software allows analyzing qualitative data, in the form of unstructured texts, combining quantitative analysis techniques, natural language processing and data mining [Péladeau, 2021]. For the present study, we mostly used the following

resources: topic modeling; weighted frequency of words and phrases (TFxIDF); longitudinal analysis through cross-tabulation between the incidence rate of the topics during all the 12 months of 2020.

First, we carried out the text pre-processing, excluding words with no semantic value (stopwords), and the post-processing, with the establishment of a minimum frequency for inclusion in the analysis (we included words that appeared at least 30 times in the corpus). Next, the primary tool used for the analysis was topic modeling. The program calculates a word frequency matrix per document (which, in our corpus, is equivalent to each story) and applies a factor analysis to extract a small number of factors — the topics [Provalis Research, n.d.]. Each topic is composed of words and n-grams (or phrases, as WordStat calls it)[2] that have an underlying statistical relationship [Cogburn et al., 2020]. Furthermore, depending on the context, words and phrases can be included in more than one topic and given different weights, according to how much they contribute to a given topic [Péladeau, 2021].

Although it seems like an automated task, it requires exhaustive tests to verify the consistency of topics and the context of each word or phrase in the corpus. Therefore, researchers must be extremely familiar with the content, which was possible with the daily reading of the main news of the four analyzed newspapers. After the tests, the software was adjusted to calculate 30 topics from each newspaper (coherence > 0.3). Then, we analyzed in detail each of the 120 resulting topics (30 from each newspaper) and the more than 1,600 words and phrases that comprise them. Through the Keyword-in-Context tool, each word and phrase can be visualized precisely in the paragraph and in the story in which they are found, which allows a better understanding of their meaning. As an example, we present the result of the processing and the decisions made by us in two topics of *The New York Times* (Table 1). Although the software algorithm offers a label suggestion, we renamed them according to the scope of our study into a general topic and specific topics, some of them adapted from Mach et al. [2021] and Fitera et al. [2021]. The software also allows to exclude topics with no semantic value and to merge topics of approximate semantic value (see Data Availability for the full topic modeling).

After applying this protocol to the corpus of each newspaper, we obtained five general topics and their respective specific topics (or subtopics):

a) EPIDEMIOLOGY & SCIENCE: topic related to the several health and scientific aspects linked to the COVID-19 pandemic. Specific topics: disease spreading; mortality; preventive measures; health care; vaccine; research, institutions and experts.

b) GOVERNMENT & POLITICS: we include in this topic references to the governments and political authorities of the respective countries, in their different spheres (executive, legislative and judiciary). Specific topics: US government; UK government; Chinese government; Brazilian government; elections; congress; judiciary.

---

[2]N-gram is a sequence of n items identified in a textual sample. In this case, an n-gram is a grouping of two or more words (e.g., *coronavirus pandemic*), thus different from isolated words (*coronavirus* or *pandemic*). In WordStat, the n-gram is called a phrase.

Table 1. Example of topic modeling in *The New York Times*.

KEYWORDS	tested; positive; test; tests; testing; symptoms; infected; results; virus			
	tested positive; public health; health officials; disease control; centers for disease control; tested positive for the coronavirus; tested positive for the virus; health experts; coronavirus cases; infectious disease.			
TOPIC (software suggestion)	**COHERENCE**	**% CASES**	**GENERAL TOPIC (assigned by researchers)**	**SPECIFIC TOPIC (assigned by researchers)**
Tested Positive	0.514	36.59%	Epidemiology & Science	Disease spreading
KEYWORDS	democrats; democratic; biden; republican; election; republicans; trump; voters; senator; presidential; campaign; senate; president; vote; policy; party; white; voting; race; house; sanders; congress; administration; ballots; washington; americans.			
	president trump; vice president.			
TOPIC (software suggestion)	**COHERENCE**	**% CASES**	**GENERAL TOPIC (assigned by researchers)**	**SPECIFIC TOPIC (assigned by researchers)**
Democrats	0.668	60.81%	Government & Politics	Elections

KEYWORDS: set of words and phrases related to the topic, presented in descending order of relevance (the first words have greater weight than the last ones).
TOPIC: label suggested by the software algorithm.
COHERENCE: weighted mean of the correlation between the topic words.
% CASES: percentage of stories with at least one of the keywords.
GENERAL TOPIC: general label assigned by researchers after analyzing the topic.
SPECIFIC TOPIC: specific label assigned by researchers after analyzing the topic.

c) ECONOMY: topic composed of issues related to the economic impacts of the pandemic and fluctuations in market and industry. Subtopics related to unemployment and assistance measures taken by different countries were also included here. Specific topics: economic impacts; commerce; private sector; unemployment benefits; supply chain; international trade; economic recovery.

d) SOCIETY: in this topic, we gathered subtopics that emerged in the modeling, about societal issues that were directly or indirectly affected by the pandemic — mainly the challenges of remote learning, in addition to events that are tangential to the health issue but that are intertwined with the context of crisis. Specific topics: educational impacts; sports; art and entertainment; George Floyd; climate change.

e) EXTERNAL AFFAIRS: here, we grouped the external references to each analyzed country (e.g., references to the US presidential elections in the *China Daily* corpus were classified as external affairs). Specific topics: China; United States; United Kingdom; Europe; Hong Kong; Australia; African countries.

As a complementary analytical strategy, we also proceeded with an ordering of the most frequent words and phrases in the corpus. They were ordered according to the TFxIDF value, a statistical relative measure that weights the frequency of a given word with the number of stories in which it appears. It differs from absolute frequency, which simply sums up the number of times a term appears in the corpus. TFxIDF reduces the weight of words that do not contribute to interpreting possible implicit meanings because they are very frequent and already expected. In this way, the calculation reveals the most "important" terms in the corpus

[Cogburn, 2019]. Therefore, whenever we refer to the frequency of a word or phrase, it should emphasize that it is not the absolute frequency (simple counting the number of times a term appears in the corpus), but the weighted frequency. Although topic modeling already provides a list of words and phrases related to each topic, TFxIDF helped us to see more clearly the emphasis given to certain terms to the detriment of others, or that perhaps were not included in the topics by the factor analysis.

Finally, to visualize the results and proceed with the longitudinal analysis, we cross-tabulated the topics and the 12 months of 2020 in order to verify the incidence of each of them throughout a full period of time of one year. Following Cogburn et al. [2020], the topics were arranged according to the rate per 10,000 words — a relative measure adopted to avoid biasing the results due to stories of very different sizes and the discrepancy between months with more and fewer stories.

Results

For an initial overview of the media attention generated by the health emergency caused by the SARS-CoV-2 coronavirus throughout 2020, Figure 1 presents the total number of publications found with the keywords *coronavirus* and *covid* on the websites of the newspapers *The New York Times*, *The Guardian*, *China Daily* and *Folha de S. Paulo* — a total of 116,044 URLs. There is a clear contrast between the coverage of the Chinese newspaper, which has the highest number of publications in February, and that of the other newspapers, which present curves with similar evolutions, with a jump in the number of publications between February and March, and peak in April (Figure 1). Jointly, all four newspapers provided a monthly average of 9,600 URLs that point to content with mentions of the new coronavirus.
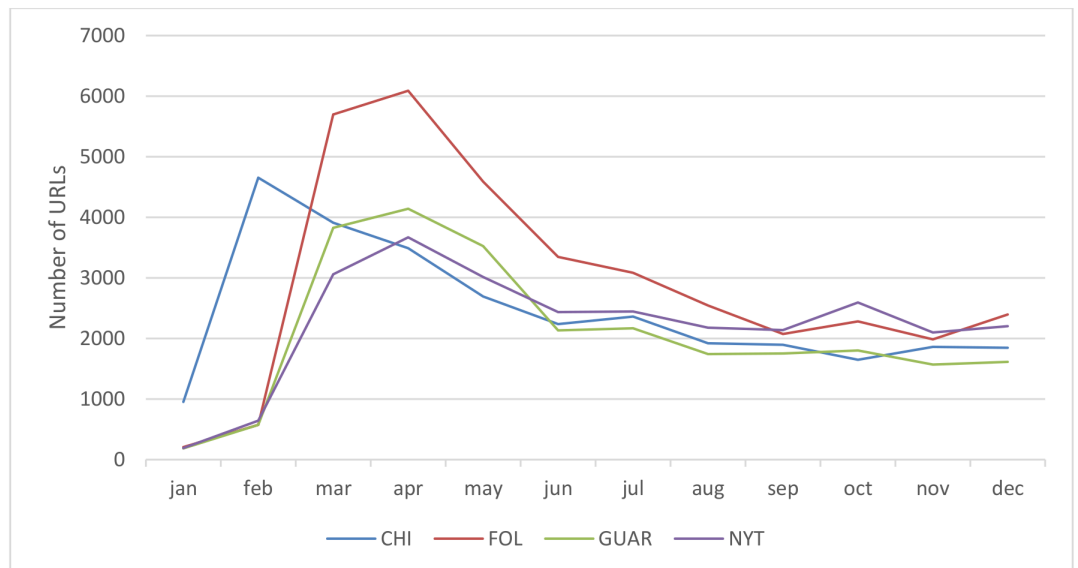
Figure 1. URLs with COVID-19 mentions (2020).

Figures 2–5 show the incidence (rate per 10,000 words) of the different topics in each newspaper. In *The New York Times* (Figure 2), the topic Epidemiology & Science gradually gives way to an emphasis on political issues from July onwards. The change coincides with the strengthening of the electoral dispute in the United

States between then President Donald Trump, candidate for reelection, and Joe Biden, in a process marked by the suspension of in-person campaign events, the anticipation of votes (part of them by mail) and record voters turnout, culminating in the Democrat's victory in November 2020. Only from then on does health coverage stand out again. The calculation of the weighted frequency of words and phrases confirms the presence of political issues in the coverage of the United States newspaper, with an emphasis on the elections. *Trump* and *biden*, for example, are the two words with the highest TFxIDF value in this corpus. Also highlighted are *election*, *voters* and *campaign*. The list of phrases highlights *white house*, *president trump* and *trump administration* (Table 2 — here, we only provided the ten words and phrases with the highest TFxIDF from each newspaper; see Data Availability for the full list).

Next, the topics Economy and Society trace competing trajectories. This last topic, which in all newspapers is essentially composed of the educational impacts caused by the pandemic, experienced a slight increase from May onwards in *The New York Times* as a reflection of the racial issue in the United States news. This is basically due to the coverage of the death of George Floyd, a black man who was asphyxiated by a police officer in the city of Minneapolis in May 2020, which sparked several protests in the United States, mostly gathered under the slogan of the Black Lives Matter movement. Consequently, this fact was inserted in the electoral debate, in the concern of health authorities regarding the risk of contagion due to the protests, and in the discussions about the unequal impact of the COVID-19 pandemic on minority communities. The corpus of the United States newspaper is the only one in which words and phrases such as *black*, *george floyd*, *black lives*, *black people* and *black lives matter* appear among the 100 most frequently appearing terms.

The low frequency of the topic External Affairs shows how the coverage of the United States newspaper is essentially focused on the country's internal issues. This topic only appears prominently in early 2020, when the health crisis was restricted to China. Still, it is observed that the antagonism with the Chinese government was constantly — and politically — exploited by Donald Trump, which, in the coverage of *The New York Times*, may be associated with the fact that the word *china* is the fourth highest frequency.

Issues related to governmental and political affairs, authorities and political institutions are strongly prominent in the Brazilian newspaper. It was the most frequently appearing topic in *Folha de S. Paulo's* coverage, in direct competition with the topic Epidemiology & Science (Figure 3). It is important to remember that Brazil also went through an electoral process in 2020, but for mayors at the municipal level. The election, held in November, coincides with the peak incidence of the topic Government & Politics. The frequency of words follows what was verified in the United States newspaper: the concentration of most mentions in the figure of the country's president. The words *bolsonaro*, *president* and *government* have the three highest TFxIDF values. *Jair bolsonaro* is also the most important phrase in this corpus, with emphasis on *president jair bolsonaro* and *federal government*.

Despite being at different levels, the other topics maintain a regular presence in the Brazilian newspaper. The topic Economy is formed by the combination of four
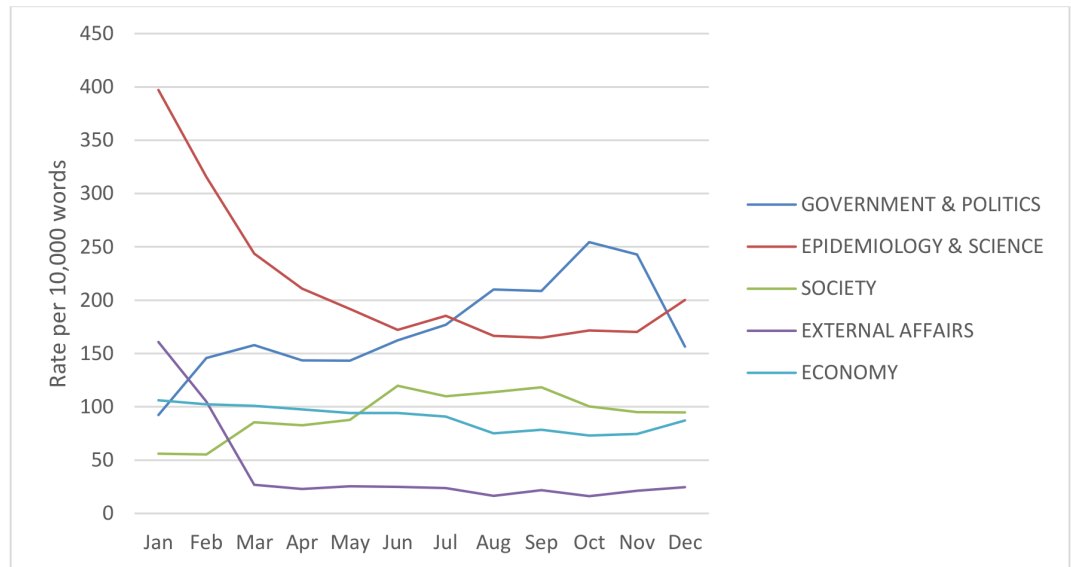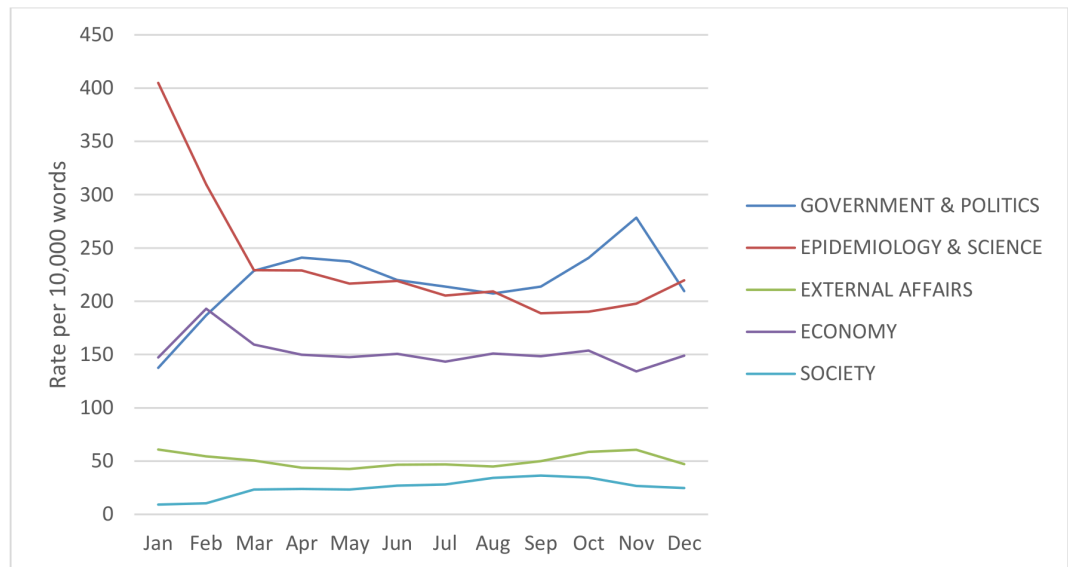
Figure 2. Distribution of topics in *The New York Times* coverage (2020).

specific topics with words and phrases related to the performance of the economy, fluctuations in the financial market and the impacts of the pandemic, especially on the labor market, which gave rise to a great public debate on the need for government assistance to workers affected by the closure of their businesses and unemployment. In this sense, in the set of most relevant words and phrases in the corpus of *Folha de S. Paulo*, we note terms commonly linked to this topic, such as *million*, *billion*, *companies*, *economy*, *emergency aid*, *central bank* and *paulo guedes* (Brazilian Minister of Economy).

In the lower portion of the Figure 3, there are the less prominent themes. The topic External Affairs consists mainly of references to the United States, China and the United Kingdom. The emphasis, however, is on the first — a finding corroborated by the emphasis on the terms *trump* and *united states* in the coverage of *Folha de S. Paulo*. The topic Society is the least frequently occurring, focused almost exclusively on the impacts of the pandemic on education.

The British newspaper is characterized by coverage essentially focused on health issues, observed by the predominance of this topic throughout the year, in the upper portion of the Figure 4). It is worth noting that, in comparison with the others, the corpus of *The Guardian* was the one that brought together the most subtopics specifically aimed at preventing the disease, such as wearing masks, hygiene measures, social distancing, attention to symptoms, flight restrictions, quarantine and lockdowns. The more focused attention on health aspects can also be seen in the most relevant words and phrases, such as *care, cases, lockdown, testing, social distancing, stay at home, contact tracing, face masks, test and trace* and *protective equipment*. It should also be mentioned that the British newspaper is the only one in which the word *vaccine* is among the top ten TFxIDF.

Competition with the topic Epidemiology & Science only takes place in the first month of 2020, when news about the new coronavirus was still an external issue. As in previous newspapers, the topic External Affairs gradually decreases. In *The*

Figure 3. Distribution of topics in *Folha de S. Paulo* coverage (2020).

Guardian, this external coverage mainly focuses on the United States and Australia. This can be explained by the editorial division of the newspaper, which has the UK, US, Australia and International editions. It is interesting to note that, in the corpus of the British newspaper, the word *trump* has the highest TDxIDF value, well above the word *johnson*, which appears in the 24th position. *Prime minister* and *boris johnson* appear prominently only in the set of most frequent phrases.

With comparatively less attention in the figure of the British prime minister and without the need to cover internal electoral processes, as in the United States and Brazil, in *The Guardian* the topic Government & Politics is among the least frequently occurring, just above the topic Society. Economic coverage regularly appears during the months, consisting of subtopics more focused on local issues, such as labor market and unemployment.

In the context of the Chinese newspaper, the situation is even more peculiar (Figure 5). During most of the year, the coverage focuses more frequently on the Economic topic, which even starts to dominate from April on. It is important to highlight that, at that time, the pandemic was already under control in the country. The topics Epidemiology & Science and Economy trace symmetrically opposite curves — the first, descending, and the second, ascending. Not coincidentally, in *China Daily*, topic modeling revealed a greater number of specific topics related to economic issues and on more diverse subjects than in other newspapers, such as the global supply chain, international trade and the resumption of economic growth. As in the Brazilian newspaper, this perception is reinforced by the more frequent presence of terms such as *percent*, *economic*, *trade*, *economy*, *million* and *percent year on year*.

The other topics sustain a regular presence in the Chinese newspaper during the year. In the same way as in the economic news, the topic External Affairs, although it has more references to the United States, also groups more diverse subtopics,
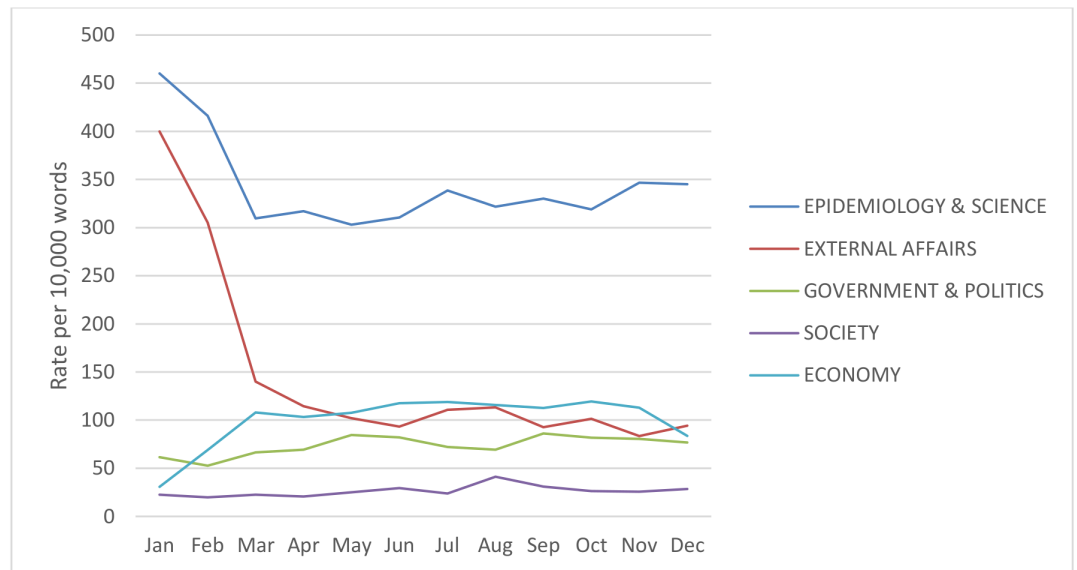
Figure 4. Distribution of topics in *The Guardian* coverage (2020).

with mentions of Hong Kong,[3] the United Kingdom, and European and African countries. Here we see an editorial similarity with *The Guardian*, as *China Daily* has the China, Asia, Hong Kong and Global editions. Finally, Figure 5 shows the low occurrence, in relation to other subjects, of the topics Government & Politics and Society, with stable evolution and practically overlapping in 2020. Unlike other newspapers, the name of the Chinese President is not among the most frequent — the word *xi* is in the 79th position and the phrase *president xi* is in the 31st place. Nevertheless, it is important to note that *China Daily* has a specific section called Xi's Moments, featured on the cover of the website, intended exclusively for personal coverage of the leader.

Discussion

The progressive coverage related to COVID-19 of the four newspapers clearly shows that the production of content, especially in the early months of 2020, followed the dynamics of the spread of the disease from China — where the first cases were recorded — to other countries. To a certain extent, the advance in coverage also coincided with the WHO's warnings for the escalation of infection on a global scale, with the declaration of a public health emergency of international concern on January 30, 2020, and a pandemic on March 11, 2020 [WHO, 2020b]. Our results corroborate what was observed in other contexts [Ducharme, 2020; Liu et al., 2020; Mach et al., 2021] and can be framed in the proposal by Ng, Chow and Yang [2021], who classified the initial phase of media coverage into the pre-pandemic, early pandemic, and peak-pandemic.

With unprecedented duration and severity in recent history, the COVID-19 pandemic was not restricted to the coverage of health aspects, being addressed by the media in its other dimensions. In this sense, our results align with those of Mach et al. [2021] and Fitera et al. [2021], who also observed the gradual

---

[3]Although Hong Kong is a Special Administrative Region of China, we classify the *China Daily* references to this territory under the topic External Affairs, as there is a specific edition of the newspaper for it.
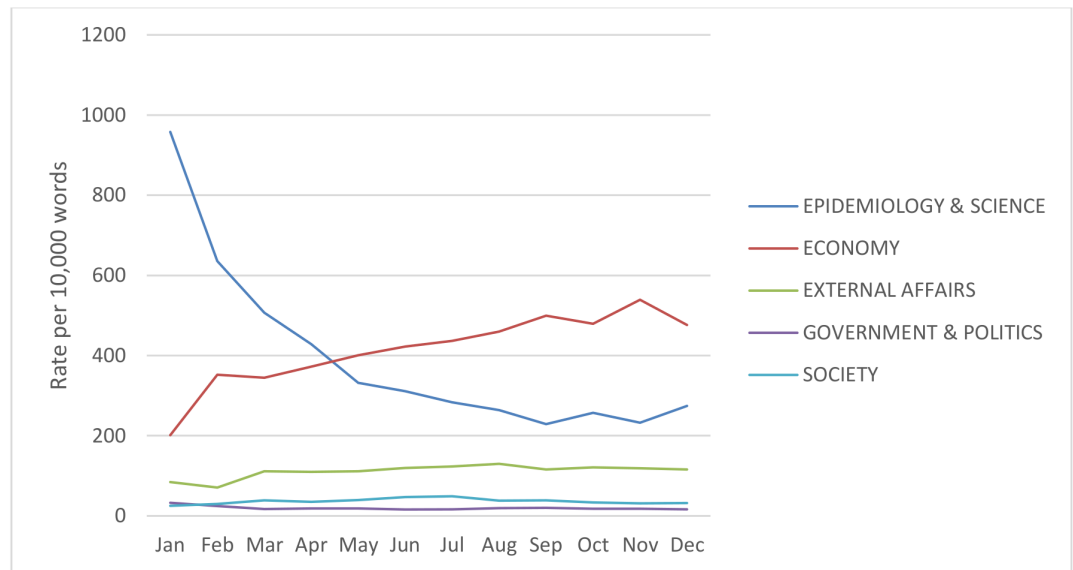
Figure 5. Distribution of topics in *China Daily* coverage (2020).

Table 2. Ten words and phrases with the highest TFxIDF value, per newspaper (2020).

NYT		GUAR		CHI		FOL	
WORD	**TFxIDF**	**WORD**	**TFxIDF**	**WORD**	**TFxIDF**	**WORD**	**TFxIDF**
trump	24795.8	trump	11160.5	china	18288.8	bolsonaro	15156.5
biden	18936.2	care	10910.7	kong	14787.8	president	12743.3
dr	16813.3	cases	10304.4	hong	14461.5	government	11249.1
china	13968.0	vaccine	9475.1	percent	14228.3	health	10880.2
city	13128.8	government	9285.5	countries	12146.7	cases	10008.6
virus	12389.8	health	9277.9	chinese	11998.5	brazil	9694.3
president	12339.2	uk	9168.3	year	11923.4	year	9225.2
state	12162.4	workers	9075.6	cases	11774.8	minister	9067.8
health	12108.6	australia	9044.3	global	11012.5	country	8797.9
percent	11865.2	man	8888.1	world	10636.9	trump	8790.5
PHRASE		**PHRASE**		**PHRASE**		**PHRASE**	
united states	8341.9	public health	5539.4	hong kong	14787.0	jair bolsonaro	4797.0
white house	6816.5	prime minister	5256.1	united states	4943.3	united states	4634.4
public health	5954.2	social distancing	3989.9	prevention and control	4788.0	ministry of health	4512.4
health care	5228.6	care homes	3878.7	public health	4585.4	social media	4500.2
york city	5160.7	stay at home	3531.9	hubei province	4397.7	president jair bolsonaro	4260.1
hong kong	4921.1	tested positive	3516.9	covid pandemic	4390.4	last year	3691.4
social distancing	4610.6	aged care	3446.1	confirmed cases	4102.8	social isolation	3492.7
president trump	4518.7	boris johnson	3239.8	year on year	3964.7	new coronavirus pandemic	3416.7
tested positive	4362.0	chief executive	3141.2	coronavirus outbreak	3886.2	federal government	3415.4
los angeles	4152.9	mental health	3101.3	Hong Kong	14787.0	united kingdom	3325.9

diversification of media coverage in the United States and Italy, respectively. In *The New York Times*, for example, we can see the newspaper's attention to the impacts of the pandemic on the electoral process (*Coronavirus and 2020 elections: what happens to voting in an outbreak* — 03/09/2020) and the appropriation of the health issue in the discourse of the then-presidential candidates (*With the debates over, Biden assails Trump's coronavirus response* — 10/23/2020).

However, a particular aspect of the insertion of the COVID-19 pandemic in the political context is the personalization that it gains in the United States and

Brazilian newspapers, with an even greater emphasis on the latter. The results show the names of former President Donald Trump and President Jair Bolsonaro as some of the most prominent in the coverage of *The New York Times* and *Folha de S. Paulo*, alongside other political actors or authorities linked to the government apparatus. In this respect, we raise here the discussion proposed by Hart, Chinn and Soroka [2020] on the degree of media amplification of the voices of political actors in an already highly politicized and polarized context. A study by the National Bureau of Economic Research [Sacerdote et al., 2020] shows that, during the first seven months of 2020, potentially positive issues, such as vaccine development, received less attention from US media than Trump's support to hydroxychloroquine, a drug used to treat malaria, whose effectiveness against COVID-19 has not been proven [Self et al., 2020]. The Brazilian government also constantly defended the medication and discursively appropriated it as the solution for the resumption of face-to-face and economic activities [Saldaña & Machado, 2020].

We agree that the media must address the political dimension of science and health, but we cannot fail to notice that editorial choices shape different approaches. For example, in Brazil, Bolsonaro made many of his denialist statements in front of his official residence, where he daily meets with supporters (*'So what? I'm sorry, what do you want me to do?', says Bolsonaro on record of coronavirus deaths* — 04/28/2020). These meetings have always received media coverage. But in May 2020, *Folha de S. Paulo* decided to suspend this coverage due to the recurring attacks on journalists (*Folha temporarily suspends coverage at Alvorada due to lack of security* — 05/25/2020), which means that at least some statements like that would no longer be echoed by the newspaper and, consequently, for its readers.

This personalized coverage was less visible in British and Chinese newspapers. In *The Guardian*, the prevalence of the topic Epidemiology & Science may reflect, in part, an editorial line focused more intensively on service journalism [Eide & Knight, 1999]. Many of their stories included a Quick Guide or a Question and Answer (Q&A) with guidance from health authorities regarding COVID-19 prevention measures and guidance on what to do when symptoms arise. It is also worth noting that, although Prime Minister Boris Johnson made statements minimizing the beginning of the crisis (*Boris Johnson boasted of shaking hands on day Sage warned not to* — 05/05/2020), the official discourse soon turned to the scientific evidence (*How coronavirus advice from Boris Johnson has changed* — 03/23/2020). Still, the British government came under fire when scientific adviser Patrick Vallance championed "herd immunity" — a phrase that appears among the most frequent only in *The Guardian's* corpus. The scientific community has strongly refuted the effectiveness of this strategy [Altmann, Douek & Boyton, 2020].

In the four newspapers, we also observed that part of the coverage focused on health issues was dedicated to monitoring the spread of the disease through the numbers of infected and deceased people (see subtopic Disease Spreading in the full topic modeling). Future research that delves further into our data may reveal how each media outlet addressed these subtopics. However, based on what was exposed by Ancker [2020], we were able to identify positive and negative examples of the so-called "quantitative communication" in the corpus — firstly, online animations and engaging interactive simulations; secondly, figures without context and widespread reports of raw numbers rather than rates, a practice already

detected in studies of past outbreaks and epidemics [da Silva Medeiros & Massarani, 2010; Roche & Muskavitch, 2003]. An example of well-contextualized coverage was when the United States reached the mark of 100,000 lives lost, in May 2020. In the interactive material *An incalculable loss*, by *The New York Times*, when scrolling the page, the user can locate, among the crowd, the names and characteristics of some of the victims (e.g., *Marion Krueger, 85, great-grandmother with an easy laugh*), which added a humanized dimension to the coverage.

The markedly different pattern of coverage by the Chinese newspaper, with a sharp drop in the topic Epidemiology & Science at the beginning of the year, may reflect the country's health crisis trajectory. Apparently, once the pandemic was controlled, the country started to prioritize economic recovery, which also corroborates studies that suggest that the behavior of investors is affected by the tone of media coverage [C.-H. Wu & Lin, 2017]. In the story *Xi: Balance epidemic, economy* (02/04/2020), *China Daily* reports that President Xi Jinping has instructed government officials to take more measures to coordinate disease prevention and the resumption of economic production, in an effort to meet the development goals for that year. As a result, in a year in which the International Monetary Fund (IMF) predicted a 4.4% retraction in the world economy, China's Gross Domestic Product (GDP) grew by 2.3% — a result six percentage points higher than the global average [Yuyu, 2021]. This change of approach in *China Daily* was also observed by Gong and Firdaus [2022], who identified that the Chinese newspaper started to address more positive topics, such as prosociality, cooperation, pandemic recovery and informativeness.

Our data do not allow us to list the reasons that caused this change, but we bring to the discussion researchers who point out that Chinese state communication — which controls most of the country's newspapers, including *China Daily* — was used to maintain and increase its political legitimacy [Lemus-Delgado, 2020]. As the first English-language Chinese newspaper, and distributed in several countries, *China Daily*'s mission is to brand China to the outside world [Zhang & Wu, 2017]. In this sense, this newspaper and the Chinese state press as a whole were also inserted in a dispute of narratives regarding the origin of the new coronavirus, including allegations by the US government that China manufactured it (*Beijing says Pompeo lied about virus' origin* — 05/08/2020).

Furthermore, according to Ang [2020], the timeline from the detection of the first case of SARS-CoV-2 in Wuhan, in December 2019, to the adoption of effective containment measures, on January 22, the State apparatus and the media acted in accordance with President Xi Jinping's directives. This helps to explain the two extremes of the Chinese government in combating the pandemic: on the one hand, it was successful in mobilizing a strong national response to the crisis; on the other hand, there was a lack of transparency and delay in adopting initial measures to contain the outbreak [Ang, 2020; J. T. Wu, Leung & Leung, 2020].

The results also show that the pandemic has crossed over into other sectors of society, although with less media attention. We observed that the coverage had a more practical character, such as informing the situation of sporting, artistic and cultural events in the face of the restrictions caused by the pandemic. A frequent topic common to the four newspapers was the impacts of COVID-19 on education, which affected more than 1.6 billion students worldwide [Unesco, Unicef & World

Bank, 2021]. In general, coverage followed a similar path: the closing of nurseries, schools and universities; the alternatives and challenges of remote teaching (especially in countries like Brazil, where there is a deficit in access to computing resources for a large part of the population); and the resumption of face-to-face activities.

Finally, only in *The New York Times* did the racial issue form a consistent — but tangential –subtopic due to the death of George Floyd and the subsequent protests. This is not a fortuitous presence, as the fact was politically exploited (*The George Floyd election* — 03/06/2020), it involved broad sectors of the United States society, including health professionals (*George Floyd protests add new frontline for coronavirus doctors* — 07/06/2020), generated concern among health authorities due to the possibility of an increase in infection as a consequence of the protests, and sparked debates about the unequal effects of a health crisis. As Jean [2020] points out, although police brutality and COVID-19 are distinct problems, they intersect insofar as they affect the black population more intensely [Garg et al., 2020].

Final considerations

The current study allowed us to identify the general characteristics of the extensive coverage of the COVID-19 pandemic by important elite newspapers in the United States, the United Kingdom, China and Brazil during 2020. With an exploratory text mining approach, we identified topics, subjects and actors present in the coverage of *The New York Times*, *The Guardian*, *China Daily* and *Folha de S. Paulo* during the first year of the pandemic, and how attention to them varied over the months.

Although it is expected that different contexts generate distinct media coverage, our study contributes to the existing research in science communication by pointing out, using a statistical relative measure, which elements marked these differences, and which situations and voices were more reverberated or silenced by essential outlets of the communication environment of those countries. In quantitative terms, the fact that the increase in journalistic production in the United States, British and Brazilian newspapers followed a similar curve shows that, for a while, the disease was seen as a local problem in "distant" China. Regarding science reporting practices, it will be interesting to see whether this stance — and this antagonism between the local and the global — holds up with new health emergencies such as the 2022 monkeypox outbreak [Kumar, Acharya, Gendelman & Byrareddy, 2022].

In qualitative terms, it was possible to observe the focus of each coverage: more politicized and personalized in *The New York Times* and *Folha de S. Paulo*; more focused on health aspects in *The Guardian*; and with an emphasis on the economic situation in a post-crisis context in *China Daily*. In this respect, the pandemic has motivated a deeper approach to the multiple dimensions of science and health, pointing to a broader perspective of science communication. Sometimes focused on the uncontrolled spread of the disease, sometimes eclipsed by political actors who acted as public disseminators of denialism and conspiracy theories, the coverage also reflected some evidence of what can be considered distinct news values. Evidently, several and complex factors are at stake in the process of making media content, but part of them is linked to what the outlet understands as newsworthy

and how much a subject will occupy (or can occupy, in the case of a controlled media system) in its coverage.

To take a few current examples, the British press was actively at the forefront of the dissemination of information that resulted in the resignation of Health Secretary Matt Hancock, in 2021, and of Prime Minister Boris Johnson, in 2022, with accusations of violating social isolation rules [Helm, Savage & Walker, 2021; Walker & Stewart, 2022]. In China, the state press was and remains responsible for detailing the strict official measures to combat the pandemic, such as social isolation, traffic control, screening, testing and monitoring ["Report: China's fight against COVID-19", 2020]. In addition, our analysis contributed to revealing the complexity of the crisis in its most diverse dimensions, including topics such as the challenges of remote teaching and learning and the social determinants of health, such as racial impacts.

Finally, this study also makes a methodological contribution when using text mining techniques in science communication research. Although this methodological approach is extensively applied in other research areas, it has only recently been adopted by researchers in science communication and science journalism studies. By allowing us to deal with a complex and large amount of unstructured data, we were able to analyze almost all the news, reports, articles, columns and editorials published in text, during a year, by the four selected newspapers. Therefore, our results may also motivate future studies, with a deepening of each topic or expanding it to other media outlets.

Data availability

The full dataset, containing the 95,970 collected texts; the topic modeling processing, with the assignment to the resulting 120 topics; and the full list of the most frequent words and phrases in the corpus of each newspaper are available at https://bityli.com/mXjmm.

References

Altmann, D. M., Douek, D. C. & Boyton, R. J. (2020). What policy makers need to know about COVID-19 protective immunity. *The Lancet 395* (10236), 1527–1529. doi:10.1016/s0140-6736(20)30985-5

Ancker, J. (2020). The COVID-19 Pandemic and the Power of Numbers. *Numeracy 13* (2). doi:10.5038/1936-4660.13.2.1358

Ang, Y. Y. (2020). When COVID-19 meets centralized, personalized power. *Nature Human Behaviour 4* (5), 445–447. doi:10.1038/s41562-020-0872-3

Aslam, F., Awan, T. M., Syed, J. H., Kashif, A. & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications 7* (1). doi:10.1057/s41599-020-0523-3

Asogwa, C. E. (2021). Nigerian media coverage of medical progress on the development of COVID-19 vaccine. *Human Vaccines & Immunotherapeutics 17* (8), 2421–2426. doi:10.1080/21645515.2021.1882282

Atri, H., Kouki, S. & imen Gallali, M. (2021). The impact of COVID-19 news, panic and media coverage on the oil and gold prices: An ARDL approach. *Resources Policy 72*, 102061. doi:10.1016/j.resourpol.2021.102061

Aweisi, A., Arora, D., Emby, R., Rehman, M., Tanev, G. & Tanev, S. (2021). Using Web Text Analytics to Categorize the Business Focus of Innovative Digital Health Companies. *Technology Innovation Management Review 11* (7/8), 65–78. doi:10.22215/timreview/1457

Bauer, M. W. & Gregory, J. (2007). PUS Inc.: from journalism to corporate communication in post-War Britain. In M. W. Bauer & M. Bucchi (Eds.), *Journalism, Science and Society. Science Communication between News and Public Relations* (1st ed., pp. 33–52). doi:10.4324/9780203942314

Becker, B. J. (2011). *Unravelling Starlight: William and Margaret Huggins and the Rise of the New Astronomy*. Cambridge University Press.

Bendau, A., Petzold, M. B., Pyrkosch, L., Maricic, L. M., Betzler, F., Rogoll, J., … Plag, J. (2020). Associations between COVID-19 related media consumption and symptoms of anxiety, depression and COVID-19 related fear in the general population in Germany. *European Archives of Psychiatry and Clinical Neuroscience 271* (2), 283–291. doi:10.1007/s00406-020-01171-6

Bonneux, L. & Van Damme, W. (2006). An iatrogenic pandemic of panic. *BMJ 332* (7544), 786–788. doi:10.1136/bmj.332.7544.786

Carpenter, S. (2007). U.S. Elite and Non-elite Newspapers' Portrayal of the Iraq War: A Comparison of Frames and Source Use. *Journalism & Mass Communication Quarterly 84* (4), 761–776. doi:10.1177/107769900708400407

Casero-Ripollés, A. (2020). Impact of COVID-19 on the media system. Communicative and democratic consequences of news consumption during the outbreak. *El Profesional de la Información 29* (2). doi:10.3145/epi.2020.mar.23

Chadwick, A. (2013). *The hybrid media system: politics and power*. doi:10.1093/acprof:oso/9780199759477.001.0001

Chen, H., Huang, X. & Li, Z. (2020). A content analysis of Chinese news coverage on COVID-19 and tourism. *Current Issues in Tourism 25* (2), 198–205. doi:10.1080/13683500.2020.1763269

Chen, W. & Stoecker, C. (2020). Mass media coverage and influenza vaccine uptake. *Vaccine 38* (2), 271–277. doi:10.1016/j.vaccine.2019.10.019

Cogburn, D. L. (2019). Analyzing Trends and Topics in Internet Governance and Cybersecurity Debates Found in Twelve Years of IGF Transcripts. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Cogburn, D. L., Thomas, D. K. S. & Lai, J. (2020). Using Computational Text Mining to Understand Public Priorities for Disability Policy Towards Children in Canadian National Consultations. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

da Silva Medeiros, F. N. & Massarani, L. (2010). Pandemic on the air: a case study on the coverage of new influenza A/H1N1 by Brazilian prime time TV news. *JCOM 09* (03), A03. doi:10.22323/2.09030203

de Castro Moreira, I. & Massarani, L. (2002). Aspectos históricos da divulgação científica no Brasil. In L. Massarani, I. d. C. Moreira & F. Britto (Eds.), *Ciência e Público: caminhos da divulgação científica no Brasil* (pp. 43–64). Rio de Janeiro, Brazil: Casa da Ciência — Centro Cultural de Ciência e Tecnologia da UFRJ.

Retrieved from https: //www.museudavida.fiocruz.br/index.php/publicacoes/livros/747-tcc-54

Ducharme, J. (2020). News Coverage of Coronavirus in 2020 is Very Different Than it Was For Ebola in 2018. *Time*. Retrieved from https://time.com/5779872/coronavirus-ebola-news-coverage/

Dunwoody, S. (2008). Science journalism. In M. Bucchi & B. Trench (Eds.), *Handbook of public communication of science and technology* (1st ed., pp. 15–26). London, U.K.: Routledge.

Eide, M. & Knight, G. (1999). Public/Private Service. *European Journal of Communication 14* (4), 525–547. doi:10.1177/0267323199014004004

Fahy, D. & Nisbet, M. C. (2011). The science journalist online: Shifting roles and emerging practices. *Journalism 12* (7), 778–793. doi:10.1177/1464884911412697

Fitera, J. M., Abuín-Vences, N. & Sánchez, J. S. (2021). The coronavirus pandemic narratives in Italian digital media. *JCOM 20* (02), A07. doi:10.22323/2.20020207

Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., . . . Fry, A. (2020). *Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020*. US Department of Health, Human Services/Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm

Giri, S. P. & Maurya, A. K. (2021). A neglected reality of mass media during COVID-19: Effect of pandemic news on individual's positive and negative emotion and psychological resilience. *Personality and Individual Differences 180*, 110962. doi:10.1016/j.paid.2021.110962

Gong, J. & Firdaus, A. (2022). Is the Pandemic a Boon or a Bane? News Media Coverage of COVID-19 in 'China Daily'. *Journalism Practice*, 1–21. doi:10.1080/17512786.2022.2043766

Hart, P. S., Chinn, S. & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science Communication 42* (5), 679–697. doi:10.1177/1075547020950735

Helm, T., Savage, M. & Walker, P. (2021). Matt Hancock resigns as health secretary after day of humiliation. *The Guadian*. Retrieved from https://www.theguardian.com/politics/2021/jun/26/matt-hancock-resigns-after-questions-over-relationship-with-aide

Ihekweazu, C. (2016). Ebola in Prime Time: A Content Analysis of Sensationalism and Efficacy Information in U.S. Nightly News Coverage of the Ebola Outbreaks. *Health Communication 32* (6), 741–748. doi:10.1080/10410236.2016.1172287

Jean, T. (2020). Black Lives Matter: Police Brutality in the Era of COVID-19. *Syracuse University Lerner Center for Public Health Promotion 31*, 1–4.

Katz, C. & Cohen, N. (2021). Invisible children and non-essential workers: Child protection during COVID-19 in Israel according to policy documents and media coverage. *Child Abuse & Neglect 116*, 104770. doi:10.1016/j.chiabu.2020.104770

Kumar, N., Acharya, A., Gendelman, H. E. & Byrareddy, S. N. (2022). The 2022 outbreak and the pathobiology of the monkeypox virus. *Journal of Autoimmunity 131*, 102855. doi:10.1016/j.jaut.2022.102855

Lacy, S., Fico, F. & Simon, T. F. (1991). Fairness and Balance in the Prestige Press. *Journalism Quarterly 68* (3), 363–370. doi:10.1177/107769909106800306

Lemus-Delgado, D. (2020). China and the battle to win the scientific narrative about the origin of COVID-19. *JCOM 19* (05), A06. doi:10.22323/2.19050206

Liu, Q., Zheng, Z., Zheng, J., Chen, Q., Liu, G., Chen, S., ... Ming, W.-K. (2020). Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. *Journal of Medical Internet Research 22* (4), e19118. doi:10.2196/19118

Lynch, J., Evans, N., Ice, E. & Costa, D. K. (2021). Ignoring Nurses: Media Coverage during the COVID-19 Pandemic. *Annals of the American Thoracic Society 18* (8), 1278–1282. doi:10.1513/annalsats.202010-1293ps

Mach, K. J., Reyes, R. S., Pentz, B., Taylor, J., Costa, C. A., Cruz, S. G., ... Klenk, N. (2021). News media coverage of COVID-19 public health and policy information. *Humanities and Social Sciences Communications 8* (1). doi:10.1057/s41599-021-00900-z

Massarani, L. & Neves, L. F. F. (2021). Communicating the "Race" for the COVID-19 Vaccine: An Exploratory Study in Newspapers in the United States, the United Kingdom, and Brazil. *Frontiers in Communication 6*. doi:10.3389/fcomm.2021.643895

Morgan, T., Wiles, J., Williams, L. & Gott, M. (2021). COVID-19 and the portrayal of older people in New Zealand news media. *Journal of the Royal Society of New Zealand 51* (sup1), S127–S142. doi:10.1080/03036758.2021.1884098

Ng, R., Chow, T. Y. J. & Yang, W. (2021). News media narratives of Covid-19 across 20 countries: Early global convergence and later regional divergence. *PLOS ONE 16* (9), e0256358. doi:10.1371/journal.pone.0256358

Nielsen, R. K., Kalogeropoulos, A. & Fletcher, R. (2020). *Most in the UK Say News Media Have Helped Them Respond to COVID-19, but a Third Say News Coverage Has Made the Crisis Worse*.

Park, S., Fisher, C., Lee, J. Y. & McGuinness, K. (2020). COVID-19: Australian news and misinformation. *News and Media Research Centre*.

Patairiya, M. (2004). Science Journalism in India. *Pataneto Press*. Retrieved from http://pantaneto.co.uk/science-journalism-in-india-manoj-patairiya/

Péladeau, N. (2021). Mixing Beyond Mixed Methods: QDA Miner, SimStat, and WordStat. In A. J. Onwuegbuzie & B. Johnson (Eds.), *The Routledge Reviewer's Guide to Mixed Methods Analysis* (pp. 291–303). Routledge.

Pollock, A. M., Roderick, P., Cheng, K. K. & Pankhania, B. (2020). Covid-19: why is the UK government ignoring WHO's advice? *BMJ*, m1284. doi:10.1136/bmj.m1284

Provalis Research (n.d.). WordStat 8 User's Guide.

Al-Rawi, A., Grepin, K., Li, X., Morgan, R., Wenham, C. & Smith, J. (2021). Investigating Public Discourses Around Gender and COVID-19: a Social Media Analysis of Twitter Data. *Journal of Healthcare Informatics Research 5* (3), 249–269. doi:10.1007/s41666-021-00102-x

Al-Rawi, A., Kane, O. & Bizimana, A.-J. (2021). Topic modelling of public Twitter discourses, part bot, part active human user, on climate change and global warming. *Journal of Environmental Media 2* (1), 31–53. doi:10.1386/jem_00039_1

Al-Rawi, A., Al-Musalli, A. & Fakida, A. (2021). News Values on Instagram: A Comparative Study of International News. *Journalism and Media 2* (2), 305–320. doi:10.3390/journalmedia2020018

Report: China's fight against COVID-19. (2020). *China Daily*. Retrieved from https://covid-19.chinadaily.com.cn/a/202004/21/WS5e9e2c62a3105d50a3d17880.html

Roche, J. P. & Muskavitch, M. A. T. (2003). Limited Precision in Print Media Communication of West Nile Virus Risks. *Science Communication 24* (3), 353–365. doi:10.1177/1075547002250300

Sacerdote, B., Sehgal, R. & Cook, M. (2020). Why is all COVID-19 news bad news? *NBER Working Paper Series*.

Saldaña, P. & Machado, R. (2020). Enquanto Mandetta reitera cuidado com cloroquina, Bolsonaro volta a defender remédio contra coronavírus. *Folha de S. Paulo*. Retrieved from https://www1.folha.uol.com.br/equilibrioesaude/2020/04/enquanto-mandetta-reitera-cuidado-com-cloroquina-bolsonaro-volta-a-defender-remedio-contra-coronavirus.shtml

Self, W. H., Semler, M. W., Leither, L. M., Casey, J. D., Angus, D. C., Brower, R. G., . . . Brown, S. M. (2020). Effect of Hydroxychloroquine on Clinical Status at 14 Days in Hospitalized Patients With COVID-19. *JAMA 324* (21), 2165–2176. doi:10.1001/jama.2020.22240

Shomron, B. (2021). The Capability 'To Be Secure': Media Coverage of African Asylum Seekers During Covid-19 in Israel. *Journal of Refugee Studies 34* (4), 4361–4380. doi:10.1093/jrs/feaa131

The Lancet (2020). COVID-19 in Brazil: "so what?" *The Lancet 395* (10235), 1461. doi:10.1016/s0140-6736(20)31095-3

Tian, Y. & Stewart, C. M. (2005). Framing the SARS Crisis: A Computer-Assisted Text Analysis of CNN and BBC Online News Reports of SARS. *Asian Journal of Communication 15* (3), 289–301. doi:10.1080/01292980500261605

Unesco, Unicef & World Bank (2021). *The State of the Global Education Crisis*. Unesco.

Van den Bulck, J. & Custers, K. (2009). Television exposure is related to fear of avian flu, an Ecological Study across 23 member states of the European Union. *The European Journal of Public Health 19* (4), 370–374. doi:10.1093/eurpub/ckp061

Vasterman, P. L. M. & Ruigrok, N. (2013). Pandemic alarm in the Dutch media: media coverage of the 2009 influenza A (H1N1) pandemic and the role of the expert sources. *European Journal of Communication 28* (4), 436–453. doi:10.1177/0267323113486235

Wahl-Jorgensen, K. (2020). Coronavirus: how media coverage of epidemics often stokes fear and panic. *The Conversation*. Retrieved from https://theconversation.com/coronavirus-how-media-coverage-of-epidemics-often-stokes-fear-and-panic-131844

Walker, P. & Stewart, H. (2022). Boris Johnson admits attending Downing Street party during lockdown. *The Guardian*. Retrieved from https://www.theguardian.com/politics/2022/jan/12/boris-johnson-admits-attending-downing-street-party-during-lockdown

Wallis, P. & Nerlich, B. (2005). Disease metaphors in new epidemics: the UK media framing of the 2003 SARS epidemic. *Social Science & Medicine 60* (11), 2629–2639. doi:10.1016/j.socscimed.2004.11.031

WHO (2020a). COVID-19 - China. Retrieved from https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229

WHO (2020b). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Retrieved from https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Wu, C.-H. & Lin, C.-J. (2017). The impact of media coverage on investor trading behavior and stock returns. *Pacific-Basin Finance Journal 43*, 151–172. doi:10.1016/j.pacfin.2017.04.001

Wu, J. T., Leung, K. & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet 395* (10225), 689–697. doi:10.1016/s0140-6736(20)30260-9

Wu, K. (2021). Agenda-Setting in Cross-National Coverage of COVID-19: An Analysis of Elite Newspapers in US and China with Topic Modeling. *Online Journal of Communication and Media Technologies 11* (4), e202116. doi:10.30935/ojcmt/11083

Yamey, G. & Gonsalves, G. (2020). Donald Trump: a political determinant of covid-19. *BMJ 369*, m1643. doi:10.1136/bmj.m1643

Yuyu, C. (2021). All-round effect of growth on economy. *China Daily*. Retrieved from https://global.chinadaily.com.cn/a/202101/21/WS6008c314a31024ad0baa405e.html

Al-Zaman, M. S. & Khan, T. (2021). Framing environmental news in Bangladesh. *Media Asia 49* (2), 98–110. doi:10.1080/01296612.2021.1997526

Zhang, L. & Wu, D. (2017). Media Representations of China: A Comparison of 'China Daily' and 'Financial Times' in Reporting on the Belt and Road Initiative. *Critical Arts 31* (6), 29–43. doi:10.1080/02560046.2017.1408132

Authors

Luiz Felipe Fernandes Neves is a PhD candidate in Biosciences and Health at Oswaldo Cruz Foundation (Fiocruz), Brazil. He is a journalist with a master in Communication, and works at the Communication Department of the Federal University of Goiás (UFG), Brazil. The professional activities and academic research include science communication, science journalism, organizational communication, and digital social networks.
luiz.felipe@ufg.br.

Luisa Massarani is a Brazilian science communicator who carries out both practical and research activities in the field. She is the coordinator of the Brazil's National Institute of Public Communication of Science and Technology and for Latin American SciDev.Net. She is a researcher at House of Oswaldo Cruz/Fiocruz (Brazil). Fellow Productivity of the National Council for Scientific and Technological Development (CNPq) 1B, Our State Scientist of the Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (Faperj). She is recipient of the Mercosur Award for Science and Technology (2014), the Brazilian Award for Science Communication (2016) and the Literature Jabuti Award (2017).
luisa.massarani@fiocruz.br.