

# The beginnings of the COVID19 pandemic on Twitter. Computational Analysis of Public Conversation in Spanish Language

Los inicios de la pandemia de COVID19 en Twitter. Análisis  
computacional de la conversación pública en lengua española

*O início da pandemia COVID19 no Twitter. Análise computacional de  
conversas públicas na língua espanhola*

**Manuel Cebal-Loureda**, Tecnológico de Monterrey, Monterrey, México  
(manuel.cebral@tec.mx)

**Gabriela Elisa Sued-Palmeiro**, Investigadora Postdoctoral Conacyt Instituto de  
Investigaciones Sociales - Universidad Nacional Autónoma de México, Ciudad de  
México, México (gabriela.sued@sociales.unam.mx)

**ABSTRACT** | At the beginning of the COVID19 pandemic, social platforms played a crucial role in the production and access to information. This study aims to identify the topics of most significant interest and their associated feelings during the onset of the pandemic on Spanish-language tweets. In addition, we analyzed the role of Twitter as a social platform involved in the public conversation, both as a means for mass self-communication and for amplifying the voice of a reduced set of high visibility actors. 231,375 tweets were collected in Spain and Latin America over two months. Then, the sample was analyzed with digital methods and techniques through computer programming in R. Frequency and sentiment indicators were measured, and terms were grouped to identify topics and determine users' interests. The frequency of the main terms is dynamic throughout the period studied, suggesting different perceptions of the pandemic. The main topics refer to conversations around the number of cases, deaths, and infections. Sentiment analysis shows the prevalence of negative feelings. The analyzed sample corresponds to ordinary users' messages for the great majority, but a part of it has been amplified on a large scale through retweets and bookmarks.

**KEYWORDS:** COVID-19; health; Twitter; social media; digital methods; text mining; Spain; Latin America.

## HOW TO CITE

Cebal-Loureda, M. & Sued-Palmeiro, G. E. (2021). Análisis computacional de la conversación pública en lengua española. *Cuadernos.info*, (49), 1-25. <https://doi.org/10.7764/cdi.49.27467>

---

**RESUMEN** | Cuando comenzó la pandemia de COVID 19, las plataformas sociales tuvieron un rol central en la producción y acceso a la información. Este estudio identifica los tópicos de mayor interés y sus sentimientos asociados en Twitter en la conversación pública en lengua española en ese periodo. Asimismo, analiza el rol de Twitter como plataforma social involucrada en la conversación pública, como medio para la autocomunicación de masas y para amplificar la voz de un conjunto reducido de actores de alta visibilidad. Se recolectaron 231.375 tuits en España y América Latina durante dos meses. Se midieron indicadores de frecuencia y sentimientos, y se agruparon términos para identificar tópicos y determinar el interés de los usuarios sobre estos, mediante métodos digitales y de lenguaje computacional en R. La frecuencia de los principales términos es dinámica a lo largo del período estudiado, lo que sugiere diferentes percepciones de la pandemia. Los tópicos principales refieren a conversaciones en torno a la cantidad de casos, muertos y contagiados, con prevalencia de sentimientos negativos. La muestra analizada corresponde a mensajes generados por usuarios comunes en su gran mayoría, pero una parte de ella ha sido amplificada a gran escala mediante retuits y marcas de favoritos.

**PALABRAS CLAVE:** COVID-19; salud; Twitter; percepciones; métodos digitales; minería de textos; España; Latinoamérica.

---

**RESUMO** | No início da pandemia COVID-19, as plataformas sociais desempenharam um papel central na produção e no acesso à informação. Este estudo tem como objetivo identificar na conversa pública os temas de maior interesse e seus sentimentos associados no Twitter em língua espanhola nesse período. Além disso, foi analisado o papel do Twitter como plataforma social envolvida na conversação pública, tanto como meio de autocomunicação de massa quanto para amplificar a voz de um conjunto reduzido de atores altamente visíveis. Foram recoletados 231.375 tuites na Espanha e na América Latina, durante dois meses. A amostra foi analisada medindo indicadores de frequência e sentimentos, e os termos foram agrupados para identificar temas e determinar o interesse dos usuários sobre eles através de métodos digitais e da linguagem computacional R. A frequência dos termos principais é dinâmica durante o período estudado, sugerindo diferentes percepções sobre a pandemia. Os principais temas referem-se a conversas sobre o número de casos, mortos e contagiados. A análise mostra a prevalência de sentimentos negativos. A amostra analisada corresponde a mensagens geradas por usuários comuns em sua maior parte, mas parte dela foi ampliada em grande escala por meio de retuites e marcações.

**PALAVRAS CHAVE:** COVID-19; saúde; Twitter; mídia social; métodos digitais; mineração de texto; Espanha; América latina.

## INTRODUCTION

The COVID-19 pandemic can be considered a critical and unforeseen event, with consequences that transcend the moment (Pantojas-García, n.d.), from public health decisions to changes in economic production that affect the core of social and individual life (UNDP, 2020). Its rapid irruption in the global context, the general ignorance of its evolution, and the radical changes it has introduced in the daily lives of millions of people increased consumption and the circulation of information in the initial phase of the pandemic (“¿Cómo es el consumo?”, 2020). According to the Reuters digital news consumption report, more than half of global respondents accessed news about the pandemic from social platforms (Newman et al., 2020).

Social networks have been considered by journalism and consultancy as the gateway to news about the pandemic broadcast by traditional media (Vega, 2020). However, Twitter’s uses extend beyond being an intermediary network between press and audience as a space for ordinary users to express opinions, ideas, and feelings about public issues of global importance (Rogers, 2019). Because of its central role in constructing conversations and trends in public opinion (van Dijck, 2016) Twitter is, among social platforms, the best suited to study how the COVID-19 pandemic has been socially signified.

A previous study (Sued Palmeiro & Cebal Loureda, 2020) analyzed the functioning of authoritative voices on Twitter at the beginning of the pandemic, defining them as users with a public profile and many followers. In this study, we propose to identify the most relevant topics of interest and the feelings linked to them in the public conversation at the beginning of the COVID-19 pandemic on Twitter in Ibero-America, as well as to determine whether the messages studied are representative of the public opinion of ordinary users who use the platform for self-expression, or reproduce few messages posted by high-visibility users, such as media and government and opinion leaders, among others.

## DIGITAL METHODS AND COMPUTATIONAL TEXT ANALYSIS APPLIED TO PLATFORM STUDIES

This work is situated in the interrelation of four areas of knowledge: digital methods (Rogers, 2019), computational text analysis (Castelfranchi, 2017), platform studies (Plantin et al., 2018), and health communication studies (Rodríguez Roura et al., 2018).

Digital methods make use of digital objects available on the web, such as metrics, geolocations, and publications, and reorient them to social research through a set of techniques that are also digital. They are posited as a research strategy that

studies social change and cultural conditions with online data (Rogers, 2019). This is not to say that digital platforms reflect offline social life, but they are indicators of social interest and trends in public opinion (Rogers, 2018).

Platform studies are conceived as a subfield of digital methods that, through the collection of posts, metrics, and metadata, critically study different dispositions, forms of use, and cultural patterns that distinguish each social platform (Burgess & Green, 2019).

In the last decade, the web's social study, focused first on its users and practices, has shifted to the analysis of online data. This innovation requires incorporating digital technologies into the research process (Marres, 2017). Among them, computational text analysis techniques, a subfield of natural language processing, allow the discovery of new information from the automatic extraction of patterns and trends in written sources (Moreno & Redondo, 2016).

Both digital methods and platform studies represent an interesting challenge for the so-called Global South (de Souza Santos, 2014), since there is an imbalance between the data-centered production of American and European literature and that produced in Ibero-American countries (Milan & Treré, 2017). Differences in languages, academic schools of social communication, topics of interest, and conjunctures regarding technological uses mean that digital methods and platform studies need to be readapted in the literature in Spanish, bringing new topics, perspectives, and procedures to the international field (Meneses Rocha, 2018).

This study adopts the research strategy of digital methods to study a sociocultural condition, the public conversation, and its social significance at the beginning of the COVID-19 pandemic, exploiting data extracted from Twitter. Simultaneously, it attempts to shed light on communicative patterns on that platform, understanding that it plays a specific role in the predisposition to the production and circulation of messages.

### **TWITTER AS A SOURCE FOR THE STUDY OF PUBLIC EXPRESSIONS**

Ranked fourteenth most used social network in the world, Twitter is relevant for the Global South, since Spanish is the third most used language on the platform, after English and Japanese ("2018 Research on 100 Million Tweets", 2018). Its structure focuses on the communication of instant and fast-moving messages about relevant current events, among an extensive international community of users who share diverse thematic interests (Bruns & Weller, 2016). Twitter works as a platform where ordinary users share their opinions on various issues and as a medium to disseminate messages produced by high-visibility users (Kwak et al., 2010; Colleoni et al., 2014), defined as users who have many followers and achieve

a large amount of adherence, so their messages have a greater chance of spreading (Omena et al., 2020). The connection between high visibility and ordinary users is made possible through a series of technical operations and metrics, such as the number of retweets and favs, or the use of hashtags to follow public conversations in real time, without geographical barriers (Bruns & Highfield, 2016).

The contents created and distributed by Twitter users leave digital traces that can be recovered for social research (Marres, 2017). Research with data collected from Twitter has contributed to study opinion trends in public conversation and political debate, as well as to shed light on its use in crisis contexts, among other important topics (Burgess & Bruns, 2015).

### **Health and COVID-19 on Twitter**

In the last decade, health communication has been conceived as a set of interaction processes and social construction of meaning by the different actors involved (Rodríguez Roura et al., 2018). The field incorporates into its research the collection of data produced in social networks, supervised sentiment analysis, and network analysis (Arcila-Calderón et al., 2018). During the COVID-19 pandemic, several papers have studied the role of social platforms, especially Twitter; a large part presents automated collection and analysis methods, highlighting the reference to recurring themes, frequent words, retweets, and hashtags.

There is precedent for the usefulness of Twitter as a platform for communicating health news. Kullar and colleagues (2020) refer to its function as a platform for disseminating health messages to both medical professionals and the public in the pandemics of avian flu in 2009, Ebola in 2014 and COVID-19 in 2020. The platform has been useful for disseminating and accessing scientific news, as well as for patients and their families, who often group together to share information. On the other hand, Rufai and Bunce (2020) analyze the use of the platform by G7 political leaders in the context of the pandemic. Their content analysis revealed three types of use: informational, which was predominant, morale booster for the population, and political discussion seeking to raise points of debate. Along the same vein, Haman (2020) reviews a larger corpus of Twitter expressions of government leaders, finding that 65% of the leaders of countries belonging to the United Nations have tweeted at some point about the pandemic, increasing their number of followers.

Another line of work, closer to that of this study, addresses the expression of opinions, perceptions, and feelings by analyzing messages, especially lexical references. Han and colleagues (2020) study the expression of opinions and feelings in posts on the Chinese network Sina Weibo, the most important microblogging site in that country, during the early stages of COVID-19, and classify the most

frequently referred topics and subtopics: recommendations for social isolation, blessings and prayers, objective comments on the disease, protective measures, willingness to return to work, among others. The authors note that the importance of these topics changes as coronavirus events evolve. Thelwall and Thelwall (2020) focus on the gender perspective of the pandemic by comparing the frequency of words used by men and women. Sports and politics are frequent topics among the former, while social distance, home and family are frequent among the latter. Finally, Thelwall and Levitt (2020) study the demands and issues surrounding people with disabilities in a small but widely retweeted set of publications. The main themes are the need for support, the possibilities of discrimination, and difficulties in handling technologies.

Two relevant studies address the relationship between everyday experiences of pandemics and media framing of the disease. Davis and colleagues (2014) demonstrated, in relation to the A1HN1 pandemic, that everyday perceptions cannot be separated from their media framing. Towers and colleagues (2015) found evidence that the media influenced most of the tweets produced during the Ebola pandemic in Africa.

Another important line of work addresses pandemic disinformation on social platforms. Regarding Twitter, Pérez-Dasilva, Meso-Ayerdi and Mendiguren-Galdospín (2020) study actors and semantic networks of disinformation through network analysis. The authors refer to infodemic in the context of political debates between Republicans and conservatives in the United States, and between critics and pro-government in China. They highlight how, at first, false arguments are introduced that exaggerate the dangers of the coronavirus to exacerbate the perception of economic crisis –in the case of the United States, in the context of the imminence of presidential elections– while, at a second stage, measures of care in the face of the disease take on greater importance.

Finally, in one of the few works focused from the Global South, Gutiérrez and colleagues (2020) study the use of Twitter in Bahía Blanca, Argentina, through an analysis of social networks that allows them to identify that the conversations present a low degree of interaction and the most retweeted actors do not coincide with their representativeness in the public sphere. According to these studies, Twitter is a legitimate medium for distributing relevant information about COVID-19, as well as for audiences to share opinions, feelings, and perceptions.

## **MATERIALS AND METHODS**

Data was collected, processed, and visualized through the R programming language. We used text mining techniques to discover new information (Moreno

& Redondo, 2016). The main libraries used were *rtweet* (Kearney, 2020), *tidytext* (Robinson & Silge, 2020), *topicmodels* (Grün et al., 2020), *widyr* (Robinson, 2020), *ggraph* (Lin Pedersen, 2020), and *igraph* (Csardi & Nepusz, 2006).

The collection was conducted between March 19 and May 16, 2020, the initial phase of the pandemic. This period made it possible to obtain publications of the beginning of social isolation. A total of 231,375 tweets in Spanish were collected using the keyword *coronavirus*, accessing the Rest API (Application Program Interface) of Twitter through the *tweet* library. We avoided collecting answers to retweets. To capture how the pandemic has been assimilated over the weeks and to emphasize the temporal variable, we made 13 collections, 12 of them of 17,999 tweets and one of 15,392, with an interval of between three and four days. The size of the collections responds to the maximum allowed by Twitter for each search with the used method.

To meet the proposed objectives, data processing had two main orientations: first, the messages posted were studied in terms of lexical frequency, topic modeling, and sentiment and emotional analysis. Then, we retrieved metrics of message reactions that were quantified to determine whether they correspond to the expression of opinion of a set of ordinary users or, on the contrary, if the messages of a few high-visibility users are amplified.

To identify relevant topics through lexical frequency analysis, tweets were disaggregated into minimal linguistic terms, a process called tokenization and included in *tidytext* R library. The removal of context-dependent terms –such as articles, prepositions, and pronouns– was applied from the extensive list provided by Díaz (2016). To avoid non-significant repetitions, the term *coronavirus*, contained in the totality of the messages, was removed. We proceeded in the same way with other equivalent terms. Thus, the results of frequent words are those that appear most often together with *coronavirus*, obviating this term and its direct equivalents. We performed most of the text mining operations using these linguistic terms or tokens as the minimum unit of computation, starting with the word count (Robichaud & Blevins, 2011).

In addition, we recovered the tweet as a unit of meaning to apply the topic modeling technique, using the Latent Dirichlet Allocation (LDA) algorithm, included in the *topicmodels* library, also used in the study by Ghan and colleagues (2020) mentioned above. This algorithm classifies every document as a set of topics and each topic as a set of terms (Blei et al., 2003). With these references, the LDA algorithm creates an unsupervised vector space in which it generates the topics according to the vectors that relate terms and documents.

LDA is a complex algorithm that supports different configurations. One of them is the alpha value, which makes the algorithm look for more separated groups –for alpha values less than one– or more combined groups, in which there are more terms common to several topics –for alpha values greater than one. In this case, alpha was set to a value lower than 1, specifically 0.4, so that LDA generates more diverse topics. It is assumed that the lexical variety on Twitter is quite limited, and the topics are already highly related.

Another configurable value of LDA is  $k$ , which indicates the number of topics generated. This value has been established by a previous study to calculate the most efficient number of topics for a dataset. This is the perplexity measurement (Blei et al., 2003), available in the topicmodels library via the perplexity function (Grün et al., 2020). Due to the extraordinary computational cost of this calculation, it was performed on a subsample of 10,000 tweets. The results for  $k$  values between 2 and 11 were: 1970,461, 1791,831, 1732,825, 1725,598, 1760,093, 1781,241, 1817,099, 1851,048, 1894,675, 1943,602; the lowest value was 1725,598, which corresponds to a value of 5 for the  $k$  variable. Thus, the study indicates that topical modeling is optimized for five topics. It should be noted that the resulting numbering of the topics is completely random.

Finally, we applied two different sentiment analysis methods to the most frequent words in the corpus: the AFINN classifier (Nielsen, 2011), translated into Spanish (Mendoza Vega, 2018), and the NRC -strictly for emotion recognition- in its Spanish version (Mohammad et al., 2013). The former identifies whether interactions contain expressions of mood states and whether they express positive or negative sentiment. The second assigns values with respect to various emotions: fear, confidence, sadness, anger, expectation, disgust, joy, and surprise. Using two classifiers, complementary perspectives are obtained and detected trends, confirmed.

Respecting the ethical considerations of AoIR (Franzke et al., 2020), the study is based on data published online and presented in an aggregated form, and therefore does not cause harm to third parties. The databases were anonymized respecting the users' privacy.

## RESULTS

This section presents the findings obtained through automatic text discovery. The first three sections respond to the first objective, as they identify the most relevant topics of the public conversation. The fourth responds to the second objective since it determines Twitter's role in the public conversation about



COVID-19 with respect to its two possible uses: the amplification of messages from highly visible users or the expression of opinions of ordinary users.

### **An Alarmist Narrative: Lexical Frequency and its Evolution**

The most frequent words in the sample construct a narrative of alarm based on the priority given to cases, deaths, and pandemic. This is intensified if words of close meaning, such as contagion, virus and dead, are included in this group. On the contrary, terms referring to palliatives and health measures, such as quarantine, health, house, doctors, and measures, are of less interest, both in frequency and retweets. Thirdly, another grouping stands out, which refers to the coronavirus in more structural and general terms, where government, country, crisis, and world can be grouped, and which perceives the pandemic as a disruptive fact that encompasses public and private spaces, directly affecting people. Among these three different possibilities –a life-ending disease, a set of health measures to contain it, and an unforeseen global situation affecting all people and territories– the first one receives more circulation and adherence.

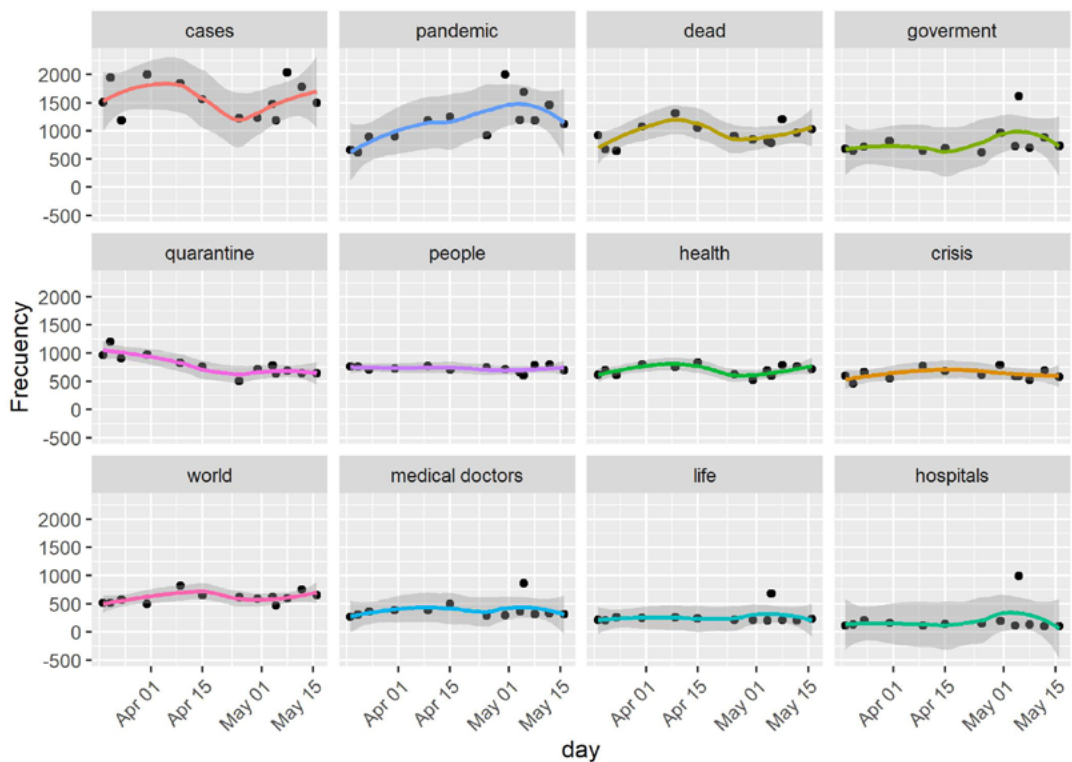
Figure 1 shows the temporal evolution of the frequency of some terms. The frequency of the term pandemic increases progressively from mid-March to the end of the sample in May, reaching its peak at the beginning of that month. Along with this term, cases and deaths also increase in frequency. The term government has a similar evolution, although with a smoother progression. Other terms such as health or world follow a more irregular evolution, although with a pattern similar to that of cases and deaths: higher frequency at the beginning and middle of April and at the end of May. Finally, the terms crisis, people, hospitals, life, or doctors remain constant, which does not mean that their presence is not important.

The evolution of the term quarantine is noteworthy. During the first days of COVID-19, the new situation of confinement seems to affect the population the most: it is during the last two weeks of March, a time when social isolation measures are taken worldwide, and people's lives change radically in a few days. While it starts out as the second most frequent term in the first data collections, its presence in the public conversation declines as the weeks go by: it drops from over a thousand occurrences in the March 31 collection to half, 504 occurrences, on April 24.

Along with the information provided by the frequency of words in the tweets' bodies, hashtags agglutinate the public conversation around a topic. As stated by Wikström (2014), they can be classified according to communicative functions. Hashtags such as #cuarentena (quarantine) refers to a topic of interest that happens rather in the realm of private life. In contrast, the hashtag #quedateencasa (stay at home), also referring to a private space, installs a collaborative, empathetic, and performative action that seeks a change of habits and community care. Figure 2 shows that it is the most

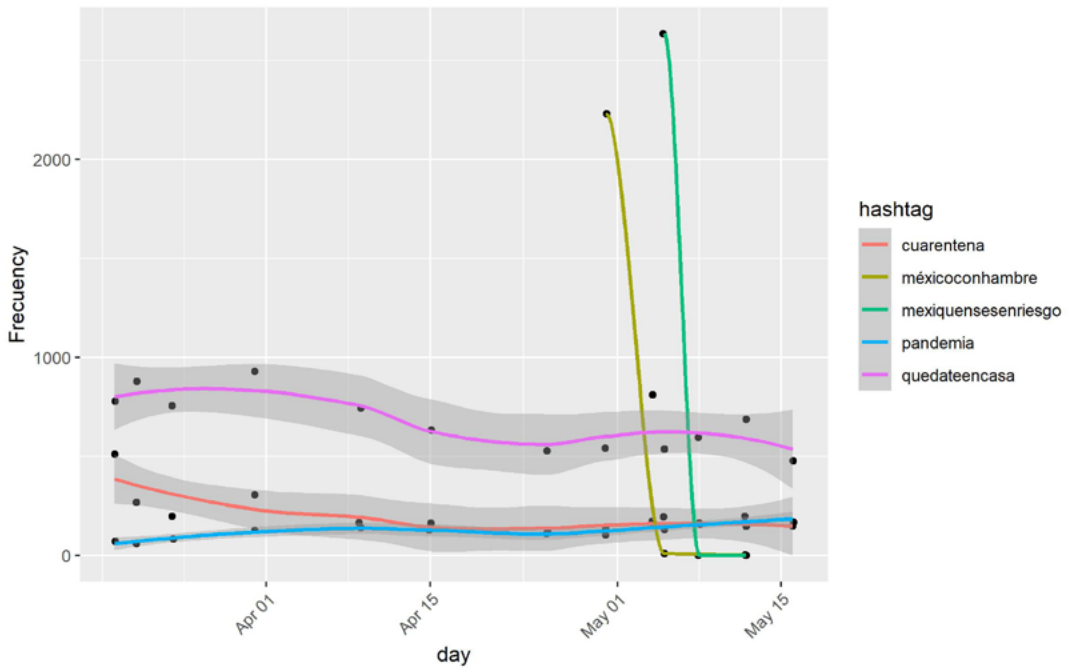
frequent hashtag in this sample –always after coronavirus and its direct equivalents–, remaining in this position over time, although with some decline since the beginning of April. The hashtag #cuarentena, which remains the second most frequent overall, also declines. In contrast, the hashtag #pandemia follows an upward trend, even surpassing #cuarentena in frequency, emphasizing a different conversation, no longer about care but about concern. Both trends corroborate the findings regarding the most frequent terms, although in this case the variations in frequency are more attenuated, proving that the use of hashtags remains more constant.

There are important exceptions to this behavior. In the captures of May 5 and April 30, two hashtags related to risk and hunger suddenly appear in Mexico: #mexiquensesenriesgo (Mexicans at risk) and #méxicoconhambre (Mexico has hunger). Given the volume of their publications and their ephemeral duration, it is possible to conjecture that they are the result of a local and concerted strategy to insert the issue of disease in a context of political debate. In fact, these hashtags coincide in time with increases in the frequency of terms such as government in figure 1. However, it should also be inferred that, while the population gets used to the confinement and the debate about quarantine falls, concern about the economic issue burst in force at the end of April and the beginning of May.



**Figure 1. Evolution over time of a selection of words that most frequently accompany the term coronavirus**

*Source: Own elaboration.*



**Figure 2. Detail of the evolution of the most used hashtags in the sample along with the strong irruption of two hashtags located in Mexico related to the crisis**

*Source: Own elaboration.*

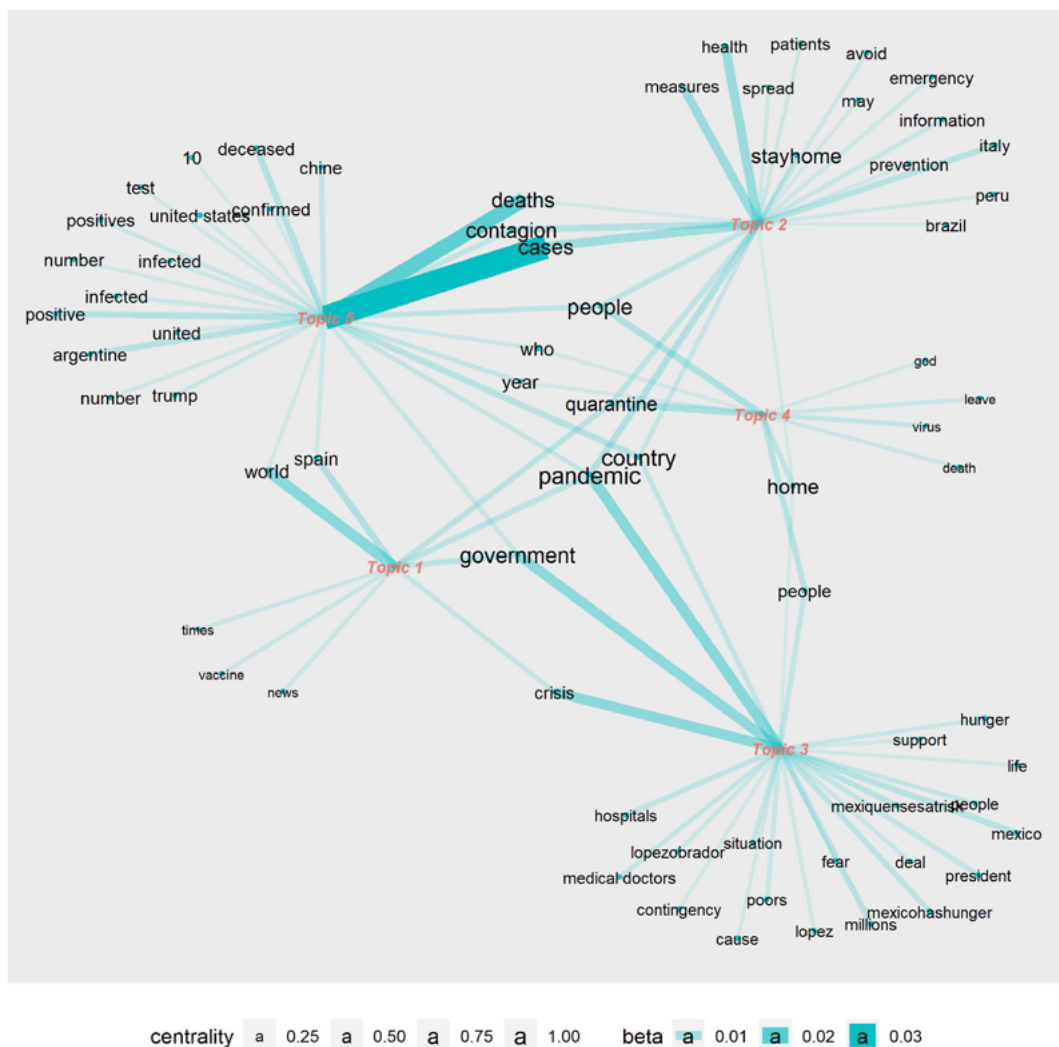
The evolution in the hashtags’ frequency shows the variation in public conversations: from a more private one, centered on quarantine and care, to a more public and pessimistic one, focused on the pandemic and oriented towards political and economic aspects. This trend confirms the previous lexical frequency analysis.

**Pandemic Outlook: Topic Modeling**

To determine the conversation topics regarding COVID-19 in the selected sample, we conducted a synchronic analysis by means of topic modeling. For this purpose, we used the LDA algorithm, previously explained, with the values  $\alpha=0.4$  and  $k=5$ .

The results can be seen in figure 3, where the nodes corresponding to each topic are shown in red. The words are linked to these nodes according to the beta value, which expresses the degree of belonging of a term to each topic. This value has been emphasized in the figure by thicker and more opaque lines. The centrality of each word, i.e., the greater number of topics it connects, has also been indicated by the size of the labels. The measure of centrality is fundamental to correctly read the network of relationships, since, as it can be seen, a higher value of beta does not always imply greater centrality. Finally, due to the spatial and legibility limitations of this type of figure, we only show terms with a beta value above 0.0026.

The most characterized topic is number 4, as it has more terms above the threshold established for beta in the figure and, in general, the terms appear more strongly linked.



**Figure 3. Topic modeling using the LDA function of the topicmodels library, and figures using igraph and ggraph. The beta values indicate the degree to which each term belongs to the topic plotted with the thickness and opacity of the lines. The centrality of each term in the network is indicated by the size of the label**

*Source: Own elaboration.*

This topic would be defined mainly by terms such as cases, deaths, pandemic, world, or country. Overall, it provides an international or global perspective on the problem, as most of the countries present in the conversation, including China, and entities such as the WHO appear. Many of the terms in this topic refer to political leaders, country names, and numerical terms, such as figure or number. Therefore, it can be assumed that it is linked to media messages, which produce precisely the most characteristic topic of the sample, from which a narrative based on the number of infections and deaths is amplified.

The second most characterized topic is 3, whose main terms are government, pandemic, health, and crisis. It calls for greater action in the face of the economic situation. It also has some terms that refer to collective response and solidarity, such as support or confrontation. The message as a whole is linked to the two hashtags that are installed briefly but intensely at the beginning of May, visible in figure 2.

The next most characterized topic is 2, through the terms cases, persons, people, dead, years, and quarantine. Semantically, it has an affective register, referring to frustration and pain due to the personal and deadly reach of the pandemic to people of different ages. It is a topic located in Argentina.

Topic 5, with fewer words above the value established for beta, would be characterized by the terms quarantine, pandemic, information, health, positive, or measures. It gathers transversal concerns related to medical aspects and the way to deal with the virus, through a more rational and less tense conversation, emphasizing preventive measures such as washing hands and staying at home to avoid contagion. This conversation is not located in any specific country.

Finally, topic 1 shows a conversation based on the terms government, quarantine, home, pandemic, and crisis. Although in this topic the terms less strongly define a specific theme, there is a concern for the government's work and that of the president to face the new situation. The conversation is located in Spain.

The LDA algorithm also detected semantic centrality values with respect to the conversation as a whole. The two terms with the highest centrality in the sample are positive: persons and health. The term persons is also the most retweeted term in the sample. It is followed by pandemic, contagion, country, and crisis.

### **Affective Repercussions: Sentiment and Emotional Analysis**

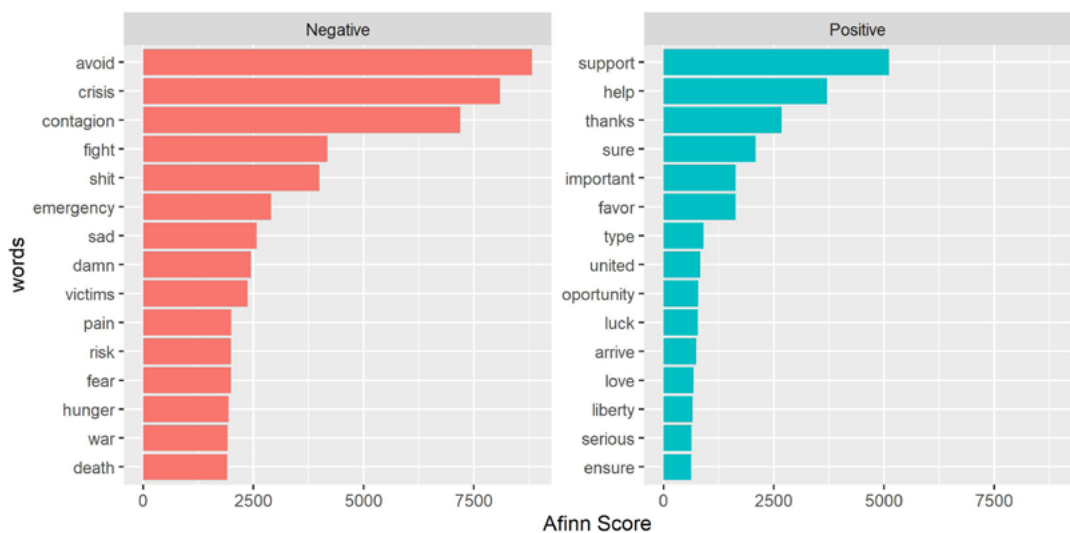
In addition to the word counts and their modeling into topics, we performed a word analysis using sentiment and emotional load classifiers. The overall balance of the analysis is negative: adding all the terms that score negatively in the AFINN sentiment classifier, a value of -141253 is reached, compared to a value of 66641 obtained through the sum of all the positive terms. By distributing this load along the sample's time axis, we found that this negative value is constant and offers hardly any temporal variation.

Figure 4 details the ten terms that contribute to the positive and negative scores in the corpus studied. Thus, a more semantic perspective is possible. According to the AFINN classifier, the most negative contributing terms are avoid, crisis, and contagion, while the positive ones, with much less total impact, are support, help, and thanks.

By analyzing semantically the words that contribute most to the AFINN score in the sample, it can be observed that the negative charge has to do with a set of reactions to the pandemic, which has brought struggle, pain, boredom, frustration, fear, hunger, risk, and death. All these feelings and sensations have almost twice as much score in the sample than the semantics of positive reactions, related to mutual support, cooperation between united citizens, gratitude to health authorities, and the longing for safety, freedom, or love that the Twitter users share.

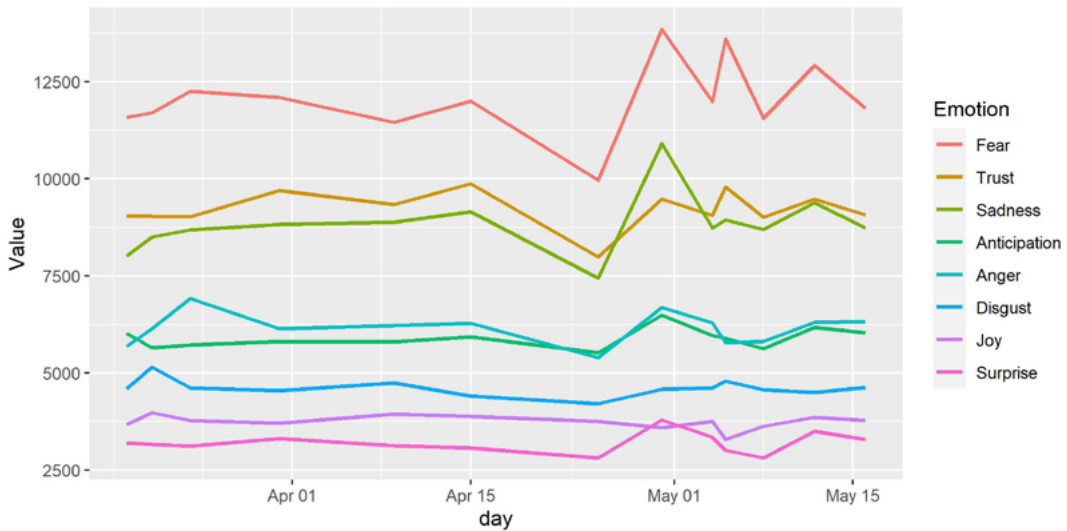
This negative load of the sample is confirmed by the analysis using the NRC classifier (figure 5). Although this classifier is more specific when evaluating emotions, there are those with negative connotations as opposed to positive ones. In this case, high values predominate in the negative emotion of fear, with a rating above 12,000 points. This is followed by the negative emotion of sadness, with a rating close to 10,000 points. However, with a similar rating is the score for the positive emotion of confidence, which is also around 10,000 points, making it the predominant positive emotion in the sample. This is followed by the emotions of anger and expectation – negative and positive, respectively– with a rating slightly above 6,000 points.

Although for the AFINN classifier the negative load is consistently predominant throughout the period studied, it is interesting to note the variability over time of the results obtained using the NRC classifier. As can be seen in figure 5, the emotional load in NRC drops ostensibly in mid-April, but rises again, even more drastically, at the end of that same month, with the highest values in early May, especially the emotion of fear. This moment coincides with the peak detected in figure 2 of the hashtags #mexiquensesenriesgo and #m xicoconhambre.



**Figure 4. Words that contribute most to negative and positive sentiment according to the AFINN classifier. In absolute terms, the count of the negative terms is much higher**

*Source: Own elaboration.*



**Figure 5. Feelings' evolution according to the NRC classifier in the longitudinal analysis throughout the period studied**

*Source: Own elaboration.*

Based on the analyses it can be affirmed that, rather than repudiation, anger, or sadness, the COVID-19 pandemic has provoked fear in the population; at the other extreme, there are very low scores for emotions such as surprise and joy. Semantically, the terms in the sample that contribute most to the high fear score are pandemic, quarantine, avoid, fight, hospital, or emergency, some of which also produce high values for sadness. On the other hand, the sample terms that contribute to trust are president, hospital –which appears in the classifier also scoring as fear–, personal, economy, or important.

### **Conversation and Echo Chamber around COVID-19**

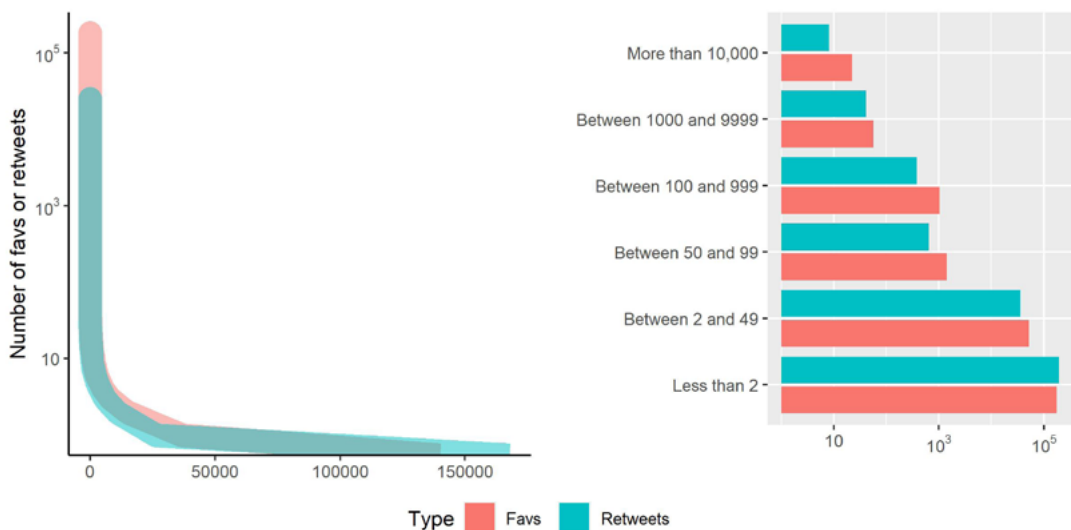
Kwak, Lee, Park, and Moon (2010) and Colleoni, Rozza, and Arvidsson (2014) identify two functions for Twitter: on the one hand, it is considered a platform where public conversation takes place, sustained by participatory audiences that create, share, and distribute content, which Castells (2009) identifies as mass self-communication; on the other, it is seen as a powerful medium for disseminating news created by traditional media and as an instrument of expression for politics. In the latter case, tweets work as an echo chamber by amplifying and disseminating on a large scale a reduced set of public voices. In the expression about the coronavirus, both instances are present.

As we work with a sample without retweets, it can be argued that the lexical analysis sufficiently represents the expression of ordinary users, since it contains at least 99% of posts that are not assumed to be amplifications of high-visibility

users. Thus, the sample has the form of a long-tail graph: a head of few highly amplified publications –left of figure 6– and a tail of many publications without amplification, which respond to the expression of individual users. On the right of figure 6, it can be seen that the tweets in the sample that received more than one thousand retweets are less than fifty, that those that received between one hundred and one thousand retweets are less than five hundred, and that almost the entire sample received less than two retweets.

For all these reasons, the amplification of voices is greater than self-communication. In fact, if we consider the sum of retweets of the 49 most replicated tweets, it practically doubles the total of the sample studied: only eight publications have been replicated more than ten thousand times. Their voices are divided between the president of Argentina, the governor of the state of Jalisco, Mexico, the president of the community of Madrid, two popular journalists from Spain and Mexico, a youtuber, and two Spanish media, *El País* and *ABC*, which have played a fundamental role in the dissemination of news about the coronavirus (Lázaro-Rodríguez & Herrera-Viedma, 2020).

The topic modeling previously conducted, especially the characterization of topic 3, along with the frequency analysis, allows us to assume that the messages produced by the media are the ones that have an impact on the production and circulation of messages among ordinary users.



**Figure 6. Two ways of showing the large number of tweets that have few interactions and the large number of interactions that very few tweets have**

*Source: Own elaboration.*



## DISCUSSION AND CONCLUSIONS

Regarding the first objective –to determine the issues of importance in the public conversation– several relevant topics have been identified for the initial phase of the pandemic. Although the most frequently used has also been the one with the most negative connotations, because of its reference to terms such as pandemic, deaths, cases, and contagions, the topic modeling has shown that it is possible to find other themes. For example, the second most recurrent topic refers to healthcare, preventive measures, and the importance of staying at home. This positive aspect is reinforced by the large use of the hashtag #quedateencasa, which is proactive and, in contrast to the negative diagnosis offered by the frequent words, performatively proposes the execution of the only palliative in the initial phase of the pandemic.

Through topic modeling, it has also been found that the words persons and health are the most central, even though they are not among the most frequent in the sample. This indicates that, even if the frequency of the conversation is semantically marked by cases, government, pandemic, and quarantine, what is at the center of the conversation are people and health. This observation is reinforced by the extraordinary adhesion that the term people obtains in number of retweets. Finally, topic modeling also allowed us to identify different perspectives on the pandemic according to geographical locations: in Mexico, the economic crisis, in Argentina, government measures, and in Spain, the quarantine carried out in private spaces.

The sentiment and emotional analysis clearly shows a balance in favor of negative feelings. It is observed that fear is the first emotion in the sample, followed by confidence and sadness. Especially in the second half of April, the emotion of sadness exceeds that of trust by a peak. This may be due to two factors: on the one hand, in mid-April there is a main rise in deaths worldwide ; in addition, the hashtags #mexicanosconhambre and #mexiquensesenrisgo –the only two detected as orchestrated in the sample– have a rapid rise in frequency of use in the same period, at the end of April and the beginning of March. It can be conjectured that they contribute, intentionally, to install discomfort among users. Further study is needed to corroborate this hypothesis.

In a longitudinal approach, it can be concluded that the narrative tends to become more concerned with economic and political aspects, as the use of the hashtag #quedateencasa assumes a downward curve after the first few weeks, coinciding with the use of hashtags with a tone of political discussion and complaint –even when they may respond to an orchestrated strategy–, as well as with the rise of the hashtag #pandemia, which clearly connotes an increase in concern.

In general, data collection techniques and computational text analysis allow reorganizing the information in a chronological logic: it was possible to identify the evolution of perceptions about COVID-19 in its early days for a global Spanish Twitter audience, and even perceptions about the different reception of the pandemic by country. Nevertheless, a discussion remains about how findings produced with online data can be linked to offline experiences and data (Rogers, 2019). It is understood that, due to the impact that platforms currently have, both in their number of users and the close links they sustain with other media such as the press, television, or other widely used social platforms such as Facebook, they can generate public opinion trends, install issues on the public agenda and promote feelings in the population. The great emotional depth generated in social networks is probably one of the best examples that these messages affect in a way that is not merely informational.

As for the second objective –to determine whether the amplification of the word of high-visibility users or the self-expression of ordinary users prevails on Twitter– the analysis of the interactions reveals that they correspond to a sample made up of 99% of tweets that are not assumed to be retweets or exact replicas of authoritative voices. However, the amplification aspect of high-visibility users should not be underestimated, since the sum of retweets of the 50 most disseminated tweets in the sample more than doubles the number of tweets with two retweets or less. These findings help to reveal that, for the sample studied, the amplification of messages produced by high-visibility users is quantitatively more important than the production of messages by ordinary users. Thus, although it is possible to study public opinion trends, it is evident that the platform has been used to a greater extent to reproduce messages than to produce them.

The Twitter platform goes beyond being a means of reproducing media messages that the public opinion accesses for consultation; it becomes a central element in the production of interactions that influence the public opinions of ordinary users. When observed massively, they indicate general trends of perspectives and feelings regarding events that are collectively constructed within the platform. This relationship between the amplification of public voices and the expression of ordinary users, as well as the content analysis of the most replicated posts in the case of COVID-19, have already been addressed in a complementary study that identifies which are the relevant high-visibility actors and which topics are expressed in them (Sued & Cebal, 2020).

This study has some limitations. First, by collecting data based on Spanish, the particularities that the pandemic assumes in each country are left aside. Given that political and public health decisions regarding this issue vary, future studies

are needed to develop this perspective more specifically. Second, the paper only analyzes Twitter texts; other digital objects, such as images and hyperlinks, could be relevant. Third, part of the sample may respond to positioning strategies of political hashtags; further research is needed to know how they could affect the generality of the sample. Fourth, the findings obtained are validated for the universe of tweets collected in the period studied, the first phase of the pandemic. To deepen the findings and understand their dynamism, more extensive monitoring of the subject would be necessary.

Although it is to be expected that at a time of pandemic and uncertainty, frequent words, topics, and feelings tend to be negative, other possible reactions have been shown throughout the article, in agreement with the study by Han and colleagues (2020) on the Chinese network Sina Weibo, which shows a different perception of the pandemic, focused on topics such as epidemic control and dissemination of government information, opinions on the closure of cities, health protection measures or the importance of anti-pandemic knowledge. These differences reaffirm the relevance of cultural and geographic conditioning when accounting for experiences during the pandemic and further amplify the article's findings.

The literature (Davis et al., 2014; Towers et al., 2015) indicates that media frames are closely related to everyday perceptions of the pandemic. This previous finding is confirmed in the topic modeling analysis, especially the characterization of topic 3 which, along with the frequency analysis, allows us to assume that media messages impact the production and circulation of messages among ordinary users. This link between the amplification of public voices and user expression, as well as the content analysis of the most replicated publications in the case of COVID-19, should be addressed in future research.

This study also corroborates that, at least in health crisis conditions, the meaning of health communication is constructed among a heterogeneous set of actors and expressions (Rodríguez Roura et al., 2018). In this regard, the findings of Kullar and colleagues (2020) -who highlight the usefulness of Twitter to study health communication, but restrict it to specialists, professionals, and patients- and those of Rufai and Bunce (2020) on its use by political leaders to disseminate information on care are extended. This study adds a dimension associated with public opinion, less specialized, but more linked to the perspectives on the pandemic expressed by ordinary users.

The findings of this study are of interest to the area of health communication, as they show a type of public conversation in which different publicly available narratives about the pandemic converge. Although topic modeling suggests that the narrative linked to cases, deaths, and contagions is the most promoted and

generates the most negative feelings, other narratives of care, prevention, and medical attention are present. This diagnosis can be used to insist on the importance of the presence of opinion leaders, health institutions, and authorities in social networks to disseminate messages linked to care and prevention. The findings can be considered to understand how social networks intervene in the collective construction of topics associated with health, in this case, distributing pessimistic narratives over proposals for care, prevention, and action by health professionals. Subsequent studies will be able to identify which social actors promote one or the other, what is their level of acceptance by users, and, above all, what reinforcements are needed to use social platforms as a means of disseminating care and raising awareness of health measures for the general population, rather than as a means of spreading fear.

## REFERENCIAS

- 2018 Research on 100 Million Tweets: What it Means for Your Social Media Strategy for Twitter. (2018, March 12). Vicinitas. Retrieved from <https://www.vicinitas.io/blog/twitter-social-media-strategy-2018-research-100-million-tweets>
- Arcila-Calderón, C., Alonso-Palacio, L. M., & García-Jiménez, A. (2018). Enfoques big data para la comunicación en salud: Análisis de redes y de sentimiento a gran escala (Big data approaches in health communication: Network analysis and large—scale supervised sentiment analysis). *Revista Científica Salud Uninorte*, 34(1). Retrieved from <http://rcientificas.uninorte.edu.co/index.php/salud/article/view/10544>
- Blei, D. M., Ng, Y. A., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, (3), 993-1022. Retrieved from <https://www.jmlr.org/papers/v3/blei03a.html>
- Bruns, A. & Highfield, T. (2016). Is Habermas on Twitter? Social Media and the Public Sphere. In A. Bruns, G. Enli, E. Skogerbø, A. O. Larsson, & C. Christensen (Eds.), *The Routledge Companion to Social Media and Politics* (pp. 56-73). New York, NY: Routledge.
- Bruns, A. & Weller, K. (2016). Twitter as a first draft of the present: And the challenges of preserving it for the future. In *Proceedings of the 8th ACM Conference on Web Science* (pp. 183-189). <https://doi.org/10.1145/2908131.2908174>
- Burgess, J. & Bruns, A. (2015). Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research after the Computational Turn. In G. Elmer, G. Langlois, & J. Redden (Eds.), *Compromised Data. From social Media to Big Data* (pp. 93-111). New York, NY: Bloomsbury Academic.
- Burgess, J. & Green, J. (2019). *YouTube: Online Video and Participatory Culture. Second Edition*. Cambridge, United Kingdom: Polity.
- Castelfranchi, Y. (2017). Computer-Aided Text Analysis: An Open-Aired Laboratory for Social Sciences. *Journal of Science Communication*, 16(2), C04. <https://doi.org/10.22323/2.16020304>
- Castells, M. (2009). *Comunicación y poder (Communication Power)*. Madrid, Spain: Alianza.

- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily on Twitter Using Big Data: Political Homophily on Twitter. *Journal of Communication*, 64(2), 317-332. <https://doi.org/10.1111/jcom.12084>
- Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. In *InterJournal: Vol. Complex Systems* (p. 1695). Retrieved from <https://igraph.org>
- ¿Cómo es el consumo digital en México en el marco de la pandemia? (How is Digital Consumption in Mexico in the Context of the Pandemic?). (2020, May 5). *El Universal*. Retrieved from <https://www.eluniversal.com.mx/techbit/como-es-el-consumo-digital-en-mexico-en-el-marco-de-la-pandemia>
- Davis, M., Lohm, D., Flowers, P., Waller, E., & Stephenson, N. (2014). “We Became Sceptics”: Fear and Media Hype in General Public Narrative on the Advent of Pandemic Influenza. *Sociological Inquiry*, 84(9), 499-518. <https://doi.org/10.1111/soin.12058>
- De Souza Santos, B. & Meneses, M. P. (2014). Prólogo (Foreword). In B. de Sousa Santos & M. P. Meneses (Eds.), *Epistemologías del Sur: Perspectivas* (Epistemologies of the South: Perspectives) (pp. 7-17). Madrid, Spain: Akal.
- Díaz, G. (2016). Stopwords-es. Retrieved from <https://github.com/stopwords-iso/stopwords-es/blob/master/stopwords-es.txt>
- franzke, a. s., Bechman, A., Zimmer, M., & Ess, C. (2020). *Internet Research: Ethical Guidelines 3.0*. Retrieved from <https://aoir.org/reports/ethics3.pdf>
- Grün, B., Hornik, K., & Blei, D. (2020). Topic Models (0.2-11). Retrieved from <https://cran.r-project.org/package=topicmodels>
- Gutiérrez, E. M., Viridis, J. M., Meller, L., Gurovich, C., Simón, F. D., & Leyes, L. F. (2020). ¿Retuitear o responder? Covid-19 y Twitter. El caso de la ciudad de Bahía Blanca - Argentina (Retweet or reply? Covid-19 and Twitter. The case of the city of Bahía Blanca, Argentina). *Awari*, 1(2). <https://doi.org/10.13140/RG.2.2.31471.87203>
- Haman, M. (2020). The use of Twitter by state leaders and its impact on the public during the COVID-19 pandemic. *Heliyon*, 6(11), e05540. <https://doi.org/10.1016/j.heliyon.2020.e05540>
- Han, X., Wang, J., Zhang, M., & Wang, X. (2020). Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. *International Journal of Environmental Research and Public Health*, 17(8), 2788. <https://doi.org/10.3390/ijerph17082788>
- Kearney, M. W. (2020). rtweet: Collecting Twitter Data. (1.2.5). Retrieved from <https://cran.r-project.org/package=rtweet>
- Kullar, R., Goff, D. A., Gauthier, T. P., & Smith, T. C. (2020). To Tweet or Not to Tweet—A Review of the Viral Power of Twitter for Infectious Diseases. *Current Infectious Disease Reports*, 22, 14. <https://doi.org/10.1007/s11908-020-00723-0>
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591–600). New York, NY: ACM. <https://doi.org/10.1145/1772690.1772751>

- Lázaro-Rodríguez, P. & Herrera-Viedma, E. (2020). Noticias sobre Covid-19 y 2019-nCoV en medios de comunicación de España: El papel de los medios digitales en tiempos de confinamiento (News about Covid-19 and 2019-nCoV in the Spanish Media: the Role of digital media in times of confinement). *El Profesional de la Información*, 29(3), e290302. <https://doi.org/10.3145/epi.2020.may.02>
- Lin Pedersen, T. (2020). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks (2.0.2)*. Retrieved from <https://cran.r-project.org/package=ggraph>
- Marres, N. (2017). *Digital sociology: The reinvention of social research*. Malden, MA: Polity.
- Mendoza Vega, J. B. (2018, April 15). Análisis de sentimientos con R - Léxico Afinn (Sentiment Analysis with R - Afinn Lexicon). *R Pubs*. Retrieved from [https://rpubs.com/jboscomendoza/analisis\\_sentimientos\\_lexico\\_afinn](https://rpubs.com/jboscomendoza/analisis_sentimientos_lexico_afinn)
- Meneses Rocha, M. E. (2018). Grandes datos, grandes desafíos para las ciencias sociales Big Data, Big challenges for social sciences). *Revista Mexicana de Sociología*, 80(2), 415-444. Retrieved from <http://revistamexicanadesociologia.unam.mx/index.php/rms/article/view/57723>
- Milan, S. & Treré, E. (2017, October 16). Big Data from the South: The beginning of a Conversation We Must Have. *DATACTIVE*. Retrieved from <https://data-activism.net/2017/10/bigdatasur/>
- Mohammad, S., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises* (arXiv:1308.6242). Retrieved from <https://arxiv.org/abs/1308.6242>
- Moreno, A. & Redondo, T. (2016). Text Analytics: The Convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(6), 57. <https://doi.org/10.9781/ijimai.2016.369>
- Newman, N., Fletcher, R., Schulz, A., Simge, A., & Kleis Nielsen, C. (2020). *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism. Retrieved from [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR\\_2020\\_FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf)
- Nielsen, F. (2011). A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages* (arXiv:1103.2903). Retrieved from <https://arxiv.org/abs/1103.2903>
- Omena, J. J., Rabello, E. T., & Mintz, A. G. (2020). Digital Methods for Hashtag Engagement Research. *Social Media + Society*, 6(3), 2056305120940697. <https://doi.org/10.1177/2056305120940697>
- Pantojas García, E. (n.d.). La sociedad translocal: Notas para entender el cambio de época (The Translocal Society: Notes to Understand the Change of Age). *CLACSO*. Retrieved from <https://www.clacso.org/la-sociedad-translocal-notas-para-entender-el-cambio-de-epoca/>
- Pérez-Dasilva, J.-A., Meso-Ayerdi, K., & Mendiguren-Galdospín, T. (2020). Fake news y coronavirus: detección de los principales actores y tendencias a través del análisis de las conversaciones en Twitter (Fake news and coronavirus: Detecting key players and trends through analysis of Twitter conversations). *El Profesional de la Información*, 29(3), e290308. <https://doi.org/10.3145/epi.2020.may.08>
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook. *New Media & Society*, 20(1), 293-310. <https://doi.org/10.1177/1461444816661553>

- Robichaud, A. & Blevins, C. (2011). Content-Based Analysis. *Tooling Up Digital Humanities*. Retrieved from [http://toolingup.stanford.edu/?page\\_id=3](http://toolingup.stanford.edu/?page_id=3)
- Robinson, D. (2020). widyr: Widen, Process, then Re-Tidy Data (0.1.3). Retrieved from <https://cran.r-project.org/package=widyr>
- Robinson, D. & Silge, J. (2020). tidytext: Text Mining Using «dplyr», «ggplot2», and Other Tidy Tools (0.2.4). Retrieved from <https://cran.r-project.org/package=tidytext>
- Rodríguez Roura, S. C., Cabrera Reyes, L. de la C., & Calero Yera, E. (2018). La comunicación social en salud para la prevención de enfermedades en la comunidad (Social communication in health for disease prevention in the community). *Humanidades Médicas*, 18(2), 384-404. Retrieved from <http://www.humanidadesmedicas.sld.cu/index.php/hm/article/view/1302>
- Rogers, R. (2018). Otherwise Engaged: Social Media from Vanity Metrics to Critical Analytics. *International Journal of Communication*, 12, 23. Retrieved from <https://ijoc.org/index.php/ijoc/article/view/6407>
- Rogers, R. (2019). *Doing Digital Methods*. London, United Kingdom: SAGE.
- Rufai, S. R. & Bunce, C. (2020). World Leaders' Usage of Twitter in Response to the COVID-19 Pandemic: A content analysis. *Journal of Public Health*, 42(3), 510-516. <https://doi.org/10.1093/pubmed/fdaa049>
- Sued Palmeiro, G. E. & Cebal Loureda, M. (2020). Voces autorizadas en Twitter durante la pandemia de COVID-19: Actores, léxico y sentimientos como marco interpretativo para usuarios ordinarios (Authorized Twitter voices during the COVID-19 pandemic: actors, vocabulary, and feelings as an interpretive framework for ordinary users). *Revista de Comunicación y Salud*, 10(2), 549-568. [https://doi.org/10.35669/rcys.2020.10\(2\).549-568](https://doi.org/10.35669/rcys.2020.10(2).549-568)
- Thelwall, M. & Levitt, J. M. (2020). Retweeting Covid-19 Disability Issues: Risks, Support and Outrage. *El Profesional de la Información*, 29(2). <https://doi.org/10.3145/epi.2020.mar.16>
- Thelwall, M. & Thelwall, S. (2020). Covid-19 Tweeting in English: Gender Differences. *El Profesional de la Información*, 29(3). <https://doi.org/10.3145/epi.2020.may.01>
- Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., ... Castillo-Chavez, C. (2015). Mass Media and the Contagion of Fear: The Case of Ebola in America. *PLOS ONE*, 10(6), e0129179. <https://doi.org/10.1371/journal.pone.0129179>
- UNDP. (2020). Understanding what data tell us about COVID-19's socio-economic impact. Retrieved on May 21, 2021 from <https://www.undp.org/blogs/understanding-what-data-tell-us-about-covid-19s-socio-economic-impact>
- van Dijck, J. (2016). *La cultura de la conectividad: Una historia crítica de las redes sociales* (The Culture of Connectivity: A critical History of Social networks). Buenos Aires, Argentina: Siglo XXI.
- Vega, F. (2020, March 24). Los medios tradicionales recuperan poder y credibilidad con la pandemia provocada por el Coronavirus (Traditional Media Recover Power and Credibility with the Pandemic Caused by the Coronavirus). *Comscore*. Retrieved from <https://www.comscore.com/lat/Prensa-y-Eventos/Blog/Los-medios-tradicionales-recuperan-poder-y-credibilidad-con-la-pandemia-provocada-por-el-Coronavirus>

Wikström, P. (2014). #srynotfunny: Communicative Functions of Hashtags on Twitter. *SKY Journal of Linguistics*, 27, 127-152. Retrieved from <http://www.linguistics.fi/julkaisut/SKY2014/Wikstrom.pdf>

### ACKNOWLEDGMENTS

Dr. Sued would like to thank the Postdoctoral Fellowship Program for Mexico 2020-2021 and the National System of Researchers, Conacyt México, for their support in the preparation of this paper.

### ABOUT THE AUTHORS

**MANUEL CEBRAL LOUREDA**, Ph.D. in Philosophy from the Universidad de Santiago de Compostela (Spain) and master in Data Mining from UNED (Spain). Since 2019 he has been a tenure professor at the Tecnológico de Monterrey (Mexico). His interests revolve around digital humanities and digital culture, combining the humanistic approach with data-driven methods through computational tools such as R programming. He is currently a member of the National System of Researchers (Mexico) at Candidate level.

 <https://orcid.org/0000-0001-6359-2427>

**GABRIELA ELISA SUED PALMEIRO**, Ph.D. in Humanistic Studies from the Tecnológico de Monterrey, Mexico. She is a Conacyt postdoctoral researcher at the Institute of Social Research of the Universidad Nacional Autónoma de México. She teaches undergraduate and graduate courses at the Tecnológico de Monterrey. Her lines of research focus on digital culture, social platform studies, and data-centered methods.

 <https://orcid.org/0000-0002-4516-678X>