



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Geographical analysis of the opinion and influence of users on Twitter during the coronavirus health crisis

DEGREE FINAL WORK

Degree in Computer Engineering

Author: Luis Carbonell Gironés

Tutor: Enrique Orduña Malea

Experimental tutor: Cristina I. Font

Course 2019-2020

Resumen

Desde marzo de 2020, el mundo se ha visto afectado por una pandemia causada por el COVID-19. En este estudio, analizamos el impacto de la crisis de salud pública en la opinión general y los intereses de la población. Para ello, los datos de Twitter se obtienen y procesan para mostrar los temas principales durante el período de tiempo analizado. La primera parte del estudio será un enfoque general para el análisis de los intereses y sentimientos de los usuarios de Estados Unidos y el Reino Unido, desde el 1 de diciembre de 2019 hasta el 15 de mayo de 2020. La segunda parte de este estudio será un análisis más detallado de los intereses de los usuarios de Twitter en España durante el periodo de confinamiento desde el 14 de marzo hasta el 21 de junio. El resultado muestra que los tweets en el conjunto de datos en inglés son en su mayoría endogamic, ya que su preocupación por las noticias internacionales es por su preocupación por el impacto que la pandemia puede tener en sus países. En el conjunto de datos español, sin embargo, los usuarios comentan noticias internacionales y noticias locales también. En ambos conjuntos de datos, hay una disminución en la actividad de tweeting a medida que evoluciona la pandemia.

Palabras clave: coronavirus, covid-19, análisis de sentimientos, análisis de temas, twitter

Resum

Des del març del 2020, el món s'ha vist afectat per una pandèmia provocada per la COVID-19. En aquest estudi analitzem l'impacte de la crisi de salut pública en l'opinió general i els interessos de la població. Per fer-ho, les dades de Twitter s'obtenen i es processen per mostrar els temes principals durant el termini analitzat. La primera part de l'estudi serà una aproximació general a l'anàlisi dels interessos i sentiments dels usuaris nord-americans i del Regne Unit, des de l'1 de desembre de 2019 fins al 15 de maig de 2020. La segona part d'aquest estudi serà una anàlisi més detallada dels interessos dels usuaris de Twitter a Espanya durant el període de confinament del 14 de març al 21 de juny. El resultat mostra que els tuits del conjunt de dades anglès són majoritàriament endogàmics, ja que la seva preocupació per les notícies internacionals està fora de la seva preocupació per l'impacte que la pandèmia pot tenir als seus països. En el conjunt de dades espanyol, però, els usuaris també comenten notícies internacionals i notícies locals. En ambdós conjunts de dades, hi ha una disminució de l'activitat tweeting a mesura que evoluciona la pandèmia.

Paraules clau: coronavirus, covid-19, anàlisi de sentiments, anàlisi de temes, twitter

Abstract

From march 2020, the world has been affected by a pandemic caused by the COVID-19. In this study, we analyse the public health crisis impact on the overall opinion and interests of the population. To do so, data from Twitter is fetched and processed to show the main topics during the analysed time frame. The first part of the study will be a general approach to the analysis of the interests and sentiments of the US and UK users, from the 1st of December 2019 until the 15th of May 2020. The second part of this study will be a more detailed analysis of the interests of Twitter users in Spain during the confinement period from the 14th of March until the 21st of June. The outcome shows that the tweets in the English dataset are mostly endogamic, as their concern for international news is out of their concern of the impact that the pandemic may have in their countries. In the Spanish dataset, however, the users comment on international news and local news

as well. In both datasets, there is a decrease in the tweeting activity as the pandemic evolves.

Key words: coronavirus, covid-19, sentiment analysis, topic analysis, twitter

Contents

Contents	v
List of Figures	vii
List of Tables	viii
<hr/>	
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	2
1.3 Structure	2
1.4 Research background	2
2 Methodology	5
2.1 Data gathering	5
2.2 Preprocessing	6
2.3 Topic models	7
2.4 Data visualisation	10
3 Results	15
3.1 English dataset	15
3.2 Spanish dataset	30
4 Discussion	79
5 Conclusions	81
Bibliography	82
<hr/>	
Appendix	
A Supplementary information	85

List of Figures

1.1	Network representation of related papers. Source: Scopus; powered by VOSviewer.	3
2.1	Data processing flowchart	11
2.2	US & UK users' data	12
2.3	US & UK topics' data	12
2.4	Spain users' data	13
2.5	Spain topics' data	13
3.1	English dataset choropleth map	15
3.2	Tweets per day in the English dataset	16
3.3	Timeline chart of the polarity of the tweets in the English dataset	16
3.4	Top 20 words in topic 12	23
3.5	Top 20 words in topic 19	25
3.6	Top 20 words in topic 20	26
3.7	Timeline charts of the polarity of the tweets in each topic (US UK)	27
3.8	Spanish dataset choropleth map	30
3.9	Spanish dataset users choropleth map	30
3.10	Spanish dataset tweets choropleth map	31
3.11	Top 20 words in topic 1	34
3.12	Top 20 words in topic 2	35
3.13	Top 20 words in topic 3	36
3.14	Top 20 words in topic 4	37
3.15	Top 20 words in topic 5	38
3.16	Top 20 words in topic 6	39
3.17	Top 20 words in topic 7	40
3.18	Top 20 words in topic 8	41
3.19	Top 20 words in topic 9	42
3.20	Top 20 words in topic 10	43
3.21	Top 20 words in topic 12	45
3.22	Top 20 words in topic 13	46
3.23	Top 20 words in topic 14	47
3.24	Top 20 words in topic 15	48
3.25	Top 20 words in topic 16	49
3.26	Top 20 words in topic 17	50
3.27	Top 20 words in topic 18	51
3.28	Top 20 words in topic 19	52
3.29	Top 20 words in topic 20	53
3.30	Top 20 words in topic 21	54
3.31	Top 20 words in topic 22	55
3.32	Top 20 words in topic 23	57
3.33	Top 20 words in topic 24	58
3.34	Top 20 words in topic 25	60
3.35	Top 20 words in topic 26	61

3.36	Top 20 words in topic 27	62
3.37	Top 20 words in topic 28	63
3.38	Top 20 words in topic 29	64
3.39	Top 20 words in topic 30	66
3.40	Top 20 words in topic 31	68
3.41	Top 20 words in topic 32	69
3.42	Top 20 words in topic 33	70
3.43	Top 20 words in topic 34	72
3.44	Top 20 words in topic 35	73
3.45	Timeline charts of the tweeting activity in each topic (Spain) I	74
3.46	Timeline charts of the tweeting activity in each topic (Spain) II	77

List of Tables

2.1	User geolocation distribution in the datasets	6
2.2	Bot rate of the datasets	6
2.3	Coherence scores, minimum word freq. and top words removed of each topic model	8
2.4	Coherence scores of the topic models generated	8
2.5	Initial hyperparameters of the topic model in the English dataset	8
2.6	Optimised hyperparameters of the topic model in the English dataset	9
2.7	Coherence scores of the topic models from the Spanish dataset	9
2.8	Initial hyperparameters of the topic model in the Spanish dataset	9
2.9	Optimised hyperparameters of the topic model in the Spanish dataset	10
3.1	List of topics obtained after applying LDA in the English dataset	17
3.2	List of topics obtained after applying LDA in the Spanish dataset	32

CHAPTER 1

Introduction

Nowadays, social networks are one of the largest content generators that exist¹. People share and express their ideas about almost any existing topic, and information is shared instantaneously about any event happening around the globe.

Such a vast amount of information is shared by posting on social media: concerns, reactions and interests. All this data can be useful to understand better the society and the concerns of the population regarding specific topics as well as finding issues that arise due to an event.

The challenge consists of gathering all that content from the social media platforms and processing it, so it is easier to understand. Such a large amount of data requires to use natural language processing techniques to handle all the posts gathered, a system for storing the data has to be chosen so that it allows to query the data quickly, methods that extract the essential facts from the data gathered and finally a platform to show the ideas and quantitative data obtained during the study.

Disciplines like Cybermetrics or Computational Social Science have surged due to the interest to analyse the large volumes of data that are generated daily, which are essential especially to social sciences. For such purpose, it is necessary to apply multiple research methods and analysis techniques. The results from analysing this data, can outcome useful information when making decisions, or to learn and understand social, cultural, linguistical and behavioural aspects of our society. These aspects are what makes the results so useful in certain areas, as for example public healthcare policies.

1.1 Motivation

Throughout the bachelor's degree, I have enjoyed learning methods on how to store, manage and process data from a merely entrepreneurial or business mindset: designing and building to get the desired outcome. Because my mindset has always been intrinsically curious, I did not enjoy most of the projects and tasks assigned, even though I understood its goal and usefulness, because the outcome was always predictable and did not make me learn anything new other than how to apply the theory seen in class.

At my elected class of Social Web Behaviour & Network Analysis, I found out how all my previous knowledge to that date for which I had not perceived its utility for an issue of genuine interest to me, was somehow closely integrated with a goal that made a social impact, be it for good or not so good intentions. There, I learned that techniques

¹https://web-assets.domo.com/blog/wp-content/uploads/2017/07/17_domo_data-never-sleeps-5-01.png

like those applied in previous subjects for clustering were used to analyse and eviscerate the information that we produce every second of our lives. It was the before-mentioned moment when I realised that the given approaches could be beneficial to other social issues that could make a positive effect.

This project has given me the possibility of finding a topic that captured my interest, as it was the development of a worldwide social event such as the COVID-19 pandemic.

The health pandemic brought expeditious changes to all of us, in Spain, we moved from being sceptic of the situation and thinking nothing would happen, to being continuously bombed with information about positive cases and hospitalisations. Due to the uniqueness of this tool, Twitter seemed a platform where everyone shared their concerns, interests and opinion regarding anything and everything about the pandemic, and thus the perfect platform to conduct this study.

1.2 Objectives

Given the motivation of this study, the main objective is to identify the topics that concerned the most to Twitter users during the public health crisis.

To achieve such goal, four sub-goals are taken into account:

- Determine if the topics of interest change depending on the location.
- Analyse the perception of each topic by the users.
- Quantify the evolution of the topics in the analysed time frame.
- Design an information pipeline that allows to automatise the procedure (collection, storage, analysis, visualisation) of the data from Twitter.

1.3 Structure

This document is structured in four parts:

Methodology: procedure to gather and process the data is detailed so that it can be understood how the results are obtained.

Results: the data is presented according to the findings from the methodology section, relevant news articles and tweets from the datasets are included to improve the overall understanding.

Discussion: the limitations and possible improvements or modifications that would be made in the future, as well as the findings on the accuracy of the procedure, followed.

Conclusion: answers to the proposed objectives, possible applications in other fields and how this Bachelor's Degree has helped to design and conduct this study.

1.4 Research background

During the COVID-19 pandemic, Twitter has been used in many studies to analyse public opinion and interests as shown in figure 1.1.

Most of them focused on English content except for those that gathered data in multiple languages to study the difference between communities and how their concern of the pandemic evolves [2] or those that investigate the information transmission and what words were most common when transmitting information during the health crisis in specific communities [11].

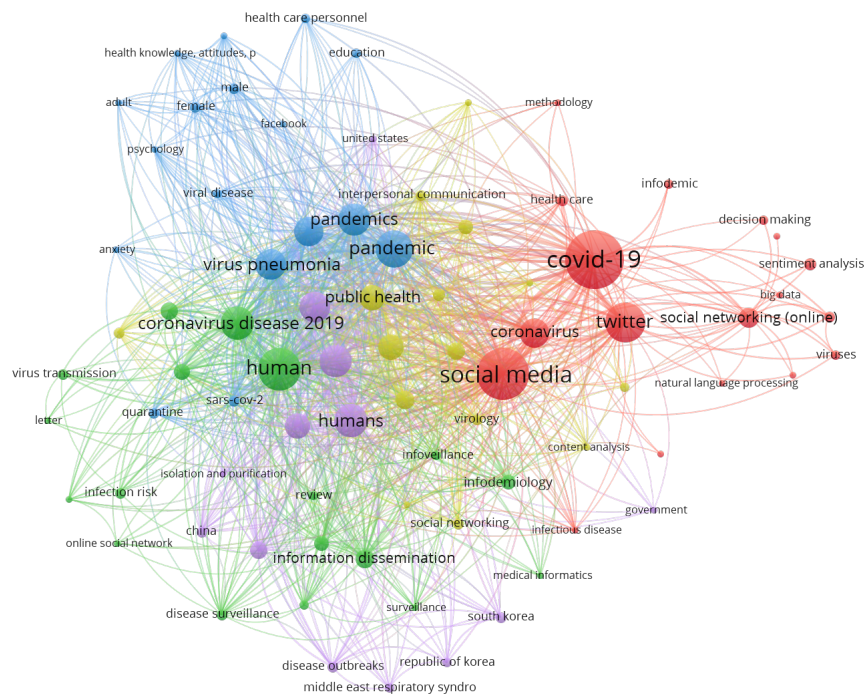


Figure 1.1: Network representation of related papers. Source: Scopus; powered by VOSviewer.

A Psychological study analysed the fear and anxiety caused by COVID-19 on Twitter[13] concluding that despite many scientists and healthcare professionals working around the clock to find a vaccine, which didn't exist at the moment, "people are tremendously living with psychological fear and anxiety throughout the world".

Jahanbin et al. worked on the prediction of the COVID-19 outbreak by using Twitter and web news mining[6] concluding that the networks used reported statistical data that could be used to predict morbidity rates in each region and help design educational programs to train communities on healthcare risks.

The analysis of tweeting trends during the pandemic by Leelawat et al.[9] brought to attention to how trending topics in English, Japanese and Chinese social media are closely related to breaking news and COVID-19 cases.

At the beginning of the pandemic, countries like Italy entered a "buffer zone", a temporal zone between normality and pandemic. During that time, a study from Vicari et al. [15] report a thrive in collective irony that allowed people to temporarily overlooking the feelings and uncertainty provoked by the health crisis.

There's been analysis of the differences between genders when sharing information and the topics of interest, laying apparent distinctions between males and females [14]. Research on how conspiracies and misinformation spread during the pandemic [8] or the prevalence of low-credibility information on Twitter [16] help understand how non-accurate information proliferates.

Political leaders are not exempt from the analysis. Some studies scrutinise the content they post on Twitter to categorise the themes of viral tweets [12] that enables further understanding of how politicians use the social network to communicate with citizens.

Conspiracies have become quite common nowadays, with the internet, the world is connected for good, but that makes any information to be quickly spread. A joint paper

from British and Spanish universities [1] analysed the 5G conspiracy during the COVID-19 pandemic, using Twitter as the source of data to analyse. The results revealed that there were no authority figures, but an isolated group and a broadcast group, and that, fake news websites played a significant role in the spread of such hoaxes and conspiracies.

It is worth noting, the uniqueness of this study regarding the geography analysed as there are no known studies to the author that use a similar technique to analyse the interests of the Spanish population during the state of alarm due to the COVID-19 pandemic.

CHAPTER 2

Methodology

2.1 Data gathering

2.1.1. Tweet download dates

When downloading the data, two datasets were designed: the English and the Spanish dataset, the former containing tweets from the US & the UK and the latter containing tweets from Spain only.

The English dataset tweets were downloaded from the 16th of May 2020 until the 19th of May 2020 with a daily limit of 200,000 tweets a day, for the period comprised between the 2nd of December 2019 and the 18th of May 2020.

The tweets in the Spanish dataset for the period comprised between the 14th of March 2020 and the 22nd of June 2020 were downloaded from the 9th of June 2020 until the 12th of July 2020 with no daily limit. The dataset contains an approximate total of 25.68 million unique tweet IDs.

In both datasets, the query sent to Twitter is an OR statement composed of the main COVID-19 pandemic words: coronavirus, covid, covid-19, covid19. Words were used instead of hashtags to capture all the possible mentions and not limiting the search to only those tweets using the hashtags.

2.1.2. TWINT

The Python toolkit Twint¹ was used to retrieve the tweets. For each tweet, all this data was also stored: tweet id, conversation id, creation time, date, time, timezone, user id, username, name, place, mentions, URLs, photos, number of replies, number of retweets, number of likes, hashtags, cashtags, links, retweet (boolean), quoted URL, video (boolean), near, geo, source, user rt id, user rt (boolean), the id of the original tweet if retweeted, data from the tweet it replies to and retweet date.

2.1.3. Geolocation

The location of the users is gathered by downloading its profile data and analysing the location field and biography looking for any geographic references. In the English dataset, the Python module GeoText is used whereas, on the Spanish dataset, the Python module Geocoder is used together with an instance of the Open Street Maps Nominatim server to

¹<https://github.com/twintproject/twint>

achieve better accuracy in the location prediction. Using different tools was needed due to the accuracy needed: country-level for the English dataset and city or province-level for the Spanish dataset. The technique used in the English dataset may geocode locations to the wrong country if the same city or state is in different countries, and is only able to geocode locations in their English form, i.e. 'Catalunya' or 'Andalucía' would not be identified. Python module Geocoder, despite offering better accuracy, also has the disadvantage of geocoding random locations like "Jupyter" to real places on the planet.

The percentage of geolocated users is shown in the table below:

Dataset	Known location	Unknown location	Total
English	640,998 (28.7%)	1,300,917 (58.23%)	2,233,831
Spanish	2,363,950 (61.78%)	1,461,919 (38.2%)	3,825,869

Table 2.1: User geolocation distribution in the datasets

2.1.4. Filters applied

Once the tweets and users are geolocated, the English dataset is filtered, so only tweets of users from the United States or the United Kingdom are used for further analysis. Likewise, the Spanish dataset only contains tweets from users located in Spain.

2.1.5. Botometer

With the aim of understanding the users base of each processed dataset, the tool Botometer rates the automation of the user accounts with a confidence level of 95% and a confidence interval of 5%. Although this score depends on factors like user activity and may include false positives in the bot rate, it gives an overall view of the number of automated accounts in the dataset. The table below shows the results for both datasets:

2.1.6. Other datasets

Other datasets included in the study of the Spanish dataset regarding COVID-19 data (hospitalisations, discharges, PCRs, tests and deaths) are from the Datadista Github repository².

Dataset	Human	Bot
English	97.00%	3.00%
Spanish	95.05%	4.95%

Table 2.2: Bot rate of the datasets

2.2 Preprocessing

Both datasets were preprocessed to ease the topic modelling process.

²<https://github.com/datadista/datasets/tree/master/COVID%2019>

2.2.1. English dataset

The tweets in the English dataset were first converted to lowercase, then the URLs, hashtags and mentions were removed, then the text was tokenised using the NLTK library, right after the stopwords and non-words were removed and finally, the words whose length is less than three words were removed, the remaining words were lemmatised.

2.2.2. Spanish dataset

The preprocessing in the Spanish dataset was slightly different, everything was the same, except for the removal of the @ and # signs but not the hashtags or account mentions, because it was considered that people use accounts and hashtags to mention places and that were being analysed. On the other hand, the Spanish dataset had a slightly higher requirement for the word length, as words had to be longer than three characters to be kept, so that articles like "de", "del", "el", "la" or "una" were automatically removed, as well as other words that do not add any meaning to the topics. This last change reduced the words bank from 46,238,487 to 38,736,492. This change was done after an initial round of LDA, where the models laid topics with words like: "per i do da as", "per i d amb els", "do da as na non". To lemmatise the words in this dataset, Stanford CoreNLP[10] was used.

2.3 Topic models

2.3.1. LDA and Tomotopy

In order to generate the topic models, the Latent Dirichlet Allocation algorithm was chosen considering the results of the survey by Jipeng, Qian et al. [7] where the tweets dataset had the highest accuracy using GPU-PDMM followed by LDA and PTM. Because of the lack of a Python library implementing the former algorithm, known to the author, LDA was chosen instead due to its popularity and wide range of libraries. Rather than using the popular libraries Gensim or Scikit to analyse the dataset, the Tomotopy toolkit implementing the LDA algorithm [3][4] is used because of its fast results despite the need for more iterations to get the results, this is because Gensim and Scikit use Variational Bayes and Tomkit uses Gibbs Sampling. The model used in the study is selected by taking into account the coherence score of each generated model and the number of topics, as per the conclusions from [5] taking into account the topics generated by each model and the social aspect of each of them.

2.3.2. English dataset

The following topics were generated in the English dataset using Tomotopy 0.9.1 with optimisation enabled and random seeds:

By observing table 2.3, it is apparent that using the minimum term frequency and top words preprocessing the results improved at least by 1%, that is why the other models were generated with those requirements, the Stanford NLP team also recommends on their Topic Modeling Toolbox ³ to remove the top words and set a minimum frequency for each word, which is why this technique is used in the study.

³<https://nlp.stanford.edu/software/tmt/tmt-0.4/>

Model no.	Topics	Min. freq.	Top words	Coherence score
1	10	0	0	0.47208244
2	15	0	0	0.48519221
3	20	0	0	0.52436863
4	10	5	30	0.47258140
5	15	5	30	0.56298295
6	20	5	30	0.58429453

Table 2.3: Coherence scores, minimum word freq. and top words removed of each topic model

Model no.	Topics	Min. freq.	Top words	Coherence score
7	25	5	30	0.59425437
8	30	5	30	0.60461300
9	35	5	30	0.60640748
10	40	5	30	0.62173954
11	45	5	30	0.62347002
12	50	5	30	0.63055895
13	55	5	30	0.63241238
14	60	5	30	0.62654774
15	65	5	30	0.63921376
16	70	5	30	0.63778834
17	75	5	30	0.64428038
18	80	5	30	0.65103415
19	85	5	30	0.64532315
20	90	5	30	0.64262086
21	95	5	30	0.64015042
22	100	5	30	0.65026269

Table 2.4: Coherence scores of the topic models generated

Model number 18 with 80 topics, had a higher coherence score, almost 7% higher than model number 6, but the topics overlapped, that and the relatively high coherence score of model 6 is the reason why a lower coherence model was chosen.

The hyperparameters are tuned by Tomotopy during the optimisation process; the seed is generated randomly unless explicitly specified. The initial parameters used in model 6 are shown in table 2.5.

Alpha	Eta	Seed
0.1	0.01	1634325128

Table 2.5: Initial hyperparameters of model 6

The optimised hyperparameters for model 6 are in table 2.6.

The optimised η is $0.00999999776482582 \approx 0.01$ and α is a vector because it determines both the concentration and the distribution of the dirichlet.

2.3.3. Spanish dataset

The topics generated in the Spanish dataset use the same initial configuration as before, and after applying LDA using Tomotopy, the following results are obtained:

Topic	α Alpha	η Eta	Topic	α Alpha	η Eta
1	0.04786995	≈ 0.01	11	0.0246688	≈ 0.01
2	0.03028768	≈ 0.01	12	0.08299106	≈ 0.01
3	0.02600852	≈ 0.01	13	0.02694346	≈ 0.01
4	0.02367546	≈ 0.01	14	0.02616131	≈ 0.01
5	0.02690362	≈ 0.01	15	0.02476281	≈ 0.01
6	0.0318893	≈ 0.01	16	0.02483426	≈ 0.01
7	0.06597356	≈ 0.01	17	0.04007532	≈ 0.01
8	0.02706708	≈ 0.01	18	0.02944445	≈ 0.01
9	0.03751139	≈ 0.01	19	0.03318707	≈ 0.01
10	0.04596604	≈ 0.01	20	0.02584574	≈ 0.01

Table 2.6: Optimised hyperparameters of model 6

Model no.	Topics	Min. freq.	Top words	Coherence score
1	10	5	30	0.43489614
2	15	5	30	0.50402074
3	20	5	30	0.48818585
4	25	5	30	0.52275900
5	30	5	30	0.53912560
6	35	5	30	0.56245883
7	40	5	30	0.56680728
8	45	5	30	0.56816329
9	50	5	30	0.58335623
10	55	5	30	0.62039953
11	60	5	30	0.63546988
12	65	5	30	0.61586836
13	70	5	30	0.62985727
14	75	5	30	0.63560033
15	80	5	30	0.65025692
16	85	5	30	0.64952227
17	90	5	30	0.63884560
18	95	5	30	0.65002834
19	100	5	30	0.64305672

Table 2.7: Coherence scores of the topic models from the Spanish dataset

The highest coherence score was obtained in model 18 with 95 topics, but as occurred in the English dataset, model 6 was chosen even though it has a lower coherence score because it is more understandable and doesn't overlap as much.

The hyperparameters and seed were set using the same procedure as in the English dataset. The initial parameters used in model 6 are shown in table 2.8.

Alpha	Eta	Seed
0.1	0.01	1095214984

Table 2.8: Initial hyperparameters of model 6

The optimised hyperparameters for model 6 are shown in figure 2.9.

The optimised η is $0.009999999776482582 \approx 0.01$ and α is a vector because it determines both the concentration and the distribution of the dirichlet.

Topic	α Alpha	η Eta	Topic	α Alpha	η Eta
1	0.00522919	≈ 0.01	19	0.01447704	≈ 0.01
2	0.03156144	≈ 0.01	20	0.01143125	≈ 0.01
3	0.01652787	≈ 0.01	21	0.01909995	≈ 0.01
4	0.01705627	≈ 0.01	22	0.02340267	≈ 0.01
5	0.01323911	≈ 0.01	23	0.01361348	≈ 0.01
6	0.00962649	≈ 0.01	24	0.0114844	≈ 0.01
7	0.00555154	≈ 0.01	25	0.02130513	≈ 0.01
8	0.03528611	≈ 0.01	26	0.01846656	≈ 0.01
9	0.03379017	≈ 0.01	27	0.01341356	≈ 0.01
10	0.0124288	≈ 0.01	28	0.01524314	≈ 0.01
11	0.01172946	≈ 0.01	29	0.01625623	≈ 0.01
12	0.03044947	≈ 0.01	30	0.0114897	≈ 0.01
13	0.00894839	≈ 0.01	31	0.01463872	≈ 0.01
14	0.01960727	≈ 0.01	32	0.01743889	≈ 0.01
15	0.01026232	≈ 0.01	33	0.02137626	≈ 0.01
16	0.01643065	≈ 0.01	34	0.0114294	≈ 0.01
17	0.00741075	≈ 0.01	35	0.01730559	≈ 0.01
18	0.0171014	≈ 0.01			

Table 2.9: Optimised hyperparameters of model 6

2.4 Data visualisation

2.4.1. Problem analysis

Because of the large amount of the data used in the study, it is almost impossible to understand the results without plotting them in charts or choropleth maps. Also, another goal of the study is making this data as open and available as possible, and thus creating a dashboard offers an easy tool to see the results of the analysis. The data by itself, are thousands and even millions of tweets and results of categorising them in topics which would have no meaning if they were not organised in a way that makes it clear and easy to understand.

To be able to find correlations in data, as for example, the change of opinion/interest depending on the cases, hospitalisations or deaths, presenting them in the same chart, improves the comprehension and results of the analysis.

2.4.2. Solution design

In order to visualise all this data, it was concluded that it was best to design a dashboard where all the data would be centralised and easily accessible, organised by sections and in a single-page view so that all data related to the same topic would be shown in the same page, structured in grids.

This decision was made because even though it would be simpler to show all the data in charts and plots in this document, it would not have any interactive components which would offer the same experience when trying to browse the data. For example, a static chart does not allow the user to carefully examine the temporal evolution, with the accuracy of the data gathered in the previous processes. On the other hand, the possibility of filtering the data shown in the charts would let the user deepen on the results that might be of greater interest.

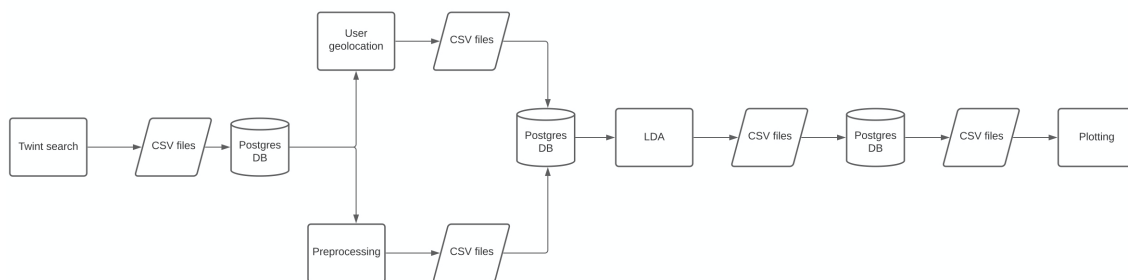


Figure 2.1: Data processing flowchart

Figure 2.1 describes the data pipeline, from the tweet search to the plotting of the charts. The data generated by the processes always resulted in CSV files because this would facilitate the saving speed for the parallel threads during the search and geolocation, it was also used as a backup method, because in the case that the database got corrupted, the CSV files would make it easy to restore the data. Using the database would also allow using fast queries during the analysis, which facilitated filtering the users and its tweets by location.

Because the main purpose of this work was to gather the data and then extract conclusions, the process of designing this interface was reactive rather than planned, that is to say: whenever new data was gathered, new charts and plots were made to see the results, and then they were structured in the dashboard.

2.4.3. Technology used

To develop the solution, two Python libraries were used: Plotly: It is a Python graphing library that allows plotting any data in many formats, allowing to customise every aspect of the resulting chart: colours, fonts, information panels... Its graphs range from bar charts, pie charts, scatter plots or bubble charts, to choropleth maps, heatmaps, contour plots, 3d charts or animations.

Dash: From the creators of Plotly, it is a framework built to create web apps that heavily depend on data visualisation. It is completely integrated with Plotly so that the developer can add any figures to the web app that can interact with forms and are updated dynamically, depending on the user response.

2.4.4. Dashboard

The final data visualisation platform is shown in figures 2.2, 2.3, 2.4 and 2.5. It can be accessed as a Heroku app on <https://upv-tfg-20.herokuapp.com>

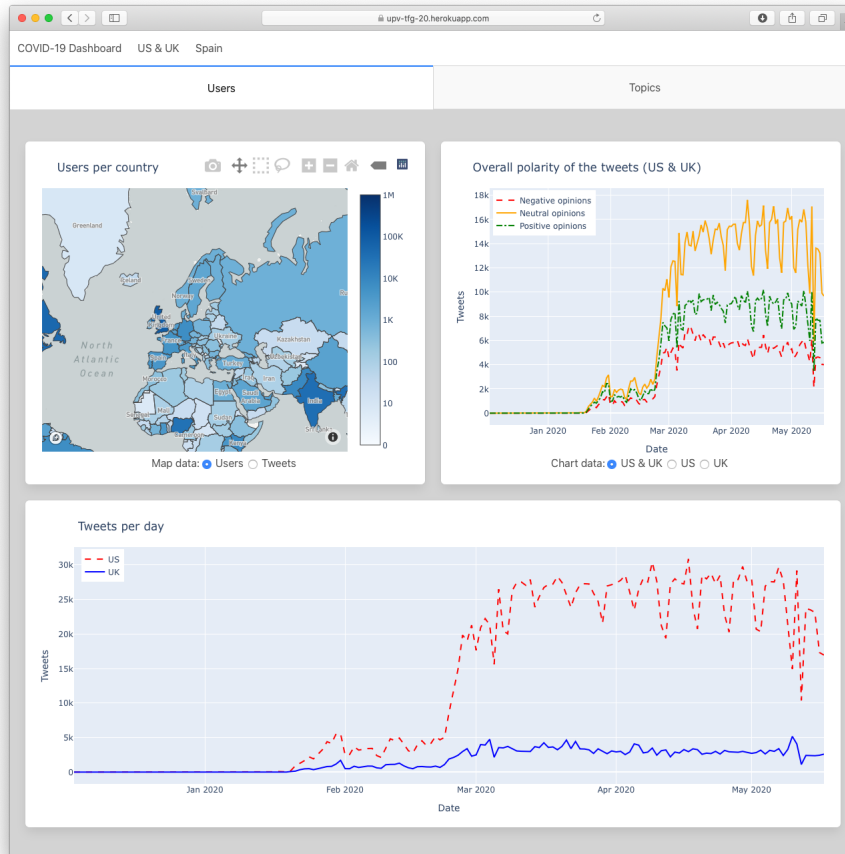


Figure 2.2: US & UK users' data

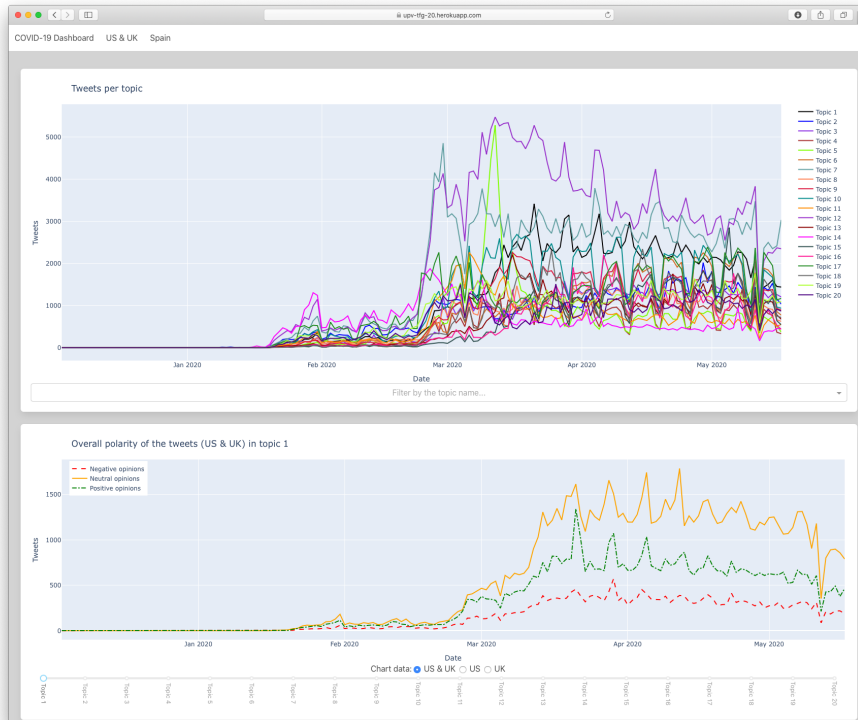


Figure 2.3: US & UK topics' data

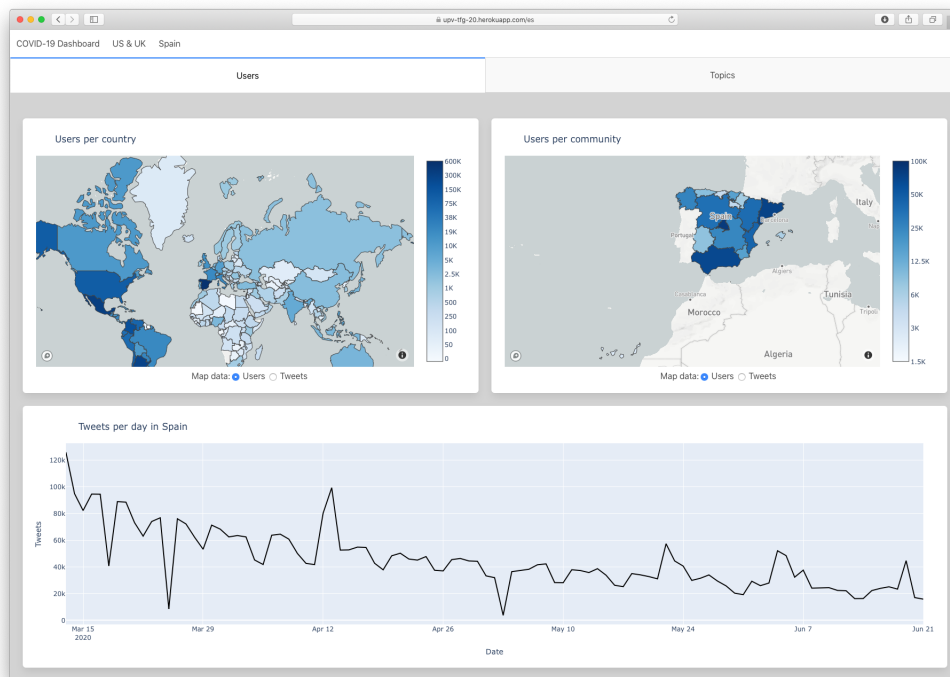


Figure 2.4: Spain users' data

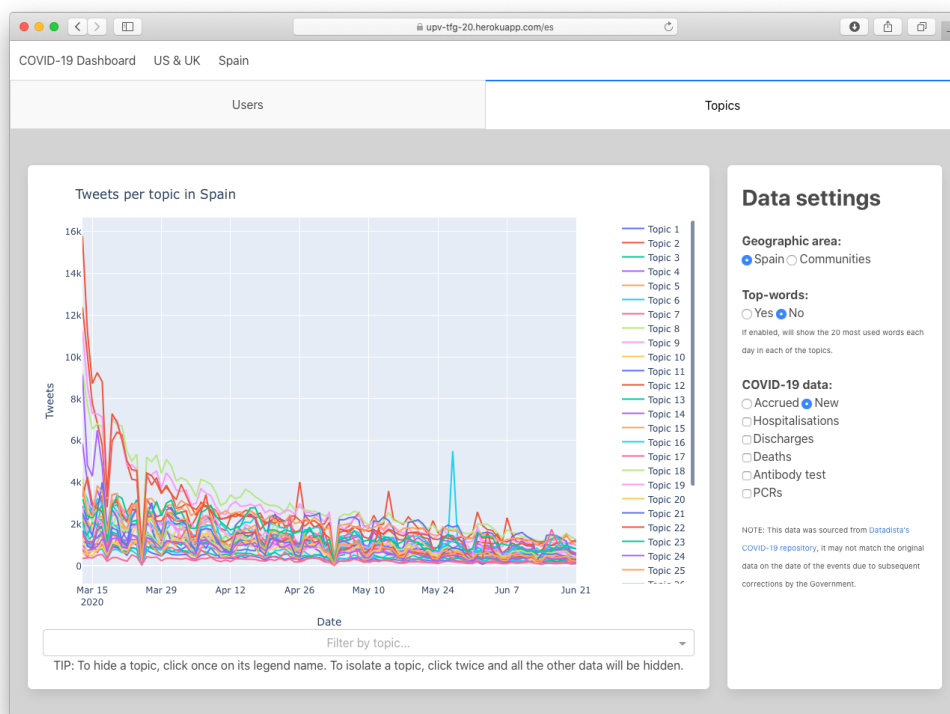


Figure 2.5: Spain topics' data

CHAPTER 3

Results

3.1 English dataset

The study on the UK and US activity on Twitter during the period comprised between the 2nd of December 2019 and the 18th of May 2020 is an overall view of the behaviour from the users in both countries; it may give a general idea of the interests, but it is clearly not an accurate representation of the concerns and the activity of the users during such period.

At first view, the dataset contains tweets from a broad range of countries, as can be seen in the choropleth map below.

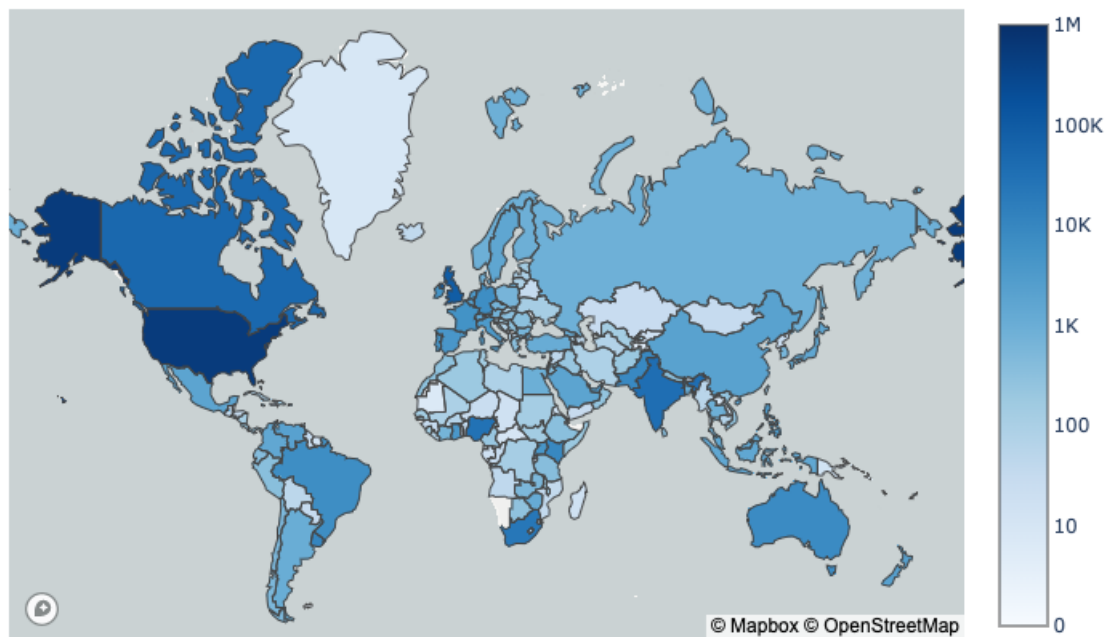


Figure 3.1: Choropleth map representing the amount of users per country in the English dataset

As per the study methodology, only tweets from the US or the UK are considered. A total of 548,520 users in the dataset are located in the US, whilst 92,477 users are located in the UK. Their tweeting activity can be seen in the timeline chart below, where it is clearly seen that the interest in COVID-19 starts to appear on Twitter in late January 2020. The peak activity starts growing in the first week of April, and something that draws attention is the fact that despite the public interest in COVID-19, in both countries the tweeting activity decreases on the weekends, as it usually happens on Twitter.

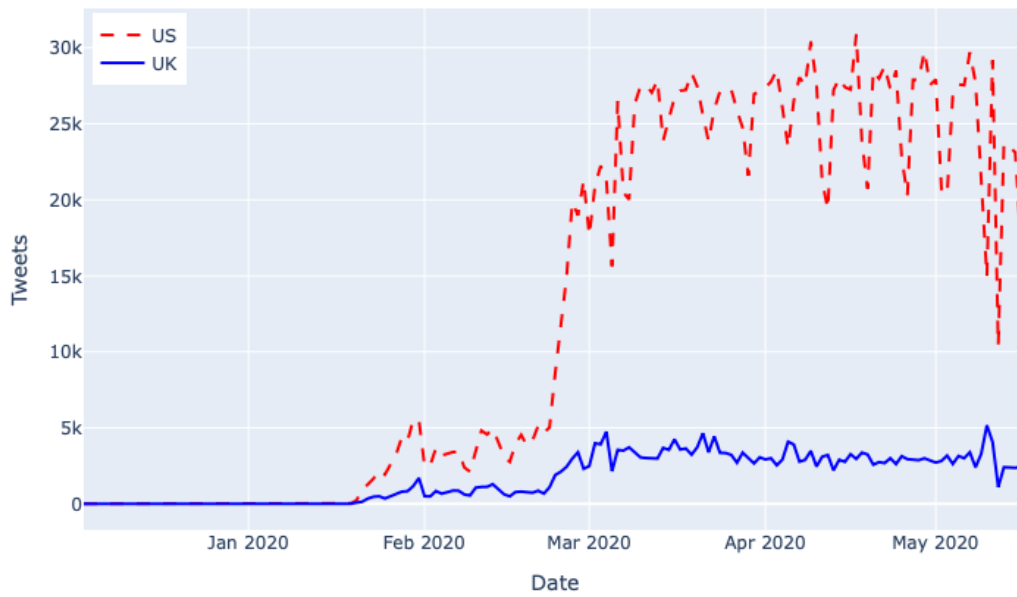


Figure 3.2: Tweets per day in the English dataset

The polarity of the tweets is mostly neutral, followed by positive tweets and negative tweets. By putting this metric together with the topics, we will be able to understand public opinion further.

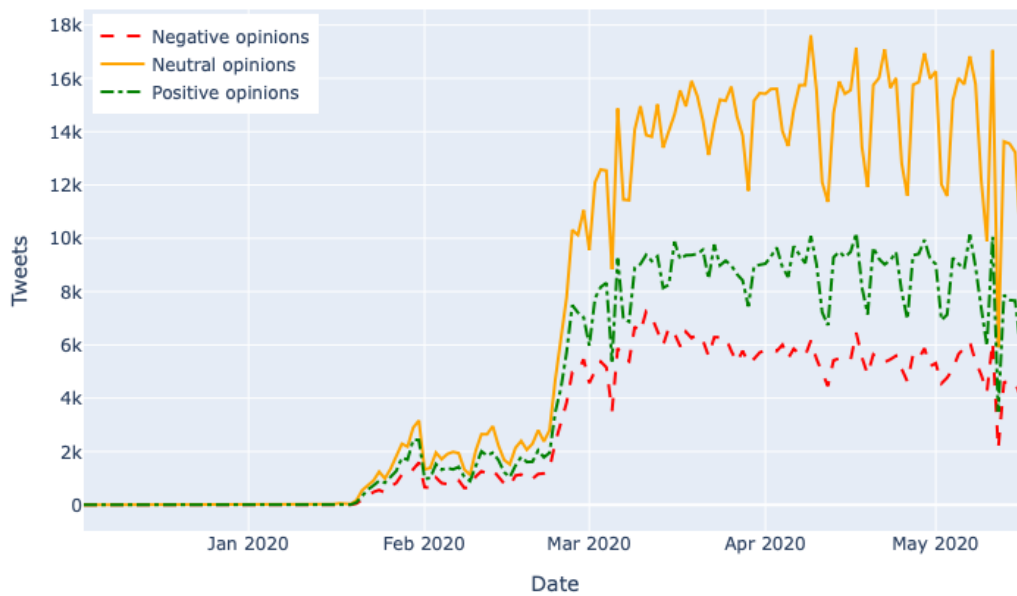


Figure 3.3: Polarity of the tweets per day.

The sixth topic model lays the following topics:

#	10 most used words
1	day love thank family share watch friends great video go
2	patients vaccine use drug disease study treatment treat symptoms may
3	county deaths report confirm total positive number death first department
4	gov governor order watch brief press response conference emergency house
5	school due cancel close students 2020 march concern events postpone
6	join discuss dr. webinar register impact talk question watch tomorrow
7	president government response would lie call china think right americans
8	positive patients hospital care die nurse symptoms release patient center
9	stay social distance safe keep care important protect please learn
10	information please resources check latest read use share visit free
11	mask hand wear wash face use buy paper water clean
12	go think would covid really come want back right good
13	close open order park store due city service restaurants food
14	china outbreak travel ship quarantine wuhan cruise italy flight report
15	support fund donate response community relief food crisis thank impact
16	workers thank care fight nurse medical healthcare line staff front
17	market impact global crisis economic world change economy company business
18	pay bill relief workers fund due businesses tax small care
19	number days weeks data could may still curve show would
20	deaths death flu rate die number million report infect toll

Table 3.1: List of topics obtained after applying LDA in the English dataset

It seems that for most of the topics, the polarity follows the same trend as in the previous chart, most of the tweets are neutral, followed by positive and then negative tweets, but somehow topic 7, 12 and 20 manage to disrupt the trends mentioned above by spiking on negative polarity in some days. In the positive polarity spectrum, topics 4, 14, 16 and 18 also manage to increase the positive polarity, especially at the beginning of March and by the end of April.

3.1.1. Topic 1

Topic 1 is not directly related to any specific events, but it evolves throughout the analysed time frame. If we check the existing tweets for the most active day, the 21st of March, tweets like the one below show that people react to the progression of the pandemic, some of them are related to a country's evolution, i.e. 512 tweets (15% of the total number of tweets that day) are associated with Italy, but 434 (84.75%) tweets are identical, sharing support messages for Italy with the hashtag #WeStandWithItaly. It would be sensible to consider this topic is similar to an "Other" category because it has the same activity progression as the whole dataset, and there are no specific reactions to events.

We stand by Italy during these trying times. Share your Support for our Italian friends, They are our colleagues, friends and family. Cari amici, siamo con voi. #COVID19 #WeStandWithItaly

— Cartney gmcc) [March 21, 2020](#)

Birthday celebrations #Covid_19 style - isolation doesn't have to be isolated. pic.twitter.com/iFCLEMr3i2

— Emily Khan (@EmilyKhan11) [March 21, 2020](#)

3.1.2. Topic 2

Regarding topic 2, the first significant upward trend was on the 27th of February. During those days GILEAD announced it would start two late-stage studies to test remdesivir, a drug for coronavirus. They also announced the expansion of the drug trials to other countries in Asia.

Gilead to Expand Coronavirus Drug Trials to Other Countries

The drug maker said that in March it would begin two more clinical trials of remdesivir, an experimental antiviral drug, in more nations outside China.
The NY Times - Denise Grady

Gilead starts two late-stage studies to test drug for coronavirus

Gilead Sciences Inc (GILD.O) said on Wednesday it has started two late-stage studies to test its drug in patients with severe and moderate cases of the illness caused by the coronavirus, sending its shares up 4% in extended trading.
REUTERS - Stephen Lam

On the 29th of April, there's a global peak on the trend, coinciding with the news release from the NIH stating that Remdesivir accelerated recoveries from advanced COVID-19 ¹. The WHO also released a statement a few days before, launching an initiative to share drugs, tests and vaccines ²

3.1.3. Topic 3

In topic three it's clear that people react to local events on social media: on Saturday 29th of February, the first death from COVID-19 in the US was confirmed in the state of Washington ³. The global peak is on the 28th of March, and on that day, there are 700 out of 1580 tweets (44.3%) that mention the word "death". There were 104686 cases and 1707 deaths, and it was the first death of an infant ⁴.

3.1.4. Topic 4

The fourth topic is closely related to the government press briefings, so it "reacts" to the press briefings during the analysed period. The White House released COVID-related statements starting from late February ⁵ and the surges on the chart match with the release date of such information. If tweet samples are extracted, it can be seen that the White House may not always be the source of information referenced, but other local authorities:

WATCH LIVE: [@GovHawaii](#) and Hawai'i Department of Health officials discuss the [#coronavirus](#) outbreak <https://t.co/3KPkjAAgOF> [#hinews](#) [#hawaii](#)
pic.twitter.com/EH3XYZFoy0

— Star-Advertiser (@StarAdvertiser) [February 27, 2020](#)

¹<https://www.nih.gov/news-events/news-releases/nih-clinical-trial-shows-remdesivir-accelerates-recovery-advanced-covid-19>

²<https://www.who.int/news-room/detail/24-04-2020-global-leaders-unite-to-ensure-everyone-everywhere-can-access-new-vaccines-tests-and-treatments-for-covid-19>

³<https://www.governor.wa.gov/news-media/inslee-statement-death-washington-state-covid-19>

⁴<http://dph.illinois.gov/news/public-health-officials-announce-first-death-infant-coronavirus-disease>

⁵[https://www.whitehouse.gov/briefings-statements/page/\\$29\\$/?issue_filter = healthcare](https://www.whitehouse.gov/briefings-statements/page/29/?issue_filter = healthcare)

.@GovLarryHogan holds a press conference to provide an update on Maryland's preparedness for COVID-19. <https://t.co/ZiEHkTpsSL>

— Maryland Department of Health (@MDHealthDept) February 27, 2020

3.1.5. Topic 5

The charts show a rise of interest in the topics when relevant news are released; for example, topic 5's interest grows abruptly on the 12th of March 2020. This sudden surge of interest can be understood from the news published on that day:

Millions of U.S. students now affected by school closings related to the coronavirus outbreak

The number of K-12 students in the United States whose schools have closed or are scheduled to shutter because of the spread of the novel coronavirus has surpassed 3.5 million — and more are expected. ...

Washington Post - Valerie Strauss

NBA suspends season until further notice after player tests positive for the coronavirus

The NBA is suspending game play following the conclusion of [Wednesday's] schedule of games until further notice," the league said in a statement issued shortly after 9:30 p.m. ET. "The NBA will use this hiatus to determine next steps for moving forward in regard to the coronavirus pandemic.

ESPN News

Disney World, Disneyland Paris Resort to close over coronavirus concerns

Due to concerns over the coronavirus outbreak, Walt Disney World Resort in Florida and Disneyland Paris Resort will close at the end of the business day Sunday through the end of the month, the company said in a statement tonight.

CNN - Frank Pallotta

Schools and theme parks were being closed, events were being postponed or suspended, and it seems that the 12th of March was the day that the public was most concerned about school closures and event cancellations, with over 5000 tweets during that day related to topic 5.

3.1.6. Topic 6

When the pandemic starts, webinars to share information related to COVID-19 become popular, as it is an easy and free way to share information fast and make it available to everyone, this is reflected in topic 6. This tool was being used even by the CDC ⁶ (Centers for Disease Control and Prevention) to communicate technical knowledge among the members of the medical community. The primary use of these webinars and online meetings was targeted to citizens, as local communities, experts and members of the academia organised such events to share information among the population. As it can be seen in figure 3.2, the tweeting activities drops in the weekends and resumes on Mondays.

⁶https://emergency.cdc.gov/coca/calls/2020/callinfo_032720.asp

Questions or concerns about #Coronavirus? UK HealthCare and @universityofky experts are weighing in live on YouTube. Watch: <https://t.co/ASkgPM4z9T>

— UK HealthCare (@UK_HealthCare) February 27, 2020

Join us TOMORROW (April 29) @2pm ET for a #COVID19 briefing w/ @CDC, @HopkinsMedicine & @UCDavis. Learn why people w/ chronic conditions are at higher risk for COVID-19 complications. Excited to co-host this webinar w/ @GHLForg & @ChronicCarePA. Register: <https://t.co/rkFyzVnqx6pic.twitter.com/Cauil2r4fg>

— Allergy & Asthma Network (@AllergyAsthmaHQ) April 28, 2020

Excited to have one more community call for the day, join me at 6pm to talk about #COVID19. Thank you to the coalition member's of @bciitybaltimore @noboundaries21217 @architectsofjustice @learnwithryan for inviting me to talk. #coronavirusbalt #keepyourdistance #stayhomeheros [pic.twitter.com/QIUhFrbtLb](https://t.co/QIUhFrbtLb)

— Letitia Dzirasa, MD (@DrTishCommish) April 29, 2020

Lee Kelso, host of WOWO's Health Call Live Saturdays from 9-10 a.m., joins to discuss changes at the Matthew 25 Clinic due to COVID-19 and dental hygienists are the career most at risk for COVID-19 <https://t.co/mUQQR7QQbj>

— Pat Miller (@WOWOpatmiller) May 7, 2020

3.1.7. Topic 7

In topic 7, the most active day is the 29th of February with a trend disruption on the polarity of the tweets, being negative tweets more frequent than positive ones. That same day, President Trump said at a rally "The Democrats are politicising the coronavirus". That same day, a Presidential Document by the Executive Office⁷ is issued suspending the entry of other countries than those already stated in the Proclamation issued on the 31st of January.

Coronavirus: Trump calls criticism Democrat 'hoax'

US President Donald Trump has called criticism from Democrats over his reaction to the global coronavirus outbreak as a "hoax".
BBC News

3.1.8. Topic 8

The amount of positives each day increases as time passes and on the 17th of March the positive cases detected in one day climb to 4,450, that makes topic 8 reach its tipping point, as government officials react and implement measures that encourage citizens to stay at home.

⁷<https://www.federalregister.gov/documents/2020/03/04/2020-04595/suspension-of-entry-as-immigrants-and-nonimmigrants-of-certain-additional-persons-who-pose-a-risk-of>

U.S. Health Departments Report Largest Number Of Deaths In One Day As Total Confirmed Cases Climb To 4,450

Experts say there's likely far more unreported cases because of failures to properly test potential patients. On Monday, 18 people died from the virus bringing the nation's total deaths to 85 so far.

KHN

These States Have Some Of The Most Drastic Restrictions To Combat The Spread Of Coronavirus

US coronavirus cases have now surpassed 4,475 and some officials say the US could end up with a count as high as hard-hit Italy if it doesn't make dramatic changes to the way Americans live, work and play. Governors and city leaders have begun implementing measures to encourage residents to stay home or skip nights out, in an effort to curb the spread of the virus.

CNN - Christina Maxouris

3.1.9. Topic 9

On the 16th of March 2020, the CDC ^{8 9} recommends not to assist to gatherings of more than fifty people in any events during the eight weeks after that day. During that day, in topic 9 words like safe (17.70%), stay, social (16.02%), keep (9.16%), and care (8.67%) are often used in the tweets. During the previous days to the CDC statement, the US banned all travel from EU countries to the US ¹⁰.

3.1.10. Topic 10

The tenth topic is similar to the sixth in the sense that it is also intended for sharing information regarding the pandemic. They both have a similar progression, and in topic ten the word information (17.36%) is often used together with words like "read" (10.63%) and "resources" (8.18%). This topic does not have a specific catalyst, as it evolves with the progress of the pandemic and it follows the same trends in the weekends as in figure 3.2.

3.1.11. Topic 11

At the beginning of the pandemic, there was a psychological reaction from the population to stockpile toilet paper, on the 6th of March, many supermarkets across Australia had their shelves in the toilet paper section, empty. This compulsive reaction from the citizens echoed on social media, as it can be seen on chart 3.7k of topic 11.

⁸Unfortunately the CDC page with said recommendations is no longer the original page since it has been updated with more recent information throughout the pandemic. Although, many newspapers state-wide and locally echoed these recommendations on their daily issues

⁹<https://www.cdc.gov/coronavirus/2019-ncov/community/large-events/considerations-for-events-gatherings.html>

¹⁰<https://www.whitehouse.gov/presidential-actions/proclamation-suspension-entry-immigrants-nonimmigrants-certain-additional-persons-pose-risk-transmitting-2019-novel-coronavirus/>

'It's crazy': Panic buying forces stores to limit purchases of toilet paper and masks

... Some retailers are having to take drastic action to limit the number of toilet paper rolls, face masks and hand sanitizer bottles each person can buy as customers stockpile goods over fears of the novel coronavirus outbreak. ... Australian supermarket chains Woolworths and Coles both began limiting toilet paper purchases to four packs per person this week. Costco Australia is also restricting how much toilet tissue, disinfectant, milk, eggs and rice each customer can buy.
CNN - Michelle Toh

3.1.12. Topic 12

On the 29th of February, coinciding with the surge of activity in figure ?? and ??, Trump dismissed the concerns about COVID-19 by qualifying it as a hoax during one of his rallies. The following day, there was the first death from COVID-19 in the US.

During the other days in the period, users share general thoughts of the current situation and comment on the progress of the pandemic and the measures applied.

"I just want people to be able to take a breath..." @saintlukeskc Dr. Marc Larsen talks symptoms with Coronavirus. #COVID19 "If you wouldn't have gone to the hospital six months ago with these symptoms, you probably don't need to go to the hospital today." @kmbc pic.twitter.com/ui6zkTBikj

— Matt Flener KMBC (@MattFlenerKMBC) March 12, 2020

Netflix and chill will go back to its original meaning during all this #COVID19

— Popeye the Taylor (@Taylor_Errthang) March 12, 2020

(inf) Note: In figures 3.4b and 3.4c, the word "coronavirus" is part of a hashtag, the full hashtag is "coronaviruspandemic".

Word	Count	Word	Count	Word	Count
coronavirus	2572	coronavirus	1816	covid	2008
covid	1025	covid	1762	19	1945
19	950	19	1726	covid_19	1364
covid19	876	covid_19	1401	coronavirus	1263
peopl	572	covid19	957	covid19	1082
go	563	go	798	peopl	801
get	543	get	781	get	722
trump	503	peopl	654	go	703
hoax	377	like	517	like	484
like	373	coronavirus ^{inf}	516	time	454
u	369	test	451	u	419
think	349	time	410	need	417
viru	326	one	381	test	398
know	310	think	376	think	390
say	277	u	370	work	385
one	269	know	358	coronavirus ^{inf}	382
realdonaldtrump	255	thing	355	one	381
would	241	need	345	day	368
coronaviru	236	work	341	know	364
die	218	would	325	home	353

(a) 29th of February

(b) 12th of March

(c) 21st of March

Figure 3.4: Top 20 words in topic 12

3.1.13. Topic 13

Because of the recommendations from the CDC regarding gatherings and the evolution of the pandemic, some states decided to close bars and restaurants¹¹ and that is reflected on topic 13 in the sudden surge of tweets on that day.

3.1.14. Topic 14

On the 26th of February, in topic 14, the tweeting activity peak is on that day, words like outbreak and travel are present in 30.4% of the tweets. It was the first day that there were more new cases outside of China than from China as stated by the WHO¹². On that same day, Italy reported a surge of 40% new more cases, from 229¹³ to 322.

3.1.15. Topic 15

Topic 15 peaks on the 5th of May, with over a thousand tweets mentioning the word "support" and almost nine-hundred tweets referring to donations, a total of 1,660 tweets out of 2,851 use either of these words. PEW Research published a study¹⁴ regarding the

¹¹<https://www.scrila.org/news/493793/National-Restaurant-Association-Coronavirus-Update—March-16-2020.htm>

¹²<https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200226-sitrep-37-covid-19.pdf?sfvrsn=2146841e2>

¹³<https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200225-sitrep-36-covid-19.pdf>

¹⁴<https://www.pewresearch.org/fact-tank/2020/05/05/financial-and-health-impacts-of-covid-19-vary-widely-by-race-and-ethnicity/>

financial and health impacts of COVID-19 and how they impact differently depending on race and ethnicity. This study, would help to understand the evolution of this topic, as it explains that many black and Hispanic have had wage losses or do not have financial reserves to cover expenses in case of an emergency.

3.1.16. Topic 16

On the 6th of April, Boris Johnson was moved to intensive care¹⁵ and in topic 16, around 40% of the tweets on that day mentioned the Prime Minister.

3.1.17. Topic 17

Since the 5th of February of 2018, the Dow had not had a single-day decline as large as the one on the 28th of February¹⁶, 1,190.95 points. But April was the best month since 1987 for the S&P 500, with a 12.7% change. These two highlights are reflected in topic 17's chart, as can be seen in figure 3.7q.

3.1.18. Topic 18

On the 25th of March, there is a peak in topic 18. The US Senate passed the CAREs Act¹⁷, a \$2.2 trillion economic stimulus aimed at helping Americans with one-time payments, increased unemployment benefits, forgivable loans for small businesses, large companies and state and local governments.

3.1.19. Topic 19

In this topic, people share the progress of the pandemic and COVID-related information regarding the number of cases and positives. Figures 3.5a, 3.5b and 3.5c show words like "test", "case", "people" or "number".

1/3. The exponential increase in COVID-19 cases in the United States over the last few days makes the next two weeks an important opportunity for Americans to slow the spread of disease. pic.twitter.com/vyYyoNdvhm

— Roman (@Roman) [March 18, 2020](#)

Last week, Georgia became the first state to partially re-open. Guess what? Positive COVID-19 tests are skyrocketing.

They're going to call this a "second wave" in order to remove human agency. Remember: this didn't have to happen. <https://t.co/il3IRK5gUF>

— Michael Socolow (@MichaelSocolow) [May 1, 2020](#)

¹⁵<https://www.bbc.com/news/uk-52192604>

¹⁶<https://www.nasdaq.com/articles/stock-market-news-for-feb-28-2020-2020-02-28>

¹⁷https://en.wikipedia.org/wiki/CARES_Act

Word	Count	Word	Count	Word	Count
covid19	406	covid19	436	covid	633
covid	387	covid	418	19	604
19	374	19	404	covid19	485
coronavirus	315	coronavirus	290	test	256
test	245	test	266	coronavirus	253
case	143	case	215	case	168
peopl	121	day	131	say	165
day	115	peopl	116	day	152
u	113	u	107	week	138
need	85	say	106	peopl	131
like	84	number	103	new	118
week	82	week	96	data	110
spread	80	new	94	lockdown	109
say	77	need	89	number	107
get	76	could	85	state	102
take	70	state	84	death	101
number	67	spread	81	u	100
new	64	get	81	peak	95
time	60	like	80	work	89
coronaviru	59	go	78	still	83

(a) 18th of March

(b) 25th of March

(c) 15th of April

Figure 3.5: Top 20 words in topic 19

3.1.20. Topic 20

From the beginning of the pandemic, many people wondered why COVID-19 was declared a pandemic if influenza usually kills up to 61,000 people in the US alone.¹⁸ As the pandemic evolves, people start to be more concerned about the death tolls and positive cases. Words like "flu", "rate", "infect", "die" or "million" are more frequent in figure 3.6a, whereas in figure 3.6b, "number", "china", "world", "confirm" or "surpass" are among the most used.

#2 We as Americans should worry more about influenza than #coronavirus.

Also: It's not too late to get a flu shot!

Last year's flu season there were 16.5M hospital visits, half a million hospitalizations and ~34,000 deaths.

That was a typical year. This year could be worse.

— Jonathan Frederick (@jonmikered) February 6, 2020

Global COVID-19 Update from Johns Hopkins:

1,002,159 cases
188,420 recovered
51,484 deaths

¹⁸<https://www.cdc.gov/flu/about/burden/index.html>

Since I started composing this tweet, the numbers have gone up.

God help us.

— #FreeRealityWinner (@NickLaparra) April 2, 2020

#UK will have Europe's worst #coronavirus death toll, study predicts, accounting for more than 40% of total deaths across Europe

Germany is testing about 6,000 of every million, three times the rate in the U.K

"Our key message is test, test, test" - @WHO <https://t.co/9hqLA0rhWc> pic.twitter.com/INo8VQyB1u

— Hastings & Rye Labour Party #blackouttuesday (@HastingsRyeLab) April 7, 2020

Word	Count
covid	342
19	340
coronavirus	249
case	204
peopl	204
death	196
flu	187
covid19	171
rate	147
000	127
infect	111
1	106
2	100
u	95
number	94
million	88
die	87
test	79
day	76
mortal	76

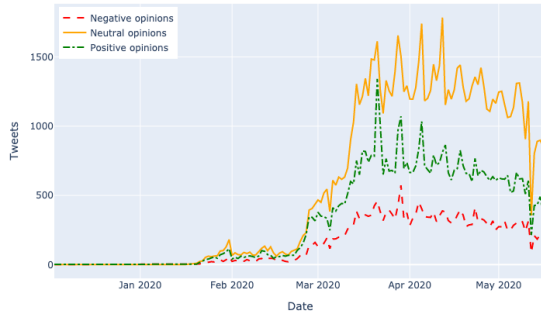
(a) 14th of March

Word	Count
case	1018
covid	738
19	705
u	553
coronavirus	530
number	470
covid19	469
china	455
death	387
world	322
000	299
1	297
confirm	288
peopl	246
surpass	232
itali	231
report	214
countri	191
rate	190
usa	188

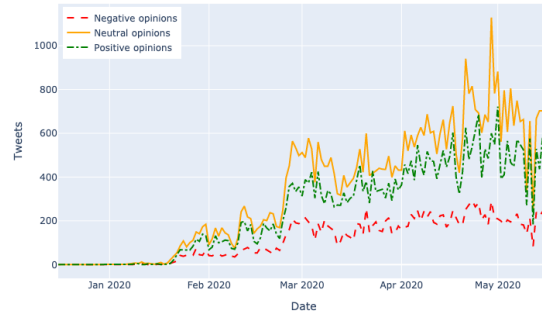
(b) 26th of March

Figure 3.6: Top 20 words in topic 20

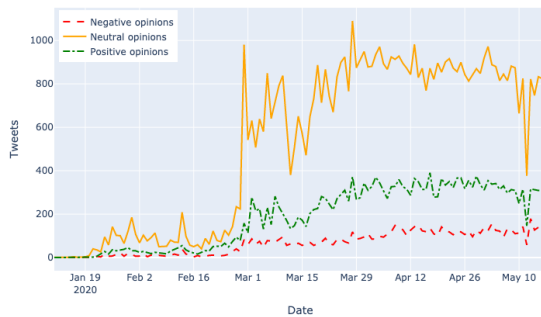
Figure 3.7: Timeline charts of the polarity of the tweets in each topic.



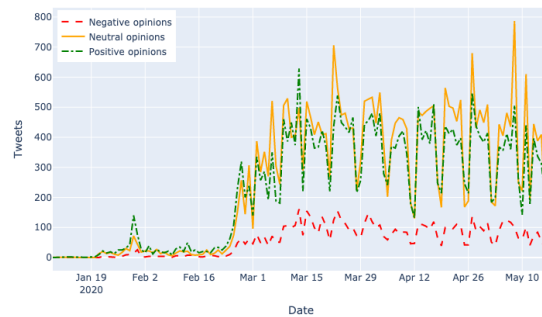
(a) Topic 1



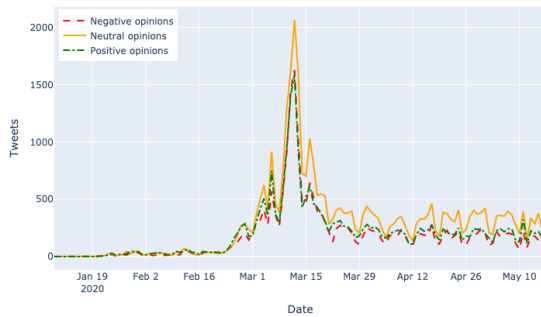
(b) Topic 2



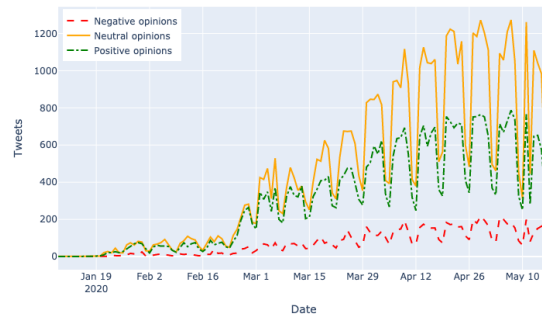
(c) Topic 3



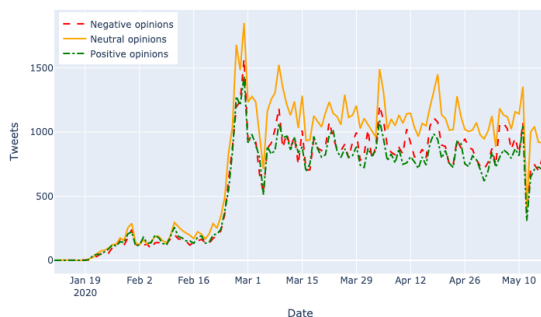
(d) Topic 4



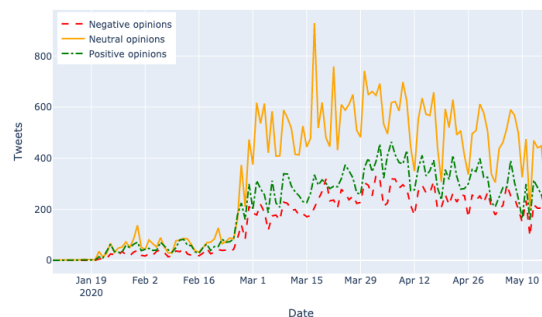
(e) Topic 5



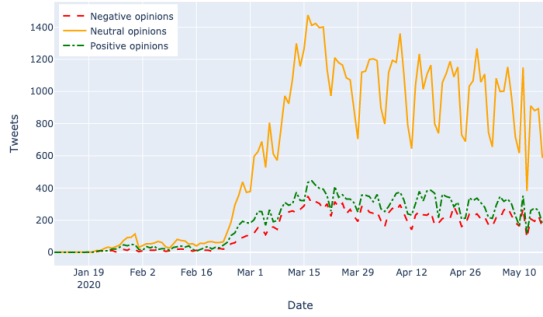
(f) Topic 6



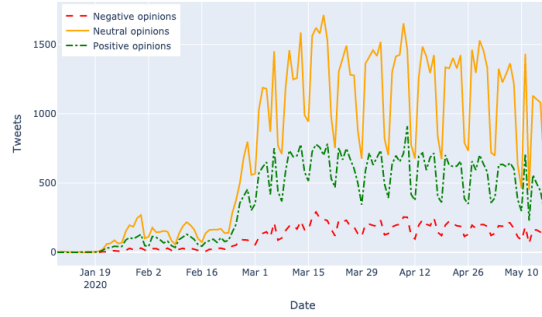
(g) Topic 7



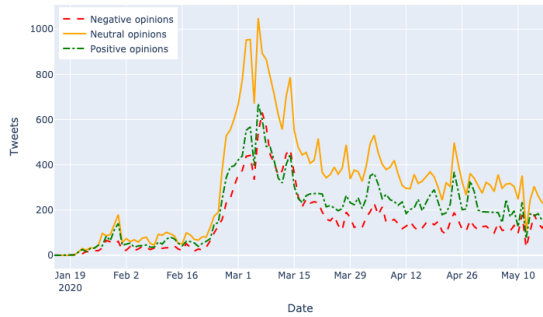
(h) Topic 8



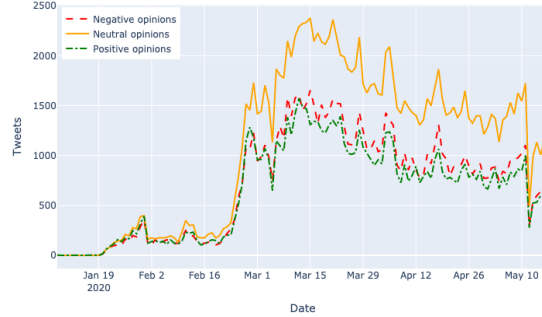
(i) Topic 9



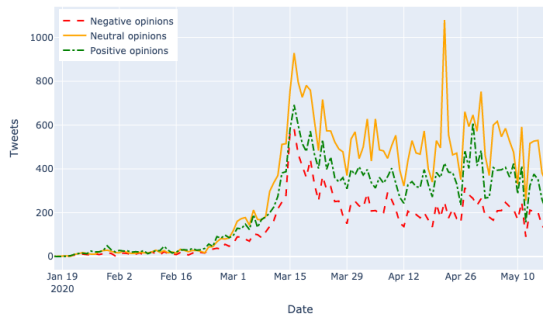
(j) Topic 10



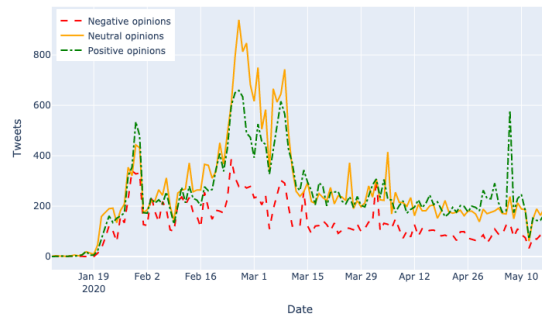
(k) Topic 11



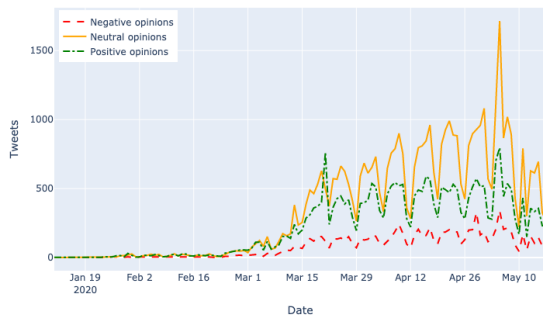
(l) Topic 12



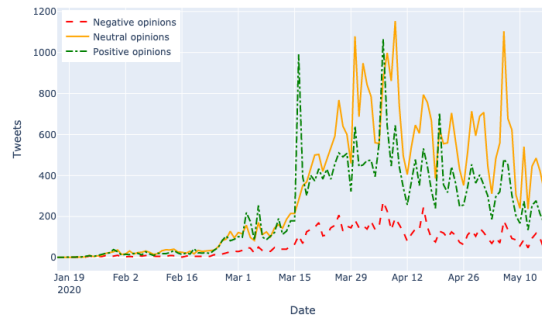
(m) Topic 13



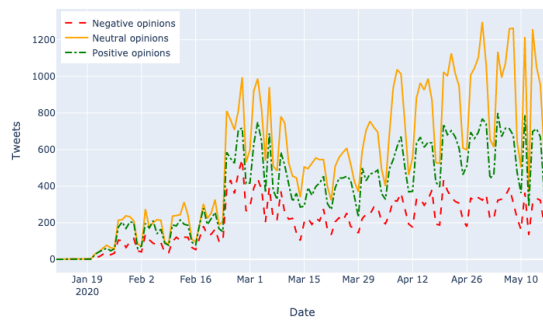
(n) Topic 14



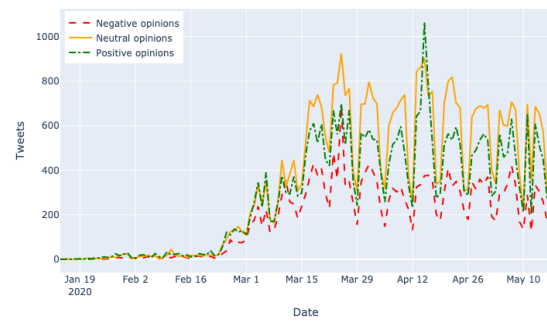
(o) Topic 15



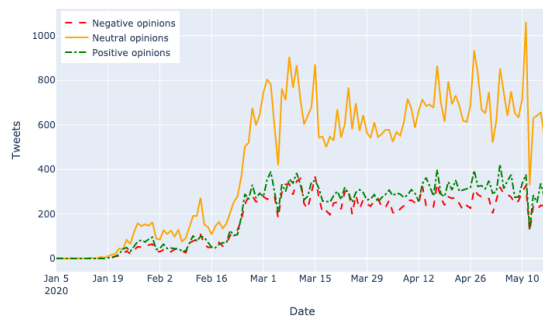
(p) Topic 16



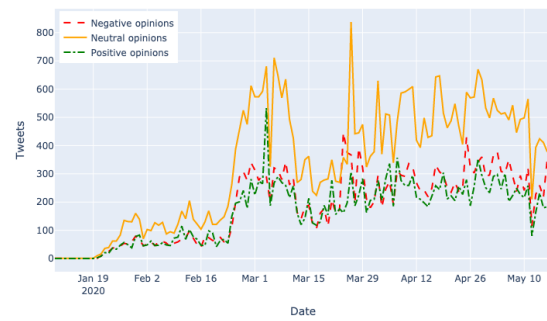
(q) Topic 17



(r) Topic 18



(s) Topic 19



(t) Topic 20

3.2 Spanish dataset

In the Spanish dataset, there are an approximate total of 27 million tweets and 3.9 million users could be fetched, of which 4.3 million and 543,724 respectively, are geolocated in Spain. The geographic distribution of geolocated users is shown in the choropleth map in figure 3.8.

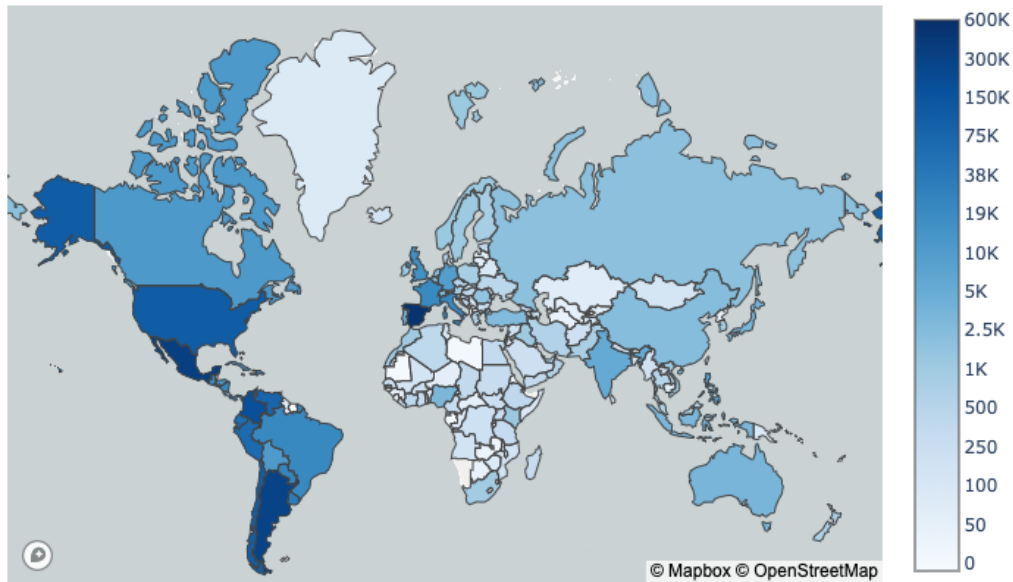


Figure 3.8: Choropleth map representing the amount of users per country in the Spanish dataset

For the users located in Spain, a precise location was obtained due to the usage of OSM geocoding, that results in the following user distribution among the Spanish communities:

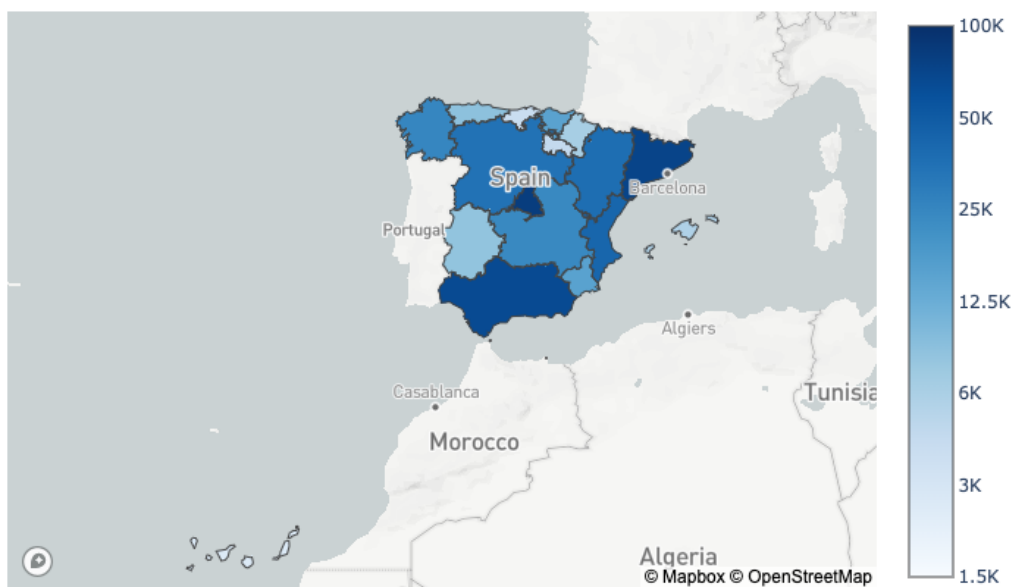


Figure 3.9: Choropleth map representing the amount of users per community

Something that stands out from the data obtained is that even though Madrid and Catalunya have a very similar amount of users in the dataset, the number of tweets is almost halved in the latter. In the case of Andalucía, something similar occurs, although the user base difference is more significant. The number of users from Madrid, Catalunya and Andalucía represented a 1.25%, 0.97% and 0.81% of the population of each community, respectively.

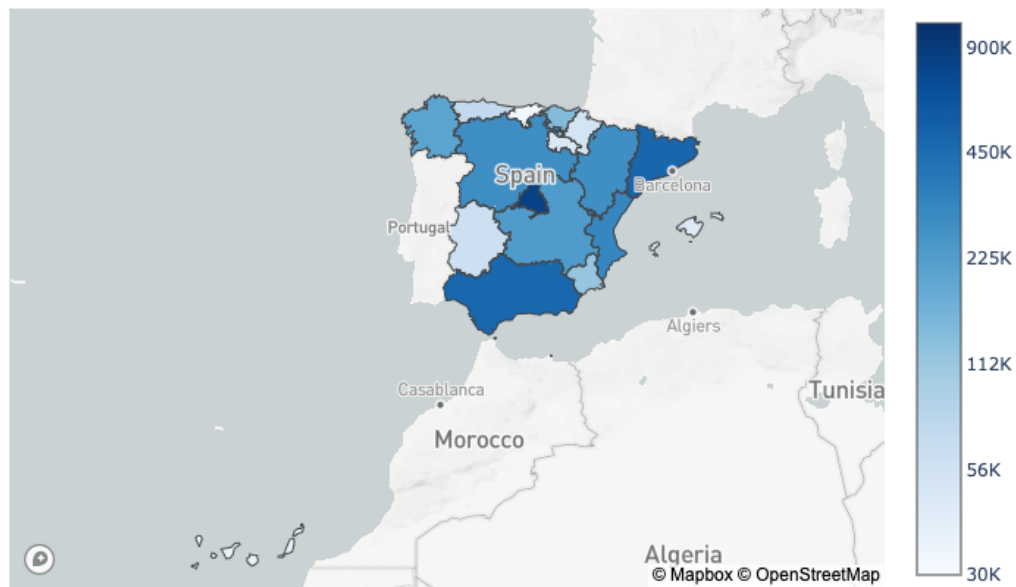


Figure 3.10: Choropleth map representing the amount of tweets per community

#	10 most used words
1	doce pingo figueira buarcos continente lidl crisi persones hora confinament
2	hace pues bien dice decir verdad parece culpa menos nadie
3	pacientes hospital atención sanitarios hospitales personal médicos profesionales sanidad sanitario
4	bares productos cierre confinamiento transporte restaurantes seguridad tiendas mascarillas abrir
5	mascarillas material millones lucha euros sanitario protección ayudar dona luchar
6	víctimas luto minuto silencio homenaje oficial todas santa memoria familias
7	galicia más crise polo xunta unha pola niño sanitaria vigo
8	bien políticos dinero hace peor mejor menos siempre vida pueblo
9	gracias tiempo cuarentena bien mejor vida confinamiento cosas vamos menos
10	2020 abril debido mayo 2021 nueva nuevo cuarentena marzo fecha
11	sánchez alarma directo pedro prensa hora última simón fernando gestión
12	cuarentena hace madre bien quiero bueno mata tener pues miedo
13	fútbol jugadores liga positivo club equipo temporada test partido confinamiento
14	prevención seguridad sanitaria evitar propagación desinfección alarma emergencia actividades trabajo
15	millones mientras pública impacto marzo petición dure economía 2020 caída
16	plan sector millones económica social ayudas efectos impacto paliar euros
17	alarma guardia policía civil información encontrar útil utilidad podréis posible
18	josé muere años juan maría entrevista hospital luis presidente médico
19	positivo años hospital caso primer paciente dado prueba mujer negativo
20	test pruebas detectar rápidos detección tests estudio sanidad diagnóstico google
21	tiempos cómo online webinar aquí impacto post sector mayo empresas
22	cómo social cambio artículo economía tiempos mundial sistema después futuro
23	pacientes síntomas riesgo enfermedad estudio infección enfermedades niños pueden podría
24	confirmados positivos total costa rica recuperado argentina contagios diario méxico
25	horas últimas muertes contagios número positivos registra total última cifra
26	china países italia nueva trump EEUU europa unido alemania unidos
27	vacuna china tratamiento laboratorio pacientes vacunas científicos trump fármaco investigación
28	evitar manos contagio mascarillas mascarilla distancia recomendaciones prevenir cómo social
29	gestión sánchez pedro iglesias pablo PSOE ayuso derecha sanidad podemos
30	residencias mayores ancianos castilla león fase comunidad residencia mancha cada
31	autónomos trabajadores empresas ayudas erte afectados pago alarma social actividad
32	gracias sanitarios lucha trabajo profesionales personal primera juntos labor apoyo
33	muertes cifras test gripe menos número hace países hecho infectados
34	niños confinamiento mujeres mayores género violencia salir cómo cuarentena tiempos
35	información aquí puedes dudas preguntas redes toda sociales noticias bulos

Table 3.2: List of topics obtained after applying LDA in the Spanish dataset

3.2.1. Topic 1

Due to the linguistic diversity in Spain, even though the tweet search was done, so only tweets in Spanish were included, some tweets in Galician or Catalan made it to the dataset.

In this topic, the first surge of activity shown in figure 3.45c is an equivalent to topic 14, in the sense that includes tweets referring to the new measures needed to limit the spread of the virus. There is also a reference to Madrid and Madrileños, as shown in figure 3.11a. Because people from Madrid foresaw these restrictions due to the progression of the pandemic, many of them decided to move to their second homes¹⁹; stirrings among the affected areas where Madrileños where moving followed immediately on Twitter.

Es una vergonya, estem a una emergencia nacional i aquí com si no passes res, Madrid encara no está confinada i va propagant el #COVID19 donde está el Gobierno? TC, TS, Audiencia, Fiscalía, Ejército, pikoletos? Si no es per anar contra els indepes no és mouen? #paisdepandereta

— Ignasi Recolons (@onlyamd) March 13, 2020

Catalan users of the social network, also echoed the news of the full attendance to a party in a club in Madrid.

#MarcaEspaña: Ple absolut a la discoteca de Pachá de Madrid que no tanca pel coronavirus <https://t.co/g41m1RlJKE>

— Toni C. (@tonicesteve) March 13, 2020

Later on, as the pandemic progresses, on the 8th of April there is a second surge of activity and "terraplanistas" (terraplanners), "verdad" (truth) and "bulos" (hoaxes) become quite frequent words in the tweets. On that day, the Spanish newspaper elDiario.es published data from researchers that showed an existing hoax network on Twitter mainly formed by right-wing personalities.

Los terraplanistas del coronavirus: la red para contarte "la verdad" está impulsada por conocidos difusores de bulos

eldiario.es, con la ayuda de investigadores y expertos independientes en difusión de mensajes, ha investigado los perfiles que están difundiendo estas ideas para contagiar las redes españolas con este tipo de desinformación.

elDiario.es - Carlos del Castillo

¹⁹<https://www.efe.com/efe/espana/portada/que-vienen-los-madrilenos/10010-4194646>

Word	Count
coronavirus	433
mesures	106
casa	76
madrid	75
hora	65
evitar	64
covid-19	62
catalunya	59
prevenció	51
propagació	49
directe	43
tanca	42
totes	40
confinament	40
abans	38
persones	38
informació	37
primer	36
feina	36
tots	35

(a) 13th of March

Word	Count
coronavirus	466
verdad	242
difusores	242
bulos	242
terraplanistas	241
contarte	241
impulsada	241
conocidos	240
análisis	150
covid-19	109
sánchez	62
fallecidos	56
fundas	55
requisado	54
envuelve	54
confinament	49
vídeo	43
persones	38
temps	35
hora	34

(b) 8th of April

Figure 3.11: Top 20 words in topic 1

3.2.2. Topic 2

More than 60% of the new deaths in Spain occurred in Madrid on the 13th of March, and this had been happening from the beginning of the week. The majority of the deaths in Spain accumulated in Madrid. These fatal rates lead to a progressive application of restrictions in the community.

Because people from Madrid foresaw the restrictions due to the progression of the pandemic, many of them decided to move to their second homes²⁰; stirrings among the affected areas where Madrileños where moving followed immediately on Twitter.

Madrid una de las comunidades españolas que más infectada está del **#coronavirus**, el gobierno español anuncia **#estadodealarma** y miles de "madrileños" se toman la semana para viajar a la playa, por lo que la **#pandemia** puede ser contagiada para otras comunidades...

SINVERGÜENZAS

— RiSe AnD sHiNe (@Eneko_SNKV) **March 13, 2020**

Word	Count
coronavirus	7811
gente	1400
casa	1033
madrid	742
ahora	737
hace	651
madrileños	649
españa	592
hacer	523
gobierno	452
virus	419
covid-19	404
solo	398
pues	398
días	397
puta	396
personas	390
bien	385
verdad	379
vacaciones	364

(a) 13th of March

Word	Count
coronavirus	3151
gente	632
casa	433
gobierno	384
ahora	381
covid-19	375
españa	366
hace	342
hacer	269
cuarentena	264
muertos	262
personas	248
puta	245
covid	230
días	226
solo	222
virus	216
pues	213
bien	213
menos	205

(b) 20th of March

Figure 3.12: Top 20 words in topic 2

3.2.3. Topic 3

This topic refers to the healthcare sector, mainly focused on healthcare professionals and the patients in bed at hospitals. The surge of tweeting activity is on the 24th of April. In

²⁰<https://www.efe.com/efe/espana/portada/que-vienen-los-madrilenos/10010-4194646>

one day the number of positive cases almost doubled²¹ with 3088 new cases in just one day, a total of 7718 people were infected.

Figure 3.13, shows that "pacientes" (patients) was among the most used during those days, due to the increasing number of positive cases.

Word	Count
coronavirus	1385
pacientes	609
covid-19	570
hospital	554
hospitales	405
sanitarios	396
personal	322
madrid	303
atención	292
protección	291
sanitario	241
sanidad	227
médicos	226
material	195
crisis	192
profesionales	190
falta	184
salud	179
enfermos	173
camas	169

Figure 3.13: Top 20 words in topic 3 on the 24th of March

3.2.4. Topic 4

The fatal rates in the health crisis, lead to the closure of bars' and restaurants' terraces in Madrid and other communities like Catalunya and Valencia. It is shown in figure that "bares" (bars), "cierre" (closing) and "restaurantes" (restaurants) were among the most used on that day.

On the 11th of May there's a second surge in figure and words like "medidas" (measures), "seguridad" (security) and "bares" (bars) reflect the new reality in Spain, the first stage would allow people to go back to bars under very strict health security measures. Many communities moved to the first stage²², in such stage, people could go to bars with terraces.

Madrid clausura las terrazas de los bares y no cobrará la tasa a los hosteleros

La ciudad de Madrid ha obligado a clausurar este viernes las terrazas de los bares y restaurantes y recomienda el cierre de estos establecimientos para evitar la concentración de personas y evitar desplazamientos innecesarios.

EFE

²¹This data was gathered from Datadista's COVID-19 repository

²²https://www.lamoncloa.gob.es/covid-19/Documents/08052020_mapa_fases.pdf

Catalunya ordena cerrar bares, restaurantes y discotecas por el coronavirus

El Diario Oficial de la Generalitat (DOGC) publica la resolución del Govern que decreta el cierre de bares, cafeterías, restaurantes y discotecas, entre otros establecimientos, a partir de esta medianoche como medida de prevención del coronavirus.

La Vanguardia - EFE

La Generalitat cierra bares y restaurantes en la Comunitat Valenciana desde esta noche

Cierran bares y restaurantes de Valencia, Alicante y Castelló desde esta noche debido a la crisis de coronavirus, que hoy suma otros 21 casos en la Comunitat Valenciana y alcanza ya un total de 119 infectados por el Covid-19...

Levante EMV - Amparo Soria

Word	Count	Word	Count
coronavirus	3018	coronavirus	418
madrid	882	covid-19	199
bares	771	fase	139
cierre	671	covid	139
restaurantes	564	medidas	126
cierra	559	seguridad	101
cerrar	474	bares	86
supermercados	331	crisis	84
españa	299	comercios	84
medidas	297	españa	83
cierran	289	terrazas	79
crisis	279	tras	74
tiendas	263	madrid	72
covid-19	251	free	71
comunidad	245	vuelo	70
terrazas	232	sello	58
parques	228	pandemia	58
gente	227	abrir	56
madrileños	213	lunes	55
farmacias	201	personas	54

(a) 13th of March

(b) 11th of May

Figure 3.14: Top 20 words in topic 4

3.2.5. Topic 5

Due to the shortage of PPE (Personal Protective Equipment) by the end of March, many companies across Spain donated masks either to the Central Government or the local authorities. On the 24th of March, there is an increase in the tweeting activity as can be seen in figure 3.45e. On that day, companies like MásMóvil, Airbus, Dulcesol or Cupa Group donated masks to health centres, city councils and the Spanish Government. In figure 3.15a, words like "mascarilla" (masks), "dona" (donate) and "lucha" (fight) are among the most common words, with the latter referencing the fight against the virus.

On the 2nd of April, there is a second surge in figure 3.45e, at that time the words "acopio" (stockpile) and "desoyó" (ignored) were quite common among the tweets as per figure 3.15b, possibly being related to the ignored recommendations from the WHO to some European countries to stockpile PPEs.

MásMóvil dona un millón de mascarillas sanitarias de alta protección para la lucha contra el coronavirus

El Grupo MásMóvil ha decidido donar al Gobierno de España un millón de mascarillas sanitarias FFP2 homologadas para su uso hospitalario para contribuir así a cubrir las necesidades del sector sanitario y la población española en la "difícil situación" a la que se enfrenta la sociedad por el brote coronavirus.

Europa Press

Airbus dona 4 millones de mascarillas a España y países europeos afectados

Airbus ha donado más de cuatro millones de mascarillas, que han llegado esta madrugada a la planta del grupo europeo en la localidad madrileña de Getafe, procedentes de China, a las autoridades sanitarias de España, Alemania, Francia y Reino Unido.

COPE - Agencia EFE

Word	Count
coronavirus	1526
material	601
sanitario	544
mascarillas	449
dona	427
euros	411
lucha	395
covid-19	351
millón	342
protección	255
españa	250
crisis	244
guardiola	221
frente	212
millones	205
personal	204
luchar	200
gobierno	182
hospitales	182
ayudar	173

(a) 24th of March

Word	Count
coronavirus	1372
material	660
sanitario	586
covid-19	453
españa	448
frente	425
hacer	372
lucha	331
millones	301
acopio	300
llamadas	299
desoyó	298
euros	290
gobierno	262
mascarillas	221
crisis	211
protección	177
diputados	170
luchar	164
parte	152

(b) 2nd of April

Figure 3.15: Top 20 words in topic 5

3.2.6. Topic 6

By the 12th of April, some Spanish cities^{23,24} had declared mourned the deaths of COVID-19 and observed a minute's of silence in tribute to the victims of the virus. However, it was not until the 27th of May that the Government officially declared²⁵ national mourning for those who died during the pandemic, for ten days, which matches with the abrupt peak in chart 3.45f. On that day, words like "misa" (mass), "luto" (mourning) or "minuto" (minute) and "silencio" (silence) referring to the moment of silence for the deceased.

España comienza el luto oficial más largo de la democracia: ¿qué supone el duelo nacional?

Pedro Sánchez anunció el pasado sábado que el Consejo de Ministros aprobaría este martes 10 días de luto nacional, el más largo de los decretados durante la democracia. La oposición lo había pedido en reiteradas ocasiones y el propio Sánchez aseguró en el Congreso de los Diputados que el Gobierno lo declararía cuando la mayor parte del país entrara en la fase 1 de la desescalada.
La Sexta

Word	Count
coronavirus	284
víctimas	157
covid-19	143
fallecidos	99
luto	84
muertos	81
misa	78
oficial	58
gobierno	55
españa	53
minuto	52
homenaje	52
silencio	49
funeral	37
ayuso	35
pandemia	34
convoca	33
cumplir	32
familiares	32
respeto	32

Figure 3.16: Top 20 words in topic 6 on the 27th of may

²³<https://www.nerja.es/crespon-negro-en-la-bandera-de-espana-por-las-victimas-del-coronavirus/>

²⁴<https://lacomarcannoticias.com/portada/el-ayuntamiento-de-lijar-decreta-luto-oficial-por-el-fallecimiento-de-su-secretario-a-causa-del-covid-19/>

²⁵<https://www.boe.es/buscar/doc.php?id=BOE-A-2020-5316>

3.2.7. Topic 7

This topic is mainly related to COVID-19 news from Galicia. On the 13th of March, the Xunta de Galicia issued an administrative decision stating that the community would thereof be in Health Emergency.

The 7th of May is a highly active day on Twitter, in figure 3.45g it shows that there are 50% more tweets than on the 13th of March. This is because "Billy el Niño" the famous torturer from the Franco dictatorship, passed away on that day. This is one of the few topics, in which the most used word, "niño" (child) is not directly related to the pandemic, as shown in figure 3.17b.

Muere por Covid-19 el ex policía nacional 'Billy el Niño', acusado de torturas en el franquismo

Antonio González Pacheco, el ex inspector de policía nacional más conocido como Billy el Niño por estar acusado de torturador en franquismo, ha fallecido por Covid-19, a la edad de 73 años, en el hospital San Francisco de Asís de Madrid. Este agente fue imputado junto a otros policías del Estado por infligir torturas salvajes a decenas de opositores a Franco. Pacheco detenía e interrogaba con formas del salvaje Oeste, presuntamente, de ahí el apodo. El Mundo - Luis F. Durán

Word	Count
coronavirus	362
medidas	178
covid-19	123
polo	120
galicia	107
situación	82
xunta	71
sanitaria	60
pola	59
concello	55
máis	54
unha	48
información	48
crise	46
recomendacións	45
casa	43
prevención	39
hoxe	38
dende	36
sanidade	36

(a) 13th of March

Word	Count
niño	948
coronavirus	735
muere	651
billy	649
expolicía	496
torturador	337
acusado	323
franquista	320
covid-19	256
torturas	237
fallece	183
policía	175
aquejado	121
franquismo	79
medallas	73
juzgado	57
muerto	48
galicia	46
covid	45
través	44

(b) 7th of May

Figure 3.17: Top 20 words in topic 7

3.2.8. Topic 8

As per the commonly used words on different dates throughout the confinement shown in figures 3.18a, 3.18b and 3.18c, it seems that this topic holds the citizens' opinion on the Government's management of the health crisis. The word "gobierno" (government) is always in the top position, whereas the words "mejor" (better) and "peor" (worse) show up as lesser used.

Word	Count	Word	Count	Word	Count
coronavirus	5113	coronavirus	1201	coronavirus	577
gente	685	covid-19	551	covid	507
gobierno	587	covid	448	covid-19	309
ahora	536	gobierno	397	gobierno	220
crisis	480	crisis	340	covid19	184
medidas	470	covid19	265	gente	138
covid-19	468	ahora	204	crisis	128
españa	450	españa	203	pandemia	116
hacer	418	mundo	199	ahora	112
país	380	gente	191	personas	99
casa	350	pandemia	182	españa	97
puede	337	muertos	168	país	96
solo	320	país	162	hacer	92
virus	294	solo	154	solo	91
mundo	291	personas	154	muertos	90
personas	289	hacer	142	puede	83
bien	273	mejor	134	madrid	81
políticos	269	políticos	126	peor	76
salud	265	peor	120	salud	75
casos	260	puede	119	hambre	73

(a) 13th of March

(b) 15th of April

(c) 15th of May

Figure 3.18: Top 20 words in topic 8

3.2.9. Topic 9

This topic characterises the tweets from the citizens at first describing their change of plans due to the state of alarm, and then expressing how they will spend their time during the quarantine. The citizens use words like "casa" (house), "hacer" (do) or "salir" (leave/go out) on the 13th of March as shown in figure 3.19a, but from a broad perspective, during the confinement, words like "gracias" (thanks), "tiempo" (time), "mejor" (better), "menos" (less) or "cosas" (things) are widely used as shown in figure 3.19b.

Word	Count	Word	Count
coronavirus	6300	coronavirus	92581
casa	1407	covid	27101
días	720	covid-19	25052
ahora	629	casa	18276
hacer	620	gracias	16093
gracias	581	días	15963
cuarentena	572	ahora	12980
gente	538	hacer	11898
tiempo	485	cuarentena	11601
mejor	468	tiempo	10961
vamos	462	bien	10482
covid-19	426	mejor	10433
bien	419	vida	10354
aquí	384	confinamiento	10200
cosas	373	tiempos	10006
semana	365	covid19	9970
vida	349	gente	9910
salir	347	menos	9527
hace	339	cosas	9034
menos	316	hace	9012

(a) 13th of March

(b) 13th of March to 21st of June

Figure 3.19: Top 20 words in topic 9

3.2.10. Topic 10

Due to the pandemic, many events had to be postponed or even cancelled due to the state of alarm restrictions. On the very same day that the restrictions were starting in some communities and the media was echoing of possible national restrictions, words like "eventos" (events), "festival" and "cancela" (cancel) were widely used among the social network users as shown in figure 3.20a. After the state of alarm is declared and the restrictions are known, on the 17th of March there is another surge in the activity as shown in figure 3.45j. The first word after the pandemic jargon, is "selectividad" the national university access tests, probably because students from 2nd of Baccalaureate were not able to assist to classes and were concerned about their future studies.

The word "eurocopa" is also in the top words because the future of all the competitions managed by UEFA was hanging on a very thin thread depending on their final decision on whether to suspend them or not.

La UEFA estudiará el 17 de marzo el futuro de la Eurocopa y resto de competiciones

La Unión de Asociaciones Europeas de Fútbol (UEFA) convocó hoy a las 55 federaciones del Viejo Continente y a los representantes de sus ligas para discutir el próximo martes 17 de marzo posibles estrategias de cara al coronavirus COVID-19, incluyendo una posible suspensión de la Eurocopa 2020.

EFE

On the 20th of April, there is a second wave of activity in figure 3.20a, words like "cortos" (short) and "celtas" are trending as shown in figure 3.20c (celts) referring to the rock

band Celtas Cortos because they released a video to collect funds for MSF (the Spanish version of Doctors Without Borders).

Celtas Cortos & Amigos: 20 de abril versión 2020 Covid-19 para MSF <https://t.co/K7mgZuxpIK#NoNospodránParar #SinFronterasEnCasa #celtas cortos #20DeAbril>

— Celtas Cortos (@CeltasCortos) April 19, 2020

Word	Count	Word	Count	Word	Count
coronavirus	2279	coronavirus	1677	abril	911
covid-19	356	2020	278	covid-19	820
debido	350	covid-19	260	versión	788
2020	318	aplaza	211	cortos	720
marzo	241	crisis	209	2020	719
casa	212	selectividad	162	celtas	716
situación	178	debido	160	coronavirus	448
eventos	175	2021	153	canción	79
fecha	163	cuarentena	135	años	78
crisis	162	pandemia	131	covid	77
festival	145	fecha	122	pandemia	71
cancela	145	marzo	108	través	71
cancelado	143	eurocopa	104	nueva	68
aplaza	139	retrasa	102	debido	66
semana	133	casa	99	artistas	62
suspende	129	abril	87	confinamiento	61
nueva	129	causa	80	crisis	57
conciertos	128	gratis	80	lucha	51
abril	125	nueva	77	2021	48
nuevo	123	películas	74	especial	44

(a) 13th of March

(b) 17th of March

(c) 20th of April

Figure 3.20: Top 20 words in topic 10

3.2.11. Topic 11

The Government made appearances and press conferences throughout the state of alarm timeline, those are mirrored in this topic, as for each time the Government appeared on TV to make a statement, there is a rise of activity in the chart 3.45k:

13th of March: Pedro Sánchez announces the declaration of the state of alarm.

15th of March: Press conference with the Minister of Defence and the Ministers of the Interior, Transport, Mobility and the Urban and Health Agenda.

22nd of March: Press conference by Pedro Sánchez after his video conference with the regional presidents.

30th of March: Press conference following the meeting of the Coronavirus Technical Management Committee.

8th, 21st and 28th of April: Press conference following the meeting of the Coronavirus Technical Management Committee.

8th of May: Press conference by Salvador Illa and Fernando Simón.

12th and 25th of May: Telematic press conference following the meeting of the De-escalation Management Committee.

17th of May: Telematic press conference by Ministers Montero and Illa.

9th of June: Telematic press conference following the meeting of the De-escalation Management Committee.

3.2.12. Topic 12

Similar to the reaction in Australia, Spaniards flocked to the supermarkets and stockpiled on toilet paper, figure 3.21a shows "papel" and "higiénico" (toilet paper) as some of the most used words on the 13th of April, matching the high tweeting activity of chart 3.45l.

¿A qué se deben las compras compulsivas de papel higiénico de estos días?

En estos últimos días de alarma por el coronavirus se han vivido escenas peculiares en algunos supermercados, sobre todo de Madrid, en las que se veía a consumidores con carritos de la compra repletos hasta arriba de rollos de papel higiénico. Estas imágenes han circulado por redes sociales y las han publicado los medios de comunicación.

El País

For a big part of the population, COVID-19 was something to fear. Data of a joint study from Spanish universities ²⁶, shows that the fear of losing someone had increased a bit on 42% and 38.9% and a lot on 29.5% and 40.5% of the male and female surveyees respectively. That matches the usage of words like "miedo" (fear), "perder" (lose) and "gente" (people) as shown in figure 3.21b. The same study states that only 3% of the surveyees required psychological care, the words "mata" (kill) and "matar" (to kill) are among the most used in figures 3.21b and 3.21c, if a search for "me va a matar" (it is going to kill me) is performed in the dataset, there are 1,696 results, whereas for the search "me mata" (it kills me) there are 2,102 results out of 262,546 tweets (1.44%), both are common negative phrases that denote stress²⁷, so it could be affirmed that the data from the study reflects on Twitter during the period.

²⁶https://www.ciencia.gob.es/stfls/MICINN/Universidades/Ficheros/Consecuencias_psicologicasCOVID19.pdf

²⁷<https://kidshealth.org/en/teens/stress-situations.html>

Word	Count	Word	Count	Word	Count
coronavirus	12617	coronavirus	7227	coronavirus	4878
casa	959	casa	624	covid-19	468
gente	947	cuarentena	483	casa	455
ahora	833	covid-19	481	cuarentena	445
papel	686	gente	467	gente	399
madre	510	puede	448	covid	329
higiénico	494	ahora	446	ahora	329
hace	457	matar	350	hace	243
tener	441	hacer	323	hacer	239
hacer	429	hace	308	puede	236
bien	413	madre	281	bien	216
cuarentena	412	días	264	virus	215
miedo	405	mata	260	madre	211
días	405	covid	259	solo	205
mierda	404	solo	254	mata	204
morir	402	bien	240	covid19	195
quiero	394	tener	238	días	188
puto	393	salir	230	matar	186
puede	388	puta	226	quiero	178
solo	385	morir	223	miedo	177

(a) 13th of March (b) 16th of March (c) 19th of march

Figure 3.21: Top 20 words in topic 12

3.2.13. Topic 13

Sports were amongst the first to take measures against COVID-19, for example by playing football matches behind closed doors as announced by La Liga. On the other hand, organisations like FIFA took more drastic measures and decided to suspend all matches that were to take place during March and April. Figures 3.22a, 3.22b and reflect those news by having the organisation names and football jargon among the most used words in those days.

La Liga se disputará las dos próximas jornadas a puerta cerrada por el coronavirus; también el Barça-Nápoles

La Liga de fútbol, como el resto de competiciones deportivas en España, se disputará al menos durante las dos próximas semanas a puerta cerrada para prevenir la expansión de la epidemia del coronavirus.
RTVE

La FIFA suspende las eliminatorias del Mundial de los meses de marzo y abril

La FIFA no vive de espaldas a todo lo que está sucediendo y ha decidido variar algunas de las normas de su reglamento, así como pedir el aplazamiento de todos los partidos internacionales de los meses de marzo y abril.
MARCA - José Félix Díaz

On the 6th of May, there is another surge in chart 3.45m. Most well-known football teams asked their players to undergo COVID-19 tests. All this news, are commented on

Twitter using words like "tests", "equipos" (teams) and "jugadores" (players) as shown in figure 3.22c.

La vuelta al fútbol de los grandes: mascarillas, policía, test para la mayoría y algunos curiosos

Este miércoles fue la fecha en la que varios equipos de LaLiga Santander regresaron al trabajo siempre cumpliendo las normas de seguridad. Repasamos cómo fue la llegada escalonada en Valdebebas, Paterna y en la Ciudad Deportiva Joan Gamper.

MARCA - D. Picó, L. F. Rojo y P. Polo

Word	Count	Word	Count	Word	Count
coronavirus	1500	coronavirus	1005	coronavirus	611
positivo	262	positivo	174	test	283
suspende	214	fútbol	162	equipos	254
fútbol	201	eurocopa	159	españa	226
liga	192	2021	148	protección	215
league	176	covid-19	139	desabastecimiento	213
covid-19	165	copa	135	pleno	211
premier	138	jugadores	129	cuba	211
entrenamientos	137	uefa	114	41.738	211
partidos	128	américa	106	exporta	207
semana	118	liga	104	sánchez	202
equipo	116	crisis	79	covid-19	201
jugadores	112	plantilla	71	jugadores	181
europa	105	junio	68	pruebas	108
champions	102	equipo	67	plantilla	101
arteta	102	pandemia	65	entrenamientos	92
crisis	88	casa	65	fútbol	82
tras	84	valencia	64	liga	80
cuarentena	81	suspende	62	madrid	65
primer	78	aplaza	61	tras	61

(a) 13th of March

(b) 17th of March

(c) 6th of may

Figure 3.22: Top 20 words in topic 13

3.2.14. Topic 14

Due to the health crisis, strict hygiene and sanitary measures were introduced in the society in a matter of days, many supermarkets established protocols and measures to avoid the spread. The tweets in this topic, use words like "recomendaciones" (tips), "propagación" (spread), "medidas" (measures) and prevención (prevention) as shown in figures 3.23a, 3.23b and 3.23c.

Horarios modificados y limitación de aforo, medidas que están tomando los supermercados en esta emergencia sanitaria

Tras decretarse el Estado de Alarma en España y el cierre de la mayoría de los comercios a excepción de los que se consideran de primera necesidad, desde los supermercados y negocios de alimentación se ha mandado un mensaje de calma, ya que el sector afirma que está asegurado el abastecimiento con normalidad
Newtral - Yolanda García

Word	Count	Word	Count	Word	Count
coronavirus	3483	coronavirus	2098	coronavirus	1001
medidas	2553	covid-19	1669	covid-19	768
covid-19	2041	medidas	1538	medidas	648
marzo	999	situación	723	prevención	274
prevención	948	alarma	676	evitar	265
situación	931	marzo	558	crisis	238
recomendaciones	867	sanitaria	545	sanitaria	219
aviso	729	crisis	513	propagación	218
evitar	725	evitar	488	seguridad	212
nuevo	709	gobierno	463	desinfección	210
propagación	707	prevención	459	salud	209
medida	697	propagación	448	marzo	198
sanitaria	640	atención	423	gobierno	192
autoridades	634	debido	410	alarma	189
gobierno	631	salud	406	situación	186
actividades	630	público	383	emergencia	169
siguiendo	614	servicios	371	frente	149
todas	598	nuevo	347	personas	145
ayuntamiento	586	servicio	347	casa	143
cierre	577	comunicado	345	servicios	143

(a) 13th of March

(b) 16th of March

(c) 20th of March

Figure 3.23: Top 20 words in topic 14

3.2.15. Topic 15

Similar to topic 14, topic 15 includes tweets regarding the stimulus packages and measures taken by the tax office. In figure 3.45o, it shows three surges of tweeting activity on the 14th of March, 2nd and 30th of April. On the former, the words "autónomos" (self-employed) and "suspensión inmediata" (immediate suspension) show as the most frequently used as shown in chart 3.24a, this is due to the measures introduced by the tax office²⁸ regarding tax flexibility for self-employees and the halt of prescriptions or expiry

²⁸https://www.agenciatributaria.es/AEAT.internet/Inicio/La_Agencia_Tributaria/Campanas/_Campanas_/Medidas_Tribu

date on any procedures. On the latter, words like "desempleados"/"parados" (unemploy-ees), "empleos" (jobs), "paro" (unemployment) or "afiliados" (affiliees) are popular due to the loss of 833,979 contributors to the social security system, which meant a rise of 9.84%²⁹ on the total number of citizens enrolled in SEPE (Public Service of State Employment).

La afiliación a la Seguridad Social perdió 833.979 cotizantes a cierre de marzo por el Covid-19

La Seguridad Social cerró el mes de marzo con un descenso de 833.979 afiliados, su mayor recorte de la historia, hasta situarse el número total de ocupados en 18.445.436, según datos del Ministerio de Inclusión, Seguridad Social y Migraciones publicados este jueves.

Europa Press

Word	Count	Word	Count	Word	Count
coronavirus	1976	coronavirus	1230	coronavirus	633
crisis	1573	marzo	769	covid-19	441
mientras	1521	crisis	646	millones	380
dure	1513	paro	544	trimestre	362
autónomos	1511	empleos	433	primer	336
pago	1509	covid-19	390	crisis	285
suspensión	1507	historia	342	caída	262
inmediata	1502	personas	235	economía	259
petición	1482	españa	229	mayor	213
hacienda	1474	mayor	219	española	201
función	1474	desempleados	218	desplome	167
pública	1474	millones	213	impacto	143
ministerio	1472	peor	210	histórico	137
través	265	subida	208	marzo	136
covid-19	74	833.979	189	españa	129
millones	54	parados	165	2020	125
españa	50	afiliados	156	hunde	121
economía	47	302.365	156	tras	121
2020	43	empleo	155	casi	113
recesión	40	datos	137	registra	90

(a) 14th of March

(b) 2nd of April

(c) 30th of April

Figure 3.24: Top 20 words in topic 15

3.2.16. Topic 16

Before the state of alarm was announced, stimulus packages were already approved. On the 17th of March, more stimulus packages were introduced to help families and self-employees. Hence the usage of the word "medidas económicas" (stimulus packages) as shown in figures 3.25a, 3.25b and 3.25c. On the 17th of April the Government approved extraordinary measures to address the impact of COVID-19 in professional training because unemployed people had signed up for SEPE courses and now the teaching meth-

²⁹<https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/trabajo14/Paginas/2020/050520-paro.aspx#:~:text=En%20abril%20se%20inscribieron%20150.461,respecto%20al%20mes%20de%20marzo.>

ods would be more flexible.³⁰ All those measures cause changes in the tweeting activity as shown in chart 3.45p.

Word	Count	Word	Count	Word	Count
coronavirus	1577	coronavirus	707	crisis	715
crisis	942	crisis	655	covid-19	604
medidas	935	covid-19	569	coronavirus	547
millones	868	medidas	376	medidas	310
gobierno	865	plan	257	gobierno	237
200.000	595	gobierno	254	plan	228
covid-19	434	sector	240	sector	224
impacto	421	ayudas	227	frente	185
euros	391	frente	205	ayudas	176
paliar	351	millones	199	millones	172
plan	332	tras	140	social	166
movilizará	312	respuesta	138	económica	157
económicas	302	económica	133	europea	145
sánchez	302	social	125	euros	132
frente	298	efectos	123	pandemia	125
anuncia	297	pandemia	119	impacto	122
efectos	294	reunión	115	pide	121
económico	292	euros	111	efectos	113
social	215	impacto	110	paliar	110
españa	210	paliar	106	unión	106

(a) 17th of March (b) 8th of April (c) 17th of April

Figure 3.25: Top 20 words in topic 16

3.2.17. Topic 17

During the state of alarm, Spanish Civil Guard had to enforce the restrictions applied during the confinement. From the 20th to the 26th of March, there is high activity in topic 17, as shown in graph 3.45q. On the 20th, the words "mujer" (woman), "guardias civiles" (Civil Guards) and "escape" are among the most used in the tweets; this matches the news from Zaragoza where a woman spat on two Civil Guards being positive on COVID-19.

Una mujer escupe a los guardias civiles que la detienen en Zaragoza y da positivo en coronavirus

Una mujer fue detenida el pasado miércoles en la localidad zaragozana de Cuarte de Huerva por efectivos de la Guardia Civil por presuntamente agredir a su compañero sentimental.

El Mundo - Javier Ortega

On the 23rd and 26th of March, there is massive spamming from an account named @A_definitivas from a legal portal offering help with anything related to COVID-19 and a FAQs section. Figures 3.26b and 3.26c include the words "información" (information), "utilidad" (useful), "dejar" (leave), "además" (furthermore)... which are most of the words in the tweet 3.2.17. This tweet was sent 559 and 436 times, accounting for 38% and 35.1% of the tweets on the 23rd and 26th, respectively.

³⁰<https://www.boe.es/boe/dias/2020/04/17/pdfs/BOE-A-2020-4506.pdf>

Nosotros hemos habilitado un apartado donde podréis encontrar información muy útil sobre el #Coronavirus y el Estado de Alarma y además podéis dejar vuestras dudas legales que contestaremos lo más rápido posible. ¡Esperamos que os sea de utilidad! <https://t.co/8dHxna908Q>

— A definitivas - Portal Jurídico (@A_definitivas) March 23, 2020

Word	Count	Word	Count	Word	Count
coronavirus	750	alarma	692	alarma	547
guardias	301	información	653	información	467
civiles	300	utilidad	639	utilidad	450
positivo	262	dejar	570	posible	444
mujer	243	útil	567	útil	441
stantibus	212	posible	566	además	440
rebus	212	además	563	dudas	439
cláusula	210	dudas	563	rápido	439
información	209	legales	562	legales	438
situación	209	podéis	562	encontrar	438
puede	209	encontrar	562	contestaremo	437
artículo	207	rápido	561	podéis	436
escupe	204	podréis	561	apartado	436
ayuda	204	habilitado	561	habilitado	436
doctrina	203	contestaremo	560	podréis	436
utilidad	203	apartado	560	dejar	436
vivimo	202	coronavirus	445	coronavirus	414
útil	201	policía	125	civil	152
mucha	201	covid-19	105	policía	141
publicado	201	alerta	91	guardia	136

(a) 20th of March

(b) 23rd of March

(c) 26th of march

Figure 3.26: Top 20 words in topic 17

3.2.18. Topic 18

In this topic, users mourn the loses of three people: Carlos Falcó, 5th Marquess of Griñón; Dolors Sala, Pep Guardiola's mother and Jose María Calleja, a Spanish journalist, writer and university professor as shown in the top words in figure 3.26.

Muere Carlos Falcó, marqués de Griñón, a causa del coronavirus

Carlos Falcó ha fallecido este viernes en Madrid a los 83 años en la Fundación Jiménez Díaz, según han informado a este diario fuentes de la familia. El marqués de Griñón había sido ingresado en este hospital madrileño a principios de semana después de haber dado positivo en el coronavirus.

El País

Muere la madre de Pep Guardiola, Dolors Sala Carrió, por coronavirus

La madre de Pep Guardiola, Dolors Sala Carrió, ha fallecido este lunes en Manresa a los 82 años de edad. Según ha informado el Manchester City, actual club del entrenador catalán, había contraído el coronavirus.

Marca

El Covid-19 se lleva a José María Calleja, el periodista valiente al que ETA no pudo matar

José María Fernández Calleja (León, 1955) había ingresado en el Hospital Clínico de Madrid el 29 de marzo. Tras agravarse su salud, fue trasladado a la UCI, donde murió.

El Mundo - Leyre Iglesias

Word	Count	Word	Count	Word	Count
coronavirus	1681	coronavirus	1218	coronavirus	1260
muere	591	muere	405	josé	939
años	530	madre	380	maría	903
falcó	381	covid-19	377	calleja	892
carlos	337	guardiola	374	periodista	812
marqués	297	años	305	muere	480
griñón	261	fallece	275	años	319
civil	253	cortina	237	fallece	290
guardia	252	alfonso	235	covid-19	276
periodista	244	médico	142	víctima	115
candela	210	fallecido	142	gran	100
segundo	191	repsol	140	columnista	100
chema	189	expresidente	130	crisis	93
covid-19	183	causa	124	escritor	91
deportivo	164	crisis	115	causa	91
fallece	157	dolors	113	lucha	77
causa	150	sala	112	hospital	69
crisis	140	josé	108	fernández	67
víctima	133	víctima	106	fallecido	67
javier	130	juan	102	antonio	59

(a) 20th of March

(b) 6th of April

(c) 21st of April

Figure 3.27: Top 20 words in topic 18

3.2.19. Topic 19

Because politicians were also affected by COVID-19, the first cases went viral in Spain, as shown in figure 3.45s. As per the top words in figures 3.28a, 3.28c, 3.28b refer to Esperanza Aguirre, Boris Johnson and Carmen Calvo respectively; the former was hospitalised together with her husband whilst the latter two were isolated.

Esperanza Aguirre y su marido, ingresados por coronavirus

La ex presidenta del Gobierno madrileño Esperanza Aguirre y su marido, Fernando Ramírez de Haro, se encuentran ingresados en la Fundación Jiménez Díaz tras haber dado positivo en coronavirus.

El Mundo - Juanma Lamet

Boris Johnson da positivo en coronavirus

Boris Johnson ha dado positivo por coronavirus, informó este viernes el político en un vídeo colgado en su cuenta de Twitter. "He decidido autoaislarme, pero seguiré dirigiendo la respuesta del Gobierno por teleconferencia para luchar contra este virus. Juntos lo derrotaremos", afirma en la grabación.

El Mundo - Carlos Fresneda

La vicepresidenta Carmen Calvo, negativo no concluyente en coronavirus

La vicepresidenta primera del Gobierno, Carmen Calvo, ha dado negativo no concluyente en las pruebas de coronavirus. Calvo ingresó el domingo por la noche en un centro privado de Madrid tras encontrarse mal.

El Mundo - Raúl Piña

Word	Count	Word	Count	Word	Count
coronavirus	2065	coronavirus	2096	coronavirus	2025
positivo	618	positivo	727	positivo	1085
años	533	calvo	457	boris	814
aguirre	521	covid-19	374	johnson	809
esperanza	495	años	371	primer	539
primer	433	negativo	365	covid-19	389
covid-19	419	carmen	361	años	374
marido	417	hospital	293	ministro	326
hospital	414	concluyente	263	dado	256
caso	353	ingresado	263	hospital	254
enfermera	330	prueba	242	británico	222
primera	315	primer	240	caso	214
tras	271	síntomas	229	tras	174
hospitalizados	224	garzón	227	paciente	173
ingresados	189	baltasar	219	síntomas	166
dado	182	tras	218	días	159
cuarentena	171	caso	214	alta	157
persona	170	dado	195	test	142
semana	163	cospedal	180	médico	140
víctima	160	marido	175	víctima	127

(a) 19th of March

(b) 24th of March

(c) 27th of March

Figure 3.28: Top 20 words in topic 19

3.2.20. Topic 20

Due to the need of COVID-19 tests, the Spanish Government decided to buy them from a company in China that sold tests approved by EU's healthcare regulations. When received, the Government returned them after considering them faulty. On chart 3.45t it shows that this news went viral and figure 3.29 shows the most commonly used words.

Word	Count
coronavirus	1660
test	1401
rápidos	1127
china	983
funcionan	780
comprados	582
bien	510
gobierno	470
españa	393
tests	377
empresa	377
compró	360
covid-19	351
licencia	310
sanidad	309
detectar	277
pruebas	206
ministerio	153
comprado	117
detección	98

Figure 3.29: Top 20 words in topic 20 on the 26th of March

3.2.21. Topic 21

In other topics in both datasets, it is shown that webinars are used to share information among the citizens from healthcare experts and Academia. In this topic, webinars were used to share information about how businesses could cope with the crisis. As shown in figures 3.30a, 3.30b and 3.30c the most common words are "curso" (course), "cómo" (how), "impacto" (impact) and "online". All these words refer to such offerings from multiple institutions trying to help businesses diminish the impact of COVID-19 on their finances, or explain the possible impacts and outcomes in their operation.

¡Hoy tenemos una cita! Te esperamos esta tarde en nuestro [#WebinarATRE-VIA](#) "Las marcas de alimentación frente al COVID-19".

¿Aún no te has apuntado? ¡Todavía estás a tiempo!

<https://t.co/3MG10QoaG4> pic.twitter.com/vMbscUyxLQ

— ATREVIA (@Atrevia) April 30, 2020

El futuro del turismo en la era post-[#coronavirus](#).

Asiste con nosotros el 6 de mayo al [#webinargratuito](#)

¡Te esperamos!

Link de inscripción: <https://t.co/o35maot6pw#QuedateEnCasa> pic.twitter.com/usZ0yWjvk1

— IUNIT (@IUNIT_madrid) April 30, 2020

Échale un vistazo la Presentación realizada por el director de @GVAivf, @MI-lueca, en el webinar "Liquidez COVID-19 en la Comunitat Valenciana" organizado por la @CEV_CV

<https://t.co/c4ArNaiUr9> pic.twitter.com/QDkNzdWJLt

— ceeielche (@ceeielche) April 30, 2020

Word	Count	Word	Count	Word	Count
covid-19	645	covid-19	652	covid-19	730
coronavirus	445	coronavirus	421	coronavirus	368
crisis	359	crisis	367	mayo	320
tiempos	239	tiempos	321	crisis	309
curso	189	mayo	216	tiempos	246
cómo	177	cómo	192	webinar	243
abril	177	online	180	sector	179
webinar	174	webinar	178	cómo	173
online	144	impacto	163	post	160
sector	121	aquí	146	aquí	147
mañana	119	tras	131	covid	141
directo	114	digital	130	empresas	125
impacto	114	covid	129	online	125
pandemia	112	pandemia	127	pandemia	123
aquí	108	situación	118	organizado	122
situación	107	sector	117	impacto	121
frente	107	directo	116	mañana	117
empresas	91	empresas	116	horas	117
covid19	89	post	115	digital	106
escolar	88	frente	110	nuevo	100

(a) 15th of April

(b) 30th of April

(c) 13th of May

Figure 3.30: Top 20 words in topic 21

3.2.22. Topic 22

ESTE TEMA NO LO TENGO MUY CLARO On the 26th of March there is a significant number of tweets suggesting to sign a petition on Avaaz,³¹ a global activist organisation, to ask the G19 to condone debt of developing countries.

Hay una cosa que los líderes del #G20 deben acordar mañana: ofrecer la condonación de deudas de los países más pobres del mundo durante el #coronavirus. ¡Firma y RT antes de que comience el #G20VirtualSummit! <https://t.co/dI7Fwfkvho>

— Yol (@Anhelamive) March 26, 2020

³¹<https://secure.avaaz.org/page/en/>

Word	Count	Word	Count	Word	Count
coronavirus	1821	coronavirus	1849	coronavirus	1296
crisis	522	crisis	561	crisis	682
medidas	279	covid-19	314	mundo	460
covid-19	248	medidas	187	covid-19	432
situación	189	pandemia	180	líderes	356
salud	177	mundo	173	países	356
españa	173	epidemia	163	deben	336
puede	168	cómo	150	pobres	322
pandemia	167	sistema	147	cosa	307
cómo	159	mundial	140	condonación	288
mundo	157	situación	136	comience	287
sistema	155	economía	130	acordar	287
ahora	153	ahora	129	pandemia	241
responsabilidad	150	gobierno	129	ofrecer	198
virus	137	días	126	deuda	158
gobierno	133	españa	125	mañana	158
mundial	131	social	122	deudas	153
hacer	130	hacer	121	tiempos	138
sociedad	121	puede	119	españa	135
social	119	virus	116	artículo	126

(a) 13th of March (b) 17th of March (c) 26th of March

Figure 3.31: Top 20 words in topic 22

3.2.23. Topic 23

At the beginning of the pandemic, many drugs were being used to palliate the effects of COVID-19 on patients ³².

On the 14th of March, Olivier Véran, the French Healthcare Minister announced on his twitter account that "Taking anti-inflammatory drugs (ibuprofen, cortisone, ...) could be a factor in aggravating the infection. In case of fever, take paracetamol.". This information quickly went viral as shown in figure 3.32a and in ??.

#COVID-19 | La prise d'anti-inflammatoires (ibuprofène, cortisone, ...) pourrait être un facteur d'aggravation de l'infection. En cas de fièvre, prenez du paracétamol.

Si vous êtes déjà sous anti-inflammatoires ou en cas de doute, demandez conseil à votre médecin.

— Olivier Véran (@olivierveran) **March 14, 2020**

Researchers proposed obtaining antibodies from the plasma of patients who had overcome the illness to see whether or not it could work as an antidote for COVID-19. On the 26th of March, words like "enfermedad" (illness), "pacientes" (patients), "sangre" (blood) and "esperanza" (hope) show as the most used words, coinciding with the release of articles similar to the one below.

³²<https://www.redaccionmedica.com/secciones/sanidad-hoy/coronavirus-tratamientos-vacunas-curacion-covid-19-7851>

¿Es el plasma de los pacientes que superan el Covid-19 un posible antídoto?

Investigadores de distintos países están valorando la extracción de plasma de pacientes que hayan superado el Covid-19 para la obtención de anticuerpos que puedan hacer frente al virus. El plasma se utilizaría para tratar a los pacientes infectados o para prevenir el desarrollo de la enfermedad en sujetos que hayan tenido contactos próximos.

Gaceta Médica - Daniela González

On the 27th of April, an internal communication from the AEP (Spanish Paediatric Association) alerting of the existence of some cases that could be related to Kawasaki's illness or toxic shock syndrome; the document also stated that some of the cases were shown in kids with either a positive PCR or a negative PCR but a positive SARS-CoV-2 serology test. Even though that communication was meant to be internal, on that day words like "niños" (kids), "síndrome" (syndrome), "casos" (cases) and "pediatras" (paediatricians) were among the most used as shown in figure 3.32c.

Los pediatras llaman a la calma a las familias respecto al shock pediátrico asociado al Covid-19

Explican que como procedimiento habitual en el marco de las enfermedades infecciosas, la Asociación Española de Pediatría ha remitido una información profesional a los pediatras y otros médicos que ven niños, ya que, "gracias a nuestro modelo español de Atención Pediátrica, en el cual los pediatras de atención primaria son el primer escalón en la atención sanitaria de los niños, es posible un diagnóstico de sospecha temprano ante patologías que requieren una intervención inmediata", afirman.

Expansión - M.C.G.

Word	Count	Word	Count	Word	Count
coronavirus	1188	coronavirus	670	coronavirus	679
ibuprofeno	764	covid-19	454	niños	674
covid-19	471	síntomas	200	covid-19	672
infección	282	enfermedad	181	síndrome	238
síntomas	243	pacientes	173	pediatras	211
infecciones	230	riesgo	120	casos	196
sanidad	223	virus	113	síntomas	170
agrave	196	personas	111	relación	167
paracetamol	178	sangre	94	shock	165
ministro	165	nueva	82	podría	163
puede	164	días	81	enfermedad	160
ningún	158	podría	79	puede	154
dato	155	superado	79	covid	146
tomar	149	nuevo	77	alertan	132
francés	145	esperanza	76	pacientes	120
antiinflamatorios	144	donantes	75	nueva	120
pacientes	135	casos	73	pediátrico	104
virus	133	puede	72	detectan	104
fiebre	132	covid	67	grave	99
riesgo	129	infección	66	relacionada	99

(a) 15th of March

(b) 26th of March

(c) 28th of April

Figure 3.32: Top 20 words in topic 23

3.2.24. Topic 24

Because of the pandemic, many countries in Latin America had to enforce confinement or declare the state of emergency. In figure it shows that countries like Venezuela, Argentina, México, Costa Rica or Ecuador were among the most commented on that day. This coincides with some of the measures taken by the countries and the news of spread all over the continent.

Venezuela inicia hoy una cuarentena en en los 23 estados del país para frenar el coronavirus y Ecuador declara el estado de excepción

El presidente de Venezuela, Nicolás Maduro, ordenó este lunes una "cuarentena total" que entrará en vigor a partir de este martes y que se aplicará en los 23 estados del país para frenar la expansión del coronavirus.

Expansión - Agencias

Venezuela inicia hoy una cuarentena en en los 23 estados del país para frenar el coronavirus y Ecuador declara el estado de excepción

El presidente de Venezuela, Nicolás Maduro, ordenó este lunes una "cuarentena total" que entrará en vigor a partir de este martes y que se aplicará en los 23 estados del país para frenar la expansión del coronavirus.

Expansión - Agencias

On the 20th of March, some countries still show up on the list: Costa Rica and Argentina. Costa Rica confirmed their second death due to COVID-19 and Argentina declared mandatory quarantine until the 31st of March.

Costa Rica confirma la segunda víctima mortal por el nuevo coronavirus

El Ministerio de Sanidad de Costa Rica ha confirmado la segunda víctima mortal por la pandemia del nuevo coronavirus originado en la ciudad china de Wuhan, un hombre de 87 años de edad que tenía problemas de salud previos.

Europa Press

Argentina entra en cuarentena obligatoria hasta el 31 de marzo

Argentina se somete a una cuarentena total. Desde la medianoche del jueves y hasta la medianoche del 31 de marzo, los ciudadanos deberán permanecer en sus casas y limitar sus salidas a la compra de alimentos y medicamentos.

El País - Enric González

By the 26th of March, the situation had worsened in those countries and the amount of positive cases and deceased had risen. Venezuela confirmed their first death due to COVID-19.

Venezuela registra la primera muerte por COVID-19 y suma 107 contagiados

Venezuela registró este jueves la primera muerte por COVID-19 y con un nuevo caso alcanzó un total de 107 personas contagiadas por el nuevo coronavirus, informó la vicepresidenta ejecutiva, Delcy Rodríguez.

La Vanguardia - EFE

Word	Count
casos	813
coronavirus	688
covid-19	197
nuevos	182
confirmados	166
total	151
salud	135
país	120
positivos	106
marzo	96
caso	94
venezuela	88
argentina	80
confirma	79
méxico	77
personas	74
cuarentena	73
infectados	71
rica	67
ecuador	64

(a) 17th of March

Word	Count
casos	671
coronavirus	579
covid-19	284
confirmados	185
salud	169
nuevos	149
total	139
costa	124
rica	123
marzo	122
personas	113
caso	112
país	109
positivos	106
cuarentena	101
argentina	97
fallecidos	80
confirma	71
nuevo	67
sospechosos	67

(b) 20th of March

Word	Count
casos	840
coronavirus	571
covid-19	377
nuevos	266
confirmados	206
positivos	165
total	161
salud	142
caso	121
país	114
marzo	102
fallecidos	88
rica	85
costa	83
argentina	80
nuevo	79
personas	78
venezuela	74
contagiados	73
pacientes	68

(c) 26th of March

Figure 3.33: Top 20 words in topic 24

3.2.25. Topic 25

On the 20th of March, Italy exceeds 4,000 deaths with coronavirus after the most massive increase in one day of 627 people. That same day, Spain exceeds 1,000 deaths and is getting near to 20,000 positive cases. Both countries show in figure ?? that users of the social network are echoing the news by using words like "italia" (Italy), "españa" (Spain), "casos" (cases) and "fallecidos" (deceased); the rounded up numbers of COVID-19 victims in Spain also shows up: 1,000 and 20,000.

Italia supera los 4.000 muertos con coronavirus tras sumar 627, el mayor aumento en un día

Italia sigue sumando muertes con el coronavirus Covid-19: ya ha superado la barrera de los 4.000 muertos tras contabilizar 675 nuevos fallecidos, el mayor aumento en 24 horas hasta la fecha.

RTVE

España supera el millar de muertos y roza los 20.000 contagiados por coronavirus

El Ministerio de Sanidad ha actualizado el balance de contagiados y víctimas mortales por el coronavirus con origen en Wuhan (China). El balance de este viernes de los fallecidos es el siguiente...

NIUS diario - Miguel Manso de Lucas

On the 27th of March among the most used words are "fallecidos" (deceased), "muertes" (deaths), "casos" (cases), "positivo" (positive) and 4,858 as shown in figure 3.34b. On that day, the deceased rate rose to 4,858 almost doubling the rate of three days before, coinciding with the third peak in figure 3.45x.

Coronavirus: España registra 64.059 casos, 4.858 muertos y 9.357 curados

El Estado de Alarma por el coronavirus cumple su decimocuarto día este viernes. Una semana "difícil", según ha señalado Fernando Simón, director del Centro de Coordinación de Emergencias Sanitarias del Ministerio de Sanidad, en la que nuestro país llegue al pico de casos de coronavirus, que ya acumula 64.059 casos, 4.858 muertos y 9.357 curados.

Redacción Médica

The third peak in figure 3.45x, again shows words like "fallecidos" (deceased), "muertes" (deaths) and "casos" (cases) in figure 3.34c. On that day the deceased rate rose to over 10,000.

Los casos por coronavirus se elevan a 110.238 personas y a más de 10.000 fallecidos

Los casos de coronavirus en España se ha elevado este jueves a 110.238 contagiados (8.102 más que el miércoles), 10.003 fallecidos (950 más en 24 horas) y 26.743 curados (4.096 más), según los últimos datos aportados por el Ministerio de Sanidad.

Europa Press

Word	Count	Word	Count	Word	Count
coronavirus	2123	coronavirus	2043	coronavirus	1875
casos	1032	casos	1096	casos	936
españa	955	españa	744	españa	890
muertos	824	fallecidos	741	fallecidos	849
fallecidos	644	muertos	648	muertes	548
italia	600	horas	501	muertos	541
supera	510	contagiados	465	horas	446
muertes	478	últimas	453	covid-19	422
positivos	390	número	419	personas	421
contagiados	369	muertes	414	10.000	385
20.000	343	nuevos	373	supera	372
1.000	332	covid-19	366	últimas	360
contagios	305	contagios	328	nuevos	355
hora	305	positivos	320	número	326
nuevos	301	personas	313	positivos	299
registra	285	4.858	306	contagios	291
covid-19	278	infectados	299	registra	282
infectados	276	registra	296	datos	272
última	271	datos	293	hora	256
horas	270	64.059	272	110.238	248

(a) 20th of March

(b) 27th of March

(c) 2nd of April

Figure 3.34: Top 20 words in topic 25

3.2.26. Topic 26

This topic embraces the conversations and tweets regarding the global impact of COVID-19. Countries show in the most used words lists: China, Italy and the US in figure 3.35a; China, Italy and France in figure ?? and Holland, Germany, Portugal or the US in figure .

On the 16th, Spain closes its land borders³³ and allows entry only to nationals and residents. That same day, the President of the European Commission announced that she would propose the immediate temporal closure of the EU borders.

La UE cerrará sus fronteras exteriores temporalmente y con excepciones

La presidente de la Comisión Europea, Ursula von der Leyen, ha anunciado este mediodía que propondrá con efecto inmediato el cierre temporal de las fronteras de la UE. Durante al menos 30 días, pero prolongables, ningún extranjero podrá acceder al territorio comunitario, salvo excepciones.

El País - Pablo R. Suanzes

It is worth noting that on the 27th of March, the day that the Netherlands shows in the list, the Spanish Foreign Affairs Ministry replied to Wopke Hoekstra on his criticism of the Spanish management of the crisis.

.@WBHoekstra estamos juntos en el mismo barco europeo. Hemos chocado contra un iceberg. Ahora todos corremos los mismos riesgos. No hay tiempo

³³<https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2020/160320-covid-fronteras.aspx>

para discusiones sobre supuestos pasajeros de 1a o 2a clase. No es momento de defraudar a nuestros ciudadanos <https://t.co/66ZDiAmNnK>

— Arancha González (@AranchaGlezLaya) March 27, 2020

Word	Count	Word	Count	Word	Count
coronavirus	2582	coronavirus	2149	coronavirus	1635
españa	485	fronteras	724	covid-19	605
china	461	españa	526	españa	538
europa	452	cierra	364	italia	386
declara	384	covid-19	348	china	314
países	367	china	301	país	297
trump	360	gobierno	259	países	285
italia	354	casos	253	holanda	284
emergencia	327	cierre	246	casos	249
casos	317	italia	223	mundo	212
covid-19	316	países	218	alemania	192
pandemia	285	días	210	portugal	183
país	227	europa	191	ancianos	180
gobierno	209	frenar	180	crisis	173
presidente	208	terrestres	177	eeuu	167
eeuu	208	país	168	pandemia	165
unidos	207	medidas	164	nueva	164
epicentro	207	francia	142	hospital	163
nacional	200	crisis	139	unidos	162
alerta	199	cuarentena	125	holandesa	160

(a) 13th of March (b) 16th of March (c) 27th of March

Figure 3.35: Top 20 words in topic 26

3.2.27. Topic 27

Topic 27 is all about the challenge to create the vaccine or finding a cure for COVID-19. On the 13th of March, China suggested that they had developed a vaccine against COVID-19; users referenced this news by using words like "éxito" (successful), "desarrollado" (created) or "producción" (production) as shown in figure .

China asegura haber desarrollado una vacuna contra el Covid-19

El Ministerio de Defensa de China ha asegurado en un comunicado que han desarrollado con éxito una vacuna contra el coronavirus, por lo que han aprobado la realización de ensayos en humanos.

Gaceta Médica - Sandra Pulido

Conversely, on the 24th of April words like "trump", "desinfectante" (disinfectant) or "inyectar" (inject) are among the most used as per figure 3.36b; that same day, Trump stated that COVID-19 patients could be treated with disinfectant shots or by using ultra-violet light.

Donald Trump sugiere tratar el Covid-19 con inyecciones de desinfectante

El presidente de Estados Unidos, Donald Trump, sugirió ayer durante una rueda de prensa en la Casa Blanca que el coronavirus podría tratarse con inyecciones de desinfectante en el cuerpo. También propuso que se irradiara a los pacientes con luz ultravioleta, ante la mirada atónita de una médica que le acompañaba en la comparecencia.

El Mundo - Agencia

On the 16th of June, Oxford University published the results from a clinical trial showing that dexamethasone could save the lives of critically ill patients with COVID-19.

WHO welcomes preliminary results about dexamethasone use in treating critically ill COVID-19 patients

The World Health Organization (WHO) welcomes the initial clinical trial results from the United Kingdom (UK) that show dexamethasone, a corticosteroid, can be lifesaving for patients who are critically ill with COVID-19. For patients on ventilators, the treatment was shown to reduce mortality by about one third, and for patients requiring only oxygen, mortality was cut by about one fifth, according to preliminary findings shared with WHO.

WHO News Release - Fadela Chaib

Word	Count	Word	Count	Word	Count
coronavirus	1444	coronavirus	1369	covid-19	683
vacuna	929	trump	946	coronavirus	578
china	586	desinfectante	867	dexametasona	440
éxito	306	covid-19	590	pacientes	389
haber	301	sugiere	382	fármaco	385
covid-19	301	inyectar	355	vacuna	372
desarrollado	293	tratar	306	universidad	314
dice	258	vacuna	277	reduce	292
humanos	205	solar	261	vidas	290
científicos	159	inyección	249	primer	279
virus	146	donald	192	mortalidad	278
ensayos	127	pacientes	187	graves	275
gran	125	enfermos	184	covid	257
escala	114	virus	170	oxford	225
producción	109	inyecciones	149	salvar	201
defensa	105	matar	143	presidente	171
pruebas	99	presidente	137	barato	163
prepara	98	españa	136	puede	159
clínicos	98	propone	134	descubierto	156
ensayo	96	plantea	127	través	149

(a) 17th of March

(b) 24th of April

(c) 16th of June

Figure 3.36: Top 20 words in topic 27

3.2.28. Topic 28

By the time COVID-19 was declared a pandemic on the 11th of March by the WHO³⁴, there had been an ongoing campaign to avoid the spread of the virus³⁵. The Spanish Government had issued a guide to good practice in the workplace with regard to COVID-19³⁶. The ongoing days and weeks, the Healthcare Ministry would share guides and advices³⁷ on how to avoid the spread of the virus.

All this information was also shared on Twitter, as shown in figures 3.37a, 3.37b and 3.37c by using the words "recomendaciones" (tips), "medidas" (measures), "lavarse" (wash yourself), "prevenir" (avoid), "propagación" (spread) or "contagio" (infection). The most active days when these recommendations were being shared more often are the 13th, 16th and 19th of March as shown in figure 3.46d, at the beginning of the state of alarm.

Word	Count	Word	Count	Word	Count
coronavirus	1672	coronavirus	1162	coronavirus	1074
manos	804	manos	442	casa	393
evitar	710	evitar	412	covid-19	387
casa	576	medidas	371	manos	330
recomendaciones	457	covid-19	369	evitar	298
medidas	450	casa	366	prevenir	215
contagio	420	contagio	327	contagio	210
responsabilidad	360	recomendaciones	270	medidas	207
covid-19	312	prevenir	219	cómo	192
lavarse	289	propagación	201	recomendaciones	178
prevenir	286	cómo	190	seguir	173
frenar	279	hacer	176	hacer	170
propagación	237	prevención	168	propagación	148
prevención	234	personas	163	prevención	147
cómo	214	consejos	160	personas	146
personas	207	agua	145	salir	131
hacer	197	importante	137	días	130
debemos	193	social	136	expansión	123
virus	190	virus	134	virus	119
social	188	salud	129	superficies	115

(a) 13th of March

(b) 16th of March

(c) 19th of March

Figure 3.37: Top 20 words in topic 28

³⁴<https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>

³⁵<https://www.who.int/westernpacific/emergencies/covid-19/information/transmission-protective-measures>

³⁶https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/presidencia/Paginas/2020/10042020_buenaspracticas.aspx

³⁷<https://www.mscbs.gob.es/ca/profesionales/saludPublica/ccayes/alertasActual/nCov/ciudadania.htm>

3.2.29. Topic 29

After the state of alarm was declared on the 13th of March, on Saturday the 14th of March, the Government met to decide on how to proceed. The meeting was marked by disagreements between the PSOE and Unidos Podemos. In figure it shows that the words "sánchez", "gobierno" (government)

Word	Count	Word	Count	Word	Count
coronavirus	2166	coronavirus	1640	coronavirus	1395
sánchez	818	sánchez	703	montero	1174
gobierno	642	gobierno	597	irene	1089
crisis	469	crisis	441	decir	889
españa	384	iglesias	437	bajada	764
york	347	gestión	365	cifras	610
iglesias	344	pedro	358	debido	551
gestión	339	españa	328	gobierno	310
pedro	306	pablo	251	través	226
times	236	cacerolada	188	españa	218
pablo	210	torra	166	vídeo	178
torra	185	covid-19	165	gestión	164
madrid	173	sanidad	157	covid-19	164
aznar	154	ahora	141	gente	143
casos	153	podemos	140	covid	143
permitir	147	toda	139	debió	117
ahora	144	fallecidos	136	ministra	116
número	140	criminal	132	menos	115
marbella	139	muertos	129	reconoce	113
país	136	nefasta	126	medidas	104

(a) 14th of March

(b) 21st of March

(c) 1st of June

Figure 3.38: Top 20 words in topic 29

3.2.30. Topic 30

Although this topic has high activity for a long period, starting from the 13th of March to the 8th of May approximately, as shown in figure 3.46f, the highest activity day was the 8th of April and thus, then analysed days are those that preceded it, to better understand how the topic evolves.

On the 1st of April, words like "residencias" (nursing homes), "mayores"/"ancianos" (elderly), "fallecidos" (deceased), "muertos" (dead), "monago" (José Antonio Monago) and "extremadura" are among the most frequently used words in the topic.

On that day, José Antonio Monago, president of Extremadura, accused the Government of "hiding the real number" of deaths due to COVID-19.

Monago acusa al Gobierno de "ocultar el número real" de fallecidos por coronavirus y Marlaska lo niega

El presidente del PP de Extremadura, José Antonio Monago, ha asegurado hoy que hay una orden expresa de "ocultar el número real" de fallecidos por el Covid-19.
El País - David Vigarío

The Community of Madrid, intervened 8 private nursing homes that were in "critical status".

La Comunidad de Madrid intervendrá ocho residencias privadas en situación "crítica"

La Consejería de Políticas Sociales de la Comunidad de Madrid ha designado a funcionarios para dirigir ocho residencias privadas donde la situación de contagios y fallecimientos por coronavirus es más "crítica", como la residencia de la Santísima Virgen y San Celedonio, donde han muerto al menos 28 personas.
ElDía - EFE

On the 6th of April, the community of Castilla-La Mancha is in the trending words list, together with the words "datos" (data), "oficiales" (official), "muertos" (deaths), "fallecidos" (deceased) and "triplican" (triple). The latter word is important to analyse the topic on that day, because there are news stating that COVID-19 burial permits almost triple the official death toll in Castilla-La Mancha.

Las licencias de enterramiento por coronavirus casi triplican el dato oficial de muertes en Castilla-La Mancha

Las licencias de enterramiento expedidas por los registros civiles de Castilla-La Mancha para personas que fallecieron por coronavirus o con sospecha de padecerlo en marzo duplican y casi triplican los datos oficiales de fallecidos a causa del virus en esta comunidad autónoma.
El País - Esther Sánchez

The last day and highest peak in figure 3.46f, corresponds to the news from the Community of Madrid reporting the death of 4,620 people, living in nursing homes, due to COVID-19 or with symptoms of the illness.

Madrid atribuye al virus la muerte de 4.260 residentes en centros de ancianos y personas con discapacidad

La Comunidad de Madrid ha informado este miércoles de que 4.260 personas usuarias de residencias de la región han muerto entre el 8 de marzo y hoy con Covid-19 (781) o teniendo síntomas de esa enfermedad (3.479). Otras 490 fallecieron por otras causas.
El País - Juan José Mateo

Word	Count	Word	Count	Word	Count
coronavirus	762	coronavirus	935	coronavirus	1029
residencias	329	mancha	421	residencias	552
covid-19	254	castilla-la	375	madrid	411
mayores	201	datos	315	ancianos	280
ancianos	181	covid-19	266	fallecidos	264
casos	165	oficiales	263	mayores	236
orden	158	casi	230	covid-19	224
españa	149	muertes	190	muertos	174
fallecidos	139	muertos	179	comunidad	153
número	132	fallecidos	175	datos	152
muertos	128	triplican	170	casos	148
real	125	dice	169	españa	148
residencia	125	casos	159	personas	133
ocultar	106	residencia	145	síntomas	124
monago	101	cifra	135	oficial	120
extremadura	90	justicia	135	mancha	116
cada	81	españa	129	cifras	115
muertes	81	enterramientos	122	oficiales	113
comunidad	79	residencias	122	calculan	111
datos	74	sospechosas	117	mapa	106

(a) 1st of April

(b) 6th of April

(c) 8th of April

Figure 3.39: Top 20 words in topic 30

3.2.31. Topic 31

This topic shows a large amount of tweets on the 13th, 17th and 31st of March as per figure 3.46g.

On the 13th of March, the words "hipoteca" (mortgage), "pago" (payment), "moratoria" (moratorium) and "autónomos" (self-employed) are among the most repeated. On that day, the Government approved a moratorium on mortgages to vulnerable groups.

El Gobierno aprueba una moratoria en el pago de hipotecas a colectivos vulnerables

El Gobierno prevé aprobar en Consejo de Ministros una moratoria en el pago de hipotecas para la población afectada por la crisis del coronavirus. La medida se incluye en un paquete de instrumentos de protección social diseñados por el Ejecutivo para mitigar el impacto del frenazo económico y la fuerte destrucción de empleo como consecuencia de la propagación de la pandemia y la declaración del Estado de alarma.

El Mundo - Víctor Martínez

That same day, the Government also announced measures to help the self-employed, but they were not popular at all because they still had to pay for the contributions but they were eligible for unemployment benefits if they closed.

El Gobierno aprueba nuevas medidas económicas contra el coronavirus que no satisfacen a los autónomos

Mientras que el Gobierno blinda a las grandes sociedades del Ibex frente a extranjeros oportunistas y las asociaciones patronales y sindicatos aplauden las medidas del Consejo de Ministros para movilizar 200.000 millones de euros para aliviar a las empresas; las asociaciones de autónomos han denunciado que el Ejecutivo de Pedro Sánchez les deja en la cuneta y que se “facilita la prestación de paro que ya tenemos y cotizamos por ella, pero no suspende las cuotas de autónomos en un momento en el que no ingresamos nada”, afirma Lorenzo Amor, el presidente de ATA, la Federación Nacional de Asociaciones de Trabajadores Autónomo. Abogacía Española - Daniel Sánchez

Due to the state of alarm, many establishments had to close and thus their employees were no longer called for work. This resulted in many businesses submitting layoffs for their employees. Following this news, the Government proceeded to temporally restrict dismissals and an announcement stating that all submitted layoffs would be revised to avoid fraud. On that day, words frequently used were: "despidos" (dismissals), "gobierno" (government), "prohíbe" (bans), "empresas" (companies), "trabajadores" (workers) and "afectado" (concerned).

El Gobierno prohíbe los despidos y revisará todos los ERTES presentados para detectar fraudes

El Gobierno ha aprobado hoy en el Consejo de Ministros extraordinario que celebra diferentes fórmulas para restringir los despidos en España de manera temporal ante la situación extraordinaria en la que ha derivado la crisis del coronavirus. El Mundo - César Urrutia

On the 31st of March, the tax office charged the contributions to 3 million self-employed citizens who outraged on this decision. Meanwhile, the Government approved measures³⁸ to ease the financial strain on businesses and the self-employed.

Estas son las nuevas medidas de ayuda para autónomos y empresas aprobadas ayer por el Gobierno

El Gobierno aprobó ayer, 31 de marzo, una serie de medidas "para aliviar la tensión financiera" de empresas y autónomos, complementarias a las de los últimos reales decretos-leyes. CdeComunicación

³⁸<https://www.boe.es/buscar/doc.php?id=BOE-A-2020-4208>

Word	Count	Word	Count	Word	Count
coronavirus	2179	coronavirus	1384	coronavirus	850
gobierno	996	gobierno	958	gobierno	529
crisis	825	despidos	753	autónomos	528
medidas	705	crisis	733	crisis	524
pago	682	covid-19	543	covid-19	445
trabajadores	552	prohíbe	539	medidas	420
hipotecas	551	autónomos	297	pago	240
autónomos	527	empresas	297	trabajadores	237
moratoria	484	erte	294	ayudas	229
empresas	459	medidas	289	moratoria	206
covid-19	413	despedir	246	afectados	205
erte	372	trabajadores	233	empresas	198
afectados	368	afectados	193	alquiler	191
paro	266	pago	158	cuota	174
alarma	218	causas	145	meses	169
españa	199	puede	134	alarma	162
prohíbe	190	mayor	132	social	151
italia	187	ayudas	131	marzo	138
impuestos	183	alarma	126	pymes	136
económicas	180	fuerza	125	cuotas	133

(a) 17th of March

(b) 27th of March

(c) 31st of March

Figure 3.40: Top 20 words in topic 31

3.2.32. Topic 32

In this topic, there are three dates marked by special events: the 14th of March, when the state of alarm was declared, the 7th of April, the World Health Day and the 12th of May, the International Nurses Day as shown in figure 3.46h.

On the 14th of April, Spanish citizens wanted to express their gratitude to the healthcare workers, and they did so by organising nationwide applause at 8 pm, which would, later on, become usual until the state of alarm finished.

Aplausos y caceroladas para sostener la "trinchera" sanitaria

Miles y miles de españoles han salido esta noche a sus ventanas y balcones para agradecer con aplausos y gritos de "viva" y "bravo" el trabajo de todos los profesionales sanitarios que atienden estos días en hospitales y en domicilios a los afectados por la pandemia de coronavirus.

El Mundo - Gonzalo Suárez

On the 7th of April, words like "gracias" (thanks), "salud" (health), "sanitarios" (health workers) and "profesionales" (professionals) are among the most used on that day, which shows that citizens wanted to thank healthcare workers for their work during the pandemic, on the World Health Day.

On the 12th of May, "enfermería" (nursing), "enfermeras" (nurses) and "gracias" (thanks) made it again to the list of the top words, which coincided on the International Nurses Day, to specifically thank them one more time for their work and effort during the pandemic.

Word	Count	Word	Count	Word	Count
coronavirus	1349	gracias	820	gracias	499
gracias	903	salud	687	enfermería	315
sanitarios	804	sanitarios	595	covid-19	312
personal	525	profesionales	567	enfermeras	283
sanitario	478	coronavirus	515	lucha	263
profesionales	394	lucha	506	trabajo	251
frenar	349	covid-19	475	labor	248
podemos	331	queremo	447	enfermeros	232
ayudar	317	personal	418	profesionales	227
balcones	313	trabajo	358	coronavirus	225
aplauaso	300	labor	344	crisis	219
casa	298	nunca	320	pandemia	214
crisis	296	sanitario	313	primera	197
vidas	284	esfuerzo	249	queremo	191
españa	280	mundial	247	línea	173
aplausos	275	agradecer	244	siempre	170
arena	273	personas	242	gran	150
compromiso	267	pandemia	240	personal	145
granito	265	primera	225	homenaje	142
salvar	263	gran	225	internacional	141

(a) 14th of March

(b) 7th of April

(c) 12th of May

Figure 3.41: Top 20 words in topic 32

3.2.33. Topic 33

From the beginning of the pandemic, users compared the progress of the health crisis in Spain to that in other countries. In this topic, the 13th, 20th and 28th of March are the days when the users are most active as shown in figure 3.46i.

On the 13th of March, the state of alarm was announced to take place at midnight on the same day. Given that Spain followed a similar evolution of the pandemic to that in Italy, users echoed the news that Spain would be in the same situation when Italy confined its population in just seven days.

En siete días, España llegará a la situación en que Italia puso todo el país en cuarentena

El Reino Unido está "cuatro semanas por detrás de Italia" en términos de la escala que puede llegar a alcanzar la pandemia del coronavirus en las islas. Así lo resumió Sir Patrick Vallance, el asesor científico del primer ministro británico, Boris Johnson, al explicar las medidas contra el brote en el país, que a última hora del jueves contaba con 590 casos de Covid-19 y 10 fallecidos por la neumonía asociada a la enfermedad.

El Condifencial - Enrique Andrés, Antonio Villarreal, Jesús Escudero

China's Health Ministry announced on the 13th of April that the day before, there were only eight new COVID cases, the lowest figure since data collection on the pandemic began in January.

Coronavirus: la extraordinaria caída en el número de nuevos contagios en China y Corea del Sur

Mientras el número de contagiados por el nuevo coronavirus se multiplica en Europa y se expande por el resto del mundo, en dos países de Asia -el continente en el que se originó- todo indica que el brote se está ralentizando.
BBC News Mundo

On the 20th of March, Fernando Simón, Director of the Coordination Centre for Health Alerts and Emergencies, said during a press conference that the number of deaths in Germany "are inexplicable". This comment went viral on Twitter that day, as shown by the most frequent words "casos" (cases), "muertos" (deaths), and "alemania" (Germany) in figure 3.42b. Similar to topic 25, users echo the news from Italy about the excess of 4,000 deaths toll.

Coronavirus: "No sabemos por qué Alemania tiene 48 muertos y España 1.000"

Fernando Simón confiesa que no puede explicar estos datos, cuando ambos países tienen los mismos contagios de Covid-19
Redacción Médica

The last peak on the chart 3.46i, corresponds to the highest death toll up to date in Spain, 832 deaths in one day according to the Government's data³⁹.

Word	Count	Word	Count	Word	Count
coronavirus	1752	coronavirus	1807	coronavirus	1473
casos	436	españa	582	muertes	627
gripe	351	casos	465	muertos	413
españa	346	muertos	455	españa	392
covid-19	237	alemania	426	covid-19	360
italia	197	muertes	352	causa	342
hace	195	covid-19	351	casos	313
personas	193	italia	266	detectadas	270
china	192	fallecidos	257	covid	266
días	191	datos	250	datos	219
infectados	191	gobierno	233	test	211
medidas	190	china	231	gobierno	198
gobierno	175	solo	215	fallecidos	197
ahora	171	gripe	202	covid19	187
gente	166	test	201	hace	187
menos	164	infectados	201	país	174
solo	141	días	198	personas	173
hacer	141	personas	194	cifras	171
virus	134	patologías	186	solo	169
muertes	128	cifras	175	infectados	168

(a) 13th of March

(b) 20th of March

(c) 28th of March

Figure 3.42: Top 20 words in topic 33

³⁹<https://www.dsn.gob.es/es/actualidad/sala-prensa/coronavirus-covid-19-28-marzo-2020>

3.2.34. Topic 34

The 13th, 17th and 19th of March are the highest tweeting days in topic 34, as shown in chart 3.46j. The most used words during those days are "niños" (children), "casa" (home), "consejos" (tips), "personas" (people), "animales" (animals), "mayores" (elderly), "ansiedad" (anxiety), "violencia" (violence) and "mujeres" (women).

All these words match the social concerns of the confinement: how to deal with kids and pets, how to take care of elderly people, psychological issues that may arise during the period, and gender-based violence, because women who were abused had to stay at home with their abuser without being able to leave the house.

Many organisations did their bit to help the citizens: advice on how to explain kids what COVID-19 is, educational TV shows for kids, sensitisation about the loneliness of the elderly and many more.

Coronavirus: cómo explicárselo a los niños para evitar el miedo <https://t.co/wjY9h3K45e> por @LauraR3_0 En @educacion3_0 #YoMeQuedoEnCasa #QuédateEnCasa pic.twitter.com/5GQJH5IJBV

— SANTILLANA España (@SANTILLANA_es) March 13, 2020

Hoy las cámaras de @telemadrid han estado en nuestras oficinas conociendo de primera mano como trabajamos en nuestra campaña gratuita, contra la soledad de las #personasmayores por aislamiento a consecuencia del #coronavirus A las 14h en @InformativosTM <https://t.co/P1efdMfvnk> pic.twitter.com/Gkr3rpzfV4

— Alares (@Alares) March 13, 2020

The Government took precautionary steps on the 13th of March to protect gender-based violence victims from their abusers, and published an action guide.⁴⁰

⁴⁰<https://violenciagenero.igualdad.gob.es/informacionUtil/covid19/GuiaVictimasVGCovid19.pdf>

Word	Count	Word	Count	Word	Count
coronavirus	902	coronavirus	797	coronavirus	665
niños	269	covid-19	191	covid-19	191
casa	206	personas	162	niños	166
crisis	146	casa	149	personas	142
covid-19	129	crisis	147	cuarentena	139
consejos	126	niños	146	crisis	134
personas	123	cuarentena	146	confinamiento	134
animales	117	confinamiento	117	casa	132
medidas	106	mayores	117	mayores	132
mayores	96	cómo	111	consejos	122
cómo	94	consejos	109	padres	79
transmiten	81	aislamiento	100	violencia	77
perros	77	animales	80	puede	74
vulnerables	75	recomendaciones	80	pueden	74
hacer	74	crecer	79	aislamiento	72
mascotas	71	casos	67	tiempos	72
padres	68	medidas	67	hacer	70
ansiedad	67	salir	65	mujeres	70
hijos	67	hijos	63	cómo	68
recomendaciones	66	violencia	61	situación	64

(a) 13th of March

(b) 17th of March

(c) 19th of March

Figure 3.43: Top 20 words in topic 34

3.2.35. Topic 35

This topic archives all the tweets offering information as well as recommendations and guides from official sources. Many accounts from official sources offered to answer questions from the citizens, or launched helplines to solve doubts about many of the aspects impacted by the health crisis.

El Gobierno de Cantabria ha habilitado un nuevo número de teléfono el 900 612 112, solo para aquellas personas que en Cantabria creen que pueden tener coronavirus.

— eldiariomontanes.es (@dmontanes) [March 13, 2020](#)

Si tienes dudas haz clic en el enlace y da respuesta a tu pregunta. Si no la encuentras escríbenos e intentaremos ir resolviendo y compartiendo todo.
#ucavilacasa #covid-19 <https://t.co/yeHW4KnI4G> pic.twitter.com/6JqH4zYIF0

— Universidad de Ávila (@UCAVILA_) [March 16, 2020](#)

On the 20th of March, the WHO launched a WhatsApp® bot aimed at answering COVID-related frequently asked questions.⁴¹

All these initiatives are referenced on Twitter by words like "teléfono" (helpline), "recomendaciones" (recommendations), "información" (information) and "Whatsapp", as shown in figures 3.44a, 3.44b and 3.44c.

⁴¹<https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp>

Word	Count	Word	Count	Word	Count
coronavirus	1506	coronavirus	1429	coronavirus	1168
información	943	información	868	información	735
covid-19	540	covid-19	693	covid-19	575
medidas	350	medidas	289	crisis	298
teléfono	254	crisis	272	aquí	224
situación	251	dudas	253	dudas	206
aquí	226	situación	226	medidas	170
recomendaciones	220	teléfono	214	situación	163
oficiales	191	través	213	preguntas	161
todas	178	aquí	212	puedes	158
dudas	171	toda	189	whatsapp	145
crisis	168	puedes	182	toda	143
toda	166	casa	168	salud	140
oficial	163	recomendaciones	168	cómo	130
importante	157	salud	167	hacer	126
puedes	157	días	162	teléfono	124
salud	151	todas	154	datos	124
redes	150	redes	153	casa	124
comunicado	140	preguntas	142	noticias	121
través	139	oficiales	142	todas	113

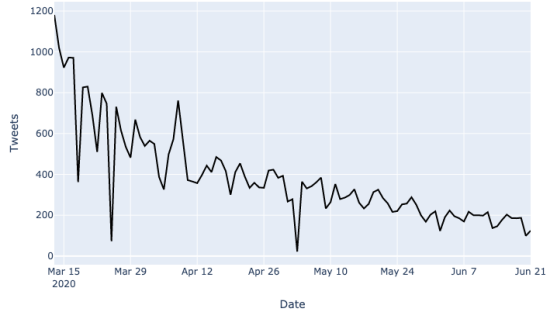
(a) 13th of March

(b) 16th of March

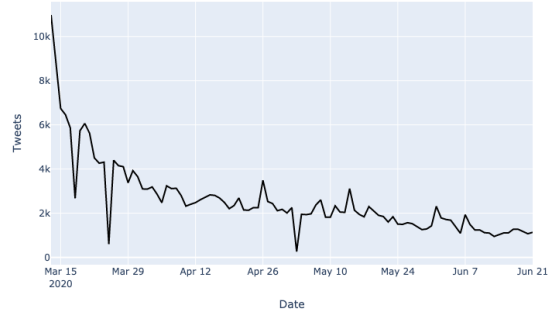
(c) 20th of March

Figure 3.44: Top 20 words in topic 35

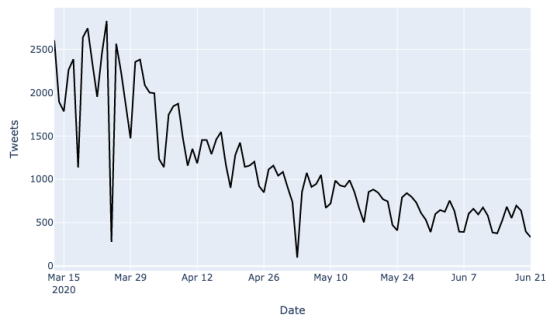
Figure 3.45: Timeline charts of the tweeting activity in each topic. I



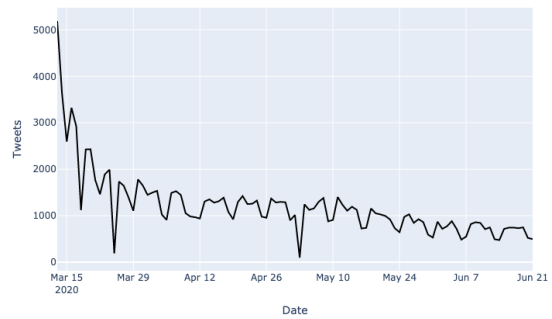
(a) Topic 1



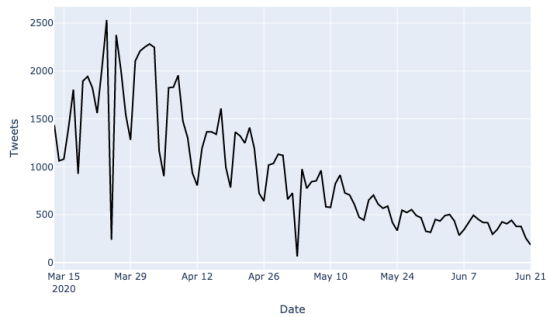
(b) Topic 2



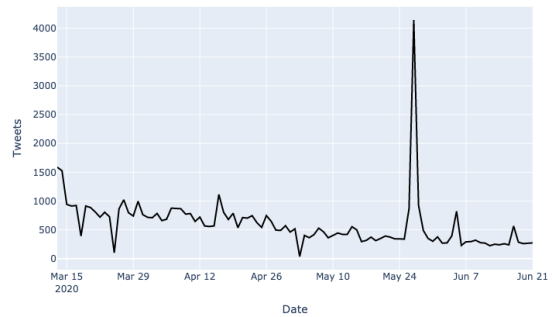
(c) Topic 3



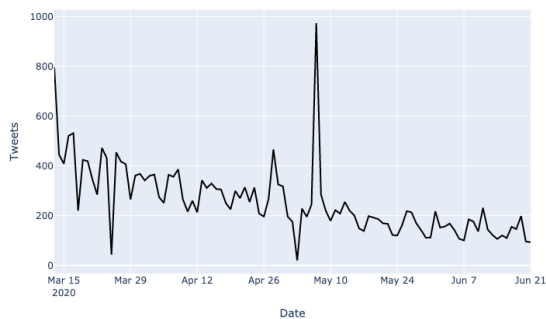
(d) Topic 4



(e) Topic 5



(f) Topic 6



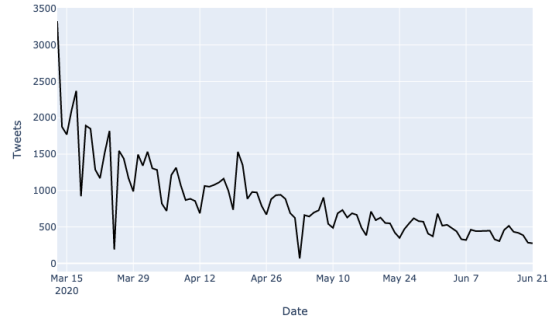
(g) Topic 7



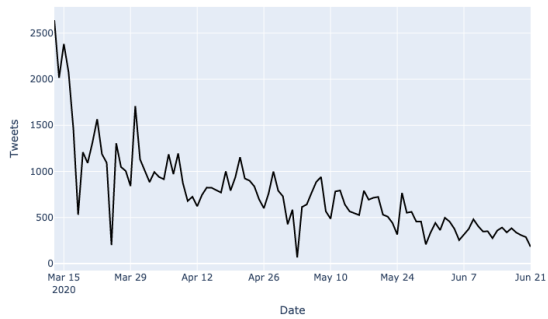
(h) Topic 8



(i) Topic 9



(j) Topic 10



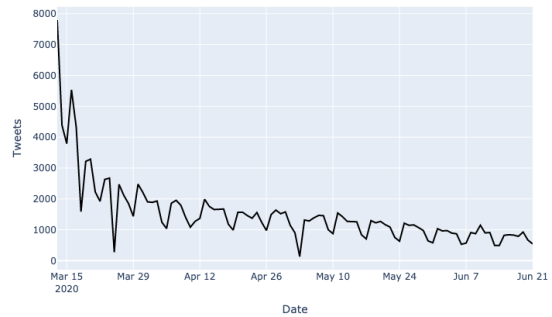
(k) Topic 11



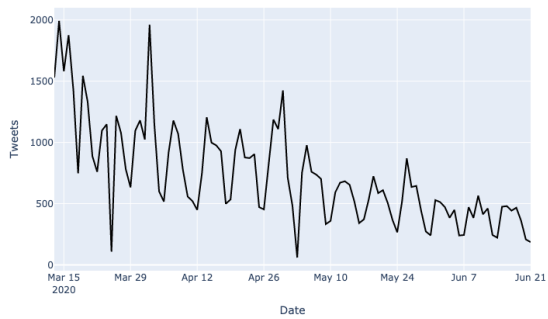
(l) Topic 12



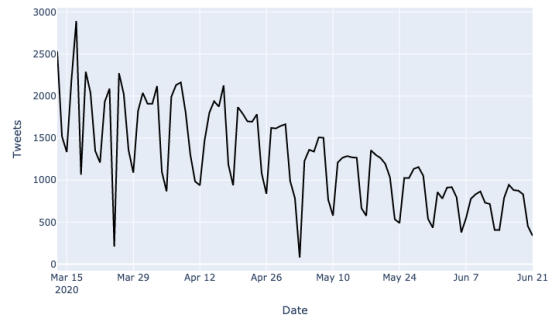
(m) Topic 13



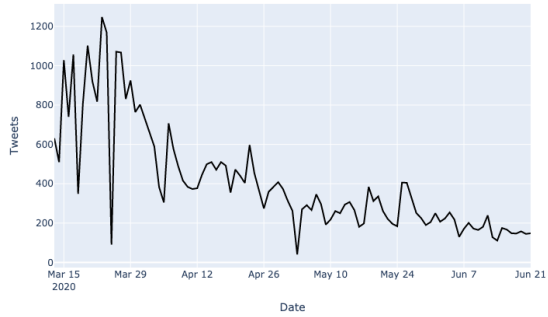
(n) Topic 14



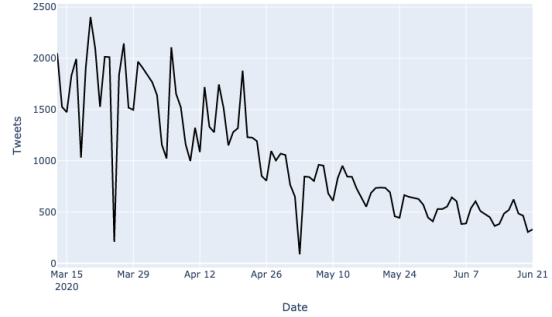
(o) Topic 15



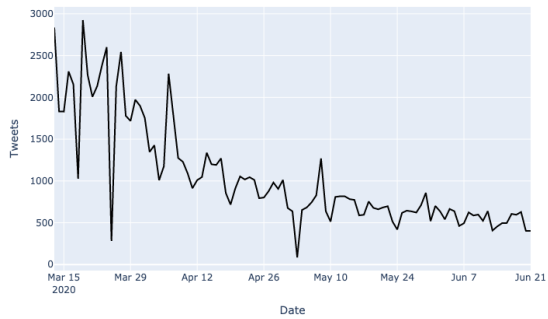
(p) Topic 16



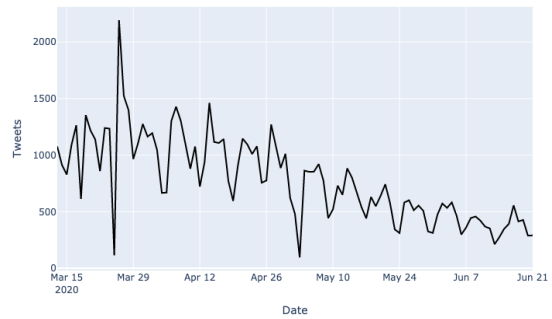
(q) Topic 17



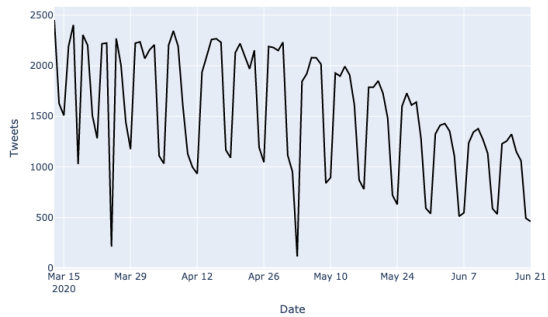
(r) Topic 18



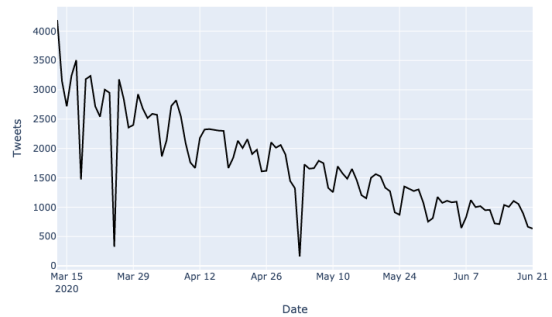
(s) Topic 19



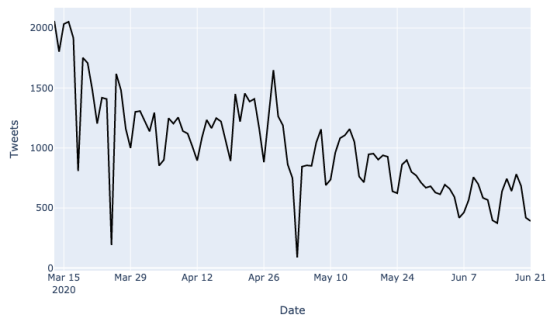
(t) Topic 20



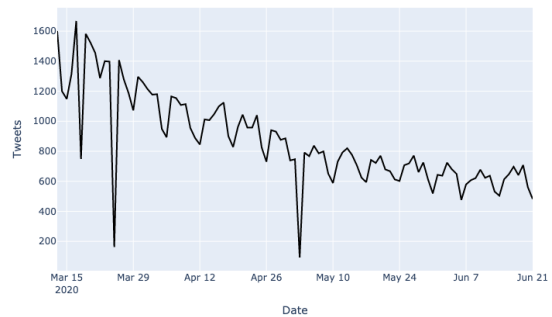
(u) Topic 21



(v) Topic 22

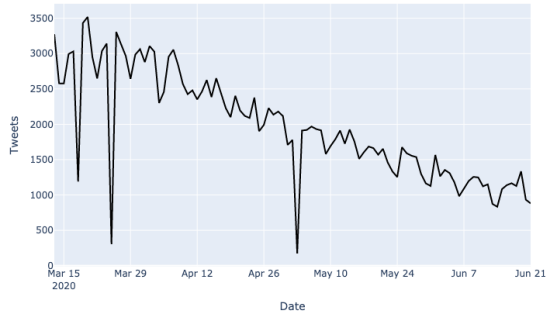


(w) Topic 23



(x) Topic 24

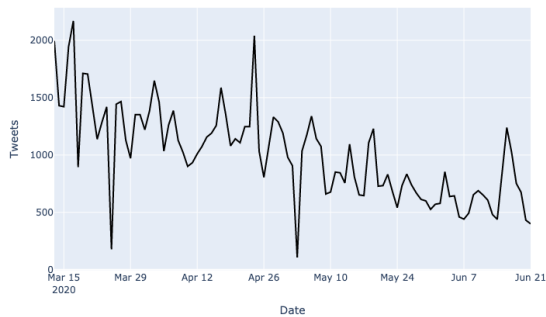
Figure 3.46: Timeline charts of the tweeting activity in each topic. II



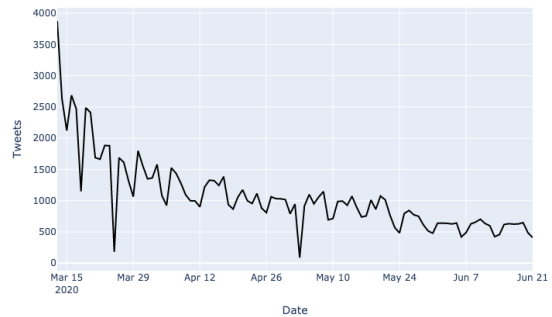
(a) Topic 25



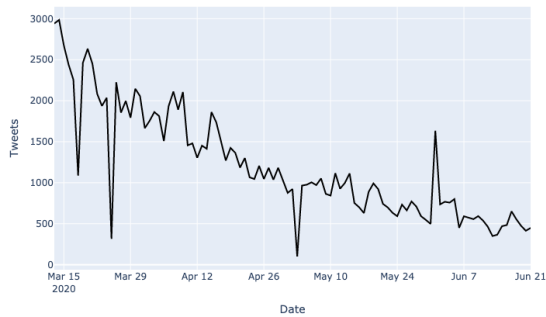
(b) Topic 26



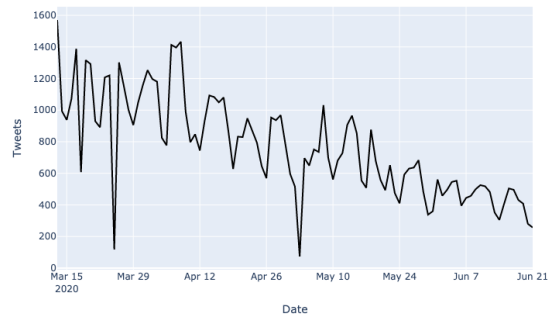
(c) Topic 27



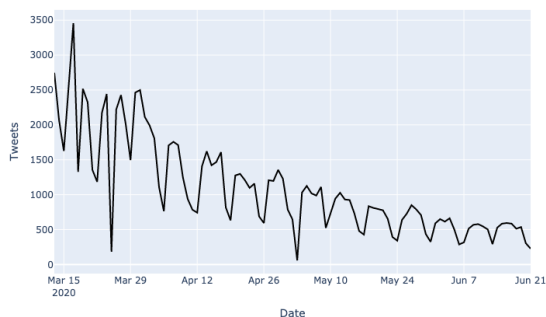
(d) Topic 28



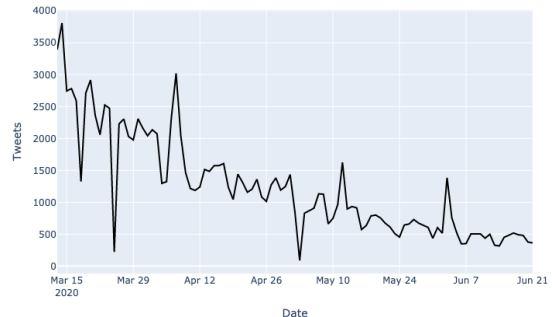
(e) Topic 29



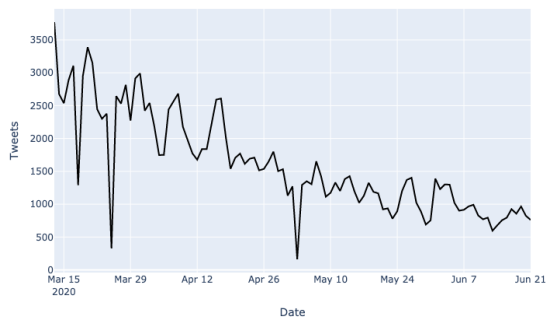
(f) Topic 30



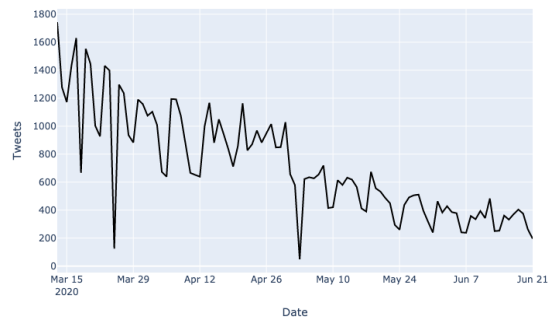
(g) Topic 31



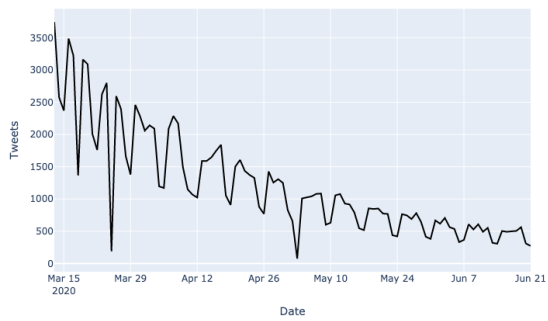
(h) Topic 32



(i) Topic 33



(j) Topic 34



(k) Topic 35

CHAPTER 4

Discussion

In this work, it is shown that with known methods and techniques, it is possible to get objective and relevant results of an analysis of the interests of users in social media, that helps to understand the progression of the views and interests of the user base throughout a time frame.

The techniques used in the study may have restricted a detailed view of certain aspects. It should be taken into consideration the limits and restrictions imposed on the English dataset because this has a significant impact on the data and the results.

The decision to use TWINT instead of Twitter's public API is entirely for economic purposes; it would have been inviable to pursue this project if TWINT or a similar alternative did not exist.

When analysing the users' location to geolocate their profiles and thus their tweets, the Spanish dataset performed well showing mostly accurate locations and helped understand further how the opinions and overall view of the crisis evolve through time. Maybe, using the same technique in the English dataset would have improved the located profiles, and could have had an impact on the generated models, as well as improving the geographical differences in opinion or interests depending on the location. After seeing the outcome, using Open Street Maps Nominatim to geocode the locations, offered good outcomes, and thus using it in the English dataset might have improved the results, but the time restrictions by the deadlines made this impossible.

Calculating the bot score with Botometer did not only make it easier to understand the dataset at first sight but when analysing the topics, since it allowed to make the assumptions that there were virtually no bots, or at least not as many as necessary to make an impact on the study results without realising. For example, in topic 17 in the Spanish dataset, the results show that the account @A_definitivas was spamming, it was easy to detect because the same tweet was posted many times. This decision is perhaps not the right one, but the time constraints paid a significant role in this decision, and using Botometer seemed a rational decision to understand the dataset's characteristics.

Due to the lack of previous specific knowledge in natural text processing, the decisions taken might not have been the best when preprocessing the dataset for further analysis. For example, in the Spanish dataset, after analysing the results, even though there was no noticeable drawback for not using such technique, it might have been interesting to see if removing any accents and special characters of the Spanish and other regional languages, would have shown any different results. Also, even though Stanford CoreNLP was used to lemmatise the Spanish text, no noticeable changes were seen; maybe it was not correctly configured or was incorrectly used.

The decision to change to a Gibbs sampling based LDA model, after analysing the results, is now well-founded because the Spanish dataset clearly shows the results achieved are more than acceptable and deliver a transparent view of the interests in differentiated topics even if some topics might not be completely intelligible. The original decision was entirely based on the time and computing restrictions imposed by the characteristics of the dataset and the delivery deadlines required by the Bachelor's Degree Final Project.

By using LDA and the most frequent words, very good results are laid if the dataset is big enough, as it is shown by the results in the Spanish dataset whilst in the English dataset, the results are affected due to its size restriction. In some topics of the latter dataset, it is difficult to understand the background and relation of the grouped tweets because the frequency of the words does not take into account its order, therefore if there are not enough close-related words, it becomes quite difficult to guess the conversation behind.

CHAPTER 5

Conclusions

From the results of the study, two main conclusions are made: in the English dataset, mainly national concerns and events are commented, whereas, in the Spanish dataset, events from all over the world concern the users, and thus they show on the tweets and words used.

The outcomes shown are useful when analysing the opinion and interests of Twitter users, as they react to news of the physical world as much as they do with those of the tweeting ecosystem: for example, the Twitter study regarding the bot network, the tweet from the Ministry of Foreign Affairs to the Belgian Minister or the death of *Billy El Niño* rampaging the social network.

The techniques used to gather and process the data have proven to be useful and to lay objective and useful insight from the data. Although, as stated in the discussion section, some improvements could be made, the results of the larger dataset are robust and what was expected when designing the strategy at the beginning.

Results show that to perform a study on the opinion and interests of users on Twitter, the bigger the dataset, the better as the technology used offers better results. Using a smaller dataset, will not lay bad results, simply put, the results will not offer detailed insight as in the former and may difficult the analysis.

After working on this project, another conclusion made is that when working with data, the more data gathered, the better; if it is not useful later, then it can be discarded, but that is better than regretting not being able to analyse data that was not stored.

The procedures and methods used in this study could be of interest to further studies in other areas. For example, in sociology, it could help better understand the opinion, interests and concerns of the population in diverse events or special situations; as shown in topic 12, with an extension to the procedures, in the field of psychology, this methodology could be of use to analyse how mental illnesses impact the user's behaviour on social media.

Some of the subjects taken in the bachelors that were most useful when making this project were Database Technologies (BDA), Data Structures (EDA), Data Storage and Retrieval (SAR) and Perception (PER). These subjects brought a broad perspective on how to approach this study: how to store the data permanently (Postgres), how to store it when processing it (chunks of the original files or database queries as RAM objects), and how to analyse it (by grouping it in clusters).

The Bachelor's End of Degree Project methodology guide states multiple skills that should be achieved after finishing the project. Looking upon the results and work in this project, it is considered that these are some of the skills that were achieved:

Ability to collect and interpret relevant data: explaining the results laid by the study, required prior comprehension as well as searching for pertinent data to explain them.

Critical, analytical and abstract thinking: in order to understand the models, and describe the results, these three skills were applied due to the need to extract objective and quantifiable facts.

Autonomous learning: even though known procedures were applied during the data processing and storage, new methods had to be learned, for example, most of the libraries used in the project were unknown and thus required previous training to be able to apply them.

Find relevant information from different sources and investigate technological developments: to be able to take on this project, functional studies and previous research had to be reviewed to understand the techniques applied in similar projects.

Apply professional knowledge following ethical principles, universal values and ethical principles: this work is carefully guided by all those principles, as the best was done to offer objective and non-biased results.

Bibliography

- [1] Wasim Ahmed et al. "COVID-19 and the 5G conspiracy theory: Social network analysis of twitter data". In: *Journal of Medical Internet Research* 22.5 (May 2020). ISSN: 14388871. DOI: [10.2196/19458](https://doi.org/10.2196/19458).
- [2] T. Alshaabi et al. "How the world's collective attention is being paid to a pandemic: COVID-19 related 1-gram time series for 24 languages on Twitter". In: (Mar. 2020). URL: <http://arxiv.org/abs/2003.12614>.
- [3] David Alvarez-Melis and Martin Saveski. *Topic Modeling in Twitter: Aggregating Tweets by Conversations*. Tech. rep. URL: <http://psiexp.ss.uci.edu/research/>.
- [4] Joshua Charles Campbell, Abram Hindle, and Eleni Stroulia. "Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data". In: *The Art and Science of Analyzing Software Data* 3 (2015), pp. 139–159. DOI: [10.1016/B978-0-12-411519-4.00006-9](https://doi.org/10.1016/B978-0-12-411519-4.00006-9).
- [5] Jonathan Chang et al. "Reading tea leaves: How humans interpret topic models". In: *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference* (2009), pp. 288–296.
- [6] Kia Jahanbin and Vahid Rahmanian. *Using twitter and web news mining to predict COVID-19 outbreak*. Aug. 2020. DOI: [10.4103/1995-7645.279651](https://doi.org/10.4103/1995-7645.279651). URL: <http://www.apjtm.org/article.asp?issn=1995-7645;year=2020;volume=13;issue=8;spage=378;epage=380;aulast=Jahanbin%20http://www.apjtm.org/article.asp?issn=1995-7645;year=2020;volume=13;issue=8;spage=378;epage=380;aulast=Jahanbin;type=0>.
- [7] Qiang Jipeng et al. "Short Text Topic Modeling Techniques, Applications, and Performance: A Survey". In: (Apr. 2019). URL: <http://arxiv.org/abs/1904.07695>.
- [8] Ramez Kouzy et al. "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter". In: *Cureus* 12.3 (Mar. 2020). ISSN: 2168-8184. DOI: [10.7759/cureus.7255](https://doi.org/10.7759/cureus.7255). URL: [/pmc/articles/PMC7152572/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7152572/](https://pmc/articles/PMC7152572/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7152572/).
- [9] Natt Leelawat et al. "Trends of Tweets on the Coronavirus Disease-2019 (COVID-19) Pandemic". In: *Journal of Disaster Research* 15.4 (June 2020), pp. 530–533. ISSN: 1883-8030. DOI: [10.20965/jdr.2020.p0530](https://doi.org/10.20965/jdr.2020.p0530). URL: <https://www.fujipress.jp/jdr/dr/dsstr001500040530>.
- [10] Christopher Manning et al. "The Stanford CoreNLP Natural Language Processing Toolkit". In: (2015), pp. 55–60. DOI: [10.3115/v1/p14-5010](https://doi.org/10.3115/v1/p14-5010).
- [11] Han Woo Park, Sejung Park, and Miyoung Chong. "Conversations and medical news frames on twitter: Infodemiological study on COVID-19 in South Korea". In: *Journal of Medical Internet Research* 22.5 (May 2020), e18897. ISSN: 14388871. DOI: [10.2196/18897](https://doi.org/10.2196/18897). URL: <https://www.jmir.org/2020/5/e18897/>.

- [12] Sohaib R Rufai and Catey Bunce. "World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis". In: *Journal of public health (Oxford, England)* (2020), pp. 1–7. ISSN: 17413850. DOI: [10 . 1093 / pubmed / fdaa049](https://doi.org/10.1093/pubmed/fdaa049). URL: <https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa049/5822639>.
- [13] P Singh et al. "Psychological fear and anxiety caused by COVID-19: Insights from Twitter analytics". In: *covid19.elsevierpure.com* (). URL: <https://covid19.elsevierpure.com/en/publications/psychological-fear-and-anxiety-caused-by-covid-19-insights-from-t>.
- [14] Mike Thelwall and Saheeda Thelwall. "Covid-19 tweeting in English: Gender differences". In: *El Profesional de la Información* 29.3 (May 2020). ISSN: 1699-2407. DOI: [10 . 3145 / epi . 2020 . may . 01](https://doi.org/10.3145/epi.2020.may.01). URL: <https://recyt.fecyt.es/index.php/EPI/article/view/epi.2020.may.01>.
- [15] Stefania Vicari and Maria Francesca Murru. "One Platform, a Thousand Worlds: On Twitter Irony in the Early Response to the COVID-19 Pandemic in Italy". In: *Social Media + Society* 6.3 (July 2020), p. 205630512094825. ISSN: 2056-3051. DOI: [10 . 1177 / 2056305120948254](https://doi.org/10.1177/2056305120948254). URL: <http://journals.sagepub.com/doi/10.1177/2056305120948254>.
- [16] Kai-Cheng Yang, Christopher Torres-Lugo, and Filippo Menczer. "Prevalence of Low-Credibility Information on Twitter During the COVID-19 Outbreak". In: (Apr. 2020). DOI: [10 . 36190 / 2020 . 16](https://doi.org/10.36190/2020.16). URL: <http://arxiv.org/abs/2004.14484><http://dx.doi.org/10.36190/2020.16><http://arxiv.org/abs/2004.14484><http://dx.doi.org/10.36190/2020.16>.

APPENDIX A

Supplementary information

The entire set of models generated can be seen on the Bachelor's End of Degree Project's repository on [Github](#).