




Article

# How Dramatic Events Can Affect Emotionality in Social Posting: The Impact of COVID-19 on Reddit

Valerio Basile <sup>1</sup>, Francesco Cauteruccio <sup>2,\*</sup> and Giorgio Terracina <sup>2</sup>

<sup>1</sup> Department of Computer Science, University of Turin, I10124 Torino, Italy; basile@di.unito.it

<sup>2</sup> Department of Mathematics and Computer Science, University of Calabria, I87036 Rende, Italy; terracina@mat.unical.it

\* Correspondence: cauteruccio@mat.unical.it

**Abstract:** The COVID-19 outbreak impacted almost all the aspects of ordinary life. In this context, social networks quickly started playing the role of a sounding board for the content produced by people. Studying how dramatic events affect the way people interact with each other and react to poorly known situations is recognized as a relevant research task. Since automatically identifying country-based COVID-19 social posts on generalized social networks, like Twitter and Facebook, is a difficult task, in this work we concentrate on Reddit megathreads, which provide a unique opportunity to study focused reactions of people by both topic and country. We analyze specific reactions and we compare them with a “normal” period, not affected by the pandemic; in particular, we consider structural variations in social posting behavior, emotional reactions under the Plutchik model of basic emotions, and emotional reactions under unconventional emotions, such as skepticism, particularly relevant in the COVID-19 context.

**Keywords:** COVID-19; emotion analysis; plutchik model; reddit; skepticism



**Citation:** Basile, V.; Cauteruccio, F.; Terracina, G. How Dramatic Events Can Affect Emotionality in Social Posting: The Impact of COVID-19 on Reddit. *Future Internet* **2021**, *13*, 29. <https://doi.org/10.3390/fi13020029>

Academic Editor: Eirini Eleni Tsiropoulou  
Received: 15 December 2020  
Accepted: 23 January 2021  
Published: 27 January 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Online information sharing and social posting are becoming more and more a common way people use to interface with each other and to gather information. Dramatic events like natural disasters, wars and pandemics affect the way people seek information, interact and express their feelings [1–3].

Studying users’ reactions to specific events is a particularly relevant task in order to understand latent feelings and, consequently, help governments and managers in adopting the best policies in order to address people satisfaction and safety. However, identifying topics from extensive text repositories, or at least identifying resources related to a specific topic, is still recognized as a difficult research task [4,5]; consequently, any analysis of user reactions to specific topics at large scale may be flawed by dirty or missing data. The recent pandemic spread of COVID-19 disease [6] impacted millions of people all around the world. At the time of writing this manuscript, almost 19 million confirmed cases have been registered worldwide, with 208 countries, areas or territories showing cases. Governments around the globe have been forced, in one way or another, to implement protective measures such as lockdowns, social distancing and travel bans to contain the spread of the disease. This resulted in more and more people increasing their online experiences of accessing information and sharing feelings. Among others, the social platform Reddit (<https://www.reddit.com/>) experienced a tremendous increase in new users and posts (<https://www.redditinc.com/>), attracting the interest of the academic community [7–9]. Reddit is a heterogeneous crowdsourced news aggregator and social platform, and at the time of writing, it is ranked 18th in the Alexa’s top 500 global websites. Reddit comprises subreddits, a set of interest-based communities where users can submit content and comments. Currently, there are more than 2.2M subreddits.

In the context of this paper, a particularly interesting feature of Reddit is the concept of a megathread, defined as a user-initiated discussion referring to a particular context or situation. Given this trait, it is fair to assume that each comment within a megathread relates to the specific topic. Thus, from this point of view, COVID-19 megathreads provide a unique opportunity to carry out clean topic-specific analyses of users' attitudes in social posting, and to point out possible changes in their stances due to variations in external factors. A further particularly interesting feature of megathreads is the fact that they can be specific for countries or areas, allowing to precisely analyze users' comments in relation to specific country-based events.

As a matter of fact, various countries approached COVID-19 with very different policies [10], and, consequently, lives of people in these countries have been affected in different ways by the COVID-19 pandemic.

In this paper we are interested in studying and analyzing if and how COVID-19 impacted users in their social posting habits and attitudes. In particular, we consider two very different aspects: structural properties of posts and emotionality expressed by the users within their posts. Moreover, we carry out our analysis by considering specific countries or areas separately, allowing us to obtain a multilingual and multicultural analysis.

The choice of countries and areas to consider has been guided by different factors. First of all, by the availability of data, we chose those countries and areas showing COVID-19 megathreads with a sufficient number of posts and users. Moreover, we considered the different policies applied by governments to face the COVID-19 outbreak, in order to build a diversified but representative group of countries. Finally, we considered those megathreads covering a sufficiently large time span to analyze temporal variations. In order to avoid bias in the analysis, we did not focus on English-based posts only, but we decided to consider megathreads using country-specific languages. It is worth noting that, in this work, we did not consider megathreads from China, since there are no COVID-19 specific megathreads from this country.

We focused on five European countries where pandemic expansion evolved first, after China. In particular, we considered megathreads from the following countries: Italy and Germany, since they are the countries whose governments operated early and firmly acted against the diffusion of COVID-19; Sweden, since it adopted a completely opposite policy to Italy and Germany; The Netherlands, since it approached the COVID-19 pandemic slightly softer than Italy and Germany; and the United Kingdom, which had a late reaction to the emergency, compared to the other countries. Moreover, besides these European countries, we considered also the COVID-19 megathreads of New York City, since this area was the first part of the United States affected by the pandemic; analyzing megathreads from New York City allows us to provide a view of the data outside Europe. In the rest of the paper, in order to simplify the presentation, with a slight abuse of terminology we refer to data from New York City as data from a country.

In summary, the main contribution of this paper are as follows:

- We provide a comparative view of country-based data about the COVID-19 pandemic relating social events, like evolution of new daily cases and government actions, to online social posting;
- We analyze differences and similarities in users' habits and attitudes in online social activities under a "normal" situation and under the COVID-19 emergency;
- We conduct a multilingual and country-based emotionality analysis of users, based on their posts, in order to detect latent feelings related both to the emergency and to the policies adopted by single countries relating them also to a "normal" situation. In particular, we focus on the eight basic emotions of the model by [11];
- We introduce a novel method to analyze skepticism, a non-basic emotion we consider particularly relevant for the analysis of COVID-19 related posts.

It is worth pointing out that the present work advances current research in several aspects.

First of all, most studies on similar topics use Twitter, or analogous social networks, as a data source; these need an automatic topic identification step as data pre-processing, which may easily introduce noise and errors in the data sources. Choosing Reddit, and specifically megathreads, allows us to focus on a well-defined, topic-based dataset; moreover, the choice of specific megathreads allows us to conduct a country-based analysis, which is quite impossible, or at least particularly difficult, with other social networks. Another innovative point of our research refers to the multilingual analysis; as a matter of fact, most related studies focus on content from a single language. Clearly, considering the original language from each country allows us to conduct a more precise analysis, to isolate cultural peculiarities and to include a wider audience of users in the analysis.

From a more technical point of view, our comparison on the structure of posts and comments in a COVID and a non-COVID period provides important insights in social posting behavior. In particular, if our intuition that posting behavior is actually affected by real-life events is confirmed, all existing approaches for user behavior classification and content recommendation should be revisited in order to take into consideration events happening outside the social network that may bias user behavior and trick automatic tools for behavior analysis.

Similarly, comparing the emotional expression in a COVID and a non-COVID period, based on countries, is a non-trivial task that allows us to reveal the importance of emotionality aspects in social posting, differentiated by country; again, if our intuition is confirmed, the present work can impact future approaches dealing with recommendations and social behavior analysis.

Last, but not least, to the best of our knowledge, no related approaches include the analysis of non-conventional emotions like skepticism. The method introduced in the present paper to deal with this emotion can be easily generalized to other ones, thus paving the way to new perspectives for emotion analysis.

In summary, understanding how social users react to dramatic events, like the COVID-19 pandemic, and the corresponding real-life events is already a research topic of interest per se; however, the outcomes of the present study may guide data managers and software developers to better cope with people interests and satisfaction even when they face extreme and non-standard real-life situations [12]. Moreover, the analyses conducted in the present paper can be well supported by automated tools; these can be put side by side with standard, but more invasive, mechanisms like questionnaires in order to help governors in monitoring a person's stance about a specific topic. In particular, automated dashboards from an emotionality analysis on social networks may overcome the intrinsic latency of questionnaires, providing timely information as soon as new data (comments) are produced.

Clearly this work is not intended to provide the ultimate analysis on COVID-19 related (online) social behavior; however, data collection and analyses provided in this work aim at yielding a wide spectrum of COVID-19 correlated information and considerations that could form the basis for more extended analyses.

The rest of the paper is organized as follows. In Section 2 we analyze related literature and point out similarities and differences with our work. Section 3 explains the methodology and the steps behind our study. Section 4 is devoted to describe the data collection phase we carried out and to specify the considered set of data, whereas in Section 5 we present methods and measures we adopted to conduct the analyses. In Section 6 we present the results of our analyses, and main findings are discussed in Section 7. Finally, in Section 8 we draw our conclusions.

## 2. Related Work

With the COVID-19 epidemic situation impacting over almost all the aspects of ordinary life, social networks quickly started playing the role of a sounding board for the content produced by people. Indeed, in a context such as the COVID-19 public health emergency, it has been shown how the social media usage increased, both in a positive

and negative way. Studying social media is interesting from several points of view [13,14]; in this particular case they have been used both as a source of information about COVID-19, and for disseminating conspiracy beliefs and theories [15]. Thus, studying social posting within a health emergency period represents an interesting research context. Moreover, the study of social networks either during or after a critical event, or a disaster, proved to be a rich source of interesting insights [12,16–18].

Starting from the very early days of COVID-19 diffusion, a plethora of different studies started to appear within the academic community, and most of them are available as online preprints. Thanks to the interdisciplinarity of the context, the studies regarding social networks and COVID-19 situation are variegated and stem from different disciplines, such as psychology [19], behavioral sciences [20] and medicine [21,22]. Among all of these studies, few of them consider the context through the lens of social network analysis facets. In this section, we mostly concentrate on these studies, analyzing and highlighting their contributions and differences with our work. Before going into the detail, we point out that, to the best of our knowledge, this is the first work studying emotionality in social posting, within Reddit, in the COVID-19 context. Nevertheless, different studies revolving around social network analysis, either on Reddit or other platforms, have been presented [23–28].

In [23], the Reddit response to COVID-19 has been characterized through the analysis of /r/China\_Flu and /r/Coronavirus. The authors studied these communities from three points of view: community genealogy, moderation and user activity. In particular, they concentrated on how these communities emerged and how they differed during the pandemic; furthermore, they also studied how people change their activity during the pandemic and how it does relate to being members of these two communities. The dataset used comprises the submissions and the comments posted on these two communities, starting from 20 January to 24 May 2020. Due to the very different aims, our proposed work is not directly comparable to the one in [23]. Even if a process to characterize language usage is also present in [23], they focused on the language distance between the two communities without considering any form of emotional or sentimental properties.

A COVID-19 social media analysis is presented in [24]. Using the term infodemic, the authors present an analysis embracing different platforms: Twitter, Instagram, Youtube, Reddit and Gab. A comparative analysis of information spreading has been carried out on a dataset including content related to COVID-19, where the relation is given by the presence of terms such as coronavirus, coronavirus outbreak, etc. In this case, no sentiment or emotion analysis was carried out, although Natural Language Processing (NLP) techniques were used to extract and analyze topics related to the COVID-19 content; in the present paper, we exploit Reddit megathreads to specifically locate COVID-19 related content.

An interesting study about media coverage and online collective attention to COVID-19 is presented in [25], where a heterogeneous dataset comprising online news articles, videos and social content was used. This dataset also comprises submissions and comments from the /r/Coronavirussubreddit. The authors studied the impact of media coverage and epidemic progression on collective attention, showing that collective attention was mainly driven by media coverage rather than epidemic progression. They illustrated how Reddit users were generally more interested in health and data regarding the disease by employing a topic modeling approach. The /r/Coronavirus subreddit was also studied in [26], in which a longitudinal topic modeling was employed to measure daily changes in the frequency of topics of discussion across COVID-19 related comments.

An observational study presented in [27] employed NLP techniques in order to reveal vulnerable mental health support groups on Reddit during COVID-19. In particular, the authors characterized changes in different mental health support subreddits, such as /r/Depression and /r/SuicideWatch, finding that a particular subreddit, /r/HealthAnxiety, showed spikes in posts about COVID-19 early on in January. Sentiment properties of comments are used as features, although they are limited to the well-known negative, positive and neutral values. The work in [28] focused on the /r/COVID19Positive subreddit, which is a subreddit where COVID-19 positive patients provide insights into their personal strug-

gles with the virus. The aim of the work was to quantify the changes in discussions about COVID-19 personal experiences for the first 14 days since symptom onset. Here, the authors exploited both topic modeling and sentiment analysis. In particular, similarly to our work, the NRC sentiment lexicon was used in order to analyze the emotional content of the text. However, there are several differences with our proposed work; these are discussed next. First of all, the sentiment analysis in [28] was only applied to a specific subreddit whose content is exclusively expressed using the English language; also, the size of the used data, consisting of 609 submissions only, imposes some limitations. Furthermore, the study only focused on experiences from COVID-19 positive patients, thus providing a non-holistic point of view of the phenomenon. On the contrary, our proposed study covers a multilingual analysis of comments from different countries, also comparing emotionality both in the COVID and non-COVID period.

The COVID-19 situation inspired researchers to conduct studies also on other popular platforms, different from Reddit, such as Twitter, Facebook, and Weibo [29–35]. In [29], the affective trajectories of users based on Twitter and Weibo were analyzed. In particular, the authors were interested in analyzing the evolution of public emotions. For the emotion classification task, they considered six major categories, i.e., anger, disgust, worry, happiness, sadness and surprise. Furthermore, they presented an approach to extract triggers for different emotions, i.e., why people are angry or disgusted. The usage of sentiment analysis is clearly similar to the one in our proposed work. However, substantial differences can be highlighted. The work in [29] was based on a heterogeneous dataset built from Twitter and Weibo, containing very different kinds of posts and considering only two different languages: English and Chinese. In our work, using megathreads, we are able to focus on a more homogeneous dataset, strongly related to COVID-19, which can be also categorized by country, and which includes English, Dutch, German, Italian and Swedish languages. The work presented in [35] was also based on Twitter. Here, the authors focused on determining emotional polarity expressed by users on Twitter, collecting comments, hashtags, posts and tweets by looking for keywords such as covid and coronavirus. Then, a recurrent neural network model was trained to classify emotional polarity in four different classes (weakly positive/negative, and strongly positive/negative). The work in [35] is not directly comparable to our proposed study, in which we focus on emotions rather than polarities. Additionally, the presence of megathreads in our study allows to collect data strongly related to the topic; also, Reddit comments allow more characters to be used, thus providing a substantially different kind of data. Moreover, although the works above analyzed some common emotions, we also propose a technique for analyzing skepticism, a non-canonical emotion that plays a relevant role in the COVID-19 context.

In the work presented in [30], the authors studied how the COVID-19 pandemic affected self-disclosure on social media, presenting a research agenda with different research questions to better characterize and examine pandemic-related self-disclosures. In [31], Twitter and 4chan served as sources of information for studying the emergence of Sinophobic behavior on social platforms during the COVID-19 pandemic, employing techniques such as word embeddings and graph analysis. In [32], a ground truth dataset of emotional responses to COVID-19 was presented. Here, several participants were asked to indicate their emotions and to express them in text. The result comprised 5000 texts, on which different analyses were performed. An interesting result is the one highlighting that a weak, positive correlation existed between self-reported emotions and psycholinguistic word lists that measure these emotions. The work in [32] is not directly comparable to our proposed approach, considering that the context therein completely differs from a social posting context. Although unrelated to COVID-19, social media analytics are also exploited on Twitter for studying pre- and post-launch emotions in new product development [36].

Finally, some works consider either Facebook or Weibo [33,34] in order to analyze different aspects of these platforms, such as the correlation of geographic spread of COVID-19 with the structure of the social networks [33], and the characterization of the propagation of situational information [34].



To provide a better overview of the related work and to better position the present work in the literature, Table 1 shows a comparison of our proposed study with the related ones along several comparison directions, namely (i) if the source of the analysis is single or multiple, (ii) how the analyzed topic has been identified, (iii) if the study considers different countries, (iv) if the study considers multiple languages, (v) if the study carries out a structural analysis, (vi) if the study carries out an emotional analysis and (vii) if the study includes a COVID vs non-COVID comparison.

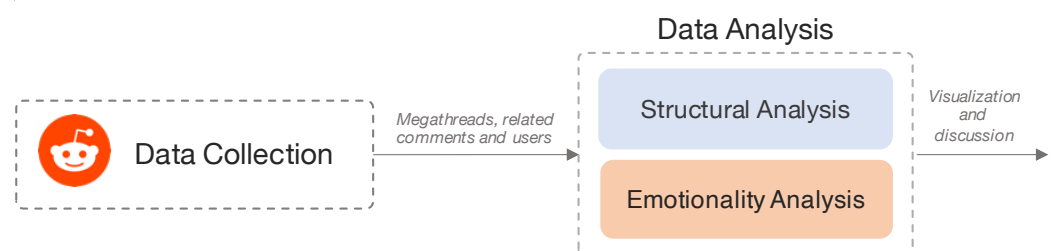
**Table 1.** Comparison properties overview of our study and most related ones; cells with N/A indicate absence of that corresponding property.

	Source	Topic Identification	Country Based	Multi-Lingual	Structural Analysis	Emotional Analysis	COVID vs. Non-COVID Comparison
Our study	Single	By megathreads	✓	✓	✓	✓	✓
[23]	Single	By subreddits	-	-	✓	-	-
[24]	Multiple	By terms	-	-	✓	-	-
[25]	Multiple	By subreddits and terms	✓	✓	✓	-	-
[26]	Single	By subreddits	-	-	-	-	-
[27]	Single	By subreddits	-	-	✓	✓	-
[28]	Single	By subreddits	-	-	-	✓	-
[29]	Multiple	By terms	✓	✓	-	✓	-
[35]	Single	By terms	-	-	-	✓	-
[31]	Multiple	By terms	-	-	✓	-	✓
[32]	N/A	N/A	-	-	-	✓	-
[33]	Single	N/A	✓	-	✓	-	-
[34]	Single	By terms	-	-	✓	-	-

In summary, COVID-19 ignited a significant amount of research studying users' reactions to the pandemic from very different points of view. In this work we address this problem from a different point of view and methodology with respect to existing literature.

### 3. Overview of the Methodology

To better understand the methodology used in our study, in this section we provide the workflow representing the tasks it consists of. The workflow is visually depicted as a pipeline in Figure 1.



**Figure 1.** The workflow representing the tasks of our investigation (Reddit logo from <https://www.redditinc.com/brand>).

The pipeline consists of two core modules, namely (i) data collection and (ii) data analysis. The former includes the process devoted to collect the data from Reddit. Starting from a list of subreddits  $S$  and a time span  $T$ , the data collection module retrieves the megathreads posted within the time range  $T$  in each subreddit  $s \in S$ , along with their related comments and users. The module is detailed further in Section 4. The collected data are then used as output for the data analysis module. It consists of two components representing the main analyses of our proposed study, which are structural analysis and emotionality analysis, and are discussed in detail in Sections 5.1 and 5.2, respectively. The output of the data analysis module can be then used in several contexts. In our study, we visualized the results by making use of several charts and discuss them. Nevertheless,

the pipeline is flexible enough to accommodate further captivating features. As an example, one could use the output of the data analysis module to create a dashboard where the interested user can analyze the data in real time.

#### 4. Data Collection

In this section we describe the data collection phase we conducted for this work, and we present the datasets we considered for the analyses. All collected data are available for download at <https://bitbucket.org/cauteruccio/reddit-dataset>. Since in this study we focused on the effect of COVID-19 on social posting, the adopted dataset should provide all of the content produced during the time span of the corresponding epidemic. This dataset should be homogeneous, in the sense that all of the content should directly refer to the COVID-19 epidemic. Moreover, the dataset should take into account the diversity in space and time: people in a country mildly affected by the epidemic might show a different behavior in social posting with respect to people in a severely affected country. As pointed out in the Introduction, in order to cope with these requirements, we focused on Reddit megathreads since they refer to particular contexts or situations, and they can be country-specific. Moreover, it is worth noting that subreddits are moderated; this ensures that only pertinent comments are included in the discussion.

In order to build the dataset, we started by searching for country-related subreddits following the wiki page of `/r/ListOfSubreddits`; this page provides a summary of several subreddits categorized from different points of view, including geography. Then, for each country-related subreddit, we manually identified each subreddit containing COVID-19 related megathreads, and we considered the time span starting from 1 January and ending to 30 June 2020.

We focused on the following subreddits, namely `/r/italy`, `/r/de`, `/r/thenetherlands`, `/r/sweden`, `/r/unitedkingdom` and `/r/nyc`. In particular, our choice was based on the fact that all of these subreddits present daily or weekly COVID-19 related megathreads. Note how each subreddit represents a country, except for `/r/nyc`; as pointed out in the Introduction, with a little abuse of terminology we refer to data from New York City as data from a country. It is important to point out that, to the best of our knowledge, a COVID-19 megathread initiative regarding the United States in its completeness is not present on Reddit; thus, we only considered the `/r/nyc` subreddit due to the presence of the megathreads in it and due to the fact that New York City was one of the first places in the United States to experience the COVID-19 epidemic; as a consequence, it provides more data in the considered time span.

For each of these selected subreddits, we retrieved all of the COVID-19 related megathreads. Megathreads are simply recognized by the title; in fact, each COVID-19 related megathread within a subreddit always reports the same title, in which only the date changes (as an example, the title “Megathread Coronavirus \* 16/05/20” represents a COVID-19 related megathread in the `/r/italy` subreddit). Then, for each of them we retrieved all of the available comments, and we discarded all the comments marked as deleted. For each megathread, we stored the author, i.e., the user who posted it, the title, the posting date and time, the score and the number of comments. For each comment, we stored the author, the content of the comment, the posting date and time, the reference to the parent entity and the score. We retrieved all the data using pushshift.io service ([37]).

Table 2 reports a summary of the main statistics for each considered subreddit. `/r/italy` and `/r/thenetherlands` contain the highest and the lowest number of megathreads, respectively. Considering the early outbreak of COVID-19 cases in Italy, this number is somewhat expected. The lowest number of megathreads in `/r/thenetherlands` is explained by the fact that they are weekly; this means that each new megathread is created at the start of the week. Moreover, it is worth pointing out that data for January and February 2020 were available only for Italy, since megathreads for the other countries started in March 2020.

In order to understand whether the COVID-19 situation really impacted the social posting behavior, we carried out a comparison between the social posting in the COVID-19

temporal range, and the social posting in a temporal range unrelated with COVID-19. In order to build this reference dataset we used Reddit data from 2019. In particular, for each country-related subreddit, we collected the whole set of submissions and comments covering the same months considered for the COVID-19 dataset, but from 2019. For instance, the COVID-19 megathreads for /r/italy covered a temporal range between January 2020 and June 2020; thus, we collected submissions and comments from /r/italy between January 2019 and June 2019. Table 3 provides an overview of collected data. A quick glance at it illustrates some differences in values with respect to data presented in Table 2; this is expected, considering that data from 2019 contain the whole number of submissions and comments made in that period and in that country, while the COVID-19 data focus only on COVID-19 megathreads and related comments in the corresponding country.

**Table 2.** Total number of megathreads, comments and distinct users for each subreddit related to COVID-19 (in the following, 2020 dataset or COVID dataset).

Subreddit	# Megathreads	# Comments	# Distinct Users
/r/italy	133	173,536	5673
/r/de	94	92,645	5815
/r/thenetherlands	19	54,756	4394
/r/sweden	113	74,015	3204
/r/unitedkingdom	121	111,317	6229
/r/nyc	117	95,532	5960

**Table 3.** Total number of submissions, comments and distinct users for each subreddit in 2019 (in the following, 2019 dataset or non-COVID dataset).

Subreddit	# Submissions	# Comments	# Distinct Users
/r/italy	12,396	661,707	13,068
/r/de	24,530	601,945	29,090
/r/thenetherlands	7677	156,731	15,111
/r/sweden	13,718	193,496	21,225
/r/unitedkingdom	15,982	489,272	30,650
/r/nyc	8780	161,391	19,256

In order to distinguish between these two datasets, in the following we refer to the first as the COVID dataset (or 2020 dataset) and to the latter as the non-COVID datasets (or 2019 dataset).

Finally, in order to correlate Reddit data with external factors possibly influencing them, with specific attention to government reactions to the COVID-19 emergency, we collected the main events and government actions strictly related to COVID-19. Tables 4–9 show a summary of the main timeline of events and policies adopted by the governments of the six countries we considered in this paper. The data in these Tables have been retrieved from the Wikipedia pages related to the pandemic situation in each country (an example is [https://en.wikipedia.org/wiki/COVID-19\\_pandemic\\_in\\_Italy](https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Italy)) and integrated with data from different other country-related sources (such as newspapers, official government announces, etc).



**Table 4.** Events and policies timeline related to the COVID-19 epidemic situation in Italy.

Date (dd/mm)	Event
30/01	First 2 confirmed cases.
31/01	Prime Minister declared state of emergency.
20/02	First outbreak in a northern region, 16 confirmed cases and first victims.
23/02	Lockdown for municipalities containing COVID-19 clusters.
25/02	First decree related to the epidemic situation (DPCM).
04/03	Closing of schools and universities.
07–08/03	DPCM to lock down northern regions.
11/03	A new DPCM extends the lockdown to all national territory.
21/03	Closing of all unessential activities.
26/04	Prime Minister announces a new DPCM lifting restrictive measures, starting from 04/05.
16/05	Starting of “phase two” (re-opening of unessential activities, households can travel within regions, etc.)
03/06	Travels between regions are allowed.

**Table 5.** Events and policies timeline related to the COVID-19 epidemic situation in Germany.

Date (dd/mm)	Event
27/01	First confirmed case.
25/02	Clusters related to Italy outbreaks are discovered.
09/03	First victim.
11/03	Chancellor of Germany announces monetary aids for companies.
13/03	Protection stage: closing of schools, visiting is not allowed, etc.
22/03	Curfew imposed over 6 states.
29/03	Finance Minister commits suicide.
20/04	Partial re-opening of stores.
21/04	Oktoberfest is cancelled.
30/04	Partial re-opening of museums, monuments, etc, under social distancing.
01/05	German Trade Union Confederation cancels its annual event after 70 years.
20/05	New outbreaks at several meat processing plants.
15/06	Travels within EU are now allowed.

**Table 6.** Events and policies timeline related to the COVID-19 epidemic situation in The Netherlands.

Date (dd/mm)	Event
27/02	Presence of the virus confirmed.
06/03	First victim.
10/03	Tighter restrictive measures in North Brabant, events cancelled.
12/03	Online classes for universities, event with more than 100 people are not allowed.
16/03	Prime Minister addresses to the nation since the oil crisis of 1973.
20/03	King of the Netherlands addresses to the nation.
31/03	All of the restrictive measures are extended to 28/04.
22/04	All of the restrictive measures are extended to 20/05.
11/05	Re-opening of indoor pools and “contact jobs”.
01/06	Restaurants, hotels and bars can open after 12:00.

**Table 7.** Events and policies timeline related to the COVID-19 epidemic situation in Sweden.

Date (dd/mm)	Event
31/01	Presence of the virus confirmed.
26/02	Clusters related to Italy outbreaks appear.
09/03	Community transmission of the virus is confirmed.
10/03	Public Health Agency suggests symptomatic people to avoid social contacts.
11/03	First victim; gathering of more than 500 people are not allowed.
18/03	Public Health Agency suggests not to travel within the country.
27/03	Gatherings of more than 50 people are not allowed.
13/05	The restrictions on domestic travel are softened.
04/06	Restriction on domestic travel to be lifted on 13/06.

**Table 8.** Events and policies timeline related to the COVID-19 epidemic situation in the UK.

Date (dd/mm)	Event
31/01	First 2 confirmed cases.
06/02	Third confirmed case.
28/02	First victim.
03/03	An action plan toward the COVID-19 situation is presented.
10/03	Health Minister tested as positive.
13/03	First victim in Scotland and Premier League suspended.
14/03	Prime Minister speaks about “heard immunity”.
20/03	Closing of schools, restaurants, bars, coffees, etc.
23/03	Prime Minister announces a partial lockdown starting from 26/03.
24/03	United Kingdom’s mobile networks send a broadcast message.
25/03	Prince of Wales tested as positive; first two doctors die from COVID-19.
26/03	Lockdown is now effective.
27/03	Prime Minister tested as positive.
05/04	Queen of the UK and the other Commonwealth realms addresses to the nation.
06/04	Prime Minister is now in intensive care.
11/04	Easter message from the Queen.
22/04	Some lockdown restrictive rules are now lifted.
28/05	Announcement of less restrictive rules starting from 01/06.
15/06	General re-opening of shops, etc.

**Table 9.** Events and policies timeline related to the COVID-19 epidemic situation in New York City.

Date (dd/mm)	Event
01/03	First case.
03/03	Governor of New York confirms the first case of person-to-person spread.
20/03	Executive order to close all of unessential business.
30/03	The USNS Comfort Hospital ship arrives in New York Harbor.
31/03	Fist child victim.
28/05	Public officials express concern about the spreading in crowded events, e.g., George Floyd protest.
08/06	Partial re-opening of different stores.

## 5. Methods for Data Analysis

In this section we describe methods and measures adopted to analyze collected posts both from the structural perspective and from the emotionality perspective. These methods and measures will be the reference for our analysis in the next sections.

### 5.1. Methods for Structural Analysis

In order to carry out the structural-based analysis, few preprocessing steps were introduced on the COVID dataset. First of all, comments having a deleted author were maintained in the dataset: in fact, in some cases the text content of a comment remains available even if the author deactivates her/his account. Moreover, in order to extract URLs present in comments we applied classic regular expressions. Additionally, comments from bots duplicating URLs (such as VideoLinkBot (<https://www.reddit.com/user/VideoLinkBot>)) were ignored, thus avoiding counting a URL more than once. Then, for each megathread  $m$  we focused on the comments belonging to  $m$  which were made in the same day (week in the case of /r/thenetherlands) of submission in the megathread. Finally, in order to carry out the structural analyses, we focused on the number of submissions, the number of comments and the corresponding scores.

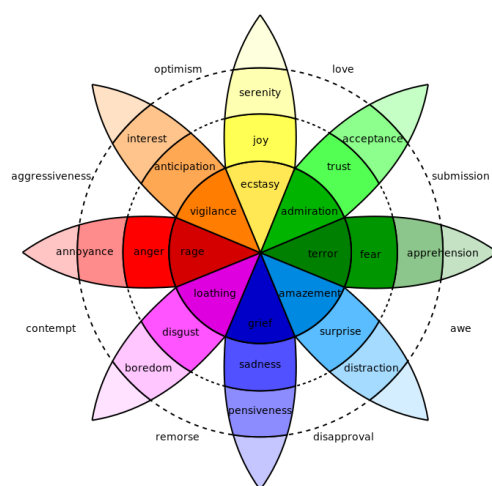
### 5.2. Methods for Emotionality Analysis

In order to complement the structural analysis, we employed a series of Natural Language Processing methodologies. Given the multilingual nature of the data, the sheer amount, and the particular linguistic characteristics (e.g., topics, tone, informal register), we relied on fully automated and unsupervised techniques to analyze natural language content.

As a first step, the text of each comment was processed with the NLP pipeline implemented in the SpaCy library for Python (<https://spacy.io/>). The texts were tok-

enized and lemmatized, and the stop words were removed. SpaCy is distributed with a set of pre-trained models for several languages, including all the languages treated in the present work except for Swedish. For this language, we performed NLP processing with the UDpipe tools ([38]), in particular using the wrapper provided for SpaCy (<https://github.com/TakeLab/spacy-udpipe>), and the Talbanken model included in the Universal Dependency repository ([39]).

In order to analyze the emotional content of the natural language expressed in the Reddit comments, we employed the NRC Emotion Intensity Lexicon (NRC-EIL) ([40]). This lexicon is a dictionary of 9922 lemmas associated with scores reflecting the emotions they convey and their intensity. The emotions in the lexicon, and therefore those considered in the present work, are the eight basic emotions of the model by Plutchik [11]: anger, fear, disgust, sadness, surprise, trust, anticipation and joy (see Figure 2). The NRC lexicon has been created by crowdsourcing judgements on the emotional content of the English words and then automatically translated in over 100 languages, including all the languages considered in this work. The scores for each emotion in the NRC lexicon range from 0 to 1.



**Figure 2.** The Wheel of Emotions by Robert Plutchik.

After a Reddit comment was tokenized and lemmatized, we looked up its constituent lemmas in the NRC lexicon, thus associating an 8-dimensional vector of 0–1 scores to each lemma. For each emotion, the sum of the scores in a comment was then considered as the final value for that comment with respect to the specific emotion. Since comparing the specific scores among emotions might be misleading and biased by specific languages, in our analysis we did not use them directly, but we measured the percentage of comments showing a specific emotion in the considered set; in some cases, we considered the most expressed emotion(s) in each comment, and we measured the percentage of comments showing them. Formally, let  $c$  be a Reddit comment, and let  $\tau(c)$  be its tokenized and lemmatized representation, i.e., the list of lemmas of which  $c$  is constituted. Then, we map each lemma  $l \in \tau(c)$  to an 8-dimensional vector  $e(l)$ , where  $e(l)[i]$  ( $1 \leq i \leq 8$ ) indicates the score of the  $i$ -th emotion for the lemma  $l$  in the NRC lexicon, with emotions indexed as follows: [anger, fear, disgust, sadness, surprise, trust, anticipation, joy]. Finally, the score of the  $i$ -th emotion for a comment  $c$  is defined as the sum of the scores of  $e(l)[i]$  for each  $l \in \tau(c)$ .

Given the specific nature of COVID-19 related posts and users' reactions to this topic, we considered particularly interesting to study, in addition to the basic emotions, the skepticism expressed by users in their comments. Here we intend "skepticism" in its common sense use, i.e., "an attitude of doubt or a disposition to incredulity either in general or toward a particular object" (from the Merriam-Webster Online Dictionary), rather than in the philosophic sense. As such, we modeled skepticism as an emotion sitting at the polar opposite of trust in the Plutchik model, following the intuition that skepticism is a

form of distrust towards particular topics. While ours is a coarse simplification, it allows us to leverage existing computational language resources to gauge the level of skepticism expressed in a natural language sentence. It is worth pointing out that, to the best of our knowledge, no computational linguistic model of skepticism has been presented so far in the literature.

In particular, we extended the dictionary-based methodology (employed for the analysis of the basic emotions) to the concept of skepticism by creating a new weighted list of words for such dimension. We looked up all the NRC terms expressing trust in Open Multilingual WordNet, a large electronic resource resulting from the alignment of many language-specific wordnets. A wordnet (of which the best known is by far the original English WordNet in [41]) is an electronic dictionary and semantic network, where words and their senses are linked in a graph structure by several semantic relations such as synonymy, hypernymy, metonymy and so on. We exploited the antonymy relation in order to find the words expressing an opposite meaning to our original “trust” words from the NRC lexicon, in all the considered languages (with the exception of German, for which no WordNet is included in the Open Multilingual WordNet.), and thus created the new skepticism lexicon. The intensity score for the skepticism terms were replicated from the corresponding trust terms from which the new terms were derived following the antonymy relation.

Formally, let  $T$  be the set of words in the NRC lexicon expressing trust emotion. Then, we search for each  $t \in T$  in the Open Multilingual WordNet, and we focus on the antonymy relation. Let  $A(t)$  be the set of words indicated in the antonymy relation for the word  $t$ . Then, we define the skepticism lexicon as  $SL = \bigcup_{t \in T} A(t)$ , where the intensity score of each word  $t \in SL$  is replicated from the corresponding trust term from which  $t$  is derived from.

### 5.3. Evaluation and Validation of the Method for Emotionality Analysis

The proposed methodology is arguably simple and prone to miss nuances in the expression of emotions in natural language, especially in the presence of negation. While such a straightforward method could lead to sub-optimal performances, the alternative would involve crafting specialized lists of polarity shifters and amplifiers (see [42] for example, for Italian). Such approach, however, would hinder the portability of the method across languages, as different hand-crafted syntactic rules for different languages would lead to a less than fair evaluation. As a sanity check for the data set used in the present work, and before carrying out our experiments, we counted how many lemmas in the datasets exploited in our experiments were directly followed or preceded by a negation cue (e.g., “not”) and corresponded to a relevant score ( $\geq 0.25$ ) for one of the emotions in the NRC lexicon. As an example, if we consider the English datasets, this number varied between 0.6% and 2.9%, depending on the emotion. Thus, on the basis of this result, we can state that the impact of negation is limited with respect to dictionary-based emotion detection, and it does not flaw the overall analysis. Moreover, we point out again that the added value provided by our multilingual analysis overcomes this limitation.

In order to validate in a more systematic way the approach for emotion analysis, we tested our dictionary-based emotion classification on a standard benchmark for emotion detection, provided by the SemEval 2018 shared task 1: “Affect in Tweets” [43]. In this open challenge, a training set of 8566 English tweets (7102 for training and 1464 for development) are provided, along with 4068 tweets for testing. The tweets are annotated with a continuous value indicating the presence and intensity of four emotions: anger, joy, fear and sadness. We applied our method to the test set, predicting a numeric score for each emotion and evaluating the performance in terms of Pearson correlation, which is the official scoring metric of the shared task. As expected, the overall performance was lower than those reported by the participant systems, which were all supervised and trained on the training set. Our unsupervised, dictionary-based method, however, obtained a score of 36.8 (averaged over the four emotions) situated roughly halfway between the top-ranking system and the baseline of the task.

The SemEval shared task also provides a data set with binary labels, in order to evaluate the performance of an emotion detection system via accuracy, precision, recall, F1-score and other classical similar measures exploited in the literature [44]. Again, we applied our method to this test set in a straightforward way, labeling a tweet with an emotion every time the score computed with the NRC lexicon for that emotion was greater than a threshold of 0.25. The results in terms of accuracy ranged between 0.53 (trust) and 0.73 (disgust). We report an average precision of 0.37, an average recall of 0.62, and consequently an average F1-score of 0.42, which again places this system around the middle of the official ranking of the 2018 shared task for this evaluation. Our results on the SemEval classification task were in line (in terms of overall performance as measured by the F1-score) with a similar monolingual experiment carried out by Kušen et al. [45], where several lexicons, including NRC, were tested with the addition of special heuristics to deal with shifters and amplifiers.

## 6. Results

In this section we present and discuss the main results of our analysis. In particular, the analysis is divided in two parts: the first part considers structural properties of megathreads and posts from several perspectives; moreover, we compare some of the structural properties registered under the COVID-19 emergency with the reference dataset collected from 2019, as described in Section 4. The second part is devoted to analyze emotionality aspects both comparing emotional expression in normal conditions with respect to emotions expressed under the COVID-19 emergency, and also analyzing in more detail changes in the stance of the users during the evolution of the emergency.

### 6.1. Structural Analysis

In this first series of analyses, we addressed the following research questions, related to structural properties of subreddits and posts:

*RQ1: Are there any differences between the COVID and the non-COVID social posting behaviors?*

*RQ2: Is the timeline of critical events in each country about COVID-19 somehow correlated with the flow of the discussions on the corresponding subreddits?*

*RQ3: Are official online resources preferred to other resources for gathering information about COVID-19?*

#### 6.1.1. Comparison of COVID and Non-COVID Social Posting

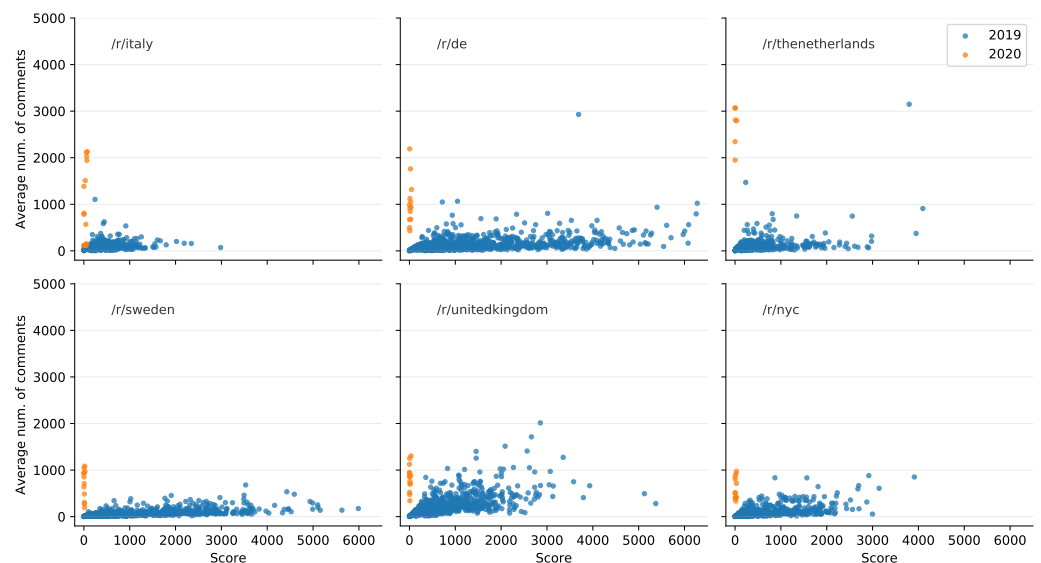
The aim of this comparison is to fulfill the research question *RQ1* introduced above. In particular, we compared some properties of posting behaviors relating the COVID and the non-COVID datasets.

First of all, note that comparing just the number of submissions makes little sense, due to the very different number of posts and comments in the two considered periods (see Tables 2 and 3). Therefore, we next focused on the distribution of comments against scores and submissions in the two scenarios.

We started by determining the distribution of the average number of comments against the scores of the submissions to which they refer. Figure 3 shows the obtained results. Without loss of accuracy, the  $x$ - and  $y$ - axes were limited in order to account for all the data while maintaining a common range for each graph. From the analysis of the figure, we observed a significant difference between 2019 and 2020 in the average scores of the submissions having a certain average number of comments. In particular, note how for the 2019 dataset the distribution skewed towards higher scores, whereas for the 2020 dataset the distribution settled for lower scores. Moreover, the number of comments was generally significantly higher in 2020 than in 2019 data. These results indicate that COVID-19 related users were more interested in discussing the topic, rather than evaluating each other's comments or resources. On the contrary, in normal situations, users were more prone to judge other comments and posts, expressing this judgment by their scores. Particularly interesting is the fact that this behavior was common to all considered countries, with just



little differences among them. As a consequence, this trait can be considered specific to users' behavior in the COVID period.

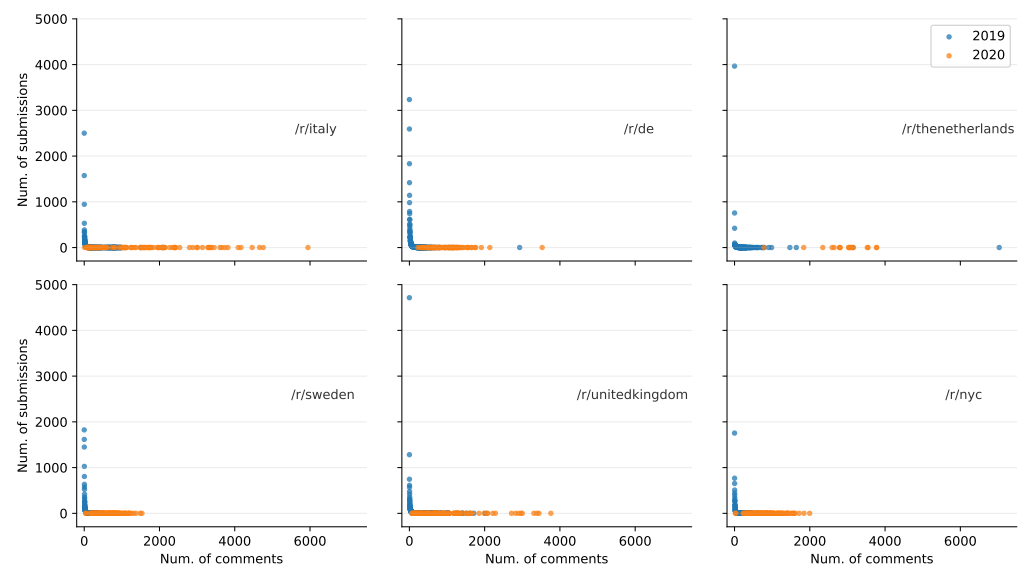


**Figure 3.** Distribution of the average number of comments against the scores of the submissions to which they refer.

Another interesting aspect to investigate is the distribution of submissions against comments. Although the submissions related to 2020 consist only on the megathreads, the number of comments may vary. In this case, it is useful to understand if this variation is similar or different between 2019 and 2020. Figure 4 illustrates this variation. In this figure, both axis were limited in order to maintain a common range for a better visualization, while preserving the precision and the completeness of the data. The analysis of this figure allows us to better understand the distribution of submissions against comments.

In particular, the 2019 distribution roughly followed a power law, exhibiting a high number of submissions with a very small number of comments. On the contrary, the 2020 distribution essentially showed that each megathread ignited a discussion between users, thus producing a very high number of comments with respect to a normal social posting situation. This means that, under COVID-19, most of the interest focused on participating in discussions rather than starting new ones. In this case, this trend was common to all considered countries, thus characterizing users' behavior in front of dramatic events like COVID-19.

Summarizing the above analysis, we can finally answer the research question: *RQ1 Are there any differences between the COVID and the non-COVID social posting behaviors?* The answer is yes, and the characterization of these differences was quite sharp.

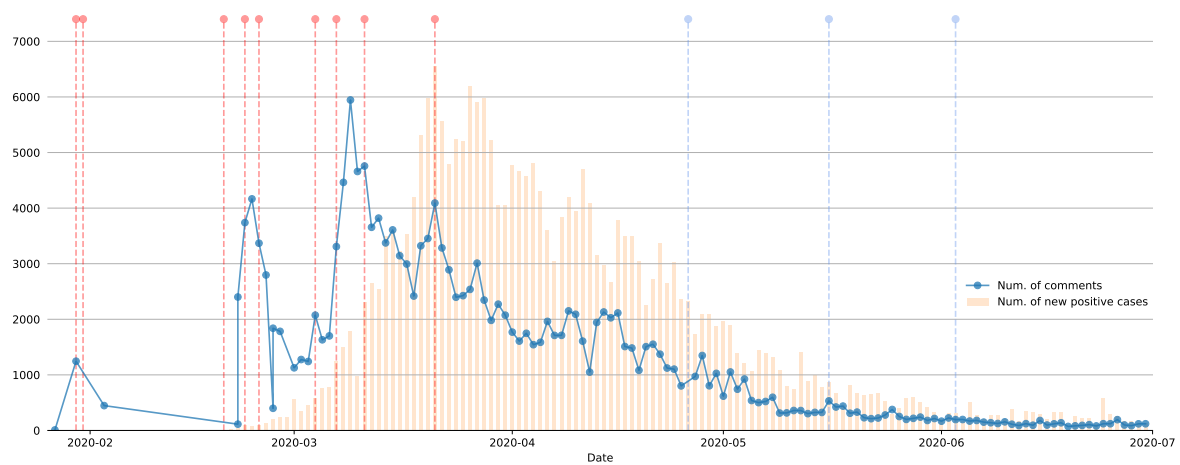


**Figure 4.** Distribution of submissions against comments.

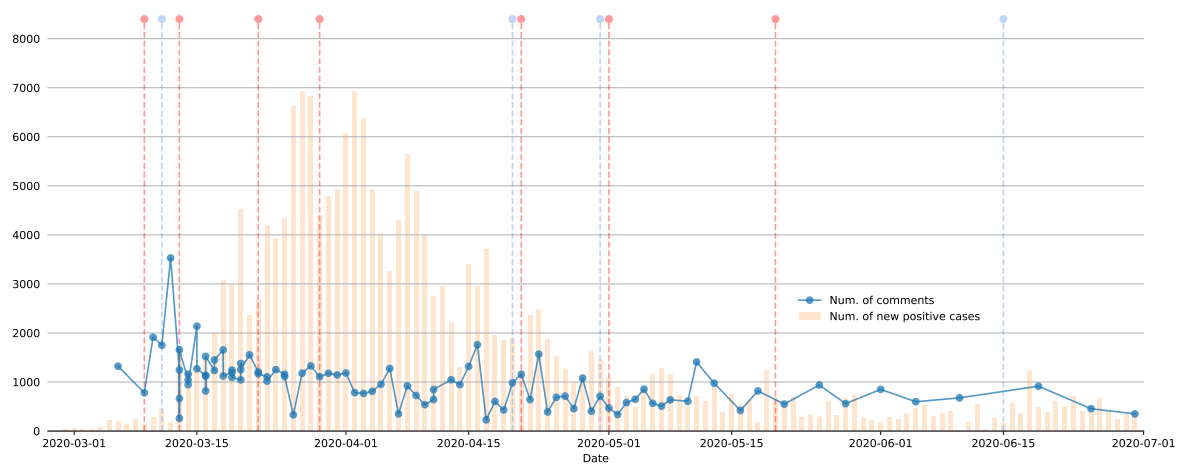
#### 6.1.2. Analysis of the Correlation between the Timeline of Critical Events and Social Posting

The aim of this comparison is to fulfill the research question *RQ2* introduced above. In particular, we analyzed the relation between the flow of the discussions on Reddit, in terms of number of comments, and the whole epidemic situation by considering the timeline of critical events in each country and the new daily registered cases of COVID-19. In particular, we plotted three different types of information, namely (i) the number of comments in COVID-19 megathreads, (ii) the number of new daily COVID-19 positive cases and (iii) the most critical events and policies related to the epidemic situation adopted by the government. For each graphical timeline, the *x*-axis starts at the date of the first confirmed positive case and ends at 1 July. Figures 5–10 show these data for each considered country. Recall that the detailed timelines of the most critical events and policies have been reported in Tables 4–9. It is worth pointing out that in Figures 5–10 events and policies have been categorized in two classes, namely closing and reaction events (identified by red vertical lines), which correspond to events and actions that impose any sort of restrictions, and opening and loosening events (identified by blue vertical lines), which correspond to events lifting up preemptive measures, such as re-opening of stores and announcement of the end of the lockdown.

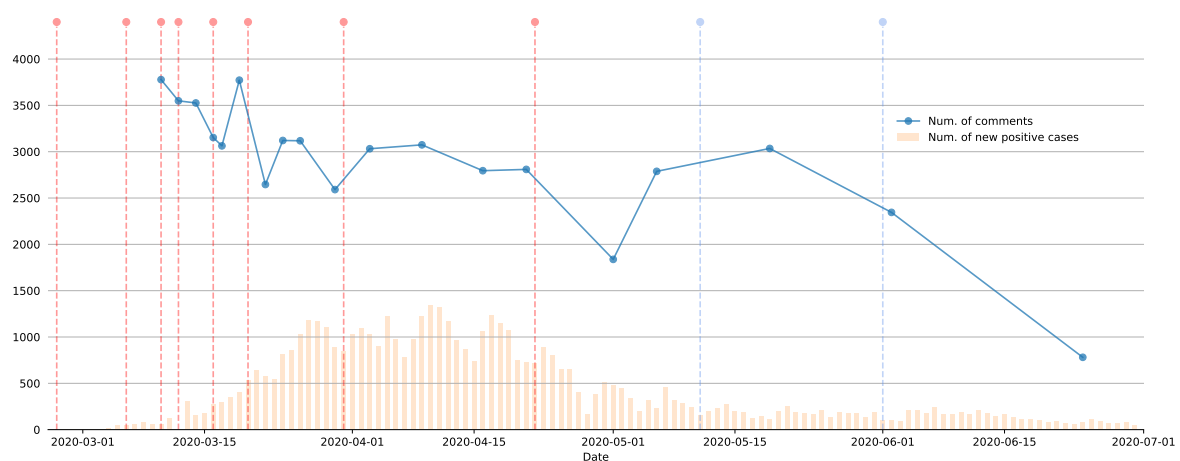
From the analysis of these figures, it is possible to observe that, generally, there was an increase in the number of comments both in correspondence of sharp increases in new daily cases and in correspondence of closing and reaction events. A reduction in daily positive cases or loosening events like lifting preemptive measures corresponded with a reduced interest in posting on COVID-19 megathreads. This is not surprising, since the drama of the emergency and the little knowledge about the disease pushed many users to gather and share information.



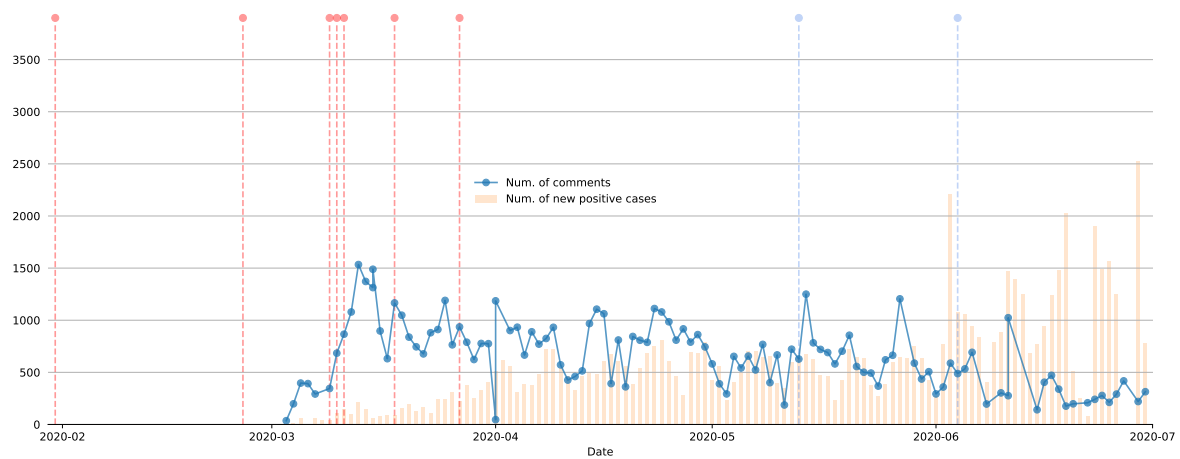
**Figure 5.** Number of comments per megathread in /r/italy compared to the number of new daily cases in Italy. Vertical lines represent adopted critical events and policies.



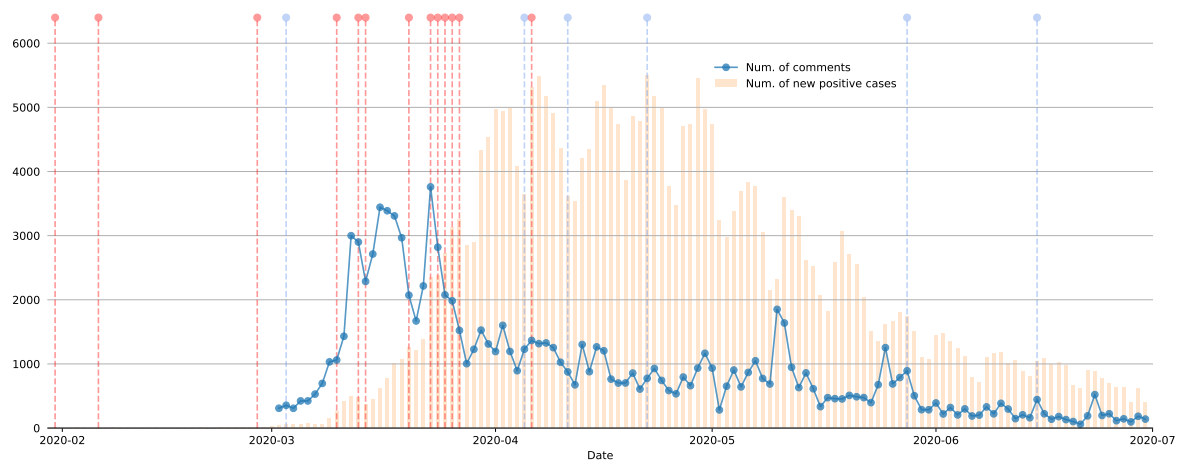
**Figure 6.** Number of comments per megathread in /r/de compared to the number of new daily cases in Germany. Vertical lines represent adopted critical events and policies.



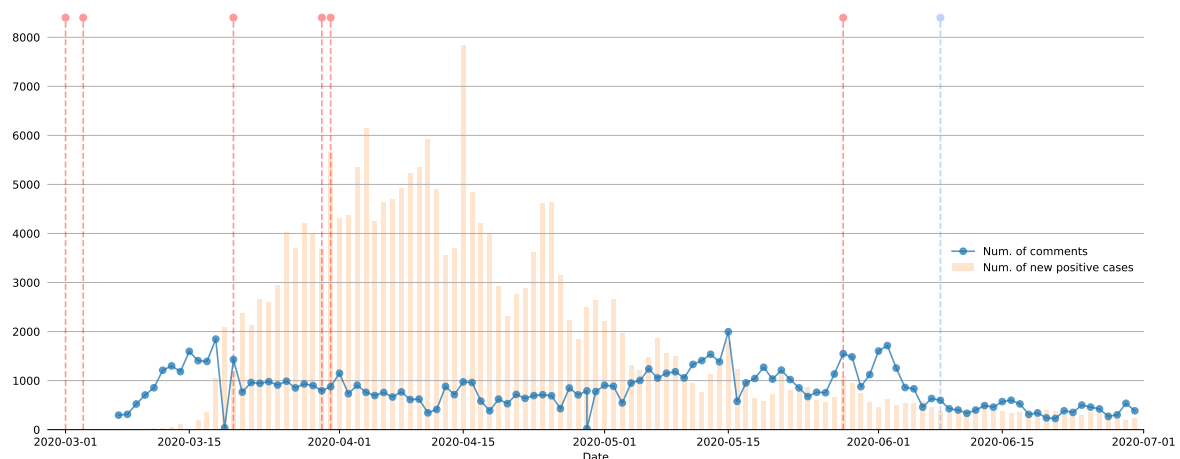
**Figure 7.** Number of comments per megathread in /r/thenetherlands compared to the number of new daily cases in The Netherlands. Vertical lines represent adopted critical events and policies.



**Figure 8.** Number of comments per megathread in /r/sweden compared to the number of new daily cases in Sweden. Vertical lines represent adopted critical events and policies.



**Figure 9.** Number of comments per megathread in /r/unitedkingdom compared to the number of new daily cases in UK. Vertical lines represent adopted critical events and policies.



**Figure 10.** Number of comments per megathread in /r/nyc compared to the number of new daily cases in New York City. Vertical lines represent adopted critical events and policies.

Particularly interesting, different trends were expressed by the different countries: Italy showed a curve of new posts which almost overlapped with the curve of new positives, thus reflecting a will of the users to discuss and inform themselves about the situation. Reactions

from the United Kingdom seemed to be mostly related to (possibly late) government actions and not to the trend of new daily cases, whereas Germany showed an almost constant number of comments, slightly reduced in the last part of the considered period both in terms of number and frequency. The Netherlands showed a quite high number of comments overall, compared to the number of inhabitants of the other countries, even if the frequency of posts was much lower than that of other countries; also in this case the high initial interest on the topic decreased quite fast when the decrease in new daily cases became stable. Sweden showed its peak of comments at the beginning of the outbreak and government actions, and the interest on the topic remained quite high over the period; most probably, this counter-trend behavior with respect to the other countries can be motivated by the fact that the number of Dutch daily cases actually increased instead of diminished, with a severe increase in the last part of the considered period. Finally, New York City showed quite a constant interest on the topic, at least for more than half of the considered period, presenting a reduction in the number of posts when the number of new daily cases decreased significantly.

From the analysis above, it is possible to answer the research question RQ2: *Is the timeline of critical events in each country about COVID-19 somehow correlated with the flow of the discussions on the corresponding subreddits?* Actually, the answer also in this case is yes, since different policies and events in the various countries affected the flow of discussion in each country.

### 6.1.3. Analysis of the Most Linked Web Domains and Subreddits

Apart from simple text, we recall that Reddit comments might also contain URLs to web resources. Thanks to this information, linked web sources are a very interesting aspect to analyze. Reddit already represents a heterogeneous source of information per se: users commonly post links to external resources, and some subreddits inherently converge towards being link-only communities. Nevertheless, the ubiquity of social networks represents an open field for the dissemination of COVID-19 news; unfortunately, some of them can be unreliable. Studying which are the most linked web resources and domains could help experts in identifying preferred news hubs, in order also to understand the inclination of users to ground on reliable sources of information.

Our analysis started by considering the URLs appearing in each comment. For each subreddit considered in our study, we parsed each comment: if a comment contained one or more URLs, we collected them. At the end of this process, we counted the occurrences of each domain. Figures 11 and 12 report the top 25 most linked web domains for each subreddit, respectively.

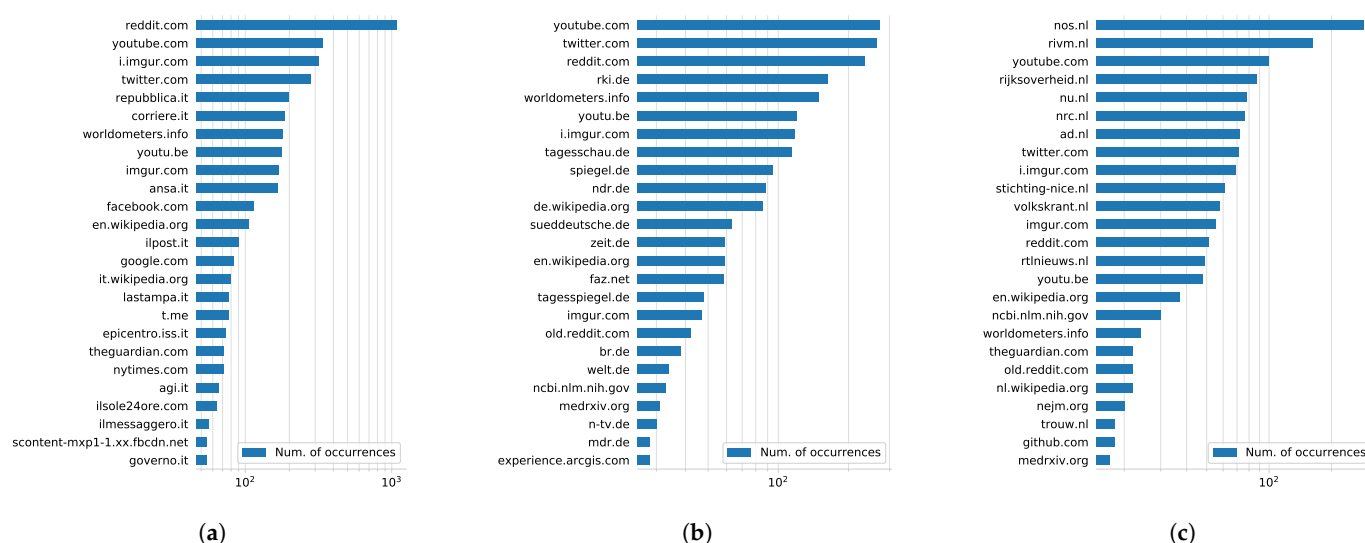
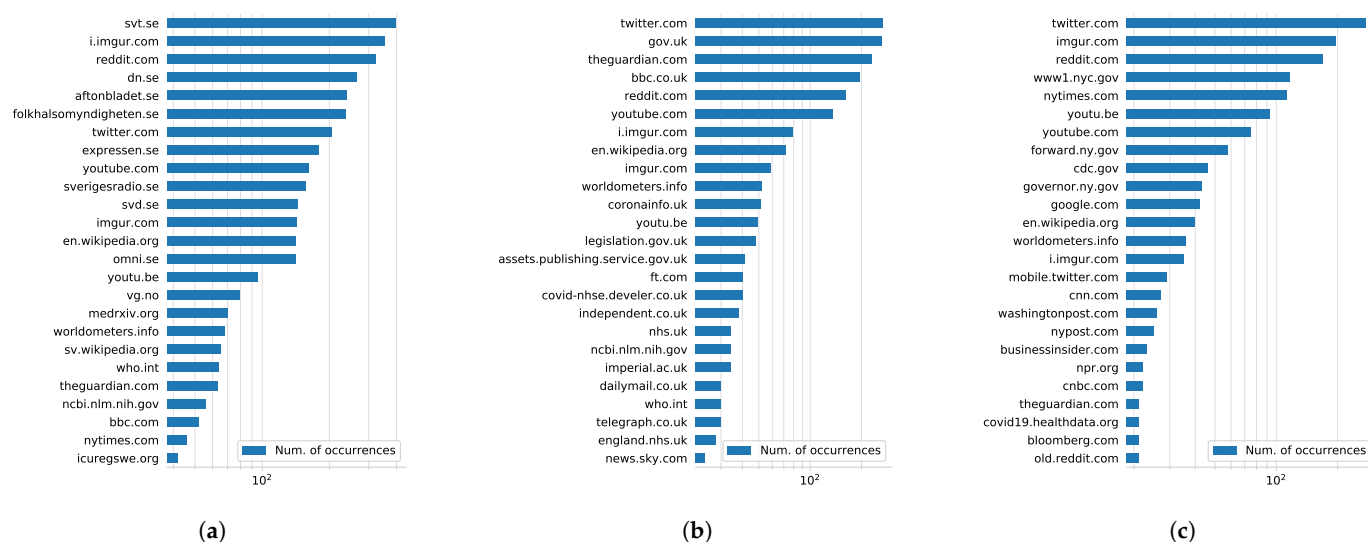


Figure 11. The top 25 most linked web domains in (a) /r/italy, (b) /r/de and (c) /r/thenetherlands.





**Figure 12.** The top 25 most linked web domains in (a) /r/sweden, (b) /r/unitedkingdom and (c) /r/nyc.

First of all, observe the raw number of occurrences: the highest number of occurrences for a domain was related to /r/italy, where the domain reddit.com occurred exactly 1082 times. The same domain did not occur as the most linked one in any other considered subreddit, implying that a high number of comments in /r/italy referred to other comments, submissions and subreddits within Reddit, thus suggesting a more Reddit-centric way of discussing in Italy. Moreover, it is interesting to observe how resources like Twitter, YouTube and Wikipedia were well represented among all the subreddits.

Instead, the case of governmental domains was very different among countries: the number of occurrences in /r/unitedkingdom and /r/nyc for governmental domains (such as gov.uk and nyc.gov) was relatively high, while in other subreddits it was relatively low. Particularly interesting, in /r/thenetherlands the domain rivm.nl was in second place; this is the domain of the The National Institute for Public Health and Environment. It is also worth noticing that different newspapers and mass media agencies were present in all subreddits. Several national newspapers and mass media agencies per subreddit were present, such as Spiegel.de for /r/de, NOS.nl for /r/thenetherlands and Repubblica.it for /r/italy; also, few resources were shared among the subreddits, such as NYTimes.com, which were referred to in /r/nyc, /r/sweden and /r/italy.

Since Reddit.com appeared in each subreddit, we studied the distribution of subreddits within the extracted URLs. Figures 13 and 14 show the top 25 most linked subreddits for each considered subreddit. A quick overview shows how the most linked subreddit was essentially the subreddit itself, thanks to a very high number of comments referring to different comments and submissions posted therein. This reinforces our hypothesis that studying these subreddits provides us a focused analysis on COVID-19 related posting habits. Moreover, the subreddits /r/Coronavirus, /r/China\_Flu and /r/COVID19 appear almost always in the top 5; we recall that /r/Coronavirus has been made the official community for COVID-related updates in Reddit, hence the relevant number of occurrences is expected.

Different insights can be drawn from observing the distribution of the linked subreddits. All countries in Europe presented the /r/europe subreddit within their top 25 subreddits, except for /r/de. Moreover, /r/italy appeared in few European subreddits: this is somehow expected given that the COVID-19 epidemic in Europe actually started in Italy, and all countries looked at it for the development of the outbreak.

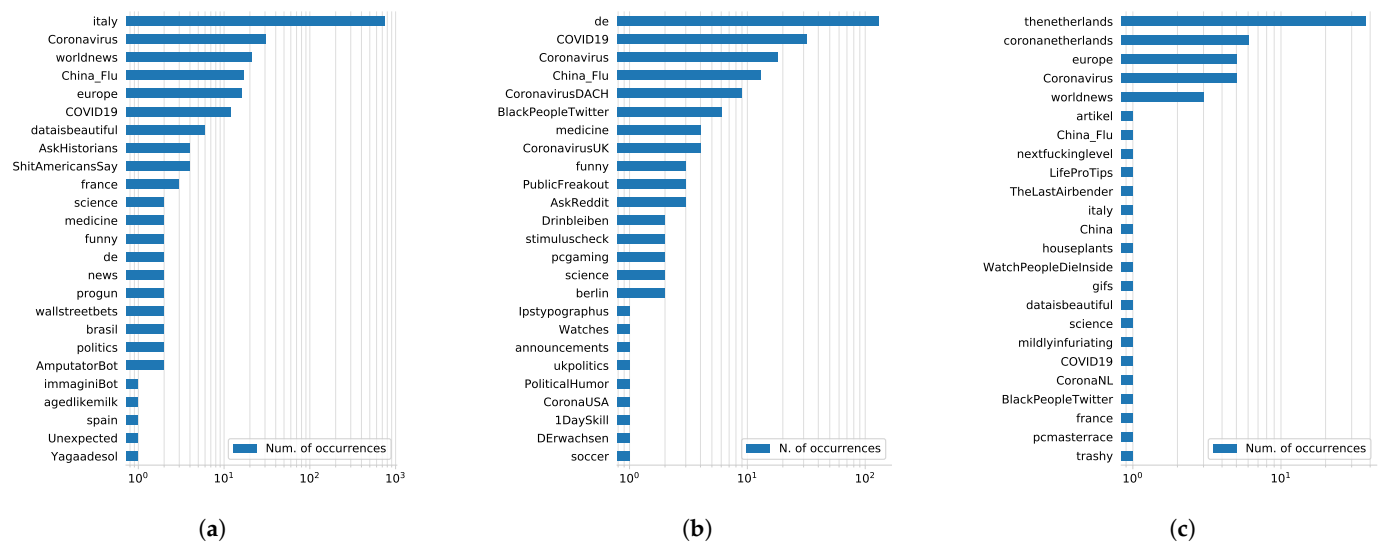


Figure 13. The top 25 most linked subreddits in (a) /r/italy, (b) /r/de and (c) /r/thenetherlands.

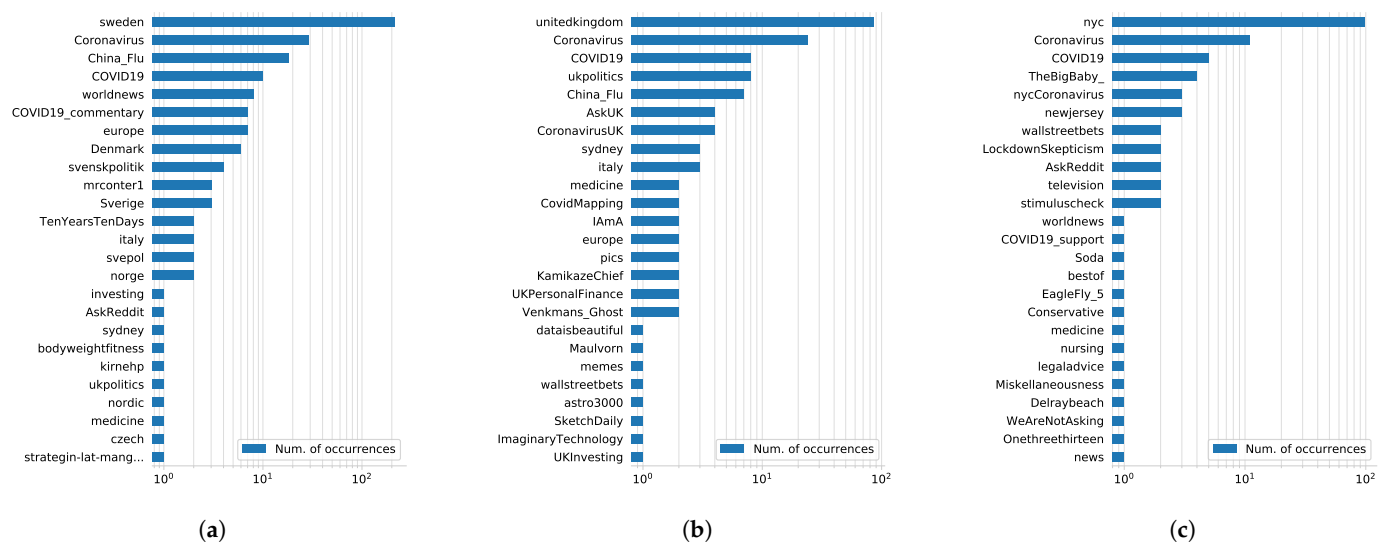


Figure 14. The top 25 most linked subreddits in (a) /r/sweden, (b) /r/unitedkingdom and (c) /r/nyc.

From the analysis presented above, we try to address the third research question *RQ3: Are official online resources preferred to other resources for gathering information about COVID-19?* In this case, there is not a clear answer. It mostly seems that Reddit users trusted other Reddit users more than official resources, given the very high number of self-referential links. Only /r/unitedkingdom, /r/nyc, and /r/thenetherlands presented official governmental domains in their top links; analogously, the massive use of YouTube and Twitter may raise some questions about the reliability of the information that users exploit to ground their opinions. In any case, the analysis above clearly shows that the attitude of people to rely on official sources of information is more about cultural factors than other considerations. As a consequence, any analysis on the usage of online resources for gathering information should take this aspect into account.

## 6.2. Emotional Analysis

In this series of analyses, we considered the emotions expressed by Reddit users in their posts. We first focused on the eight basic emotions of the model by Plutchik and on the comparison between emotions expressed in the 2019 non-COVID dataset and in the 2020

COVID dataset. Then, we considered a possible correlation between emotion expression by country and the corresponding events and government reactions related to COVID-19. Finally, we considered the expression of skepticism about COVID-19 related topics.

The main research questions addressed in this section are as follows:

*RQ4: Did the COVID-19 outbreak affect emotionality expression on social posting?*

*RQ5: Can the emotions expressed by users in each country during COVID-19 be related to government reactions to the emergency?*

*RQ6: Can the analysis of unconventional emotions, like skepticism, help in predicting potentially dangerous situations?*

In order to carry out these analyses, we focused on the top comments of each megathread; in fact, they express a direct feeling of the user who is posting and wants to share information or an opinion. Emotions expressed in replies to top comments or in comments to other comments may be reactions to very different kinds of feelings; thus, they may not be precisely focused on COVID-19. Moreover, since the same comment may express several emotions, in order to reduce noise in the data analysis, we focused on the two most expressed emotions for each comment. Finally, for each emotion, instead of evaluating their score computed by the NLP approach, we focused on the percentage of comments expressing it in each month; this allows us to compare the trends of emotion expressions in the considered period and, also, to directly evaluate possible changes in the stance of users. The same analysis was carried out both on the 2019 non-COVID dataset and on the 2020 COVID dataset.

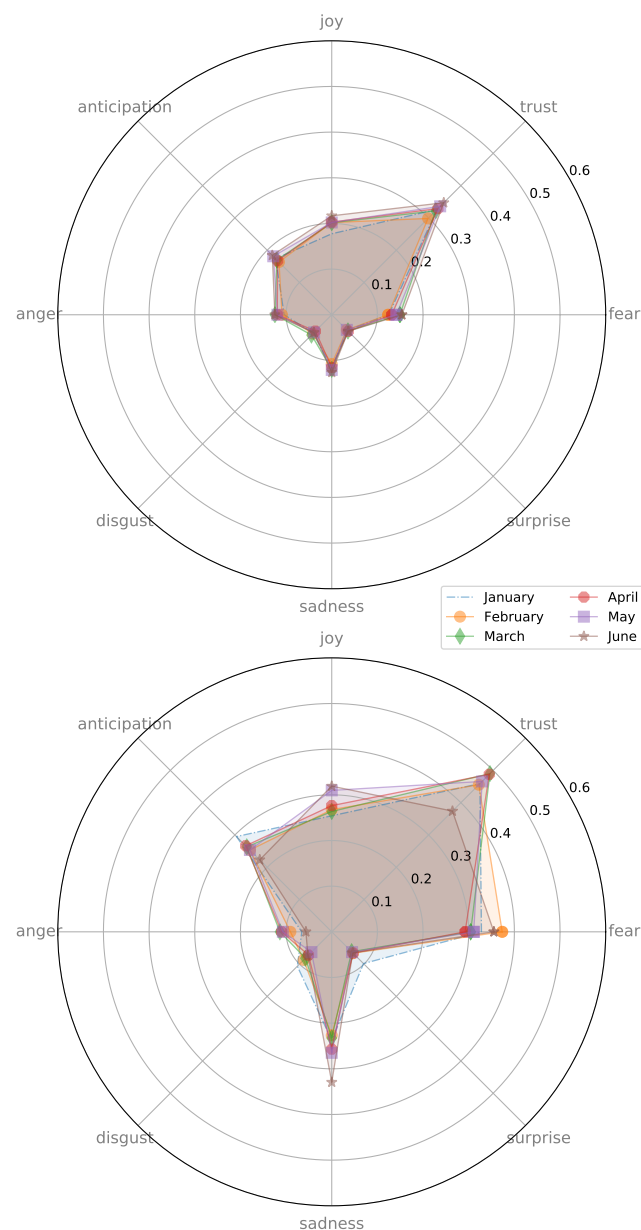
Overall results of this analysis for the eight basic emotions of the model by Plutchik are presented in Figures 15–20 for the six countries considered in this paper.

#### 6.2.1. Comparison of COVID-19 and Non COVID-19 Emotional Expression

Let us focus first on the comparison between the 2019 non-COVID and the 2020 COVID datasets by analyzing Figures 15–20. The very first and immediate consideration we can draw about these results is that, without a doubt, emotion expression was significantly much higher during COVID than the corresponding reference period in 2019. As a consequence, we can already answer the research question *RQ4: Did the COVID-19 outbreak affect emotionality expression on social posting?* The answer is yes, and we go deeper on this analysis next.

On the one side this result can be expected, since the dramatic outbreak of COVID-19 pushed users to share emotions about it; on the other side, this result gives strength to the analysis we will conduct in the next section, since it shows that emotions expressed in the COVID period were really strictly related to the topic. It is, however, interesting to observe that not all emotions obtained a substantial increase in their expression. As an example, consider Figure 15 about Italy; expressions of anger, disgust and surprise were almost similar in the two periods, whereas the main differences were related to trust, fear, anticipation and sadness. Clearly, there were differences among countries. As an example, consider Figure 16 about Germany; in this case the overall shape of the graph was similar in the two periods, but the increase in the intensity of emotion expression was generalized.

Trust and fear were the emotions that generally increased most across the evaluated countries, and they can be considered as the most characterizing emotions about COVID-19. While the increase in fear is easy to be interpreted, the stronger increase in trust with respect to fear is, in a sense, surprising: it expresses a strong desire of people across the countries to believe that everything will be alright (see for example the slogan *#andràtutto bene* that went viral in Italy, which stands for “everything will be alright”), but can be also an expression of confidence that governments, in one way or another, will be able to properly address the emergency. However, we can note that the expression of trust (as well as the other emotions) varied during the evolution of the pandemic, see for example the plots for Italy, Sweden and New York City. Some further considerations are needed about these variations and will be elaborated in the next subsection.



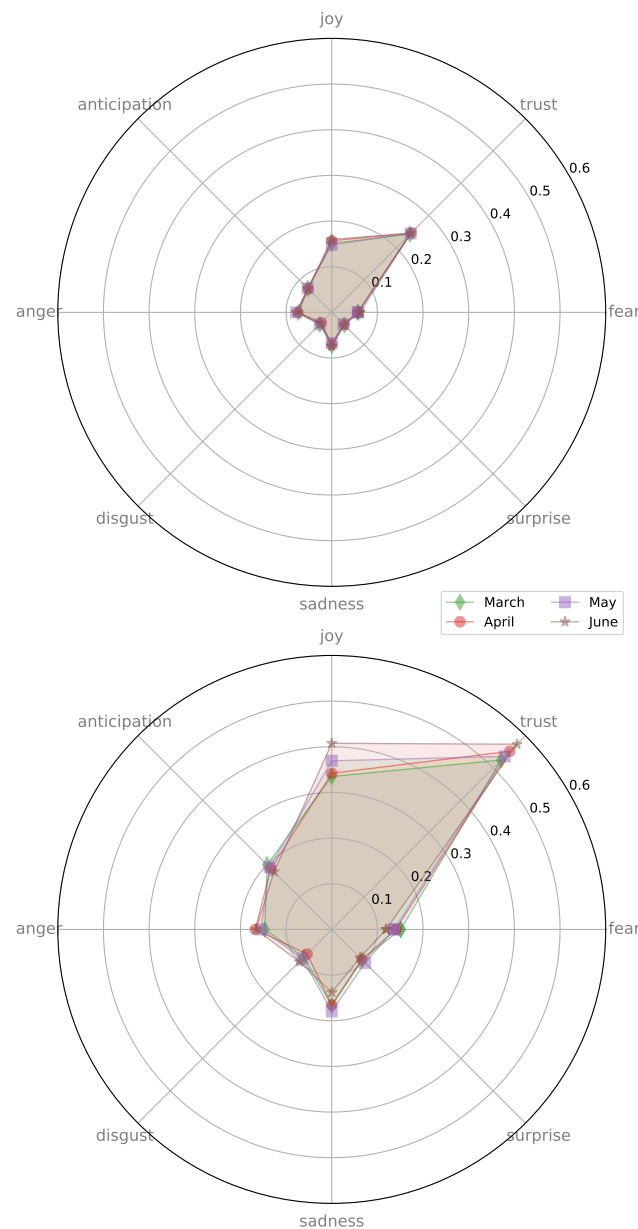
**Figure 15.** Radar plot of most expressed emotions in Italy per month for both 2019 (**top**) and 2020 (**bottom**) datasets.

#### 6.2.2. Analysis of the Most Expressed Emotions by Country and Month during the COVID-19 Outbreak

In this section we focus on the 2020 COVID dataset. Before going into the details of the analysis, it is worth pointing out again that the less expressed emotions over all the countries were surprise and disgust. This is coherent with the data since disgust has no real reference with the disease, whereas the pandemic has been well anticipated over all the countries and, consequently, it did not come unexpectedly.

Overall each country presented its characteristic shape for emotion distribution (see Figures 15–20); this is already an interesting point since it confirms that emotionality is also a cultural characteristic, with Italy showing the widest range of expressed emotions and Nordic countries showing less variability in expressed emotions. Particularly interesting are the almost overlapping shapes for users from the United Kingdom and New York City.

Let us focus now on the single radar plots of each country, focusing on the relation between the emotionality expression and the timeline of events presented in Tables 4–9 and in Figures 5–10.



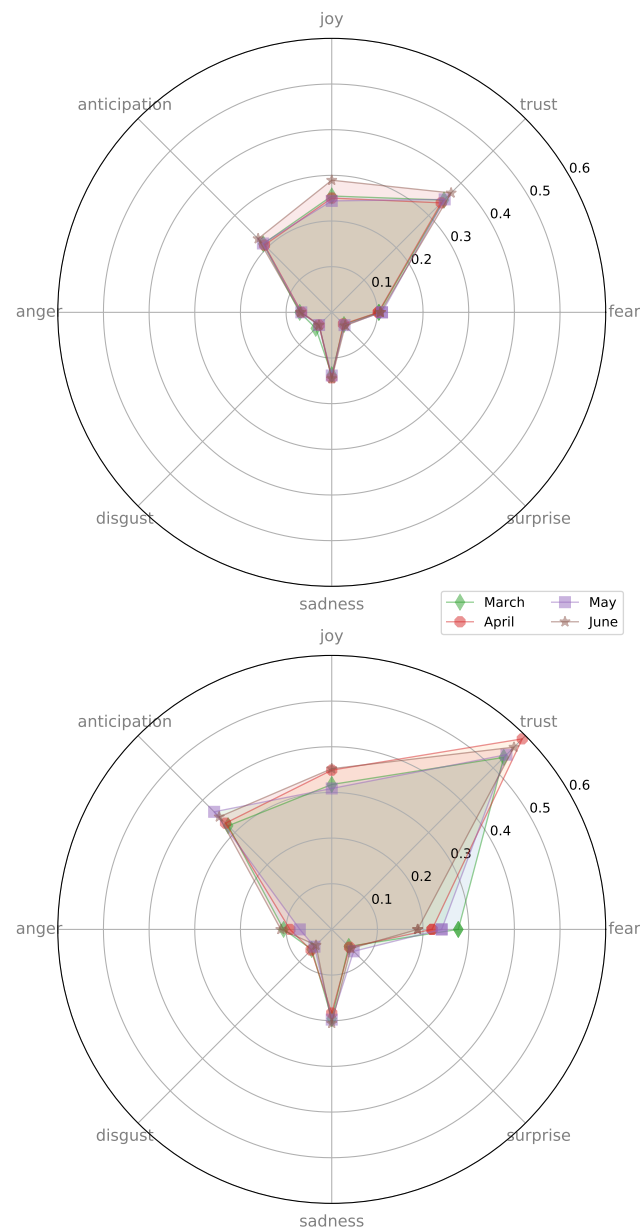
**Figure 16.** Radar plot of most expressed emotions in Germany per month for both 2019 (**top**) and 2020 (**bottom**) datasets.

### Italy

As previously pointed out, Italy showed the widest variability in expressed emotions; this is partially due to the higher availability of data both about the number of months and the number of users (see Figure 5). It is particularly interesting to observe that trust was particularly high in March and April, during the lockdown, whereas it decreased in May and dropped in June when authorities progressively reduced restrictions; conversely, fear followed an opposite trend, finding its minimum in April and its maximum expression in February and June. These data suggest that people felt safe in their home during the lockdown, whereas the fear for a second wave came up with reopening. Interestingly, joy was highest in May and June. Sadness constantly increased between February and



June when the continuation of the emergency actually showed that a solution to the emergency was far away; in fact, while anticipation was quite high in the former months, it decreased significantly in June, when people started worrying about their jobs and future. Interestingly, the low expression of anger can be interpreted as an agreement of the population with the government management of the crisis.



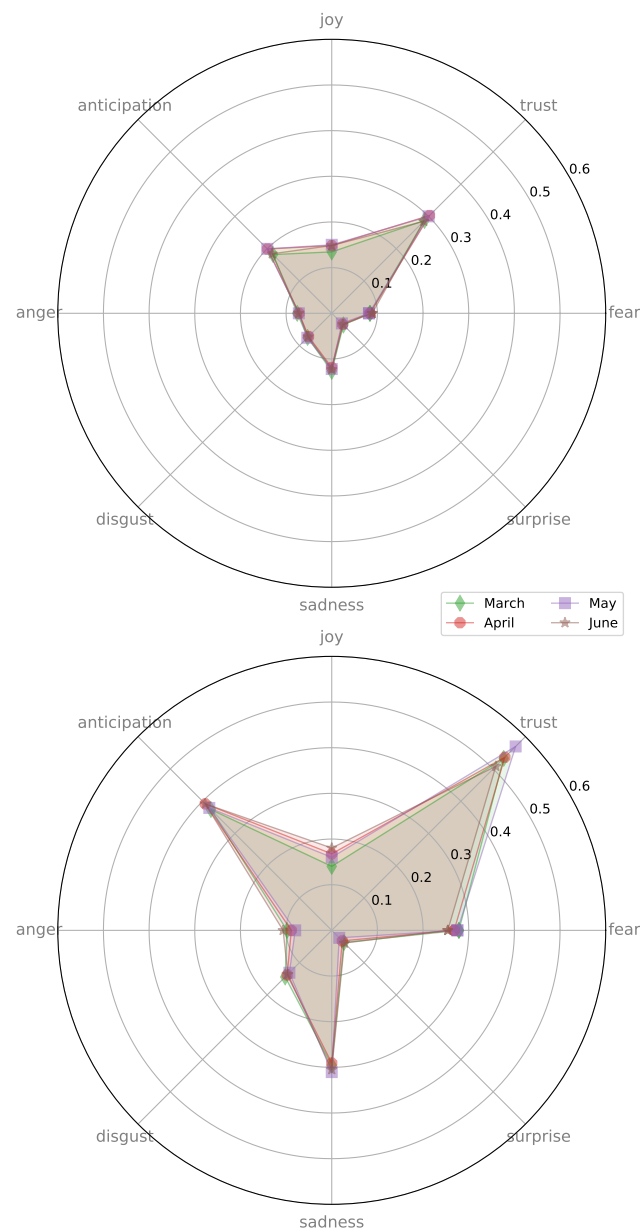
**Figure 17.** Radar plot of most expressed emotions in The Netherlands per month for both 2019 (**top**) and 2020 (**bottom**) datasets.

### Germany

As previously pointed out, the early and firm reaction of Germany to the epidemic crisis allowed it to rapidly slow down the diffusion of the virus (see Figure 6). This motivates the quite high, and increasing, values of trust along the considered months, coupled with a very low expression of fear which, by the way, constantly decreased along the considered months. Actually, among the considered countries, Germany was the most positive one, emotionally speaking. This is strongly related with the trust of population with the government reaction to the crisis.

### The Netherlands

The Netherlands approached the COVID-19 pandemic in a slightly softer way than Germany and Italy, showing lower numbers of daily infections with respect to these countries (see Figures 5–7). This motivates the values of fear which were intermediate between Germany and Italy and constantly decreased through the considered months. Emotions were mostly centered on trust, joy and anticipation. Expression of sadness was quite constant and low, as well as trust which was quite constant and high.

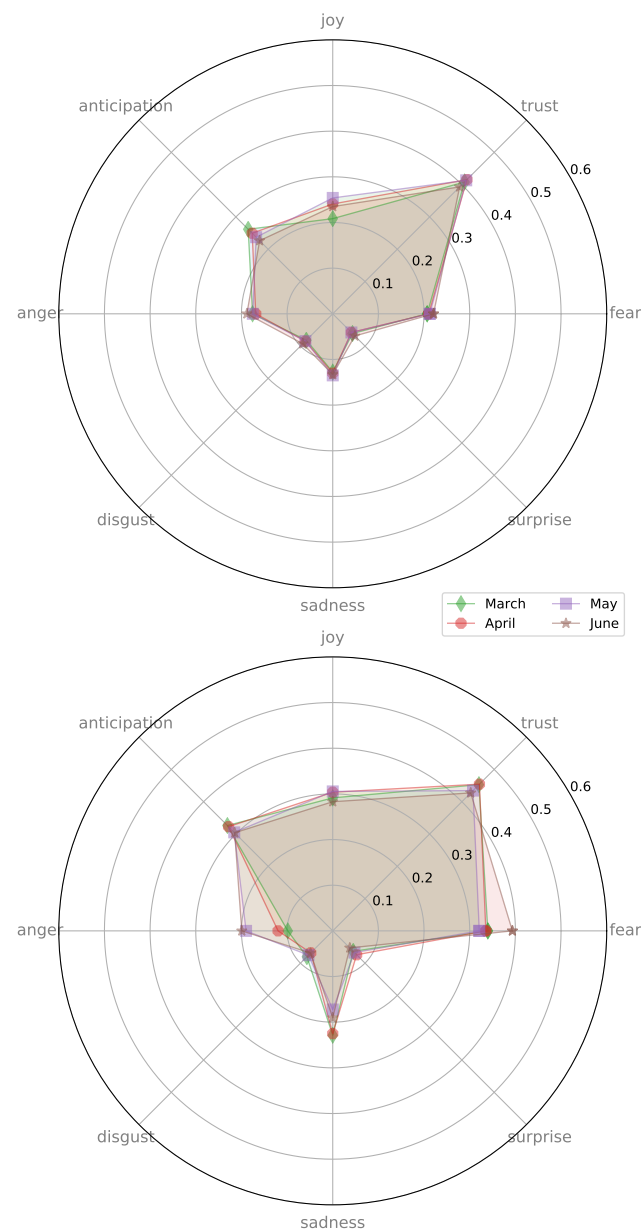


**Figure 18.** Radar plot of most expressed emotions in Sweden per month for both 2019 (**top**) and 2020 (**bottom**) datasets.

### Sweden

Sweden's response to the pandemic has been very different to other European countries. There has been no lockdown, with schools and cafes staying open, but large gatherings have been banned, and most Swedish have observed social distancing. It experienced a far higher mortality rate than its neighbors, and even if the number of new daily positive cases was lower, in absolute terms, with respect to other countries, this number continued

to increase in the considered period. Interestingly, while neighbor countries experienced a significant decrease in new daily cases in later months, Sweden observed a significant increment of infections in June (see Figure 8). As far as this aspect is concerned, it is interesting to observe that trust showed its highest expression from March to May, whereas it decreased significantly in June, with a corresponding slight increase in anger, which might be correlated to a reaction of the population to the (possibly light or absent) actions relative to the emergency. Interestingly, Sweden showed the highest expression of sadness among all considered countries, along with a quite high expression of fear.

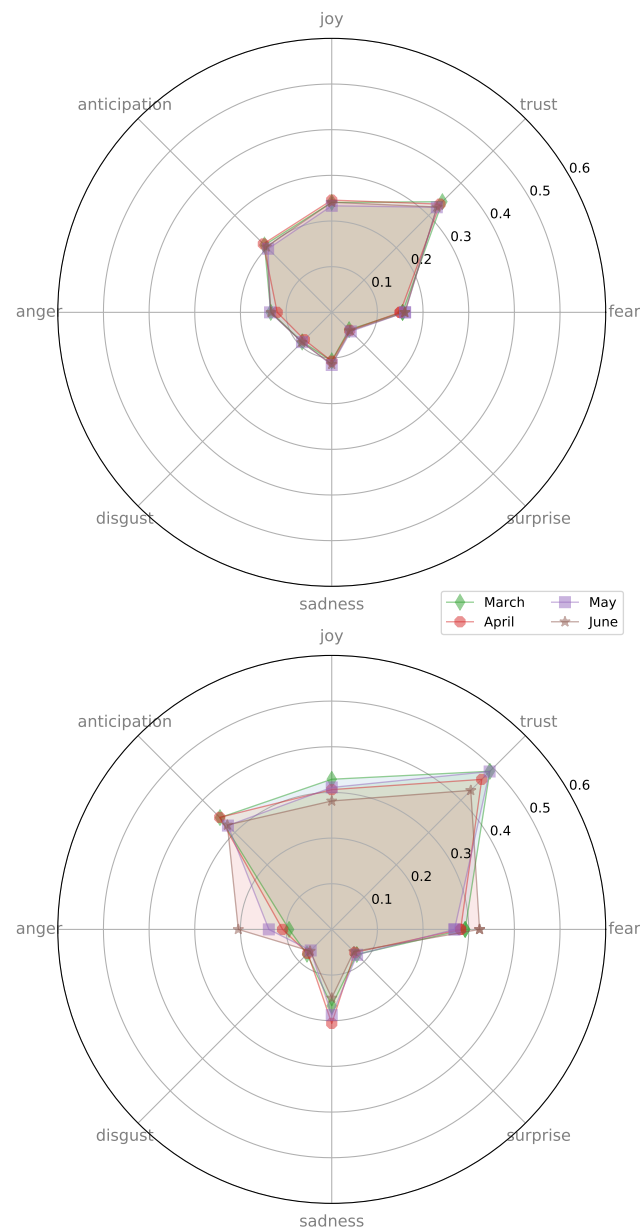


**Figure 19.** Radar plot of most expressed emotions in UK per month for both 2019 (**top**) and 2020 (**bottom**) datasets.

#### UK and New York City

We comment on data about UK and NYC together since their graphs actually overlapped, as well as the reactions of the corresponding governors. In particular, both showed a quite late and initially very soft approach to the COVID-19 crisis management. From the analysis of Figures 19 and 20 it is interesting to observe that expression of anger was the

highest among the considered countries, with values increasing over time. This can be mainly related to the late response of corresponding governments to the COVID-19 crisis. Analogously, fear presents a quite high expression, with its maximum in June, whose trend corresponds to a decrease in trust over the considered months; moreover, expression levels of trust were the lowest among all the considered countries.



**Figure 20.** Radar plot of most expressed emotions in New York City per month for both 2019 (**top**) and 2020 (**bottom**).

### Summary

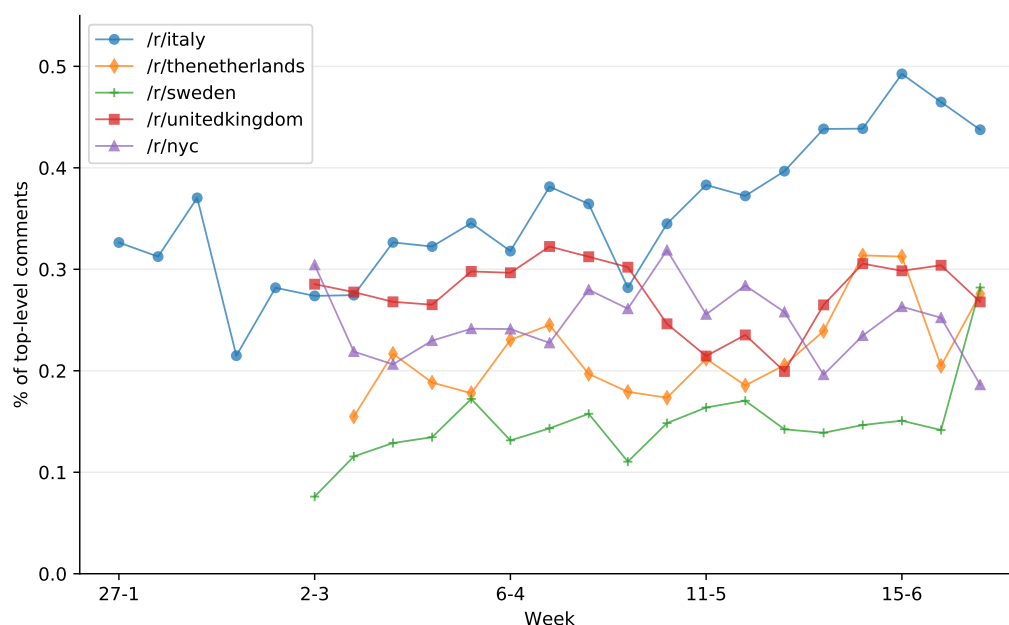
From the detailed analysis above, we can now address the research question RQ5: *Can the emotions expressed by users in each country during COVID-19 be related to governments reactions to the emergency?*. In particular, we have shown that there is a significant correlation between the way governments addressed the pandemic (firm, soft, late, medium reactions) and the emotions expressed by the users of the corresponding countries. While it is difficult in this case to state that there is an immediate correlation between specific government actions and user stances, we can certainly observe that the users have not been static in

their feelings, showing a much higher variability of emotion expressions with respect to the non-COVID period. As a consequence, we can certainly affirm that an in depth analysis of the stance of the users on specific topics, especially on critical emotions like trust, anger or sadness, can certainly help managers and governments to support their choices.

### 6.2.3. Analysis of Skepticism by Country during COVID-19

First of all, recall that skepticism is not a basic emotion, according to the Plutchik model, and that we introduced in this paper a specific method to identify it in text (Section 5.2). In our opinion, studying skepticism directly in the context of COVID-19 may provide interesting insight in the stance of users and may outline potentially dangerous situations.

Since skepticism is not one of the basic emotions, in this analysis we did not concentrate on the top expressed emotions, like in the previous sections, but we measured the percentage of comments expressing skepticism in one way or another. However, we continued to concentrate on the top comments since they specifically expressed the stance about COVID-19 of the user posting it. Finally, we conducted this analysis on a weekly basis, in order to evaluate the corresponding trend more precisely. Obtained results are depicted in Figure 21.



**Figure 21.** Timeline of skepticism expression rate during COVID-19.

From the analysis of this figure, a very first interesting observation is that, apart from UK and New York City, the expression of skepticism increased over time, even if with different trends. In particular, UK and New York City started with a quite high level of skepticism (about 30% of posts); these decreased down to 20% around the 12th week, and then raised a bit again.

Sweden and The Netherlands started quite low on skepticism, below 10% for the former and 20% for the latter, but the expression of this emotion quite constantly increased over the period up to almost 30%. Italy showed the highest values of skepticism, starting from 30% and continuously increasing up to 50% of posts showing it.

Such an increase in skepticism can become dangerous in the context of COVID-19, if associated with a decrease in attention of people to social distancing measures. Obviously, if people no longer believe in the danger of COVID-19, they can lower their level of attention and the virus can start to circulate again. Clear enough, being aware of this increasing trend in skepticism may be of strong interest for managers and governments in order to



tackle the potential risks related to it. In fact, at the time of writing this manuscript, Italy is experiencing a relatively low but yet existing increase in new daily cases.

From the analysis above, we can conclude that the answer to the research question *RQ6: Can the analysis of unconventional emotions, like skepticism, help predicting potentially dangerous situations?* is certainly yes. In fact, there are some emotions that, even if not conventional, can be strongly related to specific situations in a topic (see the increase in skepticism related to the potential increase of COVID-19 diffusion). As a consequence, this can be a research direction which deserves attention.

## 7. Discussion

The first interesting finding in our results addressing research questions *RQ1*, *RQ2* and *RQ3* is that posting behavior, in terms of submissions, comments and scores, was strongly characterized in the COVID-19 period, and it diverged from the usual behavior in normal conditions (see Figures 3 and 4). In particular, during COVID-19 users were more interested in discussing topics, rather than evaluating each other's comments or resources; each megathread ignited a significant discussion between users, thus producing a very high number of comments with respect to normal conditions. Particularly interesting is the fact that this trend was not influenced by the country at all and, consequently, it is something that characterizes generic users in front of extreme events in real life.

Moreover, extreme events, like COVID-19, particularly influenced user behavior at the very beginning of the event (see Figures 5–10) where, due to surprise and lack of knowledge about the event itself, users become bewildered. In this respect, we have shown that the moment in which countries presented determined reactions to the pandemic also had effects on user social posting behavior. This partially confirms our intuition that social posting behavior can be strongly affected by external factors, especially if they correspond to extreme events.

Our intuition is further confirmed by the emotional analysis (see Figures 15–20 and research questions *RQ4*, *RQ5*); however, in this case, it is evident that emotion expressions were strongly characterized by countries, and culture, beside governmental actions. The presence of an extreme event in real life already amplifies emotion expression; consequently, any approach exploiting emotions to characterize users must consider that specific events happening outside the social network may bias user behavior and trick the algorithms.

Finally, concerning research question *RQ6*, our results provide meaningful insights about the importance of considering non-standard, but event-specific emotions like skepticism. In particular, results show that skepticism can be used as a tool for predicting potentially dangerous situations, given that skepticism is related to a decrease in attention of people to the risks of the pandemic.

## 8. Conclusions

In this work, we presented a study focusing on how dramatic events affect social posting behaviors. In particular, we concentrated on the recent COVID-19 pandemic situation and how it affected ordinary social posting life. In order to better characterize it, we focused on social posting on Reddit, and in particular on the topic-oriented content posted by users in megathreads. Thus, we carried out a topic-centered, multilingual and country-based detailed analysis.

Three main research results have been obtained: (i) Posting behavior, in terms of submissions, comments and scores, was strongly affected by extreme events (it actually diverged from the usual behavior), and modifications were uniform across countries. (ii) Emotion expression was amplified during extreme events; the variability of single emotions was strongly related to countries and other external factors. Moreover, not all emotions were impacted by the events. (iii) Non-standard emotions, particularly related to the extreme event, can be interpreted as a forerunner for new events; in particular, as for

COVID-19, an increase in skepticism can be related to an increase in the pandemic in the corresponding country.

### 8.1. Managerial Implications

A timely, automatic and focused analysis of the stance of users, carried out on user-initiated content rather than on questionnaires, may provide managers with a powerful tool to better understand and communicate with people. As an example, monitoring skepticism in the COVID-19 pandemic may allow managers and governors to take fast countermeasures in order to anticipate a worsening of the situation. The present work can be considered as a basis in helping the academic community in the development of such tools.

### 8.2. Practical/Social Implications

Having observed that posting behavior and emotion expression were actually affected by real-life events, all approaches for user behavior classification and content recommendation should be revisited in order to avoid the bias possibly introduced by such events. Academic researchers should keep in mind that automatic tools for content recommendation must be adaptive and aware of factors external to the data itself, in order to improve the appropriateness of recommendations and the reliability of results. Moreover, the method we introduced for analyzing a non-conventional emotion, like skepticism, can be generalized to other emotions; as a consequence, new tools and new perspective studies on emotions can be devised based on our work.

### 8.3. Limitations and Future Research

The limitations of the present study are related to the size of the sample and to the number of analyzed countries. Increasing the observation period and the number of countries may yield further insights. The unavailability of linguistic corpora annotated with emotions in several languages dictates the employment of dictionary-based methods for emotion detection, which tend to perform worse than state-of-the-art machine learning approaches. Moreover, we kept the method as simple as possible in order to better compare results across the languages we considered. We plan to assess the applicability of cross-lingual zero-shot learning methods to multilingual emotion detection in order to fill such gaps in future research. There are several interesting future research directions for the present work. First of all, it would be interesting to apply topic modeling techniques to identify different facets of the main topic; as an example, identifying the stance of users about economy, health and policies within discussions about COVID-19 may allow to carry out even more refined analyses. We plan also to address, in a more systematic way, the analysis of unconventional emotions also in contexts different from COVID-19, in order to identify critical emotions directly expressing important feelings for the context. Finally, we plan to adapt anomaly detection techniques we defined in other contexts [46] in order to analyze variations in emotionality expression. In fact, they can be used to identify “triggers” in registered emotions pushing for reactions of managers and governments.

**Author Contributions:** Conceptualization, F.C. and G.T.; methodology, V.B., F.C. and G.T.; software, V.B. and F.C.; validation, G.T.; investigation, V.B., F.C. and G.T.; data curation, F.C.; writing—original draft preparation, F.C. and G.T.; writing—review and editing, V.B., F.C. and G.T.; visualization, F.C.; supervision, G.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not Applicable, the study does not report any data.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Liu, B.F.; Fraustino, J.D.; Jin, Y. Social Media Use During Disasters: How Information Form and Source Influence Intended Behavioral Responses. *Commun. Res.* **2016**, *43*, 626–646. [CrossRef]
2. Zhang, C.; Fan, C.; Yao, W.; Hu, X.; Mostafavi, A. Social media for intelligent public information and warning in disasters: An interdisciplinary review. *Int. J. Inf. Manag.* **2019**, *49*, 190–207. [CrossRef]
3. Roy, K.C.; Hasan, S.; Sadri, A.M.; Cebrian, M. Understanding the efficiency of social media based crisis communication during hurricane Sandy. *Int. J. Inf. Manag.* **2020**, *52*, 102060. [CrossRef]
4. Cordeiro, M.; Gama, J. Online social networks event detection: A survey. In *Solving Large Scale Learning Tasks. Challenges and Algorithms*; Springer: Cham, Switzerland, 2016; pp. 1–41.
5. Ling Lo, S.; Chiong, R.; Cornforth, D. An unsupervised multilingual approach for online social media topic identification. *Expert Syst. Appl.* **2017**, *81*, 282–298.
6. World Health Organization. Coronavirus Disease (COVID-19) Pandemic. 2020. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed on 6 September 2020).
7. Yoo, M.; Lee, S.; Ha, T. Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit. *Inf. Process. Manag.* **2019**, *56*, 1565–1575. [CrossRef]
8. Dutta, S.; Das, D.; Chakraborty, T. Changing views: Persuasion modeling and argument extraction from online discussions. *Inf. Process. Manag.* **2020**, *57*, 102085. [CrossRef]
9. Cauteruccio, F.; Corradini, E.; Terracina, G.; Ursino, D.; Virgili, L. Investigating Reddit to detect subreddit and author stereotypes and to evaluate author assortativity. *J. Inf. Sci.* **2020**, *28*. [CrossRef]
10. Financial Times. How Countries around the World Are Battling Coronavirus. 2020. Available online: <https://www.ft.com/content/151fa92c-5ed3-11ea-8033-fa40a0d65a98> (accessed on 6 September 2020).
11. Plutchik, R. The Nature of Emotions. *Am. Sci.* **2001**, *89*, 344. [CrossRef]
12. Kim, J.; Hastak, M. Social network analysis: Characteristics of online social networks after a disaster. *Int. J. Inf. Manag.* **2018**, *38*, 86–96. [CrossRef]
13. Saura, J.; Reyes-Menendez, A.; Palos-Sanchez, P. Are Black Friday Deals Worth It? Mining Twitter Users' Sentiment and Behavior Response. *J. Open Innov. Technol. Mark. Complex.* **2019**, *5*, 58. [CrossRef]
14. Saura, J. Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *JIK* **2020**, *11*. [CrossRef]
15. Allington, D.; Duffy, B.; Wessely, S.; Dhavan, N.; Rubin, J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychol. Med.* **2020**, 1–7. [CrossRef] [PubMed]
16. Kim, J.; Bae, B.; Hastak, M. Emergency information diffusion on online social media during storm Cindy in U.S. *Int. J. Inf. Manag.* **2018**, *40*, 153–165. [CrossRef]
17. Katembo Kavota, J.; Kala Kamdjoug, J.R.; Fosso Wamba, S. Social media and disaster management: Case of the north and south Kivu regions in the Democratic Republic of the Congo. *Int. J. Inf. Manag.* **2020**, *52*, 102068. [CrossRef]
18. Jamali, M.; Nejat, A.; Ghosh, S.; Jin, F.; Cao, G. Social media data and post-disaster recovery. *Int. J. Inf. Manag.* **2019**, *44*, 25–37. [CrossRef]
19. Pennycook, G.; McPhetree, J.; Zhang, Y.; Lu, J.G.; Rand, D.G. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychol. Sci.* **2020**, *31*, 770–780. [CrossRef]
20. Block, P.; Hoffman, M.; Raabe, I.J.; Dowd, J.B.; Rahal, C.; Kashyap, R.; Mills, M.C. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nat. Hum. Behav.* **2020**, *4*, 588–596. [CrossRef]
21. Ahmed, W.; Vidal-Alaball, J.; Downing, J.; López Seguí, F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *J. Med. Internet Res.* **2020**, *22*, e19458. [CrossRef]
22. Depoux, A.; Martin, S.; Karafillakis, E.; Preet, R.; Wilder-Smith, A.; Larson, H. The pandemic of social media panic travels faster than the COVID-19 outbreak. *J. Travel Med.* **2020**, *27*, taaa031. [CrossRef]
23. Zhang, J.S.; Keegan, B.C.; Lv, Q.; Tan, C. A Tale of Two Communities: Characterizing Reddit Response to COVID-19 through /r/China\_Flu and /r/Coronavirus. *arXiv* **2020**, arXiv:2006.04816.
24. Cinelli, M.; Quattrocioni, W.; Galeazzi, A.; Valensise, C.; Brugnoli, E.; Schmidt, A.L.; Zola, P.; Zollo, F.; Scala, A. The COVID-19 Social Media Infodemic. *arXiv* **2020**, arXiv:2003.05004.
25. Gozzi, N.; Tizzani, M.; Starnini, M.; Ciulla, F.; Paolotti, D.; Panisson, A.; Perra, N. Collective response to the media coverage of COVID-19 Pandemic on Reddit and Wikipedia. *arXiv* **2020**, arXiv:2006.06446.
26. Stokes, D.C.; Andy, A.; Guntuku, S.C.; Ungar, L.H.; Merchant, R.M. Public priorities and concerns regarding covid-19 in an online discussion forum: Longitudinal topic modeling. *J. Gen. Intern. Med.* **2020**, *35*, 2244–2247. [CrossRef] [PubMed]
27. Low, D.M.; Rumker, L.; Talkar, T.; Torous, J.; Cecchi, G.; Ghosh, S.S. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: An observational study. *J. Med. Internet Res.* **2020**, *22*, e22635. [CrossRef] [PubMed]
28. Murray, C.; Mitchell, L.; Tuke, J.; Mackay, M. Symptom extraction from the narratives of personal experiences with COVID-19 on Reddit. *arXiv* **2020**, arXiv:2005.10454.
29. Li, X.; Zhou, M.; Wu, J.; Yuan, A.; Wu, F.; Li, J. Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions. *arXiv* **2020**, arXiv:2005.14464.

30. Nability-Grover, T.; Cheung, C.M.; Thatcher, J.B. Inside out and outside in: How the COVID-19 pandemic affects self-disclosure on social media. *Int. J. Inf. Manag.* **2020**, *55*, 102188. [\[CrossRef\]](#)
31. Schild, L.; Ling, C.; Blackburn, J.; Stringhini, G.; Zhang, Y.; Zannettou, S. "Go eat a bat, Chang!": An Early Look on the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19. *arXiv* **2020**, arXiv:2004.04046.
32. Kleinberg, B.; van der Vegt, I.; Mozes, M. Measuring Emotions in the COVID-19 Real World Worry Dataset. *arXiv* **2020**, arXiv:2004.04225.
33. Kuchler, T.; Russel, D.; Stroebel, J. *The Geographic Spread of COVID-19 Correlates with Structure of Social Networks as Measured by Facebook*; Technical Report; National Bureau of Economic Research: Cambridge, MA, USA, 2020.
34. Li, L.; Zhang, Q.; Wang, X.; Zhang, J.; Wang, T.; Gao, T.; Duan, W.; Tsoi, K.K.; Wang, F. Characterizing the Propagation of Situational Information in Social Media During COVID-19 Epidemic: A Case Study on Weibo. *IEEE Trans. Computat. Soc. Syst.* **2020**, *7*, 556–562. [\[CrossRef\]](#)
35. Nemes, L.; Kiss, A. Social media sentiment analysis based on COVID-19. *J. Inf. Telecommun.* **2020**, 1–15. [\[CrossRef\]](#)
36. Kumar Rathore, A.; Vigneswara Ilavarasan, P. Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products. *Int. J. Inf. Manag.* **2020**, *50*, 111–127. [\[CrossRef\]](#)
37. Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; Blackburn, J. The Pushshift Reddit Dataset. In Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020), Atlanta, GA, USA, 8–11 June 2020; pp. 830–839.
38. Straka, M.; Straková, J. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Vancouver, BC, Canada, 3–4 August 2017; pp. 88–99.
39. Nivre, J.; de Marneffe, M.C.; Ginter, F.; Goldberg, Y.; Hajič, J.; Manning, C.D.; McDonald, R.; Petrov, S.; Pyysalo, S.; Silveira, N.; et al. Universal Dependencies v1: A Multilingual Treebank Collection. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA), Portorož, Slovenia, 23–28 May 2016; pp. 1659–1666.
40. Mohammad, S.M. Word Affect Intensities. In Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018), Miyazaki, Japan, 7–12 May 2018.
41. Miller, A. WordNet: A Lexical Database for English. *Commun. ACM* **1995**, *38*, 39–41. [\[CrossRef\]](#)
42. Gennaro, P.; Rossi, A.; Tamburini, F. The FICLIT+CS@UniBO System at the EVALITA 2014 Sentiment Polarity Classification Task. In Proceedings of the First Italian Conference on Computational Linguistics and the Fourth International Workshop EVALITA 2014, Pisa, Italy, 9–11 December 2014; pp. 93–97.
43. Mohammad, S.; Bravo-Marquez, F.; Salameh, M.; Kiritchenko, S. SemEval-2018 Task 1: Affect in Tweets. In Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, LA, USA, 5–6 June 2018; pp. 1–17.
44. Calimeri, F.; Cauteruccio, F.; Cinelli, L.; Marzullo, A.; Stamile, C.; Terracina, G.; Durand-Dubief, F.; Sappey-Marinier, D. A Logic-Based Framework Leveraging Neural Networks for Studying the Evolution of Neurological Disorders. *Theory Pract. Log. Program. (TPLP)* **2021**, *21*, 80–124. [\[CrossRef\]](#)
45. Kušen, E.; Cascavilla, G.; Figl, K.; Conti, M.; Strembeck, M. Identifying Emotions in Social Media: Comparison of Word-Emotion Lexicons. In Proceedings of the 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Prague, Czech Republic, 21–23 August 2017; pp. 132–137. [\[CrossRef\]](#)
46. Cauteruccio, F.; Cinelli, L.; Corradini, E.; Terracina, G.; Ursino, D.; Virgili, L.; Savaglio, C.; Liotta, A.; Fortino, G. A framework for anomaly detection and classification in Multiple IoT scenarios. *Future Gener. Comput. Syst.* **2021**, *114*, 322–335. [\[CrossRef\]](#)