SN Computer Science (2023) 4:11
https://doi.org/10.1007/s42979-022-01421-5

ORIGINAL RESEARCH

Using Social Media Categorical Reactions as a Gateway to Identify Hate Speech in COVID-19 News

Luciana Oliveira[1] · **Joana Azevedo**[2]

Received: 15 March 2022 / Accepted: 16 September 2022 / Published online: 16 October 2022

Abstract

The spread of COVID-19 news on social media provided a particularly prolific ground for emotional commotion, disinformation and hate speech, as uncertainty and fear grew by the day. In this paper, we examine the media coverage of the COVID-19 outburst in Portugal (March–May 2020), the subsequent emotional engagement of audiences and the entropy-based emotional controversy generated as a gateway to detect the presence of hate speech, using computer-assisted qualitative data analysis (CAQDAS) embedded in a cross-sectional descriptive methodology. Our results reveal that negative and volatile categorical emotions ("Angry", "Haha" and "Wow") serve as main engines for controversy, and that controversial news have the highest sharing ratio. Moreover, using a small sample of the most controversial news with the highest overall emotional engagement, we establish a relation between the entropy-based emotional controversy obtained from Facebook's click-based reactions and the presence of cultural and ethnic hate speech, plausibly confirming the click-based categorical emotions as a gateway to hatred comment pools. In doing so, we also reveal that negative emotions alone do not always indicate the presence of hate speech, which may sprout in seemingly humorous social media posts where irony proliferates, and negativity is not apparent. This work adds to the literature on social media categorical emotion detection and its implications for the detection of hate speech.

Keywords Social media · Facebook · COVID-19 · Emotions · Entropy · Controversy · Hate speech · Media coverage

Introduction

In December 2019, the SARS-CoV-2 virus was unleashed from Wuhan. On 12th January 2020, the World Health Organization (WHO) recorded 41 cases and one fatality, and by 11th March, a global pandemic had been declared. Since then, with the community-sustaining transmission, the globe has been changed into a heavily infected environment. Daily activities were halted or restricted worldwide, and individuals were confined to their homes in an unprecedented situation, wholly unprepared and uncertain of how the crisis would unfold. The stay-at-home movement drove news outbreaks into social media, where viewers had quick access to material that would have been otherwise unavailable via conventional means.

Social media platforms have profoundly impacted the journalism industry in recent years [1, 2]. While the traditional value creation process in the news industry has been company-centric and self-contained, with little interaction with consumers, consumer value creation in the social era is part of a bigger transformation of the media and society [3]. Network journalism is a structural concept that pervades the global journalistic sphere, affecting journalists, organizations, and audiences, as journalistic narratives began to rely on public and immediate audience participation [4]. As stated by [5], we currently live in an informational and networked society as a result of the digital and global communication era.

This article is part of the topical collection "Web Information Systems and Technologies 2021" guest edited by Joaquim Filipe, Francisco Domínguez Mayo and Massimo Marchiori.

✉ Luciana Oliveira
Lgo@eu.ipp.pt

Joana Azevedo
joanabpsazevedo@gmail.com

[1] CEOS.PP ISCAP Polytechnic of Porto, Rua Jaime Lopes Amorim, s/n, 4465-004 S. Mamede de Infesta, Portugal

[2] ISCAP Polytechnic of Porto, Rua Jaime Lopes Amorim, s/n, 4465-004 S. Mamede de Infesta, Portugal

Additionally, pandemics pose not only collective health risks, but also daily difficulties for mental and public health. Strong [6] states that, before an epidemic, fear (being a carrier of the illness), moralization (moral reactions to the epidemic itself, which may be good or bad), and action (rational or irrational changes in daily habits in response to the disease) may also spread among individuals. Additionally, he emphasizes that these are generated by language and gradually nurtured by it through the various social interactions.

The way people communicate their thoughts, their emotional state, and their reactions to a subject can be used to determine the impact of events and news on their lives. Collective emotions arise when a large number of people share one or more emotional states, which tends to happen in online communities [7], and they can spread like a virus [8]. Moreover, collective feelings tend to last longer than individual emotional responses [9], amplifying the extent of a crisis. Thus, studying the general population's behavior may aid in identifying atypical affective dynamics, which have been linked to mental diseases such as depression. [10, 11].

Social media platforms were deemed vital in this environment and quickly became a popular venue to receive and share news updates and express personal opinions about the pandemic. With the enormous influx of health, normative, political, and economic information social media quickly became the focal point for communication and engagement, facilitating the sharing of thoughts and emotions. As a result, it has developed into a thriving field for studying how people cope with crises and respond to uncertainty, providing a window into the current social landscape.

In this paper, we recover some of the findings of previous work [12] devoted to profiling news outlets in Portugal, based on the media coverage of the COVID-19 outburst (March–May 2020), the emotional engagement of audiences and the entropy-based emotional commotion generated, which is translated in the controversy produced around the phenomena. We then expand the entropy-based controversy analysis and explore its relation to the presence of hate speech in the comments to evaluate if it might consist of a relevant gateway to identify hate speech associated to the pandemic, using computer-assisted qualitative data analysis (CAQDAS), applied to a small sample of posts and comments.

Background

In this section, we briefly refer to the media coverage of COVID-19, the use of click-based reactions as proxies to analyze public emotions and emotion-based controversial news on social media.

COVID-19 Media Coverage

The media are critical for bridging the divide between science and society, as citizens rely on the media to inform their attitudes, perspectives, and behavior. The media coverage of the pandemic has been found to have a substantial impact on people's perceptions of the epidemic's origins, attitudes toward appropriate governmental measures, and general politicization of the crisis [13, 14].

Pearman and Boykoff [14] reported on a steeply rise in media coverage of COVID-19 events in 102 high-circulation newspaper sources across 50 countries around the world, as other pressing matters, such as climate change, dropped drastically. Additionally, the authors argue that, despite the fact that the COVID-19 outbreak continues to spread rapidly throughout the world, its media coverage has dwindled since the crisis's initial rush of attention in early 2020. Oliveira and Sequeira [15] confirmed the similar pattern in Portugal, claiming that media coverage remained very low following the pandemic's initial wave, even as the country experienced the second, and most severe, wave of infections. As the authors suggest, this is a reflection of the typical and expected volatility in the amount of attention paid to a public issue, as indicated by the issues-attention cycle model Downs [16].

The issue-attention cycle is a term that refers to the fluctuating level of public or media interest in a particular subject (Downs, 1972), and includes five stages. The first is the pre-problem phase, when an issue does not get much public notice. Only a few individuals, like specialists or interest groups, are aware of it. In the second phase, public awareness grows, and a time of alarming discovery may ensue. But this is frequently coupled with the idea that taking action would fix the issue. The third stage occurs when individuals realize that addressing the issue is bigger and more resource-intensive than they thought. The fourth phase is characterized by a gradual loss of public attention and a sense of detachment, even though the issue persists. In the last phase, issues are replaced by new ones, causing "spasmodic recurrences of interest" [16].

The issues-attention cycles, as developed by Downs, apply to both media coverage of news and audiences' interest and engagement with those same issues, as they can evolve at different rates.

Social Media Emotions

Social media emotions are increasingly being used to acquire a deeper understanding of audiences' behavior. Emotion detection involves categorizing text into several emotion categories. Some studies in this domain have

identified sentiment analysis and emotion detection under the umbrella of sentiment analysis, but they are different [17]. Emotions are more expressive than sentiments since they do not need a feeling to exist [18, 19].

Emotion models may be dimensional or categorical/discrete [19]. Valence, arousal, and dominance are three temporal dimensions of the dimensional models [20]. A contemporary example is Pellert, Schweighofer, and Garcia's model of emotional dynamics on social media (2020). The most well-known categorical emotion model includes the emotions anger, disgust, fear, happiness, sadness, and surprise [21]. The author sees emotions as distinct, instinctive reactions to global, cultural, and personal events [22]. Several studies have utilized Ekman's work to assess public mood by automatically classifying social media content. For example, Ofoghi and Mann [23] studied Twitter emotions linked to Ebola, and Li and Zhou [24] studied cultural emotional disparities between America and China to portray public affection dynamics during COVID-19.

Oliveira and Sequeira [15], Giuntini and Ruiz [25] believe that the attribution of emotions and polarity suggests that there may be a connection between the emotions felt and the reactions exhibited in the virtual world.

Users frequently utilize emoticons on Facebook in posts, conversations, and comments to communicate additional meaning without having to write. Emoticons are small images or combinations of diacritical symbols that serve as a substitute for nonverbal communication components [25]. Emoticons have become the most popular way to communicate feelings on social media [26], and several studies have built upon emotions and emoticon reactions on social media (such as [25, 27, 28].

Giuntini found significant links between the set of fundamental emotions and the Facebook click-based reactions set. For instance, "Angry" means angry, "Wow" means surprised, "Sorrowful" means sad, and "Love" means pleasure. “Like” is ambiguous in terms of polarity and sentiment. Fear is the only fundamental emotion that has no corresponding visible reaction [25]. However, click-based responses remain an underused resource in social media research, despite quick-draw, ready-made expressive features are becoming more common across various platforms, attracting research interest in recent years [29].

Controversy on Social Media

It is well established that media attention has been disproportionately focused on COVID-19 news, with little regard for how pandemic-related media coverage may impact people's mental health [11]. Some of the most recent risks and potential dangers of social media communication have been aggravated by the tremendous spread of COVID-19 news and information. In fact, along with a pandemic caused by a lethal virus, the globe has been experiencing an "infodemic", as defined by the World Health Organization [30, 31]. This is a reference to the epidemic of false or wrong information that is swiftly spreading via social media's fertile ground, fuelled by the fear, worry, and uncertainty caused by this new peril.

The massive effort spent to countering false news, the emergence of worldwide alliances (such as Poynter), and the increased collaboration between journalists and social media platforms have all contributed to ensuring that legacy media sources do not propagate disinformation or misinformation.

The spread of fake news, however, is not the only threat fostered by the COVID-19 infodemic. User-generated content (UCG) remains as one of the main challenges in controlling the spread of fake news [32], a challenge that escalates in the spread of hate speech, which requires a lot less creativity and effort from users. Dori-Hacohen and Sung [11] assert that good online conversation is becoming increasingly difficult to reach as the noise of dispute, mis- and disinformation, and toxic speech grows.

Controversial heated discussions are a prolific field for hate speech on social media, and according to Dori-Hacohen and Sung [11], controversy is also saliently connected with disinformation. One of the main current challenges of hate speech recognition is the automatic detection of irony [33] because people verbalize an idea while implying the opposite meaning; thus, textual features alone fail in recognizing the implicit meanings of the discourse.

Irony serves the additional social and emotional functions of projecting emotions like humor or anger, and ironic comments may provoke stronger emotional responses than literal comments [34]. In their research about irony, the authors introduce paralinguistic features (emoticons) to improve the detection of praise and criticism in written messages. Such methods had already been employed by other studies such as Carvalho and Sarmento [35] and Derks and Bos [36].

More recently, with the expansion of the Facebook “Like” button into a broad set of click-based emotional reactions to content, additional studies have emerged that take advantage of the convenience of the systematized and bulk emotional response that is immediately captured to examine the emotional irony conveyed by audiences.

This research stream is predicated on the premise that controversial posts divide a community's preferences, garnering both substantial positive and negative responses or polarized toward extremes (e.g., “Love”—“Angry”). As such, these works build on the study of social media click-based emotions such as the one conducted by Freeman and Alhoori [29], who measured the Pearson’s correlation coefficients for all reaction pairs in their dataset of scholarly articles published on Facebook; or the work of Tian and Galery [28], who used a K-means to cluster reactions and

investigate which reactions were most likely to be seen together on a post in UK, US, France and Germany.

Related research using Facebook reactions as proxies to identify controversy can be found in Sriteja and Pandey [37], who have used this method for detecting controversial topics during the US Presidential elections 2016. Basile, Caseli, and Nissim [38], also followed the same procedure to identify controversy among four major Italian newspapers and one media agency, using an entropy-based model to compute the 'disorderliness' of emotional reactions to posts. Finally, Gray [39] studied gender bias in the Facebook pages of the United States 2020 Senate candidates, using the exact same method as Basile, Caseli, and Nissim [38].

Agile strategies for detecting controversy early on may be beneficial in supporting news organizations, journalists, social media platforms, and fact-checkers in preventing hate speech and disinformation.

Emotions and Hate Speech

The precise definition of hate speech continues to be a point of contention, as it is a subjective and highly interpretable concept [40–42]. For instance, according to Nockleby [43], the term "hate speech" refers to communication that disparages an individual or a group on the basis of some attribute, like as race, color, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics. In a more systematic way, the United Nations (UN) Strategy and Plan of Action on Hate Speech defines hate speech, according to three major components, as "any kind of (1) communication in speech, writing or behavior, that (2) attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their (3) religion, ethnicity, nationality, race, color, descent, gender or other identity factor" [44]. According to the Strategy, hate speech is communication that is prejudicial, bigoted, intolerant, or prejudiced ("discriminatory"), or contemptuous or demeaning ("pejorative") of an individual or group on the basis of their identity. Yet, the UN excludes the State, its offices and symbols, the status of public affairs, religious leaders, doctrine and tenets of faiths as objects of hate, stating the only individuals or groups can be considered targets.

From the broader sense of the definition, however, it becomes quite clear that hate speech and offensive language walk alongside. In fact, most of the current automatic procedures to detect hate speech consist of Natural Langue Processing (NLP) tasks to detect cursing and prejudiced words. According to Plaza-del-Arco [45], hate speech is related to sentiment analysis because hate speech is typically negative in nature and expresses a negative opinion, and it is also related to emotion analysis because expressed hatred indicates that the author is experiencing (or pretending to experience) anger, while the addressees are experiencing (or are intended to experience) fear.

On top of the limitations associated with detecting hate speech caused by the employment of irony, to which we referred in the previous section, the absence of agreement regarding its general definition and the definition of its variations, complicates the hate speech annotation. Annotating hateful content remains subjective and culturally dependent, frequently resulting in low inter-annotator agreement and in a paucity of high-quality training data for creating supervised hate speech detection algorithms [46]. Moreover, Markov [40], while recognizing the additional challenges posed to NLP by the rapid evolution of the offensive vocabulary and keywords, identifies the need to investigate more abstract features, less susceptible to specific vocabulary, topic or corpus bias, which can be analyzed in in-domain and cross domain settings, such as different languages and cultures.

The author advanced the hypothesis that the style and emotional dimension of hateful textual content can provide useful cues for the detection of hate speech. Eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive) from the NRC emotion lexicon, were used to encode the types of emotions in a message and to capture how high-emotional or low-emotional a message is, for English and Dutch, leading to improved robustness in the detection of hateful content.

Several other authors have been incorporating discrete or categorical emotion analysis in NLP procedures do enhance the detection of hate speech.

Martins [47] enhanced the detection of hate speech using sentiment analysis with discrete emotions. The authors discovered that sentences could be grouped in groups of emotions: positive (anticipation, joy, trust) and negative (anger, disgust, fear and sadness), and that the most critical emotions to identify hate speech are anger, disgust, fear and sadness. The emotion surprise was interpreted as being a neutral emotion.

Also recurring to a ten-set discrete emotion analysis, Alorainy et al. [48] used suspended Twitter accounts, and discovered that these provide tweets with more disgust, less joy and the highest frequency of emotions in total than the active accounts. The study also concluded that suspended accounts produce a larger number of tweets containing all the ten emotions, meaning that some accounts tried to use positive emotions (e.g., trust) along with negative emotions to soothe the attitude presented in the tweets. The authors also found a higher frequency of negative emotion in neutral tweets than in hateful ones, concluding that negativity does not always indicate the presence of hate.

Rodriguez [49] applied sentiment and emotion analysis to detect clusters of Facebook posts from eight pages that contain highly negative tones, namely posts suspected to

instigate hatred, composing a set of "sensitive topics". The authors used VADER and JAMIN for sentiment and emotion analysis, and to filter texts that do not contain negative opinions. In the same way as Martins, discrete emotions were also used with the specific aim to identify anger and disgust in Facebook posts and comments, as well as bullying.

Plaza-del-Arco [45] binary sentiment analysis and emotion classification of tweets (anger, disgust, fear, joy, sadness, surprise, enthusiasm, fun, hate, neutral, love, boredom, relief, none) to detect offensive language and, thus, hate speech. The authors highlight that both sentiment and emotion analysis benefit from each other, and that there is an improvement in sarcasm detection when emotion and sentiment are both considered.

More recently, Rana et al. [50] also used emotion analysis to detect hate in multimedia content in a work that the authors define as the first multimodal deep learning framework to combine the auditory features representing emotion and the semantic features to detect hateful content. In this case, due to the challenge of discrete representation of complex emotions such as anger or fear, the authors use a dimensional representation of emotions defined by valence, arousal, and dominance attributes.

Undoubtedly, this is a growing stream of research, expanding the diversity of attributes used in the detection of hate speech while building on sentiment and discrete emotions.

However, these works build exclusively on NLP to detect emotions, leaving room to explore the click-based reactions that convey the intentional emotional state that the author/commenter wished to convey (scarce as reaction set might be). Therefore, our work focuses on establishing this connection between the entropy-based controversy generated around Facebook's click-based reactions to posts and as a gateway to identify stances in which hate speech proliferates.

Methods and Procedures

This work follows the general approach of a quantitative content analysis [51], and it consists of a cross-sectional descriptive study with an embedded component of CAQDAS. We used the Facebook Graph API to retrieve the news posted by the three major daily news providers in Portugal, Sic Notícias (1,717,794 followers), TVI 24 (1,088,453 fans) and CMTV (580,703 followers) between the 1st March and 31st May, 2020. The news outlets were chosen on the basis of three criteria: a) identified as a "TV channel" Facebook page; b) high visibility, as measured by the number of fans; and c) a generalist editorial line, covering a broad range of topics and not segregated for specialized audiences. The analysis period was defined by the government's first mandatory confinement, which included the detection of the first case of infection (first week of March 2020) and the announcement of the first measures of deconfinement (May 2020). Thus, the total duration of the analysis is three months.

The dataset, provided by the Facebook Journalism Project, is composed of 30,607 news posted on the network, for which we collected the created date and time, link (news external URL), message (text included in the post), link text (the title of the news), description (news lead), Likes, Comments, Shares, Love, Wow, Haha, Sad, Angry and Care. We refer to "Like" (somewhat a default type of interaction with content), "Comment" and "Share" as forms of interaction with content; and to "Love", "Wow", "Haha", "Sad", "Angry" and "Care" as reactions, in the sense that these convey emotional responses. The dataset of news was manually categorized into two subsets: COVID-19 news and Other news and their subdomains (e.g., politics, education, prevention, etc.). For this stage of the research, we refer only to the top-level binary categorization of COVID-19 and Other news, as our first set of goals is to a) characterize and compare the media coverage given to COVID-19 in news outlets, b) explore the public response to these news, namely their emotional state and c) identify the most controversial news and their content.

For the analysis of media attention and audiences' emotional involvement, we follow the general principles of the issues-attention cycles proposed by Downs [16] and the detection of emotions through Facebook's click-based reactions, as used by Giuntini and Ruiz [25]. For the analysis of controversial news, we follow Basile, Caseli, and Nissim [38] model and compute the entropy (quantitative measure of disorder) of the Facebook's reaction set per post as a function to determine controversy. For the analysis of the presence of hate speech in the most controversial news, we used CAQDAS in MaxQDA, and coded a random sample of four hundred user comments to the most controversial news, following the categories of hate speech identified by Guterres [44]. We have also categorized comments according to the response type, following a similar procedure to the one proposed by Zubiaga and Liakata [52]. We identify offensive language, such irony, aggression and insult according to Kreuz [53]. Finally, we identified approval (the author of the comment approves of the action reported in the news), disapproval (the author of the comment does not approve), appeal for more information (the author of the comment request additional information) and comments with not expression of taking a position toward the subject being presented on the news.

Table 1 provides an overview of the data collected, depicting the post type for each news outlet and topic of news—COVID-19 news ("COV") and Other news ("Oth").

For all news outlets, "Link" is the most frequent post type, which is consistent with the current practice of sharing

Table 1 Total posts per outlet and category [12]

Type	SICNotícias		TVI24		CMTV	
	COV	Oth	COV	Oth	COV	Oth
Link	11,866	5760	4013	3966	1935	1941
Video	3	99	30	134	281	217
Photo	0	21	10	325	0	1
Status	0	3	0	0	0	2
NSub	11,869	5883	4053	4425	2216	2161
NTot	17,752		8478		4377	
N%	66.86	33.14	47.81	52.19	50.63	49.37

news links directly from their news portals. Photos and videos are rarely posted and are more frequent for CMTV and TVI24. The news outlet with the highest number of posts, i.e., the highest communication investment, is Sic Notícias, four times higher than CMTV and two times higher than TVI24. Additionally, this is the entity with the highest rate of COVID-19 news posted in the trimester (66.86%), followed by CMTV (50.63%) and TVI 24 (47.81%).

Results and Discussion

In this section, we present the results concerning the evolution of the media attention given to COVID-19 news, the emotional response from the audiences, the detection of emotion-based controversy and its implications in the detection of hate speech.

Media Coverage

To understand how people's emotions changed over time, we look at how COVID-19 news was covered by the media.

During the fourteen weeks of the trimester, the media paid more attention to COVID-19 than to other news. This is shown in Fig. 1. For the sake of good data visualization, we show the same trio of outlets' audiences' emotional engagement with the news next to each other. This is shown in Fig. 2. Five key moments are marked to provide a clearer insight on the national context regarding the (1) first case of infection in the country (2nd March), (2) first confinement measures (12th March), (3) declaration of the State of Emergency and total lockdown (19th March), (4) declaration of the State of Calamity and the first stage of deconfinement measures (3rd May), and (5) second stage of deconfinement measures (17th May).

It is possible to notice an overall validation of Downs's hypothesized issues-attention cycles [16], which is typically represented by a bell-shaped curve with a stretched right side to imply that the subject takes longer to fade away than it did to achieve its peak of interest. This extended right side may then experience spasmodic events of interest or events that result in minor elevations (small bumps) that never reach the initial position of startling discovery. This is particularly visible in Fig. 1c, concerning CMTV, where two spasmodic occurrences happen in weeks 7 and 12, and in Fig. 1b, concerning TVI24, in weeks 8 and 9. For the trio of news outlets, the stage of alarmed discovery happens in week 4, which includes all communication and news regarding the declaration of the State of Emergency and total lockdown. The reason why the percentages of news in week 14 for all outlets is very low is that this week only refers to one day, 31st May.

Despite this broad trend, three distinct behaviors in terms of the intensity and duration of attention paid to COVID-19 news are discernible. CMTV (Fig. 1c), despite publishing the fewest COVID-19 news stories (Table 1), had the highest percentage of media coverage of the phenomenon, reaching a peak of nearly 14% in a shorter time span (seven weeks straight), before drastically decreasing coverage. This, we believe, is consistent with the outlet's reputation for sensationalism, which is bolstered further by the fact that media coverage began later and with a more dramatic increase.

Both TVI24 (Fig. 1b) and Sic Noticias (Fig. 1a) exhibit a more gradual decline in media coverage, fluctuating between 6 and 10% for nine consecutive weeks, at which point Other news exceeds the volume of COVID-19 news. Additionally, it is worth noting that TVI24 (Fig. 1b) was the only outlet with a lower discrepancy between COVID-19 and Other news coverage.

Interaction and emotional engagement

The evolution of audiences' interaction with news ("Comments" and "Shares") and emotional engagement, as computed using Facebook's click-based reaction set ("Love", "Wow", "Haha", "Sad", and "Angry"), is depicted in Fig. 2. As previously explained, we excluded "Likes" and the reaction "Care" because they were introduced mid-period in the first week of April, precluding consistent comparisons.

Prior to contextualizing the audiences' emotional engagement throughout the trimester, we analyze the audiences' emotional profiles by news outlet. Table 2 summarizes the

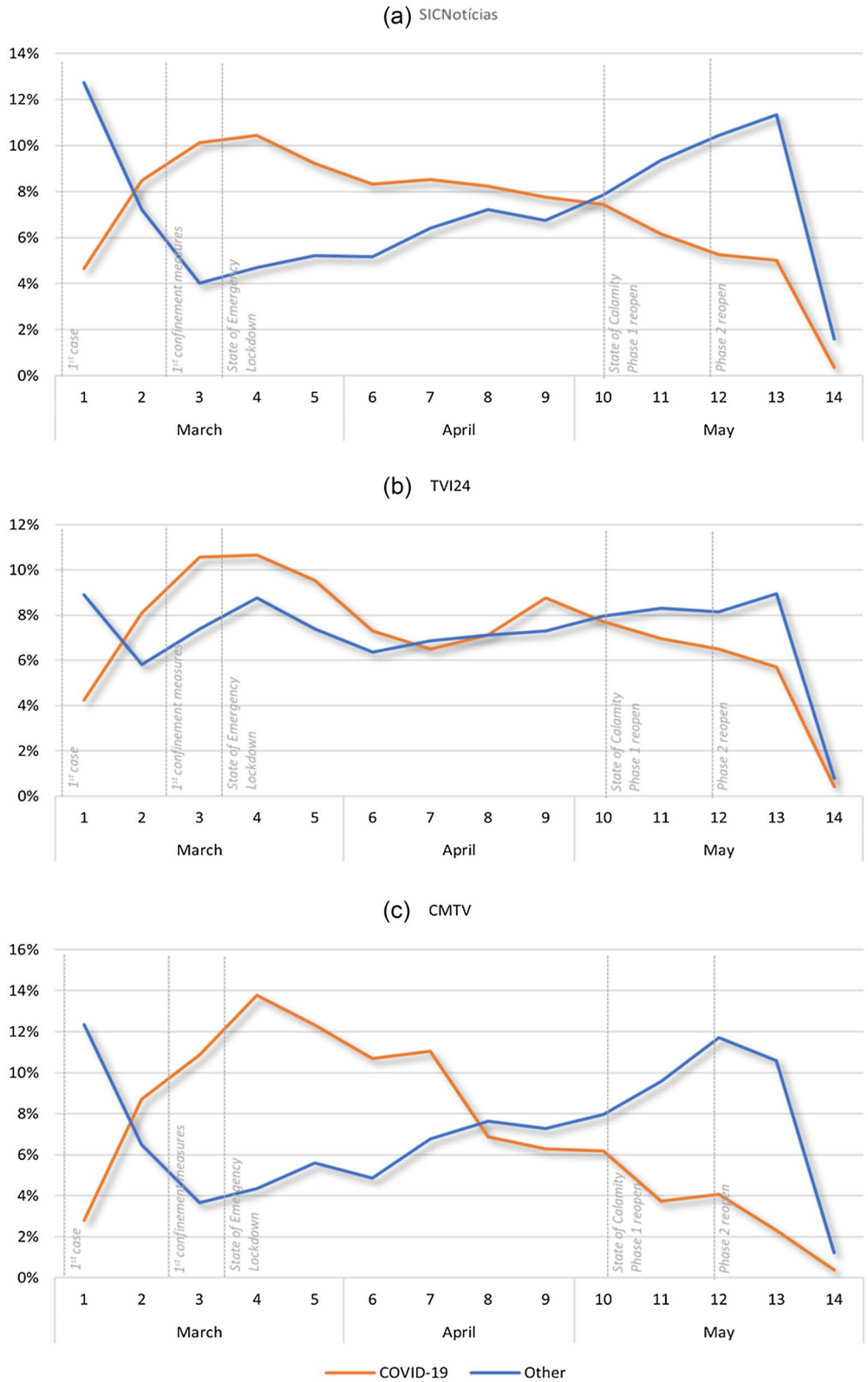

Fig. 1 Evolution of the percentage of COVID-19 news and Other news per outlet, per month and week of analysis – **a** SICNotícias, **b** TVI24, **c** CMTV [12]

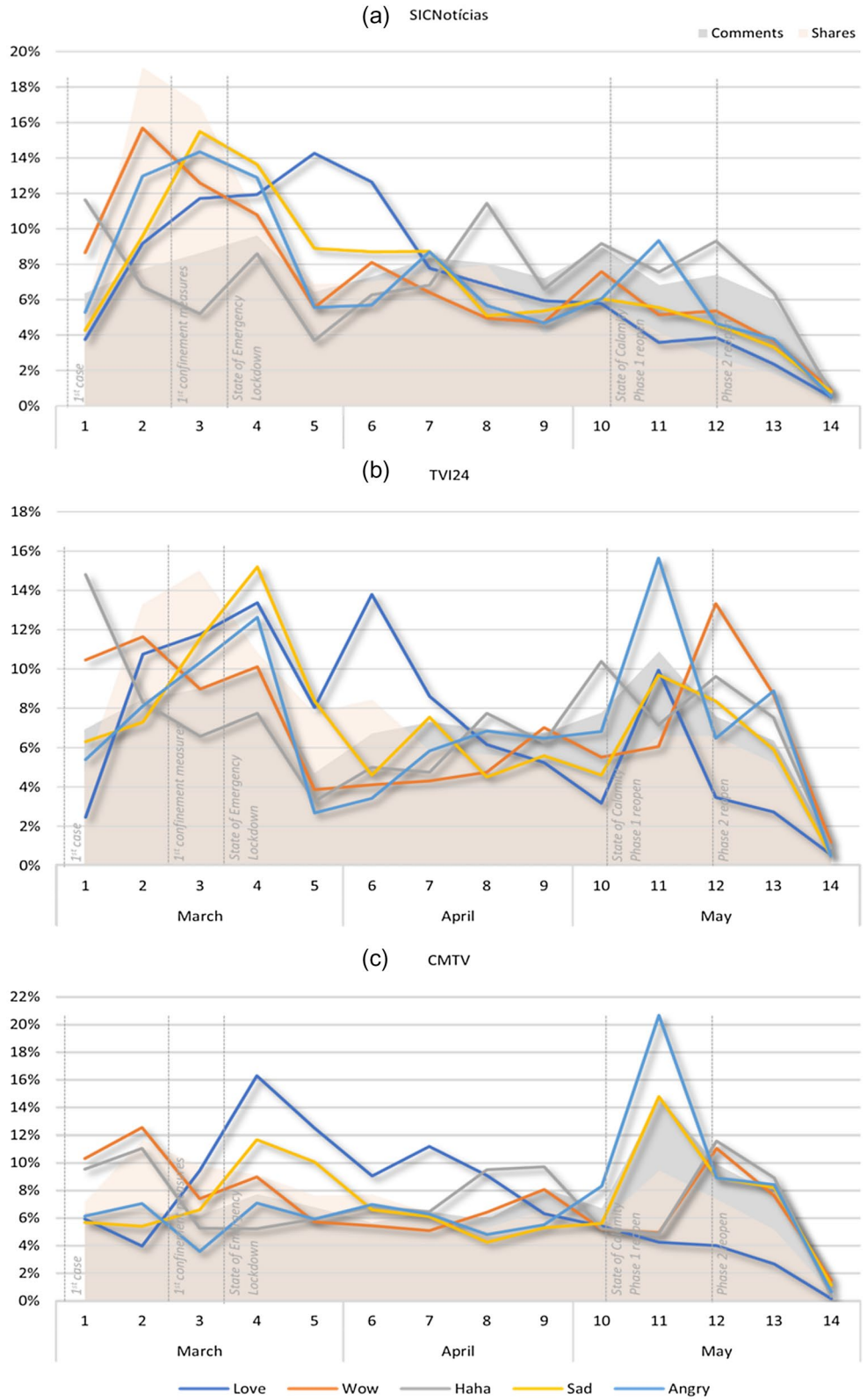

Fig. 2 Evolution of the percentage of Facebook emotions, comments and shares per outlet, per month and week of analysis – **a** SICNotícias, **b** TVI24, **c** CMTV [12]

Table 2 Average interaction per outlet and news topic

Type	SICNoticias		TVI24		CMTV	
	COV	Oth	COV	Oth	COV	Oth
Love	10	6	11	45	9	8
Wow	9	4	14	11	11	11
Haha	6	8	6	10	6	7
Sad	35	14	57	55	50	57
Angry	12	9	16	27	20	35
Com	46	32	52	86	42	58
Shares	89	28	119	149	119	98

average interaction and emotion levels for each outlet and news topic, emphasizing the statistically significant differences identified by a one-way ANOVA test. Figure 2 also demonstrates an overall prevalence of sadness and anger.

TVI24 is, undoubtedly, the news outlet generating the highest emotional commotion among audiences, for almost all types of reactions, ["Love" ($p<0.001$), "Wow" ($p<0.001$), "Haha" ($p<0.005$), "Sad" ($p<0.001$)], except for "Angry", which predominates in CMTV's audiences ($p<0.001$). This commotion of emotions is quite visible in Fig. 2b, for the entire period. The news outlet is also the one registering the highest average of "Comments" ($p<0.001$) and "Shares" ($p<0.001$) per post.

We conducted a secondary analysis to determine whether or not the variance in emotional expression is related to the presence of COVID-19 news. A series of t-tests were conducted, and the results are summarized in Table 3, which highlights significant differences.

In the case of Sic Notícias, the emotional reactions "Love", "Wow" and "Sad" are significantly associated with COVID-19 news ($p<0.001$), as well as the interactions "Comments" and "Shares" ($p<0.001$). "Haha" is more present in Other news ($p<0.001$). This is readily apparent in Figs. 1a and 2a, as audience emotional expression reflects the decline in COVID-19 media coverage. For all news organizations, sharing is more prevalent during the first six weeks of the trimester, while commenting is more prevalent during the final six, particularly for CMTV. This is consistent with the dissemination of new information about the COVID-19 outburst, followed by a period of public sharing of views following an abundance of information. In the case of TVI24, most of the emotional reactions are directed at Other news—"Love", "Haha", and "Angry" ($p<0.001$). The same occurs with "Comments" ($p<0.001$). Surprise, conveyed by "Wow", is the most common reaction to COVID-19 news ($p<0.05$).

In the case of CMTV, the only significant differences found reside in audiences sharing mainly COVID-19 posts ($p<0.05$), and commenting ($p<0.001$) and expressing anger ($p<0.001$) on Other news. For the remaining emotions, there are no statistically significant differences, as they are expressed toward both types of news.

It was only for Sic Notícias that COVID-19 news have expressively modeled the emotional attitude of audiences. The inverse is true for TVI24, where Other news is more reactive. In CMTV, emotional behavior is more diffuse, with an increased tendency toward verbalization ("Comments") and emotional expression, particularly anger toward Other news.

This, we believe, is entirely consistent with the journalistic discourse employed by each news organization. For example, in Fig. 1c, we can see that CMTV has an isolated shorter sequence of news events. On the audience side, we observe a strong preference for sharing (spreading) COVID-19 news/information, which is typical during the alarmed discovery stages. TVI24's coverage of COVID-19 news was

Table 3 Average emotions and interaction per outlet and news topic

Type	SICNoticias		TVI24		CMTV	
	COV	Oth	COV	Oth	COV	Oth
Love	10.03	6.05	11.01	44.51	9.35*	7.63
Wow	9.33	4.49	13.61	11.35	10.57	10.86*
Haha	6.40	8.32	6.21	10.22	5.50	7.40*
Sad	34.60	13.92	57.33*	54.61	50.31	56.93*
Angry	12.12	9.01*	15.71	27.22	20.46	35.39
Com	46.20	32.23	51.86	86.28	42.11	57.74
Shares	88.75	27.61	118.80	148.56*	118.81	98.28

*n.s.

not as distinct as that of the other outlets, but Other news was never completely ignored. On the audience side, the majority of reactions are directed toward Other news and "Comments," with only surprise ("Wow") directed toward COVID-19 news. In Sic Noticias, the most extensive, persistent, and sustained coverage of COVID-19 news has resulted in a significant emotional outpouring of audiences' love, surprise, and sadness for this type of news (except for laughter).

This leads us to ascertain that the media coverage and journalistic discourse significantly impact the audiences' emotions and are provided with the ability to prolong sadness or joy, hope or frustration, depression or wellbeing, in any ordinary context, but especially in periods of crisis when people are more sensitive.

Given the three distinct emotional profiles of audiences, we examined the correlations between emotions and interactions within and across news outlets to ascertain how they mutually reinforce one another and to determine their polarity. The following significant Pearson's correlations were found ($p < 0.01$).

Sic Notícias.

Moderate: Wow-Sad ($r = 0.572$).
Moderate: Haha-Comments ($r = 0.580$).
Moderate: Angry-Comments ($r = 0.481$).
Weak: Sad–Angry ($r = 0.324$).

TVI24

Strong: Love-Comments ($r = 0.724$).
Strong: Love-Shares ($r = 0.703$).
Moderate: Share-Comments ($r = 0.531$).
Moderate: Sad-Shares ($r = 0.436$).
Weak: Wow-Sad ($r = 0.375$).
Weak: Angry-Comments ($r = 0.345$).
Weak: Sad-Comments ($r = 0.312$).
Weak: Sad-Angry ($r = 0.275$).
Weak: Haha-Comments ($r = 0.263$).

CMTV

Moderate strong: Angry-Comments ($r = 0.643$).
Moderate: Wow-Shares ($r = 0.588$).
Moderate: Comments-Shares ($r = 0.558$).
Moderate: Sad-Shares ($r = 0.533$).
Moderate: Angry-Shares ($r = 0.530$).
Moderate: Sad-Angry ($r = 0.471$).
Moderate: Sad-Comments ($r = 0.445$).
Weak: Wow-Sad ($r = 0.366$).
Weak: Haha-Comments ($r = 0.342$).
Weak: Wow-Comments ($r = 0.328$).

There is a strong correlation between negative emotions (the pairs Sad-Angry and Wow-Sad) and the highest interaction rates with news across the three news organizations (comments and shares). Negativity appears to be the primary motivator for engaging with news and disseminating information. There is one notable exception in the case of TVI24, where interaction is also strongly correlated with positivity toward Other news (the pairs Love-Comments and Love-Shares). The pairs Angry-Comments and Haha-Comments are also evident among the trio of outlets.

Laughter and surprise, conveyed by the click-based reactions "Haha" and "Wow" consist of volatile emotions, as they can acquire distinct polarity according to other prevalent emotions they are paired with. For instance, the pair Haha-Love can translate into passion, affection, friendship, happiness, amusement, joy and fun.

By contrast, the pair Haha-Angry can convey rage, fury, frenzy, indignation, scorn, contempt, cynicism, and irony. The correlation analysis above reveals the prevalence of these latter types of associations, in which emotional volatility tends toward negative polarity. Additionally, we believe that this demonstrates the presence and/or prevalence of sarcasm, which is generally defined as content that elicits both positive and negative feedback [54], or, in our case, falling into two or more classes of emotion that may or may not be diametrically opposed in terms of polarity.

Since our research is aimed at revealing cues for linking controversy and hate speech, we further explore the media outlets' controversy profiles and most controversial, considering the COVID-19 and Other news.

Controversy and Hate Speech

Following Hessel and Lee's [54] methodology to determine controversy, we computed the entropy of the set of Facebook reactions per post, according to the entropy formula shown below, where x_i is the number of each reaction for a post, and $p(x_i)$ is the ratio of that reaction to the total reactions on a post.

$$H(X) = -\sum_{i=1}^{n} P\left(x_i\right) logP(x_i) \quad (1)$$

We consider that if the users' reactions fall into two or more emotion categories with a high frequency, the controversy surrounding a news item is greater; thus, the greater the entropy, the greater the controversy. Table 4 contains examples to aid in comprehension.

Users' differing responses indicate that a text is likely to be controversial, as shown by the high values of entropy (H), as demonstrated in examples b) and c). The overall profile of controversy per news outlet, based on the entropy means is presented in Table 5.

Table 4 Examples of variation of entropy per post

	Love	Wow	Haha	Sad	Angry	H
(a)	32	0	0	0	0	0
(b)	12	12	13	9	9	2.30
(c)	26	80	26	62	222	1.85

Table 5 Overall profile of controversy per news outlet, based on entropy means

	N	Mean	SD	Max
SICNoticias	17,752	0.795	0.648	2.321
TVI24	8478	0.993	0.611	2.321
CMTV	4377	0.929	0.615	2.251
Total	30,607	0.869		2.321

Table 6 Average entropy per news type and outlet

Type	Outlet	N	Mean	Max	MeanTot
COVID-19 news	SICNotícias	11,869	0.855	2.321	0.895
	TVI24	4053	0.997	2.311	
	CMTV	2216	0.924	2.252	
Other news	SICNotícias	5883	0.674	2.252	0.831
	TVI24	4425	0.989	2.322	
	CMTV	2161	0.934	2.246	

Considering the overall average of entropy for the full news dataset, and according to our previous reasoning, Sic Notícias is the entity producing the news with the least controversial potential (below average). The news outlet TVI24 has the highest overall average entropy (0.993), followed by CMTV (0.929) and SICNotícias (0.795) ($F(30{,}604) = 303.870$; $p < 0.001$). Both TVI24 and CMTV present above-average entropy values, and TVI24 leads in the amount of controversy produced.

Our total entropy average is slightly lower than the total entropy average reported ($H = 0.9386$) by Basile, Caseli, and Nissim [38], who analyzed four Italian newspapers and one news agency. The Italian newspaper with the highest average of entropy is *Il Gionale* ($H = 1.127$), an openly biased right-wing newspaper. Although this was not a feature in the detection of sarcasm in the Italian case, it is curious to notice that the two Portuguese media outlets with higher entropy averages are also (not openly) right-wing news outlets, according to the European Journalism Observatory [55].

This reality, however, may be altered by the COVID-19 phenomenon, as our dataset spans the pandemic's outbreak in Portugal. As a result, we believe it is necessary to examine the level of controversy generated specifically by COVID-19 news, as shown in Table 6.

Table 7 Average of reactions and interactions per (un)controversial news

	COVID-19 news		Other news	
	Contr	Uncont	Contr	Uncont
Love	6.35	11.11	8.70	22.43
Wow	13.27	9.74	12.30	7.10
Haha	15.02	4.08	18.79	6.66
Sad	18.22	47.36	17.02	39.91
Angry	22.68	11.79	24.47	19.08
Comments	90.08	36.33	102.81	45.59
Shares	124.78	92.81	81.99	82.95

We found statistically significant differences between the average entropy among the types of news and news outlets. On average, COVID-19 news have higher entropy (0.895) than Other news (0.831) (t (26,112) = 8.529; $p < 0{,}001$).), as depicted in Table 6. However, since we try to profile the news outlets, we analyzed these differences within their subsets of news, also included in Table 6.

A set of independent samples t test only confirms significant differences of entropy between news categories for Sic Notícias (higher in COVID-19) and CMTV (higher in Other news), although with no significant differences for CMTV.

Still, the overall averages of entropy are relevant for both categories of news and overall more prevalent on COVID-19 news, namely when considering other entropy values reported in the literature [38, 39].

Thus, we examined which Facebook reactions contributed the most to the emergence of controversy. To do so, we annotated the dataset, labeling as "Controversial" all news with entropy values greater than one standard deviation above the mean entropy value for each news outlet (c.f. Table 5). The results indicate statistically significant differences in the average distribution of Facebook reactions and interactions for controversial and noncontroversial news (t test), which we illustrate in Table 7 by news category.

Both for COVID-19 and Other controversial news, the most prevalent reactions, in decreasing order of average ($p < 0.005$):

"Angry".
"Haha".
"Wow".

The remaining emotions, "Sad" (47.36) and "Love" (11.11) are significantly associated with noncontroversial news ($p < 0.005$).

Considering the interactions with the news, "Comments" are always substantially higher in controversial news ($p < 0.001$), but the average of "Shares" is significantly higher for COVID-19 controversial news.

This means that controversy is primarily founded on negative ("Angry") and volatile emotions ("Haha", "Wow"), reinforcing the concept of irony. Considering Hessel and Lee [54] views on controversy not always being a bad thing, specifically in bringing up an issue that merits a spirited debate that can benefit community health, we believe this is not the case. Indeed, irony is rarely conducive to a civilized and constructive debate. However, this requires, for instance, content analysis over the comments posted in controversial news for further elaboration.

We also observe that the COVID-19 controversial news are the ones harvesting higher "Shares", i.e., they consist of the news with the highest reach and potential of spread of controversy on social media. This contradicts Freeman and Alhoori [29], who state that content that is more likely to inspire a negative reaction from users is less likely to be shared. Bellow, we present the top twelve most controversial COVID-19 news with the highest number of "Reactions", and the corresponding number of "Shares" and media outlet.

As it is possible to observe, these results are in line with the average entropy per news type and outlet (Table 6), in which SIC Notícias is the media outlet with the highest values in COVID-10 news, followed by CMTV in Other news.

These news titles cover, essentially, social, economic, and political issues revolving around the first measures of the control of the pandemic in the country. Provided that these consists of the news with the highest emotional arousal combined with a relevant spread on the social network, we speculated that we would find hate speech toward people not complying with confinement measures (non-compliance), toward ethnic groups and other minorities (such as criminal offenders) as well as toward politics or public figures.

As Markov et al. [40] has identified in previous research, automated hate speech detection in social media is a challenging task that has recently gained significant traction in the data mining and Natural Language Processing community; however, it has been dependent heavily on the annotated hate speech datasets, which are imbalanced and often lack training samples for hateful content. Using sentiment and emotions has become a relevant procedure in identifying hate speech. Several authors have used similar approaches [40, 47–49], focused mainly on detecting hate, anger and sadness.

Using computer-assisted qualitative data analysis (CAQDAS) with MaxQDA, we have analyzed the user's response to the top four most controversial news (sic_01, sic_02, sic_03 and sic_04 in Table 8), to characterize the approval and disapproval of actions reported, to identify if hate speech is present and which type of hate have these news produced. To do so, we selected the first one hundred comments for these news (in a subset of a total of four hundred comments), which we categorized according to hate speech, offensive language and response type [44, 52, 53]. In Table 9, we depict the results of the CAQDAS analysis, considering that content codes could be assigned to a full comment or just a part of it, thus resulting in a total of seven hundred and thirty six coded segments.

We were able to identify five types of hate speech (Ethnicity, Political, Cultural Nationality/Regional, and Religious), with the predominance of Cultural (8.15%) and Ethnicity hate (5.43%); four types of offensive language (Aggression, Insults, Sarcastic Humor, and Irony), with the prevalence of Irony (12.64%) and Insults (11.55%); and predominance of Disapproval (38.59%) as main response type to the actions reported by the news.

Examples of Cultural hate include prejudice toward youngsters, people intelligence and the media, in which aggression is visible:

This is the picture of pure stupidity. Pure ignorance? While health professionals, policy makers, businessmen, work themselves to exhaustion, these "kids", show a lack of civic duty. Lamentable. (sicnoticias02_80).

It was to break all of their legs, and one arm, so that they could not walk even on crutches (sicnoticias02_49).

Bunch of ignorant people 😠 (sicnoticias02_22).

Make news of really important things and not things that add nothing to our society. (sicnoticias04_88).

Ethnicity hate, in our particular case, is directed to Romany individuals (of the pejorative Gypsies), with relation to delinquency and abusive consumption of social support resources, while maintaining an expensive live style:

They have never contributed anything to society what do they expect. Ask the "cousins" for help. (sicnoticias04_48).

There's always a little drug to get you off the hook… (sicnoticias04_77).

The remaining forms of hate consist of attacking the Catholic church for the scarce financial support in providing medical equipment to hospitals, the government's incompetence to implement effective quarantine measures and individuals living in Portugal or in a specific Portuguese region.

Examples of Religious hate, in which aggression, insults, humor and irony are also used:

fans is what they deserved in the horns! PIMPS! (sicnoticias01_39)

WOOOOO It's the end of the world 😎 *but beware because these are fans made with the melted gold from the donations of the fools of the faithful.* (sicnoticias01_97).

Segments coded as Political hate, in which offensive language (insult) is also used:

Table 8 Top twelve most controversial news ranked by number of shares

Outlet/n.º	News	Reactions/Shares
SIC_01	**Fatima Sanctuary offers three ventilators to the National Health Service**	3028
		5047
SIC_02	**Carcavelos beach full of bathers**	2514
		1630
SIC_03	**Carriages packed on the second day of the state of emergency**	2296
		2107
SIC_04	**Without being able to sell in fairs, the gypsy community already fears hunger**	2261
		1427
SIC_05	Several people are disrespecting isolation in Felgueiras	2241
		2365
SIC_06	Naples took to the streets: what kind of quarantine is this?	1972
		716
SIC_07	Portugal available to receive a thousand migrants from Greece	1831
		1168
CMTV_01	761 prisoners released since Saturday during the State of Emergency	1749
		2645
CMTV_02	Heat takes Portuguese to the beaches on the day that was decreed pandemic due to coronavirus	1729
		4380
SIC_08	Migrants will be able to return without being quarantined	1664
		2778
SIC_09	Joacine says state of emergency 'not necessary' to fight pandemic	1589
		349
SIC_10	Trump suspends funding to WHO	1482
		1052

Table 9 Content segments categorized according to offensive language, hate speech and response type

Dimension	Code	Segments (N)	Segments (%)
Hate speech	Ethnicity	40	5.43
	Political	20	2.72
	Cultural	60	8.15
	Nationality/Regional	12	1.63
	Religious	19	2.58
Offensive language	Aggression	5	0.68
	Insult	85	11.55
	Humor	12	1.63
	Irony	93	12.64
Response type	Comment	74	10.05
	Appeal For More Information	7	0.95
	Disapproval	284	38.59
	Approval	25	3.40

And long live the state of emergency😭😭😭😭 thank you to all Portuguese politicians lost the right to criticize Adolf Hitler because they are just like him. (sicnoticias03_37).

A government with no balls to shut down and stop this country, it will get worse than Italy (sicnoticias03_05).

Finally, commenters also revolted against non-compliance, accusing and insulting the overall Portuguese people or those living in specific regions of the country:

Once again the blame for a mediocre Portugal. It's the Portuguese's fault. (sicnoticias02_81).

ahahahah…. In a country of donkeys nothing surprises me anymore ahahahah…. (sicnoticias02_86).

Shameful! Bunch of irresponsible people! Unconscious people! This is the future of Portugal.😢😢 (sicnoticias02_93).

Lisbon setting an example😠😠 (sicnoticias03_12).

For the great majority of the analyzed segments (comments), the commenters expressed disapproval for the actions reported in the news. This reinforces the previously stated notion that negativity appears as the overall main engine for interacting with news and spreading information. The relation between the content codes is presented in Fig. 3, where the strong connection between the Disapproval node with "Insults", "Irony", "Cultural hate" and "Ethnicity hate" nodes is evident. These expressions of hate are intimately linked to offensive language, as seen, and to the emotional controversy (entropy) generated around the issues. "Irony" and "Insults" are also the most common forms of offensive language and "Cultural hate" and "Ethnic hate" are the most salient forms of hate.

However, looking at the distribution of "Reactions" in the analyzed news (Table 10), we observe that for two of the four cases analyzed the prevailing emotion is not "Angry", but "Wow" and "Haha", which we have defined as volatile emotions. Our definition of volatile emotions (labeled "neutral" by Martins [47]) is, thus, reinforced by the notion that

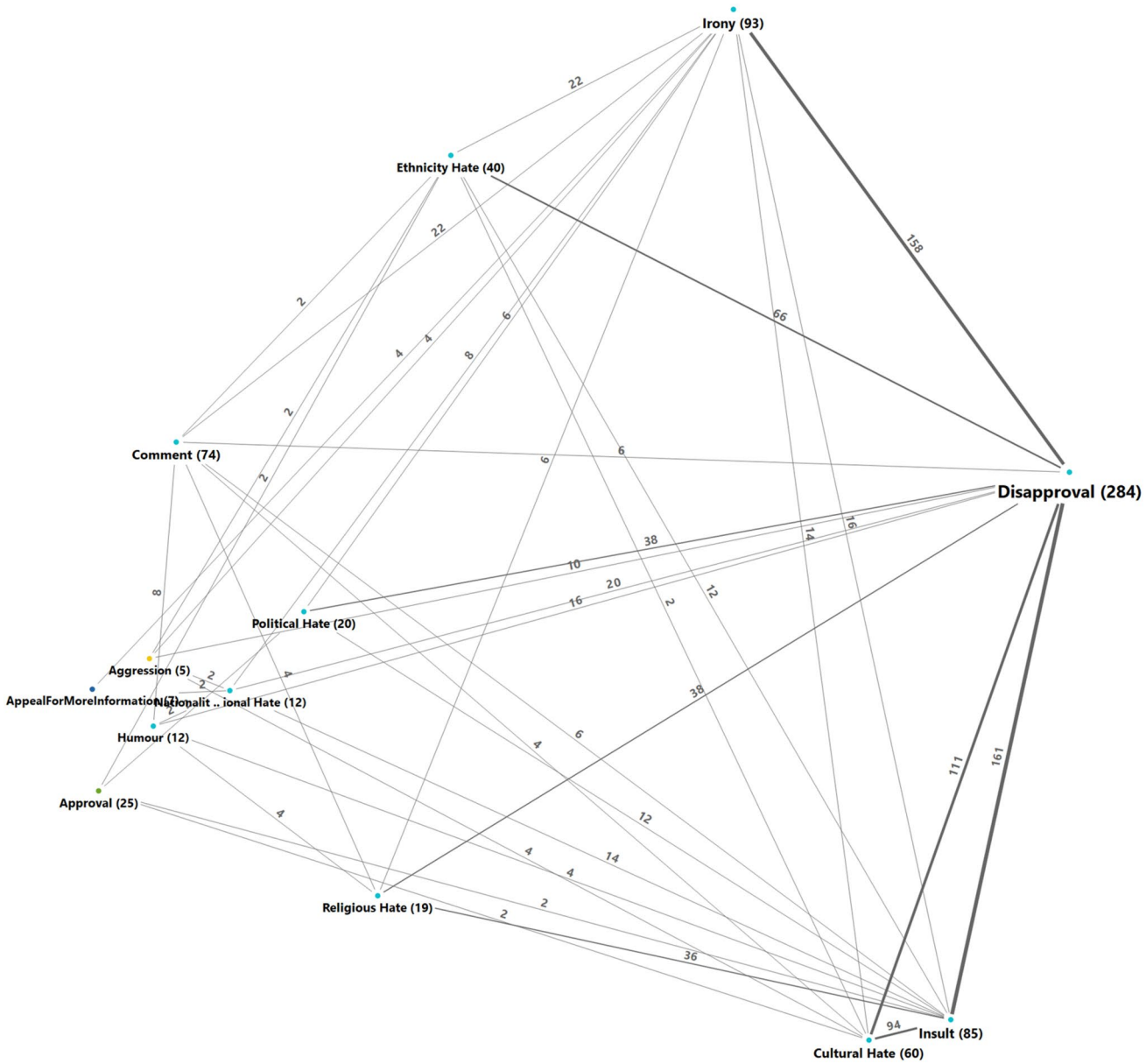

Fig. 3 Graph of the relations between content codes (hate speech, offensive language and type of response)

Table 10 Distribution of "Reactions" and controversy in the analyzed news set

	sic_01	sic_02	sic_03	sic_04
Love	291	21	6	10
Wow	248	192	265	129
Haha	1241	152	44	1210
Sad	374	696	818	770
Angry	847	1453	1163	142
Entropy	2.04	1.56	1.52	1.53

We abbreviated 'sicnoticias_01' to 'sic_01' for the economy of space.

Table 11 Percentage of hate speech, offensive language, and response type per news post

Code/post	sic_01	sic_02	sic_03	sic_04
Hate Speech				
Ethnicity	0	0	0	86.96
Political	0	1.79	60.00	2.17
Cultural	0	80.36	33.33	10.87
Nationality/Regional	0	17.86	6.67	0
Religious	100	0	0	0
Offensive language				
Aggression	0	4.05	4.17	2.56
Insult	31.03	67.57	45.83	15.38
Humor	13.79	2.70	0	5.13
Irony	55.17	25.68	50.00	76.92
Response type				
Comment	22.77	20.20	23.23	8.79
More Inf	0	2.02	2.02	3.30
Disapproval	76.24	77.78	53.54	84.62
Approval	0.99	0	21.21	3.30

irony and insults are closely linked, for instance, when the pairs of emotions "Haha" and "Angry" are dominant, but acquiring a completely different meaning when "Haha" is combined with "Love". The close link between irony and insults was also identified by Plaza-del-Arco [45]. Moreover, there is an improvement in sarcasm detection when emotion is considered, and in this case, it is intimately linked to hate speech.

The prevalence of volatile emotions (e.g., "Haha") over negative ones (e.g., "Angry") in hateful content is somewhat in line with the discoveries of Alorainy [48], who concluded that negative emotions alone do not always indicate the presence of hate speech, after having detected a full range of ten discrete emotions in a large set of hateful tweets. This is particularly relevant for the case of the fourth news post that we analyzed (sic_04), in which the prevalent emotion is "Haha", and in which the Ethnic hate proliferated (Table 11). The same applies to the news post sic_01, regarding Religious hate. The feature these posts have in common is the intensified use of irony, one of the main current challenges in automatic hate speech recognition [33], because people express an idea while implying the opposite meaning; thus, textual features alone fail in recognizing the implicit meanings of the discourse. These limitations of textual features have led several authors to introduce entropy-based emotional controversy in the detection of irony, as we have observed [35–38], although it is not common practice in the detection of hate speech.

As noted, the great majority of the research devoted to the detection of hate speech builds upon the detection of negative emotions such as "Anger", "Sadness" or "Disgust" [40, 47–49], which has patently increased accuracy in detecting hate speech, but it could benefit from considering entropy-based emotional controversy. In our analysis, similarly, we have detected relevant correlations between the pairs "Sad" and "Angry" (negative emotions), but we have also observed, from the examples in Table 10, that this correlation alone would not have allowed us to identify some of the news producing hateful comments, if that were the only the criteria adopted.

Moreover, even though our correlation analysis shows that "Anger" fosters "Comments" and "Shares", we verified that so does the emotion "Haha", which could be synonym for joy alone or irony when combined with other emotions, namely negative ones. Thus, by employing the entropy-based detection of controversy, we are not anchoring the emotional arousal of the audiences to any specific emotion, but rather evaluating disorder or dispersion.

It is also worth noticing that in the process of employing entropy-based controversy over a set of click-based emotions per post (news), we are, at first instance, attending to the emotional effect of the crowd over a specific subject as an entry point to a second instance, in which the written communication of the commenters is subjected to hate analysis (text features). In doing so, we are also capitalizing on the "social influence effect", which is known to diminish the diversity of the crowd's position toward a specific topic and to create barriers to the improvements of its collective errors [56], or to the bigoted, intolerant, prejudiced, contemptuous or demeaning speech.

We do, however, recognize that the presented results are anchored to the very small sample of news selected for analysis and, therefore, not generalizable. Also, the set of click-based Facebook reactions is smaller than others sets of emotions used by other authors [40, 47–49], which narrows the richness of the analysis of the emotional mindset of the audiences. However, at this instance, our results reveal a relevant function of the entropy-based emotion controversy based on discrete emotions in signaling stances of hate speech in social media comments.

Conclusion

In this work, we recovered some of the findings from previous work devoted to the profiling of news outlets, expanded our initial entropy-based controversy analysis by exploring its relation to the presence of hate speech in the comments to the most controversial news about COVID-19, to evaluate if it might consist of a relevant gateway to identify variations of hate.

Our results show three profiles of COVID-19 news coverage: (1) one more consistent (Sic Notícias), least controversial, with less drastic fluctuations of attention, which resulted in the significant emotional expression of audiences' love, surprise and sadness; (2) another more diffuse with approximate levels of attention to COVID-19 news and Other news (TVI24), which generated higher emotional commotion among audiences toward COVID-19-unrelated news; (3) and a more spasmodic and reactive profile of COVID-19-related and Other news, which translates into the predominance of anger among audiences (CMTV).

We have detected high levels of controversy among news outlets and among categories of news. Controversy is more prevalent in COVID-19-related news and is mostly fostered by negative and volatile Facebook reactions ("Angry", "Haha" and "Wow"). Controversial COVID-19 news was also the most shared news on Facebook during the outburst of the pandemic in Portugal.

Using a small sample of the most controversial news, with highest overall emotional engagement, we have established a relation between the computed entropy-based controversy that was generated by the Facebook's click-based reaction set and the presence of hate speech. Although we resorted to CAQDAS, instead of using automatic NLP procedures for the identification of types of hate speech, we have plausibly established that this method has potential in the detection of sarcasm, irony and hate. To the best of our knowledge, the relation between these elements and the use of entropy-based controversy supported by Facebook's click-based reactions has not been previously established in the literature.

The proposed approach consists of an agile top-down procedure for identifying potentially controversial hubs of conversation around a specific topic or news, as an entry point to detect other features, such as irony, sarcasm and hate.

These findings have ramifications for news organizations, social media managers, and society as a whole. The rapid analysis methods used in this work encourage persistent monitoring of social media to prevent the widespread spread of hate speech and unhealthy mindsets in a way that media outlets and people navigating news content on social media can immediately recognize.

This work is not without its limitations. We focused on presenting an overall overview of the main stages leading to our contribution; thus, we have favored diversity over depth in some stances. The CAQDAS was performed by a single researcher over a non-representative sample of news; thus, it is subject to a degree of bias. Moreover, our results are not generalizable, although we believe they have potential instigate other works. Future research stages are set to include a comprehensive and robust content analysis of the users' comments to the news, also considering the click-based reactions to users' comments, providing confirmation of the present tentative knowledge and effective insights on the nature of the speech surrounding the identified controversial news.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Ferrucci P. It is in the numbers: How market orientation impacts journalists' use of news metrics. Journalism. 2020;21(2):244–61.
2. Poell T. Three challenges for media studies in the age of platforms. Television & New Media. 2020;21(6):650–7.
3. Serrano, M.J.H., A. Greenhill, and G. Graham, Transforming the news value chain in the social era: a community perspective. Supply Chain Management: An International Journal, 2015.
4. Dalmaso, S.C., Jornalismo e relevância: o discurso dos leitores dos jornais de referência no Facebook. 2017.
5. Castells, M., The network society A cross-cultural perspective. 2004: Edward Elgar.
6. Strong P. Epidemic psychology: a model. Sociol Health Illn. 1990;12(3):249–59.
7. Kappas A. The psychology of (cyber) emotions. In: Cyberemotions. Springer; 2017. p. 37–52.
8. Ferrara E, Yang Z. Measuring emotional contagion in social media. PLoS ONE. 2015;10(11): e0142390.
9. Garcia D, et al. The dynamics of emotions in online interaction. Royal Society open science. 2016;3(8): 160059.
10. Koval P, et al. Affect dynamics in relation to depressive symptoms: variable, unstable or inert? Emotion. 2013;13(6):1132.
11. Dori-Hacohen, S., et al., Restoring Healthy Online Discourse by Detecting and Reducing Controversy, Misinformation, and Toxicity Online, in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021, Association for Computing Machinery: Virtual Event, Canada. p. 2627–2628.
12. Oliveira, L. and J. Azevedo, Profiling Media Outlets and Audiences on Facebook: COVID-19 Coverage, Emotions and Controversy, in Proceedings of the 17th International Conference on Web Information Systems and Technologies - WEBIST. 2021, SCITEPRESS - Science and Technology Publications.
13. Bolsen T, Palm R, Kingsland JT. <? covid19?> Framing the Origins of COVID-19. Sci Commun. 2020;42(5):562–85.

14. Pearman O, et al. COVID-19 media coverage decreasing despite deepening crisis. The Lancet Planetary Health. 2021;5(1):e6–7.
15. Oliveira, L., et al., Exploring the public reaction to COVID-19 news on social media in Portugal. arXiv preprint arXiv:2102.07689, 2021.
16. Downs, A., Up and down with ecology: The issue-attention cycle. The public, 1972: p. 462–473.
17. Balahur, A. Sentiment analysis in social media texts. in Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis. 2013.
18. Liu B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. 2012;5(1):1–167.
19. Wang, Y. and A. Pal. Detecting emotions in social media: A constrained optimization approach. in Twenty-fourth international joint conference on artificial intelligence. 2015.
20. Ekkekakis, P., The measurement of affect, mood, and emotion: A guide for health-behavioral research. 2013: Cambridge University Press.
21. Ekman P. An argument for basic emotions. Cogn Emot. 1992;6(3–4):169–200.
22. Ekman P, Cordaro D. What is meant by calling emotions basic. Emot Rev. 2011;3(4):364–70.
23. Ofoghi, B., M. Mann, and K. Verspoor. Towards early discovery of salient health threats: A social media emotion classification technique. in biocomputing 2016: proceedings of the Pacific symposium. 2016. World Scientific.
24. Li, X., et al., Analyzing Covid-19 on online social media: Trends, sentiments and emotions. arXiv preprint arXiv:2005.14464, 2020.
25. Giuntini FT, et al. How do i feel? identifying emotional expressions on facebook reactions using clustering mechanism. IEEE Access. 2019;7:53909–21.
26. Oleszkiewicz A, et al. Who uses emoticons? Data from 86 702 Facebook users. Personality Individ Differ. 2017;119:289–95.
27. Cazzolato, M.T., et al., Beyond Tears and Smiles with ReactSet: Records of Users' Emotions in Facebook Posts. 2019.
28. Tian, Y., et al. Facebook sentiment: Reactions and emojis. in Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media. 2017.
29. Freeman, C., H. Alhoori, and M. Shahzad, Measuring the diversity of facebook reactions to research. Proceedings of the ACM on Human-Computer Interaction, 2020. 4(GROUP): p. 1–17.
30. Organization, W.H. Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. 2020 23/09/2020 [cited 2020 01/12/2020]; Available from: https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation.
31. Organization, W.H. Infodemic. 2021 [cited 2021 30/04]; Available from: https://www.who.int/health-topics/infodemic#tab=tab_1
32. Ferrari, E., Sincerely Fake: Exploring User-Generated Political Fakes and Networked Publics. Social Media+ Society, 2020. 6(4): p. 2056305120963824.
33. MacAvaney S, et al. Hate speech detection: Challenges and solutions. PLoS ONE. 2019;14(8): e0221152.
34. Thompson D, et al. Emotional responses to irony and emoticons in written language: evidence from EDA and facial EMG. Psychophysiology. 2016;53(7):1054–62.
35. Carvalho, P., et al. Clues for detecting irony in user-generated contents: oh...!! it's" so easy". in Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion. 2009.
36. Derks D, Bos AE, Von Grumbkow J. Emoticons and online message interpretation. Soc Sci Comput Rev. 2008;26(3):379–88.
37. Sriteja, A., P. Pandey, and V. Pudi. Controversy Detection Using Reactions on Social Media. in 2017 IEEE International Conference on Data Mining Workshops (ICDMW). 2017.
38. Basile, A., T. Caselli, and M. Nissim. Predicting Controversial News Using Facebook Reactions. in CLiC-it. 2017.
39. Gray, L. Gender Bias Detection Using Facebook Reactions. 2020.
40. MarMarkov, I., et al. Exploring Stylometric and Emotion-Based Features for Multilingual Cross-Domain Hate Speech Detection. in Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2021.
41. Poletto F, et al. Resources and benchmark corpora for hate speech detection: a systematic review. Lang Resour Eval. 2021;55(2):477–523.
42. Waseem, Z. Are you a racist or am i seeing things? annotator influence on hate speech detection on twitter. in Proceedings of the first workshop on NLP and computational social science. 2016.
43. Nockleby JT. Hate speech. Encyclopedia of the American constitution. 2000;3(2):1277–9.
44. Guterres, A., United Nations strategy and plan of action on hate speech. Taken from: https://www.un.org/en/genocideprevention/documents/U,2019(20Strategy).
45. Plaza-del-Arco, F.M., et al., Multi-Task Learning with Sentiment, Emotion, and Target Detection to Recognize Hate Speech and Offensive Language. arXiv preprint arXiv:2109.10255, 2021.
46. Sap, M., et al. The risk of racial bias in hate speech detection. in ACL. 2019.
47. Martins, R., et al. Hate Speech Classification in Social Media Using Emotional Analysis. in 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). 2018.
48. Alorainy, W., et al. Suspended Accounts: A Source of Tweets with Disgust and Anger Emotions for Augmenting Hate Speech Data Sample. in 2018 International Conference on Machine Learning and Cybernetics (ICMLC). 2018.
49. Rodríguez, A., C. Argueta, and Y. Chen. Automatic Detection of Hate Speech on Facebook Using Sentiment and Emotion Analysis. in 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 2019.
50. Rana, A. and S. Jha, Emotion Based Hate Speech Detection using Multimodal Learning. arXiv preprint arXiv:2202.06218, 2022.
51. Bryman, A., Social research methods. 2016: Oxford university press.
52. Zubiaga A, et al. Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS ONE. 2016;11(3): e0150989.
53. Kreuz, R., Irony and sarcasm. 2020: MIT Press.
54. Hessel, J. and L. Lee, Something's Brewing! Early Prediction of Controversy-causing Posts from Discussion Features. arXiv preprint arXiv:1904.07372, 2019.
55. Cardoso, G., P. Couraceiro, and M. Ana. A esquerda no parlamento e a direita na televisão? 2019; Available from: https://pt.ejo.ch/top-stories/a-esquerda-no-parlamento-e-a-direita-na-televisao.
56. Lorenz J, et al. How social influence can undermine the wisdom of crowd effect. Proc Natl Acad Sci. 2011;108(22):9020.