Using algorithms to identify social activism and climate skepticism in user-generated content on *Twitter*

Nuria Villagra; Ana Reyes-Menéndez; Jorge Clemente-Mediavilla; Dimitrina J. Semova

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87155

Recommended citation:

Villagra, Nuria; **Reyes-Menéndez, Ana**; **Clemente-Mediavilla, Jorge**; **Semova, Dimitrina J.** (2023). "Using algorithms to identify social activism and climate skepticism in user-generated content on *Twitter*". *Profesional de la información*, v. 32, n. 3, e320315.

https://doi.org/10.3145/epi.2023.may.15

Manuscript received on 4th November 2022
Accepted on 27th April 2023

Nuria Villagra
https://orcid.org/0000-0002-2613-1530

Universidad Complutense de Madrid
Facultad de Ciencias de la Información
Avda. Complutense, 3
28040 Madrid, Spain
nuriavillagra@ucm.es

Ana Reyes-Menéndez
https://orcid.org/0000-0003-0636-9573

Universidad Rey Juan Carlos
Facultad de Ciencias Jurídicas y Sociales
Campus de Vicálvaro
Pº Artilleros, s/n
28032 Madrid, Spain
ana.reyes@urjc.es

Jorge Clemente-Mediavilla ✉
https://orcid.org/0000-0002-9819-1129

Universidad Complutense de Madrid
Facultad de Ciencias de la Información
Avda. Complutense, 3
28040 Madrid, Spain
jclemen@ucm.es

Dimitrina J. Semova
https://orcid.org/ 0000-0001-8355-5121

Universidad Complutense de Madrid
Facultad de Ciencias de la Información
Avda. Complutense, 3
28040 Madrid, Spain
djivkova@ucm.es

Abstract

Climate change has become an issue of great relevance in society in recent years, and the data provided by the scientific community recommend acting as soon as possible and forcefully. Scientists, politicians, the media, and thanks to the new media, citizens and other social agents participate in the debate on this issue. Despite the data and general consensus in the scientific community, the climate change debate is highly polarized, with skeptical voices denying or questioning climate change and using social media to amplify the reach of their message. This can encourage misinformation and polarization. This study tries to identify the key indicators of social skepticism around climate change through the analysis of users' social activism and behavioral patterns on *Twitter*. We analyze keywords, frequency, topics, and categories from a sample of 78,168 tweets. The results show, first, that there is an overlap of topics, with 24 of the 28 topics grouped in the intertopic distance map; second, that the size of the topics is relatively small and linked to specific events; and, third, that there is a significant political presence, especially from the United States. This work therefore contributes to the analysis of communication on *Twitter* about opinions against climate change.

Keywords

Climate change; Climate skeptics; Skepticism; Climate communication; Linguistic corpus; Algorithms; Social networks; Activisms; Indicators; Social media; Behavior patterns; Opinions; Politics; Polarization; *Twitter*.

Funding

This work has been carried out within the project 2021-1-ES02-KA220-YOU-000029048 *Green Artivism*. This project has been funded with support from the *European Commission* (*EC*). This publication reflects the views only of the author, and the *EC* cannot be held responsible for any use which may be made of the information contained herein.

1. Introduction

At present, social platform users actively share information about their activities, ideas, and personal experiences through their smartphones (**Dubrofsky**; **Wood**, 2014), which leads to the generation of massive amounts of data. In recent years, such user-generated content has been used extensively for research in the social sciences (**Schmid**, 2016) and has been analyzed in terms of both language aspects and contexts. The results of such analysis provide a comprehensive understanding of users' behavior (**Schmid**, 2016; **Terkourafi**; **Haugh**, 2019).

Social identity, which is an essential component of self-concept, stems from an individual's perception of their membership to a social group(s), as well as from the significance that this individual ascribes to that membership (**Tajfel**, 1974). Therefore, social identity explains how identification works from the individual, interactional, and institutional perspectives (**Jenkins**, 2014). Accordingly, social identity has been defined as the individual's concept of the self with respect to specific aspects of social behavior (**Tajfel**, 1981; **Kastanakis**; **Balabanis**, 2012; **Singh** *et al*., 2021).

Today, social activism is increasingly created through users' interactions with others via social media platforms, such as *Twitter* (**HerdaĞdelen** *et al*., 2013; **Saura**; **Rodríguez-Herráez**; **Reyes-Menéndez**, 2019).

Among the major characteristics of social media platforms is users' organization into networks, that is, communities that share common interests. This makes social platforms a valuable source of data for social scientists investigating different cultural and social issues (**Ntontis** *et al*., 2018).

Another important characteristic of social platforms is that their users collectively create and interact with content. This content, referred to as user-generated content (UGC), includes any content created by social platform users that is publicly shared with other users (**Reyes-Menéndez** *et al*., 2020) This makes social platforms such as *Twitter* structured communities where UGC offers an enriched source of users' activities (**Fujita**; **Harrigan**; **Soutar**, 2018).

Recent decades have witnessed the emergence of social activism (**Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018; **Pearce** *et al*., 2019; **Moernaut** *et al*., 2022). The rapid development of this new social activism has been supported by the emergence and rapid spread of new means of communication, such as email, websites, or social platforms. By facilitating the rapid spread of content around the globe, these new channels have proven to provide a valuable opportunity for people to shape their collective support (**Van-de-Donk** *et al.*, 2004).

Among other social platforms, *Twitter* is a very popular platform on which users can share information about their ideas, activities, opinions (**Aswani** *et al*., 2018; **Monfort**; **Villagra**; **López-Vázquez**, 2019; **Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018), and location (**HerdaĞdelen** *et al.*, 2013). Through *Twitter*, users have the opportunity to share information and their thoughts. Accordingly, there is a growing need to investigate these user interactions on *Twitter* from the perspective of social science (**Stieglitz** *et al*., 2018).

The information shared by *Twitter* users can include tweets (original content created by a user) or retweets (content that users share with others); see Table 1 for a summary of such types of interaction.

Table 1. Major types of interaction on *Twitter* by grade of engagement

Interaction	Description	Presence on Twitter	UGC type
Profile mention	User A makes a mention of User B's profile in their tweet	@[Original Profile]	Profile
Tweet	A *Twitter* user writes an online post (tweet)	[Original Tweet]	Text
Retweet	User A shares User B's tweet in their profile, thereby expanding the audience	RT: @[Original Profile] Original Tweet]	Text
Like	User A presses the "like" button on User B's tweet	Like [Original Tweet]	Action
Hashtag	User A includes a tweet with a hashtag (#). Clicking the hashtag gives access to all comments published with this hashtag	#[Hashtag]	Text

Source: **Reyes-Menéndez** *et al.* (2020).

The analysis of the network structure on social platforms enables an analysis of not only the individual activities of specific users but also their social activism. Starting social activism on a social platform is facilitated by the joint performance of the following three effects: network structure, collaboration, and the interaction of users (**Ntontis** *et al.*, 2018; **Saura**; **Reyes-Menéndez**; **Álvarez-Alonso**, 2019). Social activism can emerge around a social profile, for instance, @GretaThunberg (**Olesen**, 2022), or a hashtag (#), such as #WorldEnvironmentDay (**Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018).

The issue of climate change has grown in importance in recent decades, causing greater social concern about the effects it may have in the future.

Public interest in climate change is growing, and the European Union, for example, has earmarked significant funding within the *Horizon 2020* program for the study of this issue. The *European Green Deal Research and Innovation Program* funded a study aimed at collecting data on climate change and human opinions via *Twitter*, spanning 13 years and including more than 15 million tweets spatially distributed around the world. The variables analyzed were geolocation, user gender, climate change stance and sentiment, aggressiveness, deviations from historical temperature, topic modeling, and information about environmental disaster events. The data provided by scientists, dramatic extreme weather events, and the reports published periodically by the *Intergovernmental Panel on Climate Change* (*IPCC*), have highlighted the need to act as soon as possible and forcefully. As pointed out by **Eide** and **Kunelius** (2021), the year 2018 represented a turning point toward a general discourse on the subject, further promoting activism and conveying a message of urgency. Just like the *European Parliament*, instead of talking about climate change, they talk about a "climate emergency." This idea of urgency and risk if action is not taken quickly is also enhanced by movements such as *Fridays for Future*, which have had significant social repercussions and have demonstrated the importance of citizen mobilization, in this case led by young people. All this favors an activist stance in relation to the issue that no longer involves traditional agents such as nongovernmental organizations (NGOs) but rather citizens.

Despite the data and general consensus in the scientific community, the climate change debate is highly polarized

Despite the data and general consensus in the scientific community, the climate change debate is highly polarized (**Dunlap**; **McCright**, 2011; **Elgesem**; **Steskal**; **Diakopoulos**, 2015; **Hoggan**, 2009; **Washington**; **Cook**, 2011; **Moernaut** *et al.*, 2022; **Pearce** *et al.*, 2019), and along with the voices that promote social awareness, there is also a current of thought that denies or questions climate change and downplays its effects or the role that people themselves have on it (anthropogenic climate change theory).

Social platforms have become an environmental protest space where users express their opinions and concerns about this topic. For example, the findings of research on the #WorldEnvironmentDay tag include conclusions that, among all the Sustainable Development Goals (SDGs), those of most concern to users are related to the environment and public health, such as climate change, global warming, extreme weather, water pollution, deforestation, climate risks, acid rain, and massive industrialization.

In this sense, relying on the aforementioned research, it becomes clear that the *Twitter* platform offers an opportunity to analyze UGC related to environmental issues such as climate change (**Pearce** *et al.*, 2019; **Moernaut** *et al.*, 2022) from a user perspective by enabling an analysis of both types of interactions: those organized around a profile and those organized around a hashtag.

In this context, we seek in this work to understand the social skepticism around climate change through an analysis of users' social activism and behavioral patterns. The analysis was performed on a UGC corpus of a total of 78,168 tweets using textual analysis techniques. The first of these techniques was latent Dirichlet allocation (LDA), a machine-based technique, applied in combination with a corpus linguistic approach. We also performed discourse analysis using the log-likelihood and mutual information (MI) statistical measures.

The research question (RQ1) addressed in the present study is: What are the key indicators of social skepticism around climate change according to the analysis of users' social activism and behavioral patterns on *Twitter*?

In what follows, we explain the theoretical framework of the present study (Section 2). This is followed by a description of the data collection process and the methodology (Section 3). The results are reported in Section 4. Section 5 presents the discussion. Conclusions, limitations of the present study, and directions for further research are presented in Section 6.

2. Theoretical framework

Previous research has focused on understanding social activism around climate change (**Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018; **Pearce** *et al.*, 2019; **Moernaut** *et al.*, 2022). However, the conversation on climate change is highly polarized (**Elgesem**; **Steskal**; **Diakopoulos**, 2015; **Pearce** et *al.*, 2019; **Moernaut** *et al.*, 2022). In general, people adopting these two positions are referred to as accepters/believers and skeptics in literature. Authors such as **Washington** and **Cook** (2011) question the use of the term "skeptics" and propose that it would be more correct to call those who oppose the theory of anthropogenic climate change "deniers". However, we use the term "skeptics" herein to refer to both those who deny as well as those who question or minimize the scientific data or theories that indicate that climate change is taking place, because this term is most commonly used in previous studies (**Capstick**; **Pidgeon**, 2014; **Kaiser**; **Rhomberg**, 2016; **Moernaut** *et al.*, 2022; **Van-Eck**; **Feindt**, 2022).

To better understand which factors can influence a person to adopt one position or another, various studies have been carried out. On the one hand, the influence of political ideology on climate change opinion has been studied (**Anderson**; **Huntington**, 2017; **Van-Eck**; **Feindt**, 2022; **Whitmarsh**; **Corner**, 2017). In general, the literature that relates positions on climate change with ideology distinguishes between left and right or liberal and conservative (**Elgesem**; **Steskal**; **Diakopoulos**, 2015; **Matthews**, 2015). The results indicate a tendency for those who defend more conservative positions to show less concern about climate change than those who defend a leftist position. However, more work is needed since one study carried out in Germany by **Engels** *et al.* (2013) found a negative correlation between political participation and skepticism. Another factor that has been studied is the influence of geographical region. **Whitmarsh** and **Capstick** (2018), for example, state that there is more climate skepticism in Western countries. A study carried out by **Hagen**, **Middel** and **Pijawka***,* (2016) in different countries of the European Union (Spain, the Netherlands, the United Kingdom, and

Germany) and studies carried out in the United States (**Smith**; **Leiserowitz**, 2012) and Great Britain (**Corner**; **Markowitz**; **Pidgeon**, 2014; **Capstick** *et al.*, 2015) also highlight skepticism in public opinion and further suggest that it has become especially marked over the last two decades.

Some of the factors that are argued to be possible reasons for the greater skepticism in public opinion in recent years are

- news in the media and skeptical positions defended by politics (**Corner**; **Markowitz**; **Pidgeon**, 2014) or the scientific community (**Lahsen**, 2013);
- a lack of commitments, which were postponed to subsequent summits, at the Copenhagen *UNFCCC* in 2009 (**Van-Eck**; **Feindt**, 2022); and
- the climategate case (**Grundmann**, 2013; **Matthews**, 2015; **Van-Eck**; **Feindt**, 2022).

It is relevant that the level of education and scientific knowledge are not important factors to explain this position (**Kahan** *et al*., 2012; **Whitmarsh**, 2011), and contrary voices can even be heard within the scientific community itself (**Lahsen**, 2013), something that also has been able to contribute to increasing the level of skepticism. Additionally, opinions that deny climate change have had greater acceptance.

The documents analyzed by **McCright** and **Dunlap** (2003), produced by 14 different conservative think-tanks between 1990 and 1997, conclude that climate skeptics challenge the science of global warming by:

- treating supporting evidence as weak or nonexistent;
- highlighting the potential net benefits that might result if climate change should occur; and
- clarifying that policies designed to address climate change would be economically harmful and ineffective.

Given this increase in climate skepticism, various studies have tried to establish a categorization or typology for it, although a consensus has yet to be achieved owing to the different viewpoints and attitudes associated with climate skepticism (**Matthews**, 2015). **Capstick** and **Pidgeon** (2014) distinguish two categories:

- epistemic skepticism: those who question science; and
- response skepticism: those who question the value of acting to prevent climate change.

Lahsen (2013) analyzes the positions defended by scientists and distinguishes two types:

- mainstream scientists, who show moderate levels of skepticism; and
- contrarian scientists, who show a high level of skepticism.

On the other hand, **Matthews** (2015) analyzes the communication from climate skeptics in blogs and distinguishes three degrees of skepticism:

- lukewarmers: who believe that pollution is affecting the planet and will continue to do so but that its impact is less than what the experts predicted; therefore, these scientists do not deny climate change but understand that the generated concern is exaggerated;
- moderate skeptics: who do not consider global warming to be a problem, believe that it has been exaggerated, and distrust the scientific theories that defend it; they understand that climate change has occurred throughout history but depends more on natural processes than on human action; and
- strong skeptics: who do not believe in the opinions of climate scientists or activists and think they are dishonest and fraudulent.

Social networks have brought about a change in traditional communication structures, making it possible for messages to be spread by citizens so that they coexist alongside the messages of traditional gatekeepers (legacy news media, companies, political parties, or the scientific community). Social media promote a more interaction-oriented and open horizontal communication than legacy media (**Dahlberg**, 2001). Especially over the last decade, it has been observed that people consult information on social networks to search for information and understand and discuss different scientific topics (**Anderson**; **Huntington**, 2017; **Su** *et al*., 2015). This represents a great opportunity because it enables social debate on relevant issues such as climate change, but at the same time it can contribute to misinformation and polarization. **Williams** *et al*. (2015) propose that the online debate on climate change is polarized with each group of believers/skeptics considering the position of their opponents to be illegitimate or unnatural. Social media platforms make it easier for anti-climate-change activists to spread their ideas than it would be in legacy news media (**Moernaut** *et al*., 2022).

In their work, **Bolsen** and **Shapiro** (2017) review the climate change topic in the US news media and the emergence of related frames in the public discourse, focusing on divisions and highlighting the role that events, journalistic practices, technological changes, and individual-level factors such as ideology and identity have played in fostering polarization. They identify the core challenges facing communicators who seek to build consensus for action on climate change and highlight the most viable solutions for generating efficient messages.

> We have named the categories of topics to understand social skepticism around climate change through the analysis of users' social activism and behavioral patterns

In “The US news media, polarization on climate change, and pathways to effective communication,” **Bolsen** and **Shapiro** (2017) review the results obtained from various studies over the years regarding the debate taking place about climate change on online platforms and social networks. Regarding *YouTube* uses in the United States, they identified that post-video discussions among members of the *YouTube*-viewing public tend to debate the science of climate change regardless of its relevance to the content of the videos to which they are attached (**Bolsen**; **Shapiro**, 2017). In other words, the public is using *YouTube* –and likely other social media discussion platforms– not to deliberate but rather to campaign for increased activism or skepticism about climate change.

One of the recommendations they make is that communicating the existence of a scientific consensus about human-caused climate change shifts the public’s belief toward the scientific consensus.

There has been extensive research on social activism on social platforms (**Hardaker**; **McGlashan**, 2016; **Drakett** *et al*., 2018; **Fujita**; **Harrigan**; **Soutar**, 2018). As discussed above, an investigation of shared or social activism requires an analysis of the language used in UGC, including publications, posts, and interactions (**Hardaker**; **McGlashan**, 2016; **Kapoor** *et al*., 2018).

There are many studies and much evidence showing that *Twitter* is the platform preferred by activists or social movements, acting as a real collaborative activist arena:

- **Li** *et al*. (2021) and **Xiong** (2019) analyze its use as a tool for feminist social movements;
- **Skill**, **Passero** and **Francisco** (2021) and **Carew** (2014) emphasize *Twitter*’s use as the platform for materializing environmental activism;
- **Zoller** and **Casteel** (2021) investigate a social media campaign for health activism in *Twitter*; and
- **Sinpeng** (2021) describes young political activism.

Twitter is an emerging space with an important role in the climate change debate. It allows its users to share opinions and information about climate change. Several studies on this topic have been published in recent years (**Kirilenko**; **Stepchenkova**, 2014; **Pearce** *et al*., 2014; **Williams** *et al*., 2015; **Anderson**; **Huntington**, 2017; **Moernaut** *et al*., 2022), but given the importance and popularity of this platform for consulting and exchanging information on climate change, more work focusing on *Twitter* is needed (**Veltri**; **Atanasova**, 2017).

Although different elements of interaction can serve as objects of such analysis, the means most frequently used to identify relevant content are keywords, whether they incorporate a hashtag or not (**Zappavigna**, 2015; **Palos-Sánchez** *et al*., 2018; **Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018; **Saura**; **Reyes-Menéndez**; **Álvarez-Alonso**, 2018; **Wu** *et al*., 2021). That said, it is also possible to use other elements of interaction (Table 2).

Table 2. Previous research according to the type of social interaction element studied

Authors	Interaction type	Social network	Category	Year
Hardaker; **McGlashan** (2016)	Profile (@CCriadoPerez)	*Twitter*	Threats	2016
Karami *et al*. (2020)	Hashtag (#scflood)	*Twitter*	National disasters	2020
Muralidharan *et al*. (2011)	Comments	*Twitter*, *Facebook*	Natural disasters	2011
Reyes-Menéndez; **Saura**; **Álvarez-Alonso** (2018)	Hashtag (#WorldEnvironmentDay)	*Twitter*	Environment	2018
Singh; **Shula**; **Mishra** (2018)	Hashtag (#BeefBin, #FoodSafety)	*Twitter*	Food waste	2017
Wu *et al*. (2021)	Comments	*Sina Weibo*	Urban waste	2021

Of all the identified movements about climate change, special attention has been paid to those regarding anti-climate views or climate skepticism. The main platform from which the analyzed content has been obtained is *Twitter*. For this, user profiles (**Hardaker**; **McGlashan**, 2016) or hashtags (**Singh** *et al*., 2018) have been used.

Despite its advantages, *Twitter* can also contribute to disinformation and polarization. Williams *et al*. carried out an interesting study analyzing user opinions on *Twitter* and concluded that active users (either skeptics or believers) show strong attitudes in their discussions about climate change,

> “characterized by strong attitude-based homophily and widespread segregation of users” (**Williams** *et al*., 2015, p. 135).

Anderson and **Huntington** (2017) also analyze the sentiment of comments on *Twitter*, finding a persistent presence of incivility and sarcasm. The authors find that these characteristics are more frequent among skeptics and those who mention right-leaning politics in their profiles.

It is common to analyze comments in important moments such as weather events (**Anderson**; **Huntington**, 2017; **Capstick**; **Pidgeon**, 2014; **Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018; **Moernaut** *et al*., 2022), *Conference of the Parties* (*COP*) summits (**Kaiser**; **Rhomberg**, 2016; **Wozniak**; **Wessler**; **Lück**, 2017), the publication of *IPCC* reports (**O’Neill** *et al*., 2015; **Newman**, 2017), or *climategate* (**Porter**; **Hellsten**, 2014).

Following previous research that framed their data extraction around specific important climate dates (**Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018; **Moernaut** *et al*., 2022), we expected to find an active debate as we collected data during and from *World Environment Day* in 2022.

2.1. Hypothesis development

As argued by Lakoff (2004), a shared reality is created through words and their specific uses in a discourse. Accordingly, an analysis of language opens up a way to understand the shared reality, as well as the underlying shared identity, of its participants. As mentioned above, social identity is shaped by individuals' perceptions of their belonging to a social group or groups and by the significance they attach to this (**Tajfel**, 1974). In this respect, **Grover** *et al*. (2019) argues that users' exposure to certain *Twitter* content can reinforce their previous opinions, thus causing a polarization of such views. A parallel process that can also occur is acculturation, which is defined as adaptions of an individual's views and opinions under the influence of individuals or groups from other cultural backgrounds. This suggests that such interactions should be carefully investigated (**Stieglitz** *et al*., 2018). Interestingly, a study that used the information system success model showed that the influence of UGC can occur on not only the user but also organizational and social levels (**Alalwan** *et al*., 2017). Therefore, online social movements can be investigated through the analysis of UGC on social platforms.

Contrary to the aforementioned studies, another paradigm that can be very useful in terms of providing a holistic perspective is that of information management (**Dwivedi**; **Kapoor**; **Chen**, 2015; **Pace**; **Buzzanca**; **Fratocchi**, 2016). From this perspective, what matters the most is not the management of information but rather the ways in which information must be provided to initiate changes in individuals' behavior. Through a review of the literature on climate skeptics, we identified an important research gap in previous research, specifically regarding social activism and climate skepticism in UGC. To address this research gap, we investigate herein the association among topics that determine the social skepticism around climate change. Our aim is to identify relevant users' social activism and behavioral patterns.

To this end, in our application of the holistic approach, we focus on both differences (**Wu**; **Su**, 1993; **Arora** *et al*., 2019; **Grover** *et al*., 2019) and correlations (**Bouma**, 2009; **Iyengar**; **Sood**; **Lelkes**, 2012).

With regard to the latter, the hypothesis tested in the present study is as follows:

> (H1): There will be correlations between the UGC topics that identify the social skepticism around climate change through the analysis of users' social activism and behavioral patterns.

The data were collected from *Twitter*, with a focus on keywords related to social skepticism around climate change.

To collect keywords linked to anti-climate activism, we carried out an initial screening in which we obtained the 20 most mentioned and the 20 most relevant anti-climate-change hashtags (**Vanhala** *et al*., 2020; **Blasco-Arcas** *et al*., 2022). After analyzing these, we obtained the final hashtags that were used to extract the data: #ClimateHoax, #ClimateFraud, #ClimateBrawl, #Klimaathysterie, #ClimatePanic, #ClimateAlarm, and #ClimateFraud.

The collected tweets were then analyzed using corpus linguistics tools. In doing so, we adopted the approach previously proposed by **Fujita**, **Harrigan** and **Soutar** (2018). We also drew on the analysis of the texts about feminism using computational techniques carried out by **Al-Nakeeb** and **Mufleh** (2018) and the corpus linguistics and discourse analysis carried out by **Hardaker** and **McGlashan** (2016) regarding the social identity of users in the #MeToo movement, based on the comments published by @CCriadoPerez. To validate the existence of a social identity related to social skepticism around climate change through this analysis of users' social activism and behavioral patterns on *Twitter*, we complemented our analysis with log-likelihood statistical measures (**Iyengar**; **Sood**; **Lelkes**, 2012) and mutual information (**Wu**; **Su**, 1993; **Bouma**, 2009).

3. Methodology

3.1. Data collection

Following the work of **Reyes-Menéndez**, **Saura** and **Álvarez-Alonso** (2020), we extracted a sample of tweets to collect the data for subsequent linguistic analysis with keywords related to the climate skeptic movement between *World Environment Day* (June 5) and October 2, 2022. The optimal sample size was determined using previous studies (**Saura**; **Rodríguez-Herráez**; **Reyes-Menéndez**, 2019; **Hardaker**; **McGlashan**, 2016). The criteria used to extract the initial tweets collected in the present study are presented in Table 3, resulting in 78,168 tweets. To collect the database of tweets, we used *Python* 3.7.0.

Table 3. Sampling criteria for each type of interaction

Interaction type	Function
Tweet	Every tweet published with climate skeptic keywords
Retweet (RT) with no #	Every retweet mentioning the keywords
Retweet (RT) with #	Any retweet mentioning the keywords with hashtags

Source: Based on **Hardaker** y **McGlashan** (2016) and **Reyes-Menéndez**, **Saura** y **Álvarez-Alonso** (2020).

Next, since our aim was not to analyze multimedia content, a series of quality filters were applied to clean the data, and we eliminated images and videos (**Saura**; **Reyes-Menéndez**; **Álvarez-Alonso**, 2018). To increase the quality of the data, we also removed URLs from the tweets. The *Python* and *Pandas* software libraries were used for data cleaning. Specifically, the commands to select or replace columns and indices to reshape lost or empty values and to debug repeated or unnecessary data were run. Finally, since retweets represent users' opinions and individual behaviors, they were analyzed separately. Table 4 presents examples of the tweets included in the final sample.

Table 4. Sample tweets included in the final sample

User	Date/time	Tweet
Matrix_ backup	2022-07-20 00:21	Fecking hell!!!! It's all about FAKE man-made climate change this morning 😂 Be afraid, Be very afraid! 😂😂 😂😂 #ClimateCult #climatefraud
zoetnet	2022-06-02 02:19	Ice in #Antarctica has been increasing since 1979. #climatefraud #globalwarmingfraud

3.2. Corpus linguistics and the latent Dirichlet allocation model

Corpus linguistics (CL), a subfield of linguistics that combines both quantitative (**Jia**, 2018; **Saura**; **Reyes-Menéndez**; **Álvarez-Alonso**, 2018) and qualitative (**Baker** *et al*., 2008) research methods, focuses on the analysis of large amounts of linguistic data referred to as corpora (**McEnery**; **Hardie**, 2013; **Reyes-Menéndez**; **Saura**; **Filipe**, 2019).

The latent Dirichlet allocation (LDA) model is a widely used quantitative technique in corpus linguistics. Initially developed by **Pritchard**, **Stephens**, and **Donnelly**, (2000) as a machine-based technique, LDA was subsequently improved and expanded upon by **Blei**, **Ng**, and **Jordan** (2003). With the support of artificial intelligence (AI), this model enables the identification of both keywords and topics linked to them. The key assumption of LDA is that the topics in a database are not observable a priori and should be analyzed through a probabilistic model (**Reyes-Menéndez** *et al*., 2020). Specifically, the model aims to determine the number of times a given word is repeated in a corpus or an individual document. The mathematical model developed in *Python* establishes the variables in the sample; these variables are known as latent variables. These variables are used to determine the number of topics identified by the algorithm, considering the importance of those variables (**Jia**, 2018). In this way, once the algorithm determines the total number of words and that of repeated words, as well as the number of each of the most frequent words that occur before and after the identified words, each topic is assigned a name. The quality of the data is important for the quality of our model, so we preprocess the data by removing symbols with regular expressions and performing tokenization and delete punctuation and create *N*-grams (bigrams and trigrams), applying lemmatization and removing stop-words. Using a standardized process of the LDA model based on grounded theory studies, each topic's name is derived from the words within each of the clusters identified.

Thus, the LDA model consists of the following two steps: First, all keywords present in the corpus are obtained. Second, the topics linked to these keywords are identified (**Reyes-Menéndez** *et al*., 2020). To identify topics in a maximally objective way, the mathematical distribution shown in Equation (1) is employed.

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^{K} p(\beta_i) \cdot \prod_{d=1}^{D} p(\theta_d) \cdot \sum\nolimits_{n=1}^{N} p(z_{d,n}|\theta_d)\, p(w_{d,n}|\beta_{1:K}, z_{d,n}) \quad (1)$$

β_i is the distribution of word in topic *i* among a total of *K* topics

θ_d is the proportion of topics in document *d* among a total of *D* documents

z_d is the topic assignment in document *d*

$z_{d,n}$ is the topic assignment for the *n*th word in document *d* among a total of *N* words

w_d is the observed words for document *d*

$w_{d,n}$ is the *n*th word for document *d*

In the next step, to identify the topics that make up the dataset, we used Gibbs sampling [Equation (2); **Jia**, 2018] using the Mac version of the Python software LDA 1.0.5.

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}|w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \quad (2)$$

4. Results

This section reports the results we obtained on the keywords and frequency related to social skepticism around climate change through the analysis of users' social activism and behavioral patterns identified in our corpus (Section 4.1), the topics (Section 4.2), and the corresponding categories, social activism, and behavioral patterns (Section 4.3)

4.1. Keywords and frequency

We carried out an analysis of the keywords in the corpus, considering the importance of the fact that keywords express user behavior and the linguistic importance of the terms (**Reyes-Menéndez** *et al*., 2020).

In this same line, the frequency of a term's occurrence in a text is a key measure in corpus linguistics. Frequency is assumed to highlight users' social identity (**McEnery**; **Hardie**, 2013). Here, frequency is defined as the number of times a word appears in a given text (**Baker** *et al*., 2008). Table 5 lists the frequencies of 10 main words identified in our data.

Table 5. Frequency of 10 main words

Rank	Word	Similar words	Frequency
1	ClimateBrawl	Brawl, ClimateBrawlers	4,246
2	ClimateHoax	ClimateHoaxWe, UnrealClimateHoax	3,461
3	ClimateCrisis	NoClimateCrisis, Crisis, KlimaatCrisis	3,525
4	ClimateEmergency	NoClimateEmergency, Emergency	3,050
5	ClimateScam	Scam, Scammers, Scams	1,567
6	WEF	WorldEconomicForum, WefCrimeAgainstHumanity, FTheWef	1,205
7	Resist	TheGreatResist, Resistance	1,006
8	Science	JunkScience, Sciences	801
9	Farmers	Farmer, FarmerProtests, NoFarmersNoFood	766
10	Support	Supporting, Supports, Supporters	735

As seen in Table 5, the most frequent term in our data is the keyword "ClimateBrawl" (4,246 times) that was previously identified as a hashtag to be extracted. This was also the case with the hashtag "ClimateHoax" (3,461 times).

Other keywords that were not extracted as hashtags were "ClimateCrisis" (3,525 times), "ClimateEmergency" (3,050 times), and "ClimateScam" (1,567 times).

The fact that "WEF," which corresponds to "World Economic Forum," is present 1,205 times is interesting, as this is a global forum for economic development. Also, "science" is present 801 times, in terms such as "JunkScience." Additionally, "farmers" is mentioned 766 times, showing the interest that food production has for users.

4.2. Topics

Topics in a corpus are clusters of words linked to each other. Accordingly, topics are intrinsically related to their keywords (**Reyes-Menéndez** *et al*., 2020). To find topics in our database, the LDA model and its corresponding Equation (1) were used (Section 3.2).

Next, to evaluate our LDA model, we used the metric referred to as topic coherence, which measures the relative distance between words within a topic (**Syed**; **Spruit**, 2017; **Rama-Maneiro**; **Vidal**; **Lama**, 2020). It is rare to see a coherence of 1 or +0.9 unless the words being measured are either identical words or bigrams. The overall coherence score of a topic is the average of the distances between words. We attain a value of 0.34 in our LDAs, since there is no strong topic correlation; in other words, the distance between words within topics is not very close.

To determine whether the identified topics are relevant key indicators of social skepticism around climate change using the analysis of users' social activism and

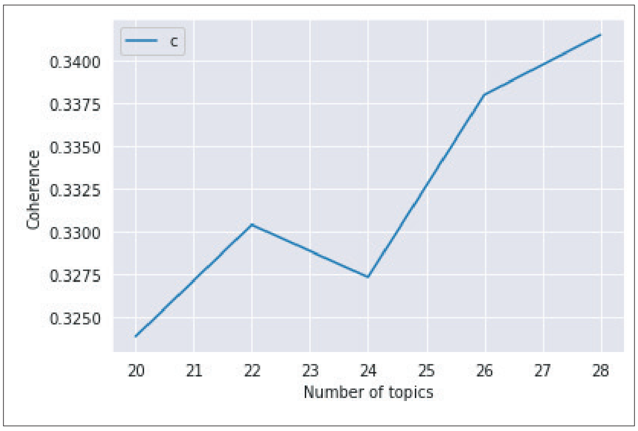

Figure 1. Coherence of topics

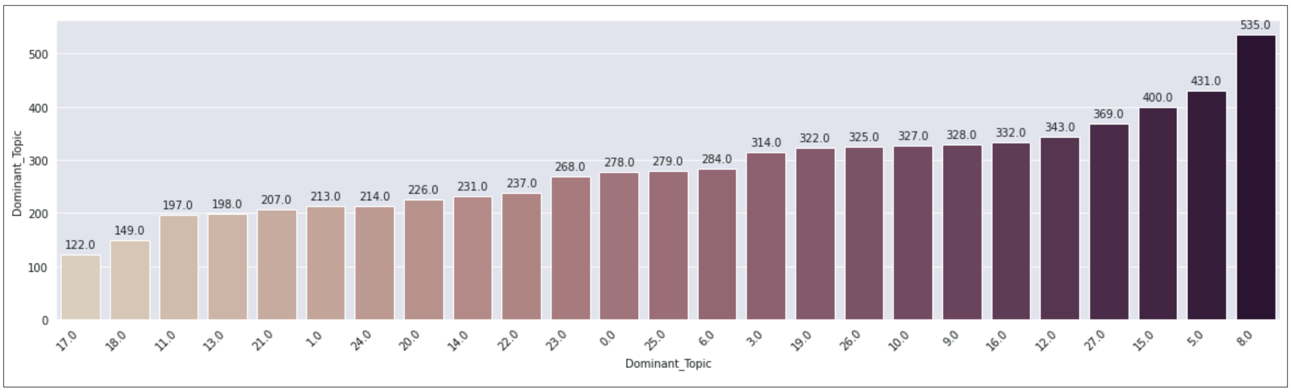

Figure 2. Bar graph of topics

behavioral patterns on *Twitter*, we relied on the measure of coherence. This function, built in *Python*, searches for an optimal number of topics in the dataset. The graph (Figure 1) shows 28 topics as optimal, with a coherence score of ~0.34 listing the ideal number of topics that will compose the social skepticism around climate change using the analysis of users' social activism and behavioral patterns.

As seen in Figure 2, the 28 identified topics have different contributions to the overall research. The topic with the greatest contribution is topic 8.0.

Below, we present the contribution of the 28 topics identified (Table 6) in the tweets database. We also highlight the main keywords that make up each topic, and each topic has been assigned a name with a randomized controlled process (**Jia**, 2018).

Table 6. Topic contribution and keywords of the topics

Topic name	Topic num.	Keywords	Topic contribution
Trump	0.0	Climatescam, climate, change, instill, fly, fear, trump, covid	0.654
Temperatura	1.0	climatefraud, colour, potus, sea_level, temperature	0.637
Story	2.0	climatehoax, climatescam, climate, change, globalwarme, week, happen, story, enjoy, destroy	0.652
Coal	3.0	climatescam, climatehoax, climate, change, eat, summer, coal, support, fart, trillion	0.649
Diesel	4.0	climatehoax, climatescam, climate, change, liberal, wef, year, people, diesel, investment	0.628
Media	5.0	climatescam, climatecrisis, climatehoax, climateaction climate, change, support, bbcnew, lie, travel	0.692
Brainwash	6.0	climatehoax, climatescam, brainwash, hear, climate, emergency, global_warme, crap, power, try	0.648
Fakenews	7.0	climatehoax, climatecrisis, climate, change, fakenew, people, life, today, save, love	0.658
Lie	8.0	climatehoax, climatescam, agenda, fear, end, support, weather, lie, climate, control	0.600
War	9.0	climatehoax, climatescam, support, joke, wake, year, war, emergency, climate, resist	0.665
Private jet	10.0	climatescam, climatecult, climatehoax, climate, change, emergency, private_jet, brain, starve, years_ago	0.607
Wef	11.0	climatehoax, climate, change, wefpuppet, ton, ocean, globalwarming, game, year, freedom	0.597
Politics	12.0	climatehoax, climatescam, climatecrisis, fuck, change, climate, retweet, politician, potus, human	0.635
Covid	13.0	climatescam, climatecrisis, climate, change, covid, basisscholen, agenda, laten, thegreatreset, charge	0.613
Dream	14.0	climatecrisis, climatehoax, climateaction, climategrifter, al_gore, climate, year, dream, support, straight	0.659
Bullshit	15.0	climatehoax, climatescam, climatecrisis, change, bullshit, listen, support, believe, climate, scam	0.630
Cows	16.0	climate, change, charge, fall, cow, carbon, alarmist, hypocrisy, weekend, work	0.659
Education	17.0	climatehoax, climatescam, climate, crisis, basisscholen, laten, thegreatreset, surprise, spot	0.611
Green new deal	18.0	climatehoax, runderen, die, geoengineering, still_legal, change, climate, support, greennewdeal, geoengineere	0.591
Biden farmers	19.0	climatescam, climatehoax, climatecrisis, climate, change, agenda, farmer, prediction, money, biden	0.625
Tax	20.0	Climatescam, tax, sky, record, gas_price, accord, warm, climate, vaccine, temperature	0.601
Narrative	21.0	climatehoax, climatecrisis, climatechange, climatescam, climate, change, narrative, speech, pay, state	0.627
Negative	22.0	climatescam, climatehoax, eat, basisscholen, laten, thegreatreset, agenda, insect, support	0.613
Support	23.0	climatehoax, climatecrisis, climatescam, climate, change, auspol, stop, die, wrong, support	0.644
Red painting rule	24.0	climatehoax, climatescam, climate, change, red, painting, rule, truth, list, night	0.638
Globalwar-ming	25.0	climatehoax, climatescam, climatecrisis, climate, change, window, sun, globalwarme, support, truth	0.618
Administration	26.0	climatecrisis, climatehoax, climate, change, hoax, buy, administration, excuse, open, support	0.631
Obama	27.0	climatescam, climatehoax, climatecult, climate, change, support, obama, control, rest, science	0.640

Figure 4 shows the intertopic distance map. This visualization represents the different topics and the distance between them. Similar topics appear closer and dissimilar topics farther away. The relative size of a topic's circle in the plot corresponds to the relative frequency of that topic in the corpus. In our case, topics 14.0 and 18.0 appear closer and not far from topic 23.0, while topic 19.0 is far away. The overlap among the remaining topics (1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 15.0, 16.0, 17.0, 20.0, 21.0, 22.0, 24.0, 25.0, 26.0, 27.0, and 28.0) is also noteworthy. Specifically, the rule of thumb was as follows: the shorter the distance between the central node and the topics, the greater the link between them (**Al-Nakeeb**; **Mufleh**, 2018).

This visualization reveals that topic 19.0 "biden farmers" in Table 7 is isolated and has a greater distance from the other topics, while topics 14 "dream," 18 "green new deal," and 23 "support" lie in the same quadrant and can form a category of topics.

The remaining topics ("trump," "temperature," "story," "coal," "diesel," "media," "brainwash," "fakenews," "lie," "war," "private jet," "wef," "politics," "covid," "dream," "bullshit," "cows," "education," "tax," "narrative," "negative," "Red painting rule," "globalwarming," "administration," and "Obama") overlap and, therefore, can form a third category of topics.

4.3. Categories of topics, social skepticism, and behavioral patterns around climate change

We have named the categories of topics to understand social skepticism around climate change using the analysis of users' social activism and behavioral patterns. Furthermore, to identify the different categories in which the topics fall, a name has been assigned through a randomized controlled process (**Jia**, 2018).

The groupings of topics explained above serve as the basis for the development of the categories of the social identity and behavioral patterns. In this way, we obtained the following three different categories:

- Biden;
- *Green New Deal*;
- Hoax.

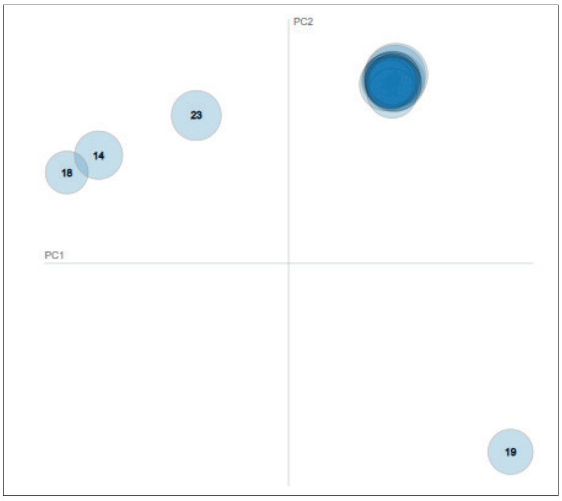

Figure 3. Intertopic distance map

Which topics correspond to which categories is explained in Table 7.

Table 7. Categories of topics

Categories	Topic num.	Topic tag	Description
Biden	19.0	Biden	This category presents the proposal developed by Biden to fund farmers' losses among other proposals. Opinions on Biden are mostly critical.
Green New Deal	14.0 18.0 23.0	Dream *Green New Deal* Support	This category shows the importance of the *Green New Deal* and the positive aspect of this proposal for those who consider it necessary. While the detractors use critical comments to express their perception of the proposal and its unsustainable nature for companies and individuals. This is especially notable in fuels, economy, supplies.
Hoax	0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 15.0 16.0 17.0 20.0 21.0 22.0 24.0 25.0 26.0 27.0	Trump Temperature Story Coal Diesel Media Brainwash Fakenews Lie War Private jet WEF Politics Covid Bullshit Cows Education Tax Narrative Negative Red painting rule Global warming Administration Obama	This category includes all the other topics that are also linked by the hashtags (#) included in the search and other hashtags such as #ClimateCrisis or #ClimateScam. The fact that these hashtags are strongly present in all the topics means that they are presented as a consolidated category determined by the strong use of hashtags linked to anti-climate-change activism.

As can be seen in Table 7, there are a series of topics that are well determined and cohesive and that are of considerable size. These topics are "dream," "green new deal," and "support," and they belong to the category *Green New Deal*.

Table 8 below shows some examples of tweets that belong to the different categories identified as Biden, *Green New Deal* and Hoax.

Table 8. Sample of tweets per category

Cat.	User	Date/time	Tweet
Biden	Jon Tveten	2022-07-11 09:44	Green tyranny has finally provoked mass reactions, and the first government has fallen after imposing insane policies that wrecked the food supply for its people. @Thomas_Lifson #ClimateFraud
	Genuinedavid	2022-06-21 22:44	Biden goes begging countries for (dirtier than ours) oil while simultaneously exporting our own clean domestic oil, thus pumping much more c02 via redundant tanker ships. #ClimateFraud Why Is The United States Still Exporting Fuel?
	Jorj X McKie	2022-06-10 18:59	More #ClimateFraud from the #BidenAdministration RT @ClimateDepot: Climate Depot's Morano: "After years of claiming you can't challenge 'the science' they now claim you can't challenge their 'solutions!'"
Green New Deal	Kevin Killough	2022-07-07 02:57	I want to set up a charity to help European victims of green energy policies. Even if I could crowdsource or something, I'd have no idea how to get it to those who are suffering under these policies. #ClimateBrawl #GreenEnergyKills
	Steve Tatum 🦅🇺🇸	2022-10-04 12:47	They're shutting down our pipelines, canceling our drilling projects, blowing up our gas prices, and destroying our economy just for their *Green New Deal* Climate Cult Hoax. #LetsGoBrandon #GasPrices #ClimateHoax #VoteGOP2022
Hoax	Mel 🇺🇸	2022-09-29 17:17	Just a friendly reminder to NEVER trust the mainstream media…#FakeNews #ClimateHoax
	Blinkered Britain	2022-09-21 02:09	Be afraid very afraid. 'My Carbon': An approach for inclusive and sustainable cities 'COVID-19 was the test of social responsibility' #WEF #TalkRadio #TalkTV #GBNews #ClimateHoax #ClimateChange #GreenAgenda

It should be highlighted that the topic "biden farmers," owing to its position in the intertopic distance map and its size, is a singleton category, while the rest of the topics overlap, composing the last category referred to as Hoax. This category includes all of the other topics that are also linked by the hashtags (#) included in the search as well as other hashtags such as #ClimateCrisis or #ClimateScam. The fact that these hashtags are strongly present in all the topics means that they are presented as a consolidated category determined by the strong use of hashtags linked to anti-climate-change activism.

5. Discussion

In the present study, we used a systematic literature review to identify, evaluate, and synthesize social skepticism around climate change indicators through an analysis of users' social activism and behavioral patterns on *Twitter*. Our study answers **Veltri** and **Atasanova** (2017) call for research efforts to better understand UGC content on Twitter.

Some previous research (**Capstick**; **Pidgeon**, 2014; **Anderson**; **Huntington**, 2017; **Wozniak**; **Wessler**; **Lück**, 2017; **Reyes-Menéndez**; **Saura**; **Álvarez-Alonso**, 2018; **Moernaut** *et al*., 2022) has been developed around events related to climate change. In the same line, this research has obtained satisfactory results with data extracted during *World Environment Day* in 2022.

Numerous previous investigations have linked climate skepticism on *Twitter* with issues of political ideology (**Whitmarsh**; **Corner**, 2017; **Van-Eck**; **Feindt**, 2022), and even with determining the discourse on the basis of the political position of the users and the parties (**Elgesem**; **Steskal**; **Diakopoulos**, 2015; **Matthews**, 2015). This is in line with the results obtained herein because the political presence is evident in the topic analysis (Section 4.2) with topics such as topic 19.0 "biden farmers," "trump," "politics," and "obama."

With respect to the categories of topics identified (Section 4.3), one of them is the "*Green New Deal*." This category identified in our results has not been analyzed in previous research dealing with social skepticism and behavioral patterns around climate change, so this opens new lines of research on skepticism about climate change, the communication that is carried out in this sense about climate change, climate on *Twitter*, and the *Green New Deal*.

> Social media platforms make it easier for anti-climate change activists to spread their ideas than it would be in legacy news media

Some of the practical implications of this work are the application of the results for the development of public and private policies by institutions, governments, or companies that are concerned about climate change. Social media platforms are a space where awareness of the climate change can be promoted, but also where the opposite effect can be achieved by amplifying anti-climate change views. As **Moernaut** *et al*. (2022) point out, in contrast to traditional media, social networks allow anti-climate change activists to easily disseminate their ideas.

Another of the practical implications would be related to the education of those who think that climate change is not a worrying issue but rather a lie. In this sense, it is also necessary to implement communication strategies based on expert opinions that reduce the polarization highlighted by several studies on this topic (**Dunlap**; **McCright**, 2011; **Moernaut** *et al.*, 2022; **Pearce** *et al.*, 2019) and fight fake news. However, two main difficulties must be taken into account:

- that *Twitter* can promote polarization and misinformation (**Williams** *et al.*, 2015; **Anderson**; **Huntington**, 2017) as users tend to search for opinions similar to their own in order to reinforce them (**Grover** *et al.*, 2019); and
- that previous studies highlight that the degree of education and scientific knowledge is not a decisive factor in explaining whether people take a position for or against.

> There will be correlations between the UGC topics that determine the social skepticism around climate change through the analysis of users' social activism and behavioral patterns

Moreover, within the scientific community there are opinions that deny or question the importance of climate change (**Lahsen**, 2013) and that, therefore, may encourage those opinions to have greater credibility.

Among the theoretical implications would be the development of new research based on the results obtained, (e.g., the relationship between political events and the polarization of opinion on climate change) and the fact that, by using data from *Twitter*, it is possible to analyze the discourse of climate change skeptics.

Among the limitations of this work are the number of data extracted, the hashtags used, the language of the extraction, the date selected, and the analysis carried out, which does not identify whether the comments are positive or negative, thus we cannot know whether they are for or against the arguments presented in the topics.

Future lines of research could include the modeling of the different topics identified and a model that integrates opinions based on political preferences, as well as longitudinal analysis using data extracted from the different editions of WED on *Twitter* to determine how the conversation and behavioral patterns evolved.

6. Conclusions

In this study, we used machine learning and artificial intelligence techniques to review 78,168 tweets to identify the keys of social skepticism around climate change indicators through an analysis of users' social activism and behavioral patterns on *Twitter*. These results were analyzed in depth to address the aim of this research.

Based on our results, we were able to answer hypothesis H1, that there will be correlations between the UGC topics that identify the social skepticism around climate change through the analysis of users' social activism and behavioral patterns. Specifically, we identified 28 topics that, in turn, could be grouped into 3 categories that identify the social skepticism around climate change through the analysis of users' social activism and behavioral patterns (Table 7).

The investigation has produced a series of results that confirm the proposed hypothesis.

In addition, some relevant conclusions have been obtained. The first is that 24 of the 28 topics are overlapping on the intertopic distance map. The second is that the size of the topics is relatively small and linked to specific events. The third is that there is a significant political presence, especially from the United States.

There is a group of topics, 24 of the 28, that appear superimposed such that, although they use different words and therefore form different clusters, they have a close relationship and, therefore, appear not only close but also superimposed. This opens up the possibility for new research focused only on these topics to better understand the reason for this overlap, although it does not permit their combination into a single larger topic.

The size of the topics is relatively small. There is no big theme, which means that attention is divided among the 28 themes developed by the skeptics opposed to climate change. This may be due to the fact that each of the user groups defends a viewpoint on a specific climate change topic, without any of them having gone viral, or because of their temporary nature. Themes arise, but none stay around for a long time. For example, in the political arena, one can mention the different political leaders in their corresponding topics ("trump", "biden", and "obama").

In relation to this point, note that the small size of the topics may be related to the relationship between the communication actions against climate change and the specific facts related to it, for example, Biden's proposal to support farmers in the topic "biden farmers", the "covid" health crisis, or the "wef" meeting. Regarding this point, the important presence of the United States, with its different presidents, also stands out, while there is no mention of other leaders of other countries.

> We identified 28 topics that, in turn, could be grouped into 3 categories that identify the social skepticism around climate change through the analysis of users' social activism and behavioral patterns

7. References

Al-Nakeeb, Ohood A. M. S.; **Mufleh, Basher A. H.** (2018). "Collocations and collocational networks of characters: A corpus-based feminist stylistic analysis". *Language in India*, v. 18, n. 9.
https://onx.la/05e12

Alalwan, Ali A.; **Rana, Nripendra P.**; **Dwivedi, Yogesh K.**; **Algharabat, Raed** (2017). "Social media in marketing: A review and analysis of the existing literature". *Telematics and informatics*, v. 34, n. 7, pp. 1177-1190.
https://doi.org/10.1016/j.tele.2017.05.008

Anderson, Ashley A.; **Huntington, Heidi E.** (2017). "Social media, science, and attack discourse: How *Twitter* discussions of climate change use sarcasm and incivility". *Science communication*, v. 39, n. 5, pp. 598-620.
https://doi.org/10.1177/1075547017735113

Arora, Anuja; **Bansal, Shivam**; **Kandpal, Chandrashekhar**; **Aswani, Reema**; **Dwivedi, Yogesh K.** (2019). "Measuring social media influencer index-insights from *Facebook*, *Twitter* and *Instagram*". *Journal of retailing and consumer services*, v. 49, pp. 86-101.
https://doi.org/10.1016/j.jretconser.2019.03.012

Aswani, Reema; **Kar, Arpan K.**; **Ilavarasan, P. Vigneswara**; **Dwivedi, Yogesh** (2018). "Search engine marketing is not all gold: Insights from *Twitter* and *SEOClerks*". *International journal of information management*, v. 38, n. 1, pp. 107-116.
https://doi.org/10.1016/j.ijinfomgt.2017.07.005

Baker, Paul; **Gabrielatos, Costas**; **Khosravinik, Majid**; **Krzyżanowski, Michal**; **McEnery, Tony**; **Wodak, Ruth** (2008). "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press". *Discourse and society*, v. 19, n. 3, pp. 273-306.
https://doi.org/10.1177/0957926508088962

Blasco-Arcas, Lorena; **Lee, Hsin-Hsuan M.**; **Kastanakis, Minas N.**; **Alcañiz, Mariano**; **Reyes-Menéndez, Ana** (2022). "The role of consumer data in marketing: A research agenda". *Journal of business research*, v. 146, pp. 436-452.
https://doi.org/10.1016/j.jbusres.2022.03.054

Blei, David M.; **Ng, Andrew Y.**; **Jordan, Michael I.** (2003). "Latent Dirichlet allocation". *Journal or machine learning research*, n. 3, pp. 993-1022.
https://dl.acm.org/doi/10.5555/944919.944937

Bolsen, Toby; **Shapiro, Mathew A.** (2018). "The US news media, polarization on climate change, and pathways to effective communication". *Environmental communication: A journal of nature and culture*, v. 12, n. 2, pp. 149-163.
https://doi.org/10.1080/17524032.2017.1397039

Bouma, Gerlof (2009). "Normalized (pointwise) mutual information in collocation extraction". *Proceedings of GSCL*, pp. 31-40.

Capstick, Stuart B.; **Pidgeon, Nicholas F.** (2014). "What is climate change scepticism? Examination of the concept using a mixed methods study of the UK public". *Global environmental change*, v. 24, pp. 389-401.
https://doi.org/10.1016/j.gloenvcha.2013.08.012

Capstick, Stuart B.; **Whitmarsh, Lorraine**; **Poortinga, Wouter**; **Pidgeon, Nicholas F.**; **Upham, Paul** (2015). "International trends in public perceptions of climate change over the past quarter century". *Wiley interdisciplinary reviews: Climate change*, v. 6, n. 1, pp. 35-61.
https://doi.org/10.1002/wcc.321

Carew, Joanne (2014). "Online environmental activism in South Africa: A case study of the #iam4rhinos *Twitter* campaign". *Global media journal: African edition*, v. 8, n. 2, pp. 207-230.
https://doi.org/10.5789/8-2-169

Corner, Adam; **Markowitz, Ezra**; **Pidgeon, Nicholas F.** (2014). "Public engagement with climate change: the role of human values". *Wiley interdisciplinary reviews: climate change*, v. 5, n. 3, pp. 411-422.
https://doi.org/10.1002/wcc.269

Dahlberg, Lincoln (2001). "The internet and democratic discourse: exploring the prospects of online deliberative forums extending the public sphere". *Information, communication & society*, v. 4, n. 4, pp. 615-633.
https://doi.org/10.1080/13691180110097030

Drakett, Jessica; **Rickett, Bridgette**; **Day, Katy**; **Milnes, Kate** (2018). "Old jokes, new media - Online sexism and constructions of gender in Internet memes". *Feminism and psychology*, v. 28, n. 1, pp. 109-127.
https://doi.org/10.1177/0959353517727560

Dubrofsky, Rachel E.; **Wood, Megan M.** (2014). "Posting racism and sexism: authenticity, agency and self-reflexivity in social media". *Communication and critical/cultural studies*, v. 11, n. 3, pp. 282-287.
https://doi.org/10.1080/14791420.2014.926247

Dunlap, Riley E.; **McCright, Aaron M.** (2011). "Climate change denial". In: Lever-Tracy, Constance (ed.). *Routledge handbook of climate change and society*. Routledge, pp. 144-160. ISBN: 978 0 203876213

Dwivedi, Yogest K.; **Kapoor, Kawaljeet K.**; **Chen, Hsin** (2015). "Social media marketing and advertising". *The marketing review*, v. 15, n. 3, pp. 289-309.
https://doi.org/10.1362/146934715X14441363377999

Eide, Elisabeth; **Kunelius, Risto** (2021). "Voices of a generation the communicative power of youth activism". *Climatic change*, v. 169, n. 6.
https://doi.org/10.1007/s10584-021-03211-z

Elgesem, Dag; **Steskal, Lubos**; **Diakopoulos, Nicholas** (2015). "Structure and content of the discourse on climate change in the blogosphere: the big picture". *Environmental communication*, v. 9, n. 2 pp. 169-188.
https://doi.org/10.1080/17524032.2014.983536

Engels, Anita; **Hüther, Otto**; **Schäfer, Mike**; **Held, Hermann** (2013). "Public climate-change skepticism, energy preferences and political participation". *Global environmental change*, v. 23, n. 5, pp. 1018-1027.
https://doi.org/10.1016/j.gloenvcha.2013.05.008

Fujita, Momoko; **Harrigan, Paul**; **Soutar, Geoffrey N.** (2018). "Capturing and co-creating student experiences in social media: a social identity theory perspective". *Journal of marketing theory and practice*, v. 26, n. 1-2, pp. 55-71.
https://doi.org/10.1080/10696679.2017.1389245

Grover, Purva; **Kar, Arpan K.**; **Dwivedi, Yogesh K.**; **Janssen, Marijn** (2019). "Polarization and acculturation in US election 2016 outcomes. Can *Twitter* analytics predict changes in voting preferences?". *Technological forecasting and social change*, v. 145, pp. 438-460.
https://doi.org/10.1016/j.techfore.2018.09.009

Grundmann, Reiner (2013). "'Climategate' and the scientific ethos". *Science, technology & human values*, v. 38, n. 1, pp. 67-93.
https://doi.org/10.1177/0162243911432318

Hagen, Bjoern; **Middel, Ariane**; **Pijawka, David** (2016). "European climate change perceptions: public support for mitigation and adaptation policies". *Environmental policy and governance*, v. 26, n. 3, pp. 170-183.
https://doi.org/10.1002/eet.1701

Hardaker, Claire; **McGlashan, Mark** (2016). "Real men don't hate women: *Twitter* rape threats and group identity". *Journal of pragmatics*, n. 91, pp. 80-93.
https://doi.org/10.1016/j.pragma.2015.11.005

HerdaĞdelen, Amaç; **Zuo, Wenyun**; **Gard-Murray, Alexander**; **Bar-Yam, Yaneer** (2013). "An exploration of social identity: the geography and politics of news-sharing communities in *Twitter*". *Complexity*, v. 19, n. 2, pp. 10-20.
https://doi.org/10.1002/cplx.21457

Hoggan, James; **Littlemore, Richard** (2009). *Climate cover-up: the crusade to deny global warming*. Greystone Books Ltd. ISBN: 1553654854

Iyengar, Shanto; **Sood, Gaurav**; **Lelkes, Yphtach** (2012). "Affect, not ideology: a social identity perspective on polarization". *Public opinion quarterly*, v. 76, n. 3, pp. 405-431.
https://doi.org/10.1093/poq/nfs038

Jenkins, Richard (2014). *Social identity*. New York: Routledge. ISBN: 978 0 415448499

Jia, Susan S. (2018). "Leisure motivation and satisfaction: a text mining of yoga centres, yoga consumers, and their interactions". *Sustainability*, v. 10, n. 12, 4458.
https://doi.org/10.3390/su10124458

Kahan, Dan M.; **Peters, Ellen**; **Wittlin, Maggie**; **Slovic, Paul**; **Ouellette, Lisa-Larrimore**; **Braman, Donald**; **Mandel, Gregory** (2012). "The polarizing impact of science literacy and numeracy on perceived climate change risks". *Nature climate change*, v. 2, n. 10, pp. 732-735.
https://doi.org/10.1038/nclimate1547

Kaiser, Jonas; **Rhomberg, Markus** (2016). "Questioning the doubt: Climate skepticism in German newspaper reporting on COP17". *Environmental communication*, v. 10, n. 5, pp. 556-574.
https://doi.org/10.1080/17524032.2015.1050435

Kapoor, Kawaljeet-Kaur; **Tamilmani, Kuttimani**; **Rana, Nripendra P.**; **Patil, Pushp**; **Dwivedi, Yogesh**; **Nerur, Sridahr** (2018). "Advances in social media research: past, present and future". *Information systems frontiers*, v. 20, n. 3, pp. 531-558.
https://doi.org/10.1007/s10796-017-9810-y

Karami, Amir; **Shah, Vishal**; **Vaezi, Reza**; **Bansal, Amit** (2020). "*Twitter* speaks: a case of national disaster situational awareness". *Journal of information science*, v. 46, n. 3, pp. 313-324.
https://doi.org/10.1177/0165551519828620

Kastanakis, Minas N.; **Balabanis, George** (2012). "Between the mass and the class: Antecedents of the 'bandwagon' luxury consumption behavior". *Journal of business research*, v. 65, n. 10, pp. 1399-1407.
https://doi.org/10.1016/j.jbusres.2011.10.005

Kirilenko, Andrei P.; **Stepchenkova, Svetlana O.** (2014). "Public microblogging on climate change: One year of *Twitter* worldwide". *Global environmental change*, v. 26, pp. 171-182.
https://doi.org/10.1016/j.gloenvcha.2014.02.008

Lahsen, Myanna (2013). "Anatomy of dissent: a cultural analysis of climate skepticism". *American behavioral scientist*, v. 57, n. 6, pp. 732-753.
https://doi.org/10.1177/0002764212469799

Lakoff, Robin T. (2004). *Language and woman's place: text and commentaries*. Oxford University Press, USA. ISBN: 0195167570

Li, Manyu; **Turki, Nadia**; **Izaguirre, Casandra R.**; **DeMahy, Chloe**; **Thibodeaux, Brooklyn L.**; **Gage, Taylor** (2021). "*Twitter* as a tool for social movement: an analysis of feminist activism on social media communities". *Journal of community psychology*, v. 49, n. 3, pp. 854-868.
https://doi.org/10.1002/jcop.22324

Matthews, Paul (2015). "Why are people skeptical about climate change? Some insights from blog comments". *Environmental communication*, v. 9, n. 2, pp. 153-168.
https://doi.org/10.1080/17524032.2014.999694

McCright, Aaron M.; **Dunlap, Riley E.** (2003). "Defeating Kyoto: the conservative movement's impact on U.S. climate change policy". *Social problems*, v. 50, n. 3, pp. 348-373.
https://doi.org/10.1525/sp.2003.50.3.348

McEnery, Tony; **Hardie, Andrew** (2013). "The history of corpus linguistics". In: Allan, Keith (ed.) *The Oxford handbook of the history of linguistics*. Oxford: Oxford University Press, pp. 727-745. ISBN: 978 0 191750434
https://doi.org/10.1093/oxfordhb/9780199585847.013.0034

Moernaut, Renée; **Mast, Jelle**; **Temmerman, Martina**; **Broersma, Marcel** (2022). "Hot weather, hot topic. Polarization and skeptical framing in the climate debate on *Twitter*". *Information, communication & society*, v. 25, n. 8, pp. 1047-1066.
https://doi.org/10.1080/1369118X.2020.1834600

Monfort, Abel; **Villagra, Nuria**; **López-Vázquez, Belén** (2019). "Exploring stakeholders' dialogue and corporate social responsibility (CSR) on *Twitter*". *Profesional de la información*, v. 28, n. 5.
https://doi.org/10.3145/epi.2019.sep.13

Muralidharan, Sidharth; **Rasmussen, Leslie**; **Patterson, Daniel**; **Shin, Jae-Hwa** (2011). "Hope for Haiti: an analysis of *Facebook* and *Twitter* usage during the earthquake relief efforts". *Public relations review*, v. 37, n. 2, pp. 175-177.
https://doi.org/10.1016/j.pubrev.2011.01.010

Newman, Tood P. (2017). "Tracking the release of IPCC AR5 on *Twitter*: users, comments, and sources following the release of the working group. Summary for policymakers". *Public understanding of science*, v. 26, n. 7, pp. 815-825.
https://doi.org/10.1177/0963662516628477

Ntontis, Evangelos; **Drury, John**; **Amlôt, Richard**; **Rubin, G. James**; **Williams, Richard** (2018). "Emergent social identities in a flood: implications for community psychosocial resilience". *Journal of community and applied social psychology*, v. 28, n. 1, pp. 3-14.
https://doi.org/10.1002/casp.2329

Olesen, Thomas (2022). "Greta Thunberg's iconicity: performance and co-performance in the social media ecology". *New media & society*, v. 24, n. 6, pp. 1325-1342.
https://doi.org/10.1177/1461444820975416

O'Neill, Safron; **Williams, Hywel T. P.**; **Kurz, Tim**; **Wiersma, Bouke**; **Boykoff, Maxwell** (2015). "Dominant frames in legacy and social media coverage of the IPCC fifth assessment report". *Nature climate change*, v. 5, n. 4, pp. 380-385.
https://doi.org/10.1038/nclimate2535

Pace, Stefano; **Buzzanca, Stefano**; **Fratocchi, Luciano** (2016). "The structure of conversations on social networks: between dialogic and dialectic threads". *International journal of information management*, v. 36, n. 6, pp. 1144-1151.
https://doi.org/10.1016/j.ijinfomgt.2016.04.001

Palos-Sánchez, Pedro; **Saura, José-Ramón**; **Reyes-Menéndez, Ana**; **Vásquez-Esquivel, Ivonne V.** (2018). "Users acceptance of location-based marketing apps in tourism sector: an exploratory analysis". *Journal of spatial and organizational dynamics*, v. 6, n. 3, pp. 258-270.
https://www.jsod-cieo.net/journal/index.php/jsod/article/view/144

Pearce, Warren; **Holmberg, Kim**; **Hellsten, Iina**; **Nerlich, Brigitte** (2014). "Climate change on *Twitter*: topics, communities and conversations about the 2013 IPCC working group 1 report". *PLoS one*, v. 9, n. 4, e94785.
https://doi.org/10.1371/journal.pone.0094785

Pearce, Warren; **Niederer, Sabine**; **Özkula, Suay M.**; **Sánchez-Querubín, Natalia** (2019). "The social media life of climate change: platforms, publics, and future imaginaries". *Wiley interdisciplinary reviews: climate change*, v. 10, n. 2, e569.
https://doi.org/10.1002/wcc.569

Porter, Amanda J.; **Hellsten, Iina** (2014). "Investigating participatory dynamics through social media using a multideterminant 'frame' approach: the case of Climategate on YouTube". *Journal of computer-mediated communication*, v. 19, n. 4, pp. 1024-1041.
https://doi.org/10.1111/jcc4.12065

Pritchard, Jonathan K.; **Stephens, Matthew**; **Donnelly, Peter** (2000). "Inference of population structure using multilocus genotype data". *Genetics*, v. 155, n. 2, pp. 945-959.
https://doi.org/10.1093/genetics/155.2.945

Rama-Maneiro, Efrén; **Vidal, Juan C.**; **Lama, Manuel** (2020). "Collective disambiguation in entity linking based on topic coherence in semantic graphs". *Knowledge-based systems*, v. 199, 105967.
https://doi.org/10.1016/j.knosys.2020.105967

Reyes-Menéndez, Ana; **Correia, Marisol B.**; **Matos, Nelson**; **Adap, Charlene** (2020). "Understanding online consumer behavior and eWOM strategies for sustainable business management in the tourism industry". *Sustainability*, v. 12, n. 21, 8972.
https://doi.org/10.3390/su12218972

Reyes-Menéndez, Ana; **Saura, José-Ramón**; **Álvarez-Alonso, César** (2018). "Understanding #WorldEnvironmentDay user opinions in *Twitter*: a topic-based sentiment analysis approach". *International journal of environmental research and public health*, v. 15, n. 11.
https://doi.org/10.3390/ijerph15112537

Reyes-Menéndez, Ana; **Saura, José-Ramón**; **Filipe, Ferrao** (2019). "The importance of behavioral data to identify online fake reviews for tourism businesses: a systematic review". *PeerJ computer science*, v. 5, e219.
https://doi.org/10.7717/peerj-cs.219

Saura, José-Ramón; **Reyes-Menéndez, Ana**; **Álvarez-Alonso, César** (2018). "Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels". *Sustainability*, v. 10, n. 9, 3016.
https://doi.org/10.3390/su10093016

Saura, José-Ramón; **Rodríguez-Herráez, Beatriz**; **Reyes-Menéndez, Ana** (2019). "Comparing a traditional approach for financial brand communication analysis with a big data analytics technique". *IEEE access*, v. 7, n. 1.
https://doi.org/10.1109/ACCESS.2019.2905301

Schmid, Hans-Jörg (2016). "Why cognitive linguistics must embrace the social and pragmatic dimensions of language and how it could do so more seriously". *Cognitive linguistics*, v. 27, n. 4, pp. 543-557.
https://doi.org/10.1515/cog-2016-0048

Singh, Akshit; **Shukla, Nagesh**; **Mishra, Nishikant** (2018). "Social media data analytics to improve supply chain management in food industries". *Transportation research part E: logistics and transportation review*, v. 114, pp. 398-415.
https://doi.org/10.1016/j.tre.2017.05.008

Singh, Devinder-Pal; **Kastanakis, Minas N.**; **Paul, Justin**; **Felix, Reto** (2021). "Non-deceptive counterfeit purchase behavior of luxury fashion products". *Journal of consumer behaviour*, v. 20, n. 5, pp. 1078-1091.
https://doi.org/10.1002/cb.1917

Sinpeng, Aim (2021). "Hashtag activism: social media and the #FreeYouth protests in Thailand". *Critical Asian studies*, v. 53, n. 2, pp. 192-205.
https://doi.org/10.1080/14672715.2021.1882866

Skill, Karin; **Passero, Sergio**; **Francisco, Marie** (2021). "Assembling *Amazon* fires through English hashtags. Materializing environmental activism within *Twitter* networks". *Computer supported cooperative work*, v. 30, n. 5-6, pp. 715-732.
https://doi.org/10.1007/s10606-021-09403-6

Smith, Nicholas; **Leiserowitz, Anthony** (2012). "The rise of global warming skepticism: exploring affective image associations in the United States over time". *Risk analysis*, v. 32, n. 6, pp. 1021-1032.
https://doi.org/10.1111/j.1539-6924.2012.01801.x

Stieglitz, Stefan; **Mirbabaie, Milad**; **Ross, Björn**; **Neuberger, Christoph** (2018). "Social media analytics-challenges in topic discovery, data collection, and data preparation". *International journal of information management*, v. 39, pp. 156-168.
https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Su, Leona-Yi-Fan; **Akin, Heather**; **Borssard, Dominique**; **Scheufele, Dietram A.**; **Xenos, Michael A.** (2015). "Science news consumption patterns and their implications for public understanding of science". *Journalism & mass communication quarterly*, v. 92, n. 3, pp. 597-616.
https://doi.org/10.1177/1077699015586415

Syed, Shaheen; **Spruit, Marco** (2017). "Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation". In: *IEEE International conference on data and advanced analytics*.
https://doi.org/10.1109/DSAA.2017.61

Tajfel, Henry (1974). "Social identity and intergroup behaviour". *Social science information*, v. 13, n. 2, pp. 65-93.
https://doi.org/10.1177/053901847401300204

Tajfel, Henry (1981). *Human groups and social categories. Studies in social psychology*. Cambridge: Cambridge University Press. ISBN: 0521280737

Terkourafi, Marina; **Haugh, Michael** (2019). "Quo vadis, pragmatics?". *Journal of pragmatics*, v. 145.
https://doi.org/10.1016/j.pragma.2019.04.005

Van-de-Donk, Wim; **Loader, Brian D.**; **Nixon, Paul G.**; **Rucht, Dieter** (2004). *Cyberprotest: new media, citizens and social movements*. Routledge. ISBN: 978 0 415297851

Van-Eck, Christel W.; **Feindt, Peter H.** (2022). "Parallel routes from Copenhagen to Paris: climate discourse in climate sceptic and climate activist blogs". *Journal of environmental policy & planning*, v. 24, n. 2, pp. 194-209.
https://doi.org/10.1080/1523908X.2021.2000376

Vanhala, Mika; **Lu, Chien**; **Peltonen, Jaakko**; **Sundqvist, Sanna**; **Nummenmaa, Jyrki**; **Järvelin, Kalervo** (2020). "The usage of large data sets in online consumer behaviour: a bibliometric and computational text-mining-driven analysis of previous research". *Journal of business research*, v. 106, pp. 46-59.
https://doi.org/10.1016/j.jbusres.2019.09.009

Veltri, Giuseppe A.; **Atanasova, Dimitrinka** (2017). "Climate change on *Twitter*: content, media ecology and information sharing behaviour". *Public understanding of science*, v. 26, n. 6, pp. 721-737.
https://doi.org/10.1177/0963662515613702

Washington, Haydn; **Cook, John** (2011). *Climate change denial*. Earthscan. ISBN: 978 1 849713368

Whitmarsh, Lorraine (2011). "Scepticism and uncertainty about climate change: dimensions, determinants and change over time". *Global environmental change*, v. 21, n. 2, pp. 690-700.
https://doi.org/10.1016/j.gloenvcha.2011.01.016

Whitmarsh, Lorraine; **Capstick, Stuart** (2018). "Perceptions of climate change". In: Clayton, Susan; Manning, Christie (eds.). *Psychology and climate change. Human perceptions, impacts and responses*. Amsterdam: Elsevier, pp. 13-33. ISBN: 978 0 128131305
https://doi.org/10.1016/B978-0-12-813130-5.00002-3

Whitmarsh, Lorraine; **Corner, Adam** (2017). "Tools for a new climate conversation: a mixed-methods study of language for public engagement across the political spectrum". *Global environmental change*, v. 42, pp. 122-135.
https://doi.org/10.1016/j.gloenvcha.2016.12.008

Williams, Hywel T. P.; **McMurray, James R.**; **Kurz, Tim**; **Lambert, F. Hugo** (2015). "Network analysis reveals open forums and echo chambers in social media discussions of climate change". *Global environmental change*, v. 32, pp. 126-138.
https://doi.org/10.1016/j.gloenvcha.2015.03.006

Wozniak, Antal; **Wessler, Harmut**; **Lück, Julia** (2017). "Who prevails in the visual framing contest about the United Nations climate change conferences?". *Journalism studies*, v. 18, n. 11, pp. 1433-1452.
https://doi.org/10.1080/1461670X.2015.1131129

Wu, Ming-Wen; **Su, Keh-Yih** (1993). "Corpus-based automatic compound extraction with mutual information and relative frequency count". In: *Proceedings of rocling VI Computational linguistics conference*, pp. 207-216.
https://aclanthology.org/O93-1009

Wu, Zezhou; **Zhang, Yan**; **Chen, Qiaohui**; **Wang, Hao** (2021). "Attitude of Chinese public towards municipal solid waste sorting policy: a text mining study". *Science of the total environment*, v. 756, 142674.
https://doi.org/10.1016/j.scitotenv.2020.142674

Xiong, Ying; **Cho, Moonhee**; **Boatwright, Brandon** (2019). "Hashtag activism and message frames among social movement organizations: semantic network analysis and thematic analysis of *Twitter* during the #MeToo movement". *Public relations review*, v. 45, n. 1, pp. 10-23.
https://doi.org/10.1016/j.pubrev.2018.10.014

Zappavigna, Michele (2015). "Searchable talk: the linguistic functions of hashtags". *Social semiotics*, v. 25, n. 3, pp. 274-291.
https://doi.org/10.1080/10350330.2014.996948

Zoller, Heather M.; **Casteel, Diana** (2021). "#March for our lives: health activism, diagnostic framing, gun control, and the gun industry". *Health communication*, v. 37, n. 4.
https://doi.org/10.1080/10410236.2020.1871167