**ORIGINAL RESEARCH**

# World Environment Day: Understanding Environmental Programs Impact on Society Using Twitter Data Mining

Enara Zarrabeitia-Bilbao[1] · Rosa-María Rio-Belver[2] · Izaskun Alvarez-Meaza[1] · Itziar Martínez de Alegría-Mancisidor[1]

**Abstract**
Environmental sustainability awareness has encouraged the promotion of a number of environmental programs and initiatives and, accordingly, the use of social networks for the dissemination and support of these initiatives has grown significantly. Thus, the purpose of the work is to understand United Nations World Environment Day (WED) programs impact on the digital public debate using Twitter data mining. For that, an ad hoc methodology is designed to provide it to authorities and organizations that wish to analyze the impact of different initiatives or programs on society. All in all, the research carried out analyzes more than 400,000 tweets sent during the 2021 edition of the WED. The tweets have been processed using Big Data techniques and Social Network Analysis. The research reveals that the WED was a trending topic initiative that was discussed in positive terms, where collective sentiment was shown. The topics covered dealt with the event day and the different initiatives related to restoration of ecosystems. However, it is noted that: there is no coordinated action by the institutions, groups or individuals involved in the conversation and the initiative tends towards homophily; digital mobilization is mostly centered in the host country (Pakistan) and, above all, in the neighboring country (India) and, the conspicuous absence of the business sphere in the discussion.

**Keywords** World Environment Day · Twitter · Social network analysis · SNA metrics · Environmental programs

✉ Enara Zarrabeitia-Bilbao
enara.zarrabeitia@ehu.eus

Rosa-María Rio-Belver
rosamaria.rio@ehu.eus

Izaskun Alvarez-Meaza
izaskun.alvarez@ehu.eus

Itziar Martínez de Alegría-Mancisidor
itziar.martinezdealegria@ehu.eus

1    Faculty of Engineering, Industrial Organization and Management Engineering Department, University of the Basque Country, Pl. Ingeniero Torres Quevedo, Bilbao, Spain

2    Faculty of Engineering, Industrial Organization and Management Engineering Department Vitoria-Gasteiz, University of the Basque Country, Nieves Cano 12, Vitoria-Gasteiz, Spain

# 1 Introduction

For decades, politicians, social movements and media have been playing an essential role in the process of shaping public opinion (Mazzoleni, 2010). However, it is an undisputed fact that the advent of the Internet and digital social platforms have reconfigured the way in which public opinion is created, altering the roles traditionally assigned to politicians, media and social movements (Lesaca, 2015). With the irruption of new communication channels, social media platforms have become a fundamental pillar for the dissemination of messages and slogans and, above all, for social mobilization and the creation of a certain state of opinion (Carrasco-Polaino et al., 2017). All in all, nowadays, institutions, different types of groups or individuals make use of digital platforms to express their opinion and influence on different topics of discussion.

The use and expansion of the main digital social networks such as Facebook, YouTube, WhatsApp, Instagram or Twitter (Statista, 2021) reinforce the idea that society is facing a form of socializing that has online interaction as one of its main supports. In this context, the microblogging social network Twitter stands out since it allows the massive exchange of interpersonal communication that can be captured, stored, represented and analyzed (Del-Fresno-García, 2014; Wu et al., 2011). Thus, Twitter has aroused great interest in the academic community due to, among other things, the type of conversation that takes place there, the impact it has had on different social movements that have emerged in recent years and the possibility of analyzing all of that (Alhindi et al., 2012; Edrington & Lee, 2018; Li et al., 2021).

In this context, a topic that has generated great interest within social networks in recent years has been the environment (Burak, 2017; Hendriks et al., 2016). The importance of the environment is obvious since all organisms obtain all the elements they need to live from the environment: from air and water, to shelter and food that allow them to grow, develop and obtain energy. Maintaining the balance of the environment is fundamental to sustaining life on Earth, as well we know (Concepto, 2021). However, the planet is suffering from a triple crisis: climate change, loss of biodiversity and pollution (UN, 2021). Thus, the importance of environmental sustainability awareness has encouraged the promotion of a number of environmental programs and initiatives.

Thereby, of late the use of social networks to disseminate and support different environmental initiatives has grown significantly. In this regard, the purpose of the work is to understand World Environment Day program impact on the digital public debate using Twitter data mining. Thus, a methodology designed ad hoc will be provided to those authorities and organizations that wish to analyze the impact of different initiatives or events on society.

# 2 World Environment Day Program

World Environment Day (WED) is a United Nations (UN) environmental initiative started back in 1974 and celebrated every year on June 5, with the participation of governments, businesses and citizens in an effort to address urgent environmental issues. The program aims to raise awareness and inspire actions on urgent issues: from marine pollution and climate change to sustainable consumption and wildlife crime. Hence, multiple activities take place in this multimedial event (WED, 2021). In many countries, this celebration is an

opportunity to sign or ratify international conventions and to establish permanent governmental structures related to environmental management and economic planning. Over all, WED provides a global platform to inspire positive environmental changes.

Each year, WED is celebrated in a different country, where the official celebrations are held. For 2021, the host country was Pakistan. In addition, every year World Environment Day revolves around a specific theme. In this case, WED (2021) had as its theme 'Ecosystem restoration' with the slogan 'Reimagine, Recreate, Restore'. In this regard, June 5, 2021 also marked the launch of the UN decade of Ecosystem Restoration, a ten-year global push to prevent, halt and reverse ecosystem degradation (WED, 2021).

All in all, under the premise of protecting and improving the environment and taking care of the health of the planet, WED has grown to be a popular and global initiative.

## 3 Literature Review and Related Work

In this section, we review previous works related to the use of Twitter as a tool for social research, in general, and for studying the environmental crisis that the planet is suffering, in particular.

### 3.1 Twitter as a Tool for Social Research

Since its launch in 2006, the microblogging platform Twitter has attracted increasing interest from different scientific fields. The first scientific studies related to Twitter appeared shortly after the launch of the platform, in the period from 2007 to 2008 (Cormode et al., 2010; Navas, 2018). These works, as expected, were descriptive and addressed basic characteristics of the tool (Java et al., 2007; Krishnamurthy et al., 2008). However, subsequently, in 2009, scientific studies began to focus on the content of tweets, and the first linguistic and semantic analyses were developed (Cormode et al., 2010; Navas, 2018); thus, the first works of content classification (Dann, 2010; Naaman et al., 2010; Pear-Analytics, 2009) and trend analysis were carried out (Cheong & Lee, 2009). Later, as of 2011, conversation definitely became the central activity of the scientific studies that took Twitter as a research tool. Scientific research shifted from a user and the tweet based perspective to one based on the interactions established between users (Navas, 2018).

Today, Twitter has become one of the virtual environments most conducive to the collection of large volumes of data at very fast speeds for subsequent data analysis through Social Network Analysis and Machine Learning techniques (Morales-i-Gras, 2020). Thus, Twitter is recognized as a leading channel in a variety of areas such as politics or business and, therefore, has taken a leading role in the scientific field of social sciences.

Several studies have analyzed political and current affair issues through Twitter. The different studies have focused on the analysis of the message senders and recipients, the political debate generated around Twitter, as well as the use of Twitter in electoral campaigns. In these works, Twitter is presented, usually, as a tool with a great functionality for both politicians and citizens (Campos-Domínguez, 2017; Marín-Dueñas et al., 2019). Furthermore, Twitter has had enormous impact on various well-known political and social mobilizations of recent years, such as the Arab Spring, Occupy Wall Street or Me Too movements.

It also highlights the role of Twitter in business studies in terms of engagement or management strategy. Due to the emergence of social networks, organizations, both public and private, are no longer the only source of information about entities, brands or products.

Thus, the communication process has ceased to be unidirectional to start the search for a continuous dialogue (Fernández-Gómez & Martín-Quevedo, 2018; Medina et al., 2018).

Likewise, the use of Twitter as a tool to predict social phenomena is recurrent in recent scientific researches. Though different machine learning techniques, successful works can be found in multiple contexts using Twitter for predictive tasks, such as: electoral or engagement predictions, stock market predictions or pandemic detection, among others (Coletto et al., 2015).

All in all, thanks to the data obtained from Twitter and the subsequent analysis of the information, it is possible to describe, explain, interpret and comprehend how and why social networks are used and which social effects they generate (Casero-Ripollés, 2018).

### 3.2 Environmental Crisis and Twitter Data Mining

There are many ways to analyze public opinion and the dynamics generated on environmental issues. However, the impact that social networks have on public perception and reaction of topics related to the environment is an established fact. Specifically, different studies have addressed issues of climate change, pollution and biodiversity loss through Twitter data mining.

Climate change is the greatest environmental threat facing humanity and the consequences could be devastating if dependence on fossil fuels is not drastically reduced (Greenpeace, 2021). In this context, numerous conversations on this topic are taking place on Twitter and being analyzed by the scientific community. Recent studies show who is behind the different discussions (politicians, activists, business, NGOs, celebrities, etc.) and what the content of the tweets is. Likewise, events related to climate change have been analyzed to understand how the public focuses attention on aspects of climate change (Fownes et al., 2018).

The topic of pollution is another recurring issue in Twitter conversations and it is reflected in the different scientific studies carried out on the subject. The research studies gather information on pollution and analyze the impact that different types of pollution, such as air, noise or plastic pollution has on citizens (Juanals & Minel, 2018; Otero et al., 2021; Peplow et al., 2021).

In the case of biodiversity conversations on Twitter, this is a less popular topic within the scientific community. However, there are works seeking to obtain information on species through Twitter (Daume, 2016) as well as considering how narrowly focused advocacy on platforms like Twitter will contribute to effective global biodiversity conservation (Barrios-O'Neill, 2020).

In addition, previous works have analyzed the discussions generated on Twitter in different environmental campaigns or programs such as 'Plastic Free July' or previous WED years (specifically, 2015 and 2018) (Heidbreder et al., 2021; Pang & Law, 2017; Reyes-Menendez et al., 2018). These studies analyze the topics of conversation, the sentiment of the discussion and the visual rhetoric used, however, none of them makes an exhaustive analysis of the impact of environmental movements on virtual public debate.

Therefore, although there are many studies that analyze different critical environmental issues, none of them goes deeper into the analysis of the impact that a targeted or planned conversation, i.e., one that does not arise spontaneously (as is the case of an initiative such as WED) has on society in order to assess whether the objectives have been achieved, both in terms of content and the public that has participated in the discussion.

## 4 Research Questions and Methodology

This study is framed within the framework of Social Network Analysis, an analytical trend focused on identifying interaction patterns between actors, based on the use of Big Data.

Specifically, this research focuses on the interactions between Twitter users in relation to WED (2021). As previously mentioned, Twitter was chosen as data source because of the potential demonstrated by this social network for variety of domains and scientific fields including recent studies about environmental issues, in general, and about WED, in particular.

In summary, this research work proposes a unique methodological design, developed ad hoc to respond to different research questions based on the data collected from Twitter.

### 4.1 Research Questions

The purpose of the study is to understand World Environment Day program impact on the digital public debate using Twitter data mining, and the approach to the main objective has been carried out based on different research questions about WED (2021), namely:

- *RQ1*: What are the main characteristics of the overall digital conversation?
- *RQ2*: What does the overall conversation revolve around?
- *RQ3*: (a) Which are the main communities generated in the conversation? (b) What is the presence of governments, business and citizens? (c) What is the presence of the most polluting countries? (d) What is the behavior of the interactions in each community (e) What does the conversation revolve around in each community?
- *RQ4*: Who are the most influential players in the overall conversation?

### 4.2 Data Extraction and Analysis Method

The workflow followed for data extraction and data analysis is summarized in Fig. 1.

- **Data extraction and preparation**

The first step in order to undertake the study was the extraction and preparation of the necessary data. In this regard, the social media used for data extraction was Twitter.

The sampling procedure adopted the criteria of data collection based on the content of the tweets: #WorldEnvironmentDay, the language used in the tweets: English, and the day of publication of the tweets: June 5, 2021.

For the collection of the tweets, the software used was T-Hoarder, an open source system that, based on Python language, is able to establish a connection with the Twitter API and retroactively download data and process it (Congosto et al., 2017; GitHub, 2021).

Subsequently, it was necessary to prepare data for further data analysis. The open source software used to clean, refine and extract the data was OpenRefine (OpenRefine, 2021) and Orange Data Mining (Ljubljana-University, 2021).
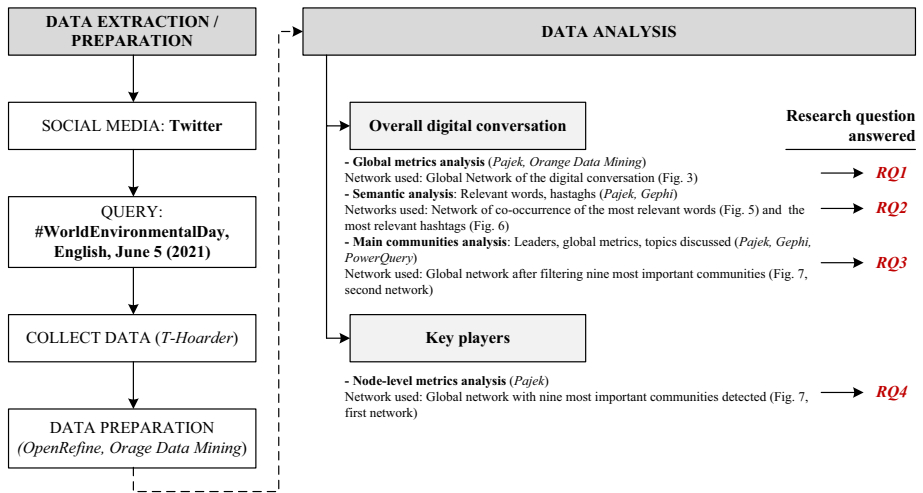
- **Data analysis**

**Fig. 1** Workflow of the scientific research

The exploratory and descriptive study was carried out in two stages. In the first stage, the overall digital conversation was analyzed (to answer RQ1, RQ2 and RQ3) and in the second stage, key players of the conversation were identified (to answer RQ4).

*RQ1*: What are the main characteristics of the overall digital conversation?

In order to answer this first question, first of all, the global network that contains the overall digital conversation about WED (2021) has been obtained. For that, once the tweets were obtained, a process of extracting mentions was carried out through the software OpenRefine to synthesize a network based on which users mention which users in the conversation itself. Thus, a graph will be obtained where each node represents a participant in the conversation (a Twitter user), and where each arc represents a Twitter mention type interaction (from one Twitter user to another Twitter user): a retweet, a direct response or an interpellation within a regular tweet.

In relation to the network obtained, information about the general morphology of the digital conversation was obtained to analyze its main characteristics. To this purpose different metrics (see Table 1) were calculated with the open source software Pajek (Mrvar & Batagelj, 2021) and Orange Data Mining.

*RQ2*: What does the overall conversation revolve around?

The topic of the digital conversation was analyzed through semantic analysis, both through the most relevant words used in the conversations established between users of the social network, and through the hashtags or keywords used by the users to tag their messages. Thus, different thematic blocks of the digital conversation have been identified through the words used in the tweets and the framework of those blocks through the hashtags.

**Table 1** Global metrics for overall digital conversation analysis (Hutto, 2022; Hutto & Gilbert, 2014; Orbegozo-Terradillos et al., 2020; Wasserman & Faust, 1994)

| Metric | Meaning in the digital conversation |
|---|---|
| Total impacts | Number of publications captured, whether tweets or retweets |
| Nodes (users) | Number of unique users who have participated in the conversation (excluding those who have not mentioned or have been mentioned by any other user) |
| Arcs (interactions) | Number of unique interactions established between users. Each arc has a weight associated with the number of mentions it represents |
| Average impacts (per user) | Number of publications that each user has published on average |
| Density | Number of connections exist between network users divided by the total possible connections between network users |
| Average degree | Number of users (nodes) that each node (user) is associated with on average |
| Maximum distance | Number of arcs between nodes separating the two farthest reachable nodes in the network |
| Average distance | Number of average arcs between nodes separating reachable nodes in the network |
| Input Degree Centralization | Number between 0 and 1, where 0 indicates total decentralization (all nodes receive mentions from the same number of unique users) and 1 indicates total centralization (a single node receives mentions from all other unique users) |
| Output Degree Centralization | Number between 0 and 1, where 0 indicates total decentralization (all nodes send mentions to the same number of unique users) and 1 indicates total centralization (a single node sends mentions to all other unique users) |
| Betweenness Centralization | Number between 0 and 1 where 0 indicates total decentralization (all nodes have the same intermediation index, all nodes are equal in intermediation) and 1 indicates total centralization (a star network) |
| Sentiment score (VADER model) | VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that performs exceptionally well in the social media domain. The compound output variable (sentiment score) is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between $-1$ (the most negative sentiment, strongly negative) and $+1$ (the most positive sentiment, strongly positive). This metric is the most appropriate if a single unidimensional measure of sentiment is desired for a given sentence |
| Modularity (Louvain Multi-Level algorithm) | Metric associated with the quality of the partitioning of the network into different clusters or communities produced by the Louvain algorithm with Multi-Level Refinement (values equal to or greater than 0.3 indicate mathematical significance) |
| Number of Clusters | Total number of clusters or communities identified by the Louvain Multi-Level Algorithm |

Hence, two networks were generated: the first one with the tweets and the most highlighted words (the 150 most relevant that appear in the tweets, obtained through TF-IDF, Term Frequency-Inverse Document Frequency correction procedure), and the second one with the tweets and the most highlighted hashtags (the 150 that most frequently appear in the tweets). Afterwards, once the weakest relations have been removed, one graph of most relevant words and other graph of most relevant hashtags in the digital conversation based on how many tweets co-occur were generated (see Fig. 2).

The software Pajek was used to transform the two-mode network (when in the same network the nodes represent different ontological entities) into a single-mode network (when in the same network the nodes represent the same ontological entity) (see Fig. 2) and the software Ghepi was used for the visualization and study of the relationships between words and hashtags (Bastian et al., 2009).

*RQ3*: (a) Which are the main communities generated in the conversation? (b) What is the presence of governments, business and citizens? (c) What is the presence of the most polluting countries? (d) What is the behavior of the interactions in each community (e) What does the conversation revolve around in each community?

To finish the overall digital conversation analysis, the main communities generated around WED (2021) were identified and analyzed.

Once the main communities have been identified (to answer question *a*), to analyze each community, concretely, each main leaders (to answer questions *b* and *c*), different global metrics (to answer question *d*) and discussed topics (to answer question *e*) have been studied.

The software used to identify communities was Pajek, concretely, by applying the Louvain Multi-Level algorithm (with 10 restarts) (Blondel et al., 2008). After that, the synthesized network and the detected communities were processed through Gephi for their visualization and analysis. The tool used to analyze what type of contents have been important in each community has been PowerQuery-Excel (Microsoft, 2021).

*RQ4*: Who are the most influential players in the overall conversation?

Finally, in the second stage, key players of the overall digital conversation were analyzed through Pajek and Node-level metrics (see Table 2).
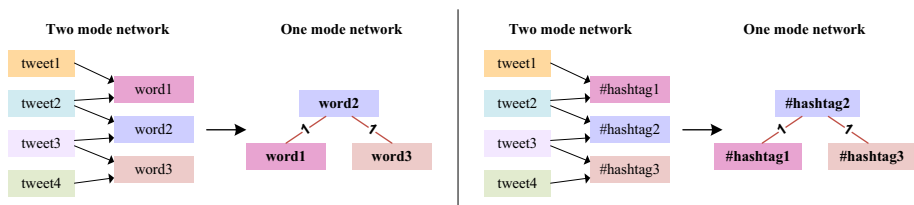


**Fig. 2** Edges and nodes for most highlighted words and hashtags networks

**Table 2** Node-level metrics to identify the key players in the network (Apodaka & Morales-i-Gras, 2016; Orbegozo-Terradillos et al., 2020; Wasserman & Faust, 1994)

| Metric | Meaning in the digital conversation |
| --- | --- |
| Input Degree Centrality | Number of mentions received by a user on the network (direct mentions, replies or retweets) |
| Output Degree Centrality | Number of mentions sent by a user on the network (direct mentions, replies or retweets) |
| Betweenness Centrality | Measure of how often a node will fall in the middle of the shortest path between two nodes |

## 5 Results and Discussion

Throughout June 5, 2021, on World Environment Day, a total of 415,121 tweets in English were sent out with the hashtag #WorldEnvironmentDay.

Moreover, considering that Twitter identifies trends in public opinion based on parameters such as the level of interaction with certain hashtags, it is worth noting that the hashtag #WorldEnvironmentDay was the most used hashtag worldwide for 7 consecutive hours and remained in the Top-10 hashtags for 10 consecutive hours. Furthermore, in the case of India, the Twitter users placed #WorldEnvironmentDay at the top of the ranking for 12 consecutive hours.

### 5.1 Overall Digital Conversation

The following three sections analyze the overall conversation around WED (2021).



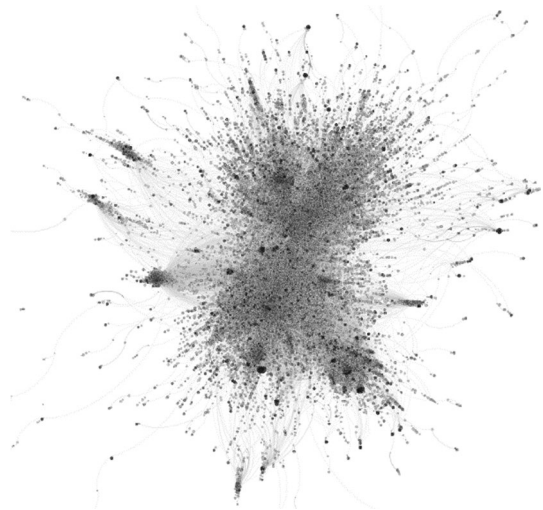**Fig. 3** Network about #WordEnvironmentDay digital conversation

**Table 3** Global metrics extracted from the overall digital conversation

| | |
|---|---|
| Total impacts | 415,121 |
| Users (nodes) | 187,322 |
| Arcs (interactions) | 470,594 |
| Average impacts (per user) | 2.22 |
| Density | 0.00001341 |
| Average degree | 5.02443920 |
| Maximum distance | 24 |
| Average distance | 9.24281 |
| Input Degree Centralization | 0.06173120 |
| Output Degree Centralization | 0.00440149 |
| Betweenness Centralization | 0.00695135 |
| Sentiment score (VADER model) | 0.366685821 |
| Modularity (Louvain Multi-Level algorithm) | 0.858381 |
| Number of Clusters | 6.089 (9 > 4%) |

### 5.1.1 Global Metrics Analysis [RQ1]

As an initial approach to the digital conversation, Fig. 3 shows the global network that contains the overall digital conversation and Table 3 shows the values of global metrics of the overall digital conversation.

The captured conversation has been transformed to a network graph containing a total of 187,322 actors and 470,594 weighted connections (see Fig. 3). Thus, the general approach to the network makes it possible to observe that each actor in the network is connected on average to 5.02 other actors, which gives a scenario of few connections. In addition, the density of the network shows that only 0.001341% of the possible links between users have occurred. It cannot be said that the members form a well-connected network; hence, it is suggested that there are still many unanswered key questions for WED (2021).

Regarding centralization metrics, the low values of input degree centralization (6.17312%) and the output degree centralization (0.440149%) show that the network is decentralized in both input and output degree. Hence, neither the reception nor the emission of mentions is dominated by a small group of actors; however, it is a network in which mentioning and being mentioned are clearly different phenomena ($C_D^{in} = 0.06173120$ vs $C_D^{out} = 0.00440149$). Betweenness centralization is also low (0.695135%), therefore, the intermediation of the network is not in the hands of a small group of actors, it is distributed in a horizontal way.

The overall sentiment found in the digital discussion is positive (0.366685821). It seems that users opt for a positive tone to celebrate a day that calls for action to prevent, halt and reverse the degradation of the planet. Figure 4 shows that 67.9% of the tweets have been positive, and specifically, 45.43% of the tweets have been categorized as very positive.

Finally, after applying the Louvain Multilevel community detection algorithm, a total of 6,089 different communities have been identified with a modularity of 0.86 (high mathematical significance). This is a relatively large number of communities, many of them composed of short interactions, which is usual in analytical strategies of massive data, with a very good value for modularity. Among the more than 6,089 communities identified, the threshold has been set for those that contain at least 4% of the nodes, a total of 9. A more in-depth analysis of the main communities detected will be made in Sect. 5.1.3.
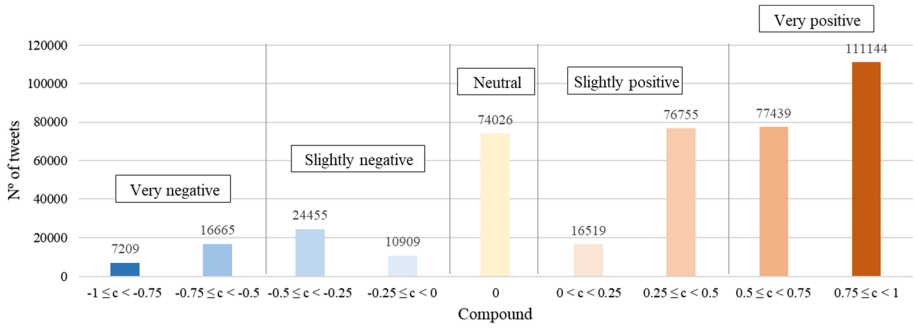
**Fig. 4** Sentiment score of the overall digital conversation

### 5.1.2 Semantic Analysis [RQ2]

Semantic analysis shows the topic hotspots of the overall digital discussion. The most relevant words network (see Fig. 5) and most relevant hashtags network (see Fig. 6) analysis makes it possible to identify key topic points around WED (2021).

On the one hand, with regard to most relevant words network, from the 150 main initial words, once stopwords, mentions, hashtags and emoji have been removed, the weakest links have been deleted. The minimum weight taken into account for the edges has been 900, i.e., 17.72% of edges were considered. Moreover, only communities grouping 2% or more of the nodes in the network have been taken into account, i.e., 97.33% of nodes were considered. Finally, in order to consider the favorable positions of the words, the nodes have been weighted according to the intermediation or betweenness centrality (Hanneman, 2001).
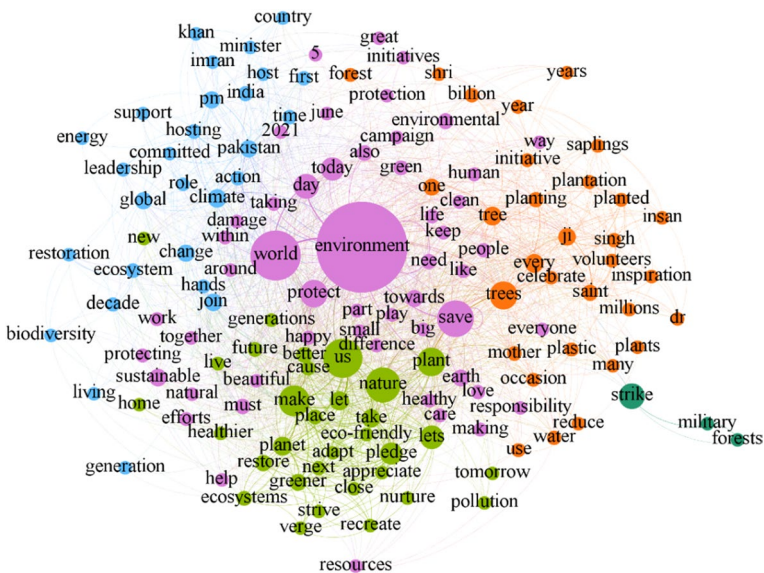


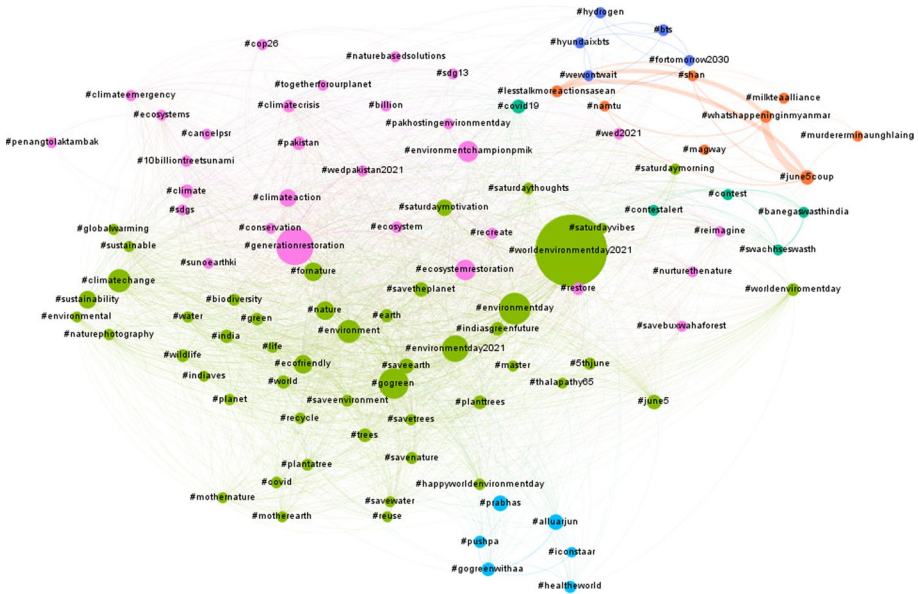**Fig. 5** Network of co-occurrence of the most relevant words

**Fig. 6** Network of co-occurrence of the most relevant hashtags

Thus, the clustering of the most relevant words defines five clusters or communities. These communities represent topics of conversation, i.e., the main discussions held about WED (2021). So, through direct observation of the network, and by searching for these words in most relevant tweets (the most retweeted), the following five communities are found:

The first community (in pink), is formed around the term 'environment' and it is related to the event day ('environment', 'world', 'day', 'today', '5', 'june', '2021'). In addition, positive communication in the conversation ('happy', 'love', 'protect', 'protection', 'protecting', 'save') and the sense of collectivity of the subject ('together', 'people', 'everyone') are observed. The second one (in light green), is slightly formed around the terms 'us' and 'nature' and mentions the central theme of WED (2021): Ecosystem Restoration ('ecosystems', 'recreate', 'restore'). In this case, the positivity and collectivity of the message ('better', 'us') are also observed. The third one (in blue) is not formed around any influential word and the topics of conversation revolve around the host country and its prime minister ('pakistan', 'host', 'country', 'khan', 'imran', 'minister', 'pm', 'leadership'). The fourth one (in orange) is formed around the term 'trees' and it is related to the nature restoration initiatives, in general ('reduce', 'use', 'water', 'plastic'), and concretely to the initiative 'Ten Billion Tree Tsunami', which aims to achieve such reforestation by 2023 ('tree', 'initiative', 'plantation', 'planted', 'planting', 'billion'). Finally, the fifth and smallest one (in dark green) is related to the protest action staged by Myitkyina's main strike group to mark WED (2021), as forests in Myanmar are being devastated by military.

After removing the hashtag '#WorldEnvironmentDay' as it is used in the query, the procedure followed with hashtags for semantic analysis has been similar to the one used with the words. The minimum weight taken into account for the edges has been 100, i.e., 19.69% of edges were considered; and only communities grouping more than 2% of the

nodes in the network have been taken into account, i.e., 68.92% of nodes were considered. The nodes have been weighted according to the intermediation or betweenness centrality.

In this case, the clustering of most important hashtags, i.e., most relevant grouping channels for discussion on specific topics, provides six clusters or communities of topics of conversation. However, of the six communities obtained, two stand out. The first one (in light green), is formed around the hashtag '#WorldEnvironmentDay2021' and the second one (in pink) is formed around the official hashtag for the main theme of the event '#GenerationRestoration'. Most of the hashtags included in the first group refer to the event itself through different names of the event, date of celebration and essence of the event ('#environmentday', '#environmentday2021', '#5thjune', '#june5', '#gogreen', '#climatechange'), but there are also specific proposals carried out in relation to WED (2021) ('#savetrees', '#plantatree', '#savewater'). In the second case, hashtags refer to ecosystem restoration ('#ecosystemrestoration', '#reimagine', '#recreate', '#restore') but also to the host country ('#pakistan', '#wedpakistan2021', '#pakhostingenvironmentday', '#environmentchampionpmik').

In the case of the smaller groups, the orange cluster refers to the coup d'état that took place in Myanmar on February 1 2021; the light blue cluster refers to India's seventh art and its celebrities; the dark blue refers to the music video launch by Hyundai Motor's, featuring pop icon BTS, to celebrate WED; and finally, the dark green cluster refers to Banega Swasth India Campaign for better hygiene and sanitation practices in India.

### 5.1.3 Main community Analysis (RQ3)

Figure 7 shows the network of the main 9 communities generated around #WordEnvironmentDay digital conversation.

In order to identify who the main communities are, their respective leaderships (top 5 of each community) were analyzed according to the input degree (see Table 4). From the data obtained, two main types of clusters can be observed: clusters belonging to the political sphere (cluster 4, 12, 14, 7, 15 and 8) and clusters belonging to the cultural and entertainment sphere (cluster 13 and 17) (cluster 22 is classified as miscellaneous). Thus, the absence of clusters formed around business leaders is notable. Although isolated actions by governments and civil society will not be enough to address the most pressing environmental challenges (UNEP, 2021), the private sector is hardly present in this discussion.

In addition, there is a strong Indian presence in the conversation. With the exception of the three clusters (which are positioned in the first half of the network: cluster 4, 7 and 8) the rest (which are positioned in the second half of the network: cluster 13, 12, 14, 15, 17 and 22) are mainly led by Indian accounts.

With China, the United States and India (DataBank, 2021) being the three countries that pollute the most through their carbon dioxide ($CO^2$) emissions, the weak presence of the U.S. in the conversation is surprising (not so in the case of China, where Twitter has been blocked since 2009 (The Guardian, 2009)); despite the United States being the second most polluting (and industrialized) country, it appears only sparsely in cluster 8.

Thus, digital mobilization is mostly focused (without taking into account the main community, cluster 4) on the host country and, above all, the neighboring country: India. The same can be seen in the coverage of the event by newspapers (news in English) on the day of the event and on the following day (June 5, 2021 and June 6, 2021): In first position, 108 news articles were published in Pakistan; in second position, India
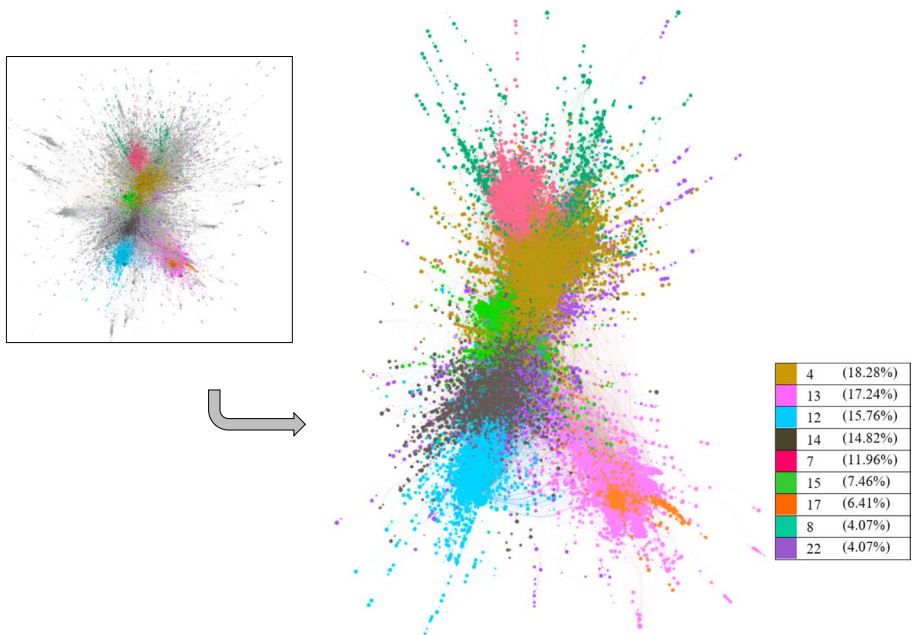
**Fig. 7** Network about #WordEnvironmentDay digital conversation with the nine most important communities and after filtering them (The community numbers have been randomly assigned by the algorithm and should not be interpreted ordinarily. The value in parentheses represents the percentage of nodes in each community)

with 98; the US ranks sixth with only 18 news articles; and China is in twelfth place with only 7 publications (ProQuest, 2021).

Table 5 shows global metrics for the main communities. This information shows that, as expected, the densities of the main communities are higher than the density of the overall network, nevertheless, they are still low i.e., connectivity in the communities is low.

In the case of input degree centralization, clusters 12, 15 and 17 stand out for being highly centralized clusters (communities with a centralization rate of higher than 50%). These clusters respond to a retweeting behavior of one or a few actors, i.e., many users mention a few users. Belonging to a community of 'high' or 'low' centralization has to be understood as a behavioral trait of an actor on Twitter. The actors of highly centralized communities participate in the conversation by exploiting more 'follower' or even 'warlordist' resources, often limiting themselves to disseminating major opinion leaders and reproducing their opinions. Moreover, there is evidence to affirm that the type of message (more or less emotional or more or less visceral) is also related to this phenomenon (Larrondo et al., 2019). In this sense, it is observed that two highly centralized communities (cluster 12 and 15) show the least positive sentiment.

Regarding output centralization, no strange behavior is detected in any community. It does not seem that there is any suspicious account from which a large number of messages have been sent. Likewise, concerning the betweenness centralization, the intermediation in the main communities is not monopolized by a few users.

Finally, the type of content that has been important in each of the communities is analyzed (see Table 6) by close reading of the most retweeted tweets of each main

**Table 4** Main community leaders

| Community /assigned name | Leaders (input degree) | Profile of leaders |
|---|---|---|
| **4 – Program organizer (UN)** | @unep (11566)<br>@un (3654)<br>@sudarsansand (2722)<br>@antonioguterres (1914)<br>@unbiodiversity (834) | UN Environment Programme<br>United Nations<br>Sand Artist (India)<br>Secretary-General of the UN<br>UN instrument for the conservation of biological diversity |
| **13 – Indian celebrities** | @urstrulymahesh (7140)<br>@imro45 (3274)<br>@sachin_rt (1435)<br>@actorvijay (1003)<br>@mohanlal (961) | Actor and producer<br>Cricketer<br>Former cricketer<br>Actor and producer<br>Actor and producer |
| **12 – Indian politicians (center-left) and activist** | @rahulgandhi (10269)<br>@srivatsayb (1569)<br>@priyankagandhi (775)<br>@srinivasiyc (755)<br>@hansrajmeena (667) | Politician, member of the parliament and former president of the INC<br>Politician (IYC)<br>Politician (INC)<br>Politician, president of the IYC<br>Activist |
| **14 – Indian institutional policy** | @narendramodi (6494)<br>@pmoindia (1386)<br>@prakashjavdekar (1384)<br>@bjp4india (1217)<br>@rashtrapatibhvn (1194) | Prime minister of India (BJP)<br>Office of the Prime Minister of India (BJP)<br>Minister of Environment, Forest & Climate Change (BJP)<br>Official Twitter account of the BJP<br>President of India |
| **7 – Pakistan and UK institutional policy** | @imrankhanpti (4547)<br>@pakpmo (3419)<br>@climatechangepk (1759)<br>@cop26 (1405)<br>@borisjohnson (1311) | Prime Minister of Pakistan<br>Official account of Prime Minister's Office (Pakistan)<br>Ministry of Climate Change (Pakistan)<br>The official account for the UN Climate Change Conference<br>Prime Minister of the UK |
| **15 – Indian and international activist and policy** | @sadhgurujv (5613)<br>@moefcc (3434)<br>@iucn (2534)<br>@unccd (2251)<br>@adgpi (675) | A Yogi, Mystic & Visionary (India)<br>Ministry of Environment, Forest and Climate Change (India)<br>International Union for Conservation of Nature<br>UN Convention to Combat Desertification<br>Additional Directorate General of Public Information (India) |
| **17 – Indian celebrities** | @alluarjun (6489)<br>@behindwoods (1082)<br>@trendsalluarjun (986)<br>@imsarathchandra (804)<br>@vamsikaka (516) | Actor<br>Digital Media organization<br>Official Twitter handle for trends related to Allu Arjun<br>Head of Content & Digital for Allu Arjun<br>Media/PR Consultant |
| **8 – US and Canadian policy** | @katieporteroc (374)<br>@justintrudeau (348)<br>@potus (183)<br>@secblinken (172)<br>@statedept (143) | Politician (member of the Democratic Party, US)<br>Prime Minister of Canada<br>President of US<br>US Secretary of State<br>US Department of State |
| **22 – Miscellaneous** | @parveenkaswan (1514)<br>@indiahistorypic (446)<br>@britishlibrary (212)<br>@listenshahid (166)<br>@arunbothra (101) | Indian Forest Service<br>Account with photos from Indian history<br>The British Library<br>Indian Administrative Service officer<br>Indian Police Service officer |

**Table 5** Main communities' global metrics

| | Density | Input Degree Centralization | Output Degree Centralization | Betweenness Centralization | Sentiment score |
|---|---|---|---|---|---|
| 4 | 0.00013732 | 0.34594762 | 0.01350035 | 0.02919706 | 0.39542769 |
| 13 | 0.00008713 | 0.39763318 | 0.00218030 | 0.00000467 | 0.293601391 |
| 12 | 0.00011473 | 0.64059298 | 0.00477769 | 0.00029309 | 0.124564 |
| 14 | 0.00016157 | 0.38198586 | 0.03605687 | 0.00022777 | 0.349070298 |
| 7 | 0.00035847 | 0.38416856 | 0.01703971 | 0.02109497 | 0.442542154 |
| 15 | 0.00034924 | 0.76423993 | 0.00333343 | 0.00246360 | 0.193382336 |
| 17 | 0.00052026 | 0.82885649 | 0.01250307 | 0.00078377 | 0.386701175 |
| 8 | 0.00036133 | 0.09080740 | 0.01210611 | 0.00025687 | 0.379348411 |
| 22 | 0.00042051 | 0.29984251 | 0.00984253 | 0.00195539 | 0.330839265 |

**Table 6** Topics discussed in the main communities

| | Main topics | Main tweets | Retweets (05-06-2021) |
|---|---|---|---|
| 14 | Global and official messages about the main slogan for 2021 UN World Environment Day: Reimagine, Recreate and Restore. The need of urgent actions for the world environment and ecosystem restoration. The launch of the UN Decade on Ecosystem Restoration. |  | 457 |
| 13 | Encourage necessary joint actions (such as tree planting) for environment and ecosystem restoration. |  | 1822 |
| 12 | Criticism and reference to social responsibility, in general, and political responsibility (Indian government), in particular, in the degradation of the environment. Encourage necessary joint actions (such as tree planting) for environment and ecosystem restoration. |  | 3077 |
| 14 | Indian government's commitment to fight for environment and ecosystem restoration and for a sustainable future. |  | 878 |
| 7 | Pakistan: host country of the 2021 UN World Environment Day. The revolutionary ten billion tree tsunami project. |  | 633 |
| 15 | Mystic criticism and reference to social responsibility in the degradation of the environment. |  | 1602 |
| 17 | Encourage necessary joint actions for environment and ecosystem restoration. The Indian actor Allu Arjun encourages netizens to plant trees. |  | 1106 |
| 8 | Criticism and reference to political responsibility (U.S. and Canadian government) in the degradation of the environment. US and Canadian government's commitment to fight for environment and ecosystem restoration and for a sustainable future. |  | 158 |
| 22 | Encourage necessary joint actions (such as tree planting) for environment and ecosystem restoration. |  | 346 |

community. Thus, it can be seen that these content are in line with the need of joint actions for environment and ecosystem restoration and the criticism and reference to political and civil society responsibility in the degradation of the environment. Once again, the absence of business or private sector responsibility in the digital conversation is conspicuous.

### 5.1.4 Key Players [RQ4]

Table 7 shows information about node-level metrics, and, therefore, about key players in the digital discussion.

In a Twitter network, as a general rule, on the one hand, a high input degree is interpreted as leadership or prescriptive capacity, i.e., as a capacity to influence in the network; and on the other hand, a high output degree is interpreted as high participation or activity in the same network. Thus, as expected, the main leader of the conversation was the organizer of the initiative: United Nations Environment Programme (@unep). Likewise, the main leaders of the conversation belong to the major communities and are well-known personalities or entities. Therefore, it can be seen that the messenger has eclipsed the message and, as expected, it was a planned conversation and not a spontaneous one. Conversely, regarding the output degree, the most active users in the network, those that have issued the most mentions, have been non-public persona or people that are more anonymous.

Another interesting indicator in communicative networks such as Twitter is the betweenness centrality. The weighted input and output degrees of a node do not provide the necessary information to interpret its centrality and the geodesic position of a node can affect its structural situation in terms of influence or power (Hanneman, 2001). Hence, assuming that the highly central nodes are located in favorable structural positions within the network, it can be seen that accounts belonging to cluster 4 (cluster related to the United Nations) control the network communication flows.

**Table 7** Node-level metrics extracted from the overall digital conversation

| Label / Community | | Input degree | Label / Community | | Output degree | Label / Community | | Betweenness centrality |
|---|---|---|---|---|---|---|---|---|
| @unep | 4 | 11566 | @sustaintrends | 9 | 827 | @unep | 4 | 0,00695175 |
| @rahulgandhi | 12 | 10269 | @1960_rani | 14 | 641 | @sudarsansand | 4 | 0,00635271 |
| @urstrulymahesh | 13 | 7140 | @kvajpayaee | 9 | 363 | @adamrogers2030 | 4 | 0,00491436 |
| @narendramodi | 14 | 6494 | @mikimiyazaki13 | 9 | 363 | @kambojananya | 4 | 0,00348857 |
| @alluarjun | 17 | 6489 | @mspadine | 9 | 363 | @wakiodorcas | 4 | 0,00298350 |
| @sadhgurujv | 15 | 5613 | @ellyannecgithae | 51 | 316 | @ellyannecgithae | 51 | 0,00282433 |
| @imrankhanpti | 7 | 4547 | @younasjavedans1 | 7 | 307 | @olumideidowu | 4 | 0,00178038 |
| @un | 4 | 3654 | @bellykempegowda | 14 | 299 | @girlupnig | 4 | 0,00161936 |
| @moefcc | 15 | 3434 | @marianatrobosky | 4 | 281 | @agnieszka85z | 4 | 0,00152694 |
| @pakpmo | 7 | 3419 | @sakshamsonu | 16 | 256 | @green_ndi | 4 | 0,00152341 |

# 6 Conclusions and Futures Lines of Work

Twitter is a social network where opinion trends are generated. Likewise, Twitter could play a fundamental role for different institutions or agents to obtain information in order to understand how to exert influence in digital public debate and, therefore, in society; what the image they transmit to public opinion is; or how a political initiative, a marketing campaign or, as in this specific case, an environmental program, is being received by the wider public.

WED (2021), subject of study of this research, deals with one of the biggest problems facing humanity today and in the future: protecting the environment. Therefore, analysis of the digital conversation held in relation to the celebration of WED (2021) provides substantial information on various issues such as the success of this public initiative, the terms under which the initiative is debated, the relationships established between the different actors taking part in the debate, etc.

Specifically, in this scientific work, through a methodology designed ad hoc, different research questions were answered in order to analyze the impact of WED (2021) program on society.

The success of the initiative was relatively demonstrated by the fact that the hashtag #WorldEnvironmentDay was a trending topic worldwide. However, this study offers other valuable and transcendent information from a social perspective, that goes beyond the mere count of tweets published worldwide about this event.

There are certain factors that indicate that the network campaign for the WED (2021), strategically designed and planned by the organizing institution (the United Nations), obtained excellent results. It can be concluded that the strengths of the event at the digital level were the following: the digital conversation was plural in terms of leadership and content; there were no groups that functioned as "trolls", feeding eventual destructive criticism towards the organizing institution or the objectives pursued by the event; participation was in positive and constructive terms; the institution controlled or directed, through the official hashtags, the frames of the digital debate and placed itself at the center, dynamizing the discussion and functioning as a link between different communities.

These types of institutions, as in the case under investigation, have to deal with similar events or initiatives every year (or every so often). This means that these institutions have to design strategic actions and communication strategies cyclically according to the established objectives. In this sense, this scientific research may help organizations to make strategic decisions, identifying the weaknesses and opportunities of the event at the digital level, which are explained as follows:

In the case analyzed, the data reveal a relative absence of global connections or coordinated actions at the digital level to address one of the biggest challenges facing society, i.e., protecting the environment. The communities are divided by country (mostly centered in the host country, Pakistan, and especially in the neighboring country, India), by political ideology and by type of actors. This means that the impact of the event is neither global nor homogeneous and that the "micro-debates" that occur in the conversation are "intra-community" and isolated. Therefore, this feeds the current theory on the influence of digital activity on public opinion in social networks, which speaks about the tendency towards homophily and the confirmation of one's own biases around public controversies.

Moreover, considering that we live in a globalized, polycentric and highly interconnected world. It is paradoxical that this type of events that address a global problem become phenomena that could be classified as 'concentric digital influence' (greater influence in

areas closer to the physical center). In this sense, the organizer of the event faces two challenges for the future: first, the decision of the host country should be taken with a highly strategic perspective; second, actions must be taken to ensure the influence of the event in the digital terrain overcomes this limitation of space-nation in the geographical domain.

In addition, it is noted that politics, culture and entertainment are the collectives that have presence in the conversation. However, there is a clear absence of the business sphere. That absence is twofold: on the one hand, they do not appear to be criticized or to be praised; and on the other hand, they do not appear as a dynamic agent of the conversation. There is practically no joint or individual strategic action on the part of the companies that also influence the real basis of the debate around World Environment Day. It is also worth noting that the second most polluting country on the planet, the United States, is also not very active in the conversation.

All in all, much of the research on the phenomenon of social networks defends the thesis that they fuel controversy and polarize society. However, the case studied in this study is a good example of non-polarized digital conversation and for the most part developed in positive terms. Consequently the actors of public communication find a forum or meeting space where they can express their wishes, proposals or even criticisms in a constructive sense.

Likewise, this study addresses a phenomenon related to mass communication. Twitter is indeed considered a communication and interconnection medium for the masses and a platform that mediates between the different actors that conform public opinion. In this sense, it is worth asking how, on the one hand, social networks have influenced the process of setting the public and media agenda and, on the other, the construction of public opinion itself. In the old paradigm, the traditional media had a monopoly on agenda setting and influenced the framework in which society should understand and interpret public controversies. Today, the traditional media have lost that exclusivity. In fact, the case analyzed in these pages is paradigmatic of this issue, since it shows that a public or private organization can set the digital agenda, without intermediaries, through a strategically planned campaign. However, in addition, and as a major difference between traditional and current communication platforms, an institution can even directly influence the setting of the subtopics that are addressed in a given conversation, the terms or frameworks in which it takes place, the agents or subjects that participate in it, etc.

Finally, despite the limitations shown by Twitter due to the bias of the data collected, bias of representation when making general assumptions and other problems, derived, for example, from the language used by the users (Ruiz-Soler, 2017), this study reinforces a valid methodology for any institution or agent wishing to analyze its impact on society through social networks.

# References

Alhindi, W. A., Talha, M., & Sulong, G. B. (2012). The role of modern technology in Arab Spring. *Archives Des Sciences, 65*(8), 101–112.

Apodaka, E., & Morales-i-Gras, J. (2016). Redes solidarias en Twitter: Un acercamiento a la estructura del independentismo catalán en base a datos capturados en Twitter. *Virtualis. Revista De Cultura Digital, 7*(14), 53–88.

Barrios-O'Neill, D. (2020). Focus and social contagion of environmentalorganization advocacy on Twitter. *Conversation Biology, 35*(1), 307–315. https://doi.org/10.1111/cobi.13564

Bastian, M., Heimann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 8*(10), 10008. https://doi.org/10.1088/1742-5468/2008/10/P10008/pdf

Burak, D. (2017). Environment as politics: Framing the Cerattepe protest in Twitter. *Environmental Communication, 13*(5), 617–632. https://doi.org/10.1080/17524032.2017.1406384

Campos-Domínguez, E. (2017). Twitter y la comunicación política. *El Profesional De La Información, 26*(5), 785–793.

Carrasco-Polaino, R., Villar-Cirujano, E., & Tejedor-Fuentes, L. (2017). Twitter como herramienta de comunicación política en el contexto del referéndum independentista catalán: Asociaciones ciudadanas frente a instituciones públicas. *ICONOS, 16*(1), 64–85. https://doi.org/10.7195/ri14.v16i1.1134

Casero-Ripollés, A. (2018). Investigación sobre información política y redes sociales: puntos clave y retos del futuro. *Profesional De La Información, 27*, 964–974.

Cheong, M., & Lee, V. (2009). Integrating web-based intelligence retrieval and decision-making from the Twitter trends knowledge base. *Proceedings of the Second ACM Workshop on Social Web Search and Mining*, 1–8.

Coletto, M., Lucchese, C., Orlando, S., & Perego, R. (2015). Electoral predictions with Twitter: A machine-learning approach. *6th Italian Information Retrieval Workshop*, 1–12.

Concepto. (2021). *Medio Ambiente—Qué es, importancia, contaminación y protección*. Medio Ambiente. https://concepto.de/medio-ambiente/

Congosto, M., Basanta-Val, P., & Sanchez-Fernandez, L. (2017). T-Hoarder: A framework to process Twitter data streams. *Journal of Network and Computer Applications, 83*, 28–39. https://doi.org/10.1016/j.jnca.2017.01.029

Cormode, G., Krishnamurthy, B., & Willinger, W. (2010). A manifesto for modeling and measurement in social media. *First Monday*. https://doi.org/10.5210/fm.v15i9.3072

Dann, S. (2010). Twitter content classification. *First Monday*. https://doi.org/10.5210/fm.v15i12.2745

DataBank. (2021). *DataBank. The World Bank*. https://databank.worldbank.org/home.aspx

Daume, S. (2016). Mining Twitter to monitor invasive alien species—An analytical framework and sample information topologies. *Ecological Informatics*. https://doi.org/10.1016/j.ecoinf.2015.11.014

Del-Fresno-García, M. (2014). Haciendo visible lo invisible: Visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales. *Profesional De La Informacion, 23*(3), 246–252. https://doi.org/10.3145/epi.2014.may.04

Edrington, C. L., & Lee, N. (2018). Tweeting a Social Movement: Black Lives Matter and its use of Twitter to Share Information, Build Community, and Promote Action. *The Journal of Public Interest Communications*, 2(2), 289. https://doi.org/10.32473/JPIC.V2.I2.P289

Fernández-Gómez, E., & Martín-Quevedo, J. (2018). La estrategia de engagement de Netflix España en Twitter. *Profesional de La Información*, 27(6), 1292–1302. https://doi.org/10.3145/epi.2018.nov.12

Fownes, J. R., Yu, C., & Margolin, D. B. (2018). Twitter and climate change. *Sociology Compass*. https://doi.org/10.1111/SOC4.12587

GitHub. (2021). *Mariluz Congosto*. https://github.com/congosto

Greenpeace. (2021). *Cambio climático*. https://es.greenpeace.org/es/trabajamos-en/cambio-climatico/

Hanneman, R. A. (2001). Introducción a los métodos del análisis de redes sociales capítulo sexto: centralidad y poder. In *Introducción a los métodos del análisis de redes sociales*.

Heidbreder, L. M., Lange, M., & Reese, G. (2021). #PlasticFreeJuly—Analyzing a worldwide campaign to reduce single-use plastic consumption with Twitter. *Environmental Communication, 15*(7), 937–953. https://doi.org/10.1080/17524032.2021.1920447

Hendriks, C. M., Duus, S., & Ercan, S. A. (2016). Performing politics on social media: The dramaturgy of an environmental controversy on Facebook. *Environmental Politics, 25*(6), 1102–1125. https://doi.org/10.1080/09644016.2016.1196967

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 216–225.

Hutto, C. J. (2022). *vaderSentiment*. https://github.com/cjhutto/vaderSentiment

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. *Proceedings of the Ninth WebKDD and First SNA–KDD 2007 Workshop on Web Mining and Social Network Analysis*, 56–65.

Juanals, B., & Minel, J.-L. (2018). Categorizing air quality information flow on Twitter using deep learning tools. In *Computational Collective Intelligence* (pp. 109–118). Springer, Cham. https://doi.org/10.1007/978-3-319-98443-8_11

Krishnamurthy, B., Gill, P., & Arlitt, M. (2008). A few chirps about Twitter. *Proceedings of the First Workshop on Online Social Networks*, 19–24.

Larrondo, A., Morales i Gras, J., & Orbegozo, J. (2019). Feminist hashtag activism in spain: Measuring the degree of politicisation of online discourse on #yosítecreo, #hermanayosítecreo, #cuéntalo y #noestássola. *Communication and Society*, *32*(4 Special Issue), 207–221. https://doi.org/10.15581/003.32.4.207-221

Lesaca, J. (2015). *Twitter como herramienta de los movimientos sociales y políticos para imponer frames en la opinión pública*. https://aecpa.es/es-es/twitter-como-herramienta-de-los-movimientos-sociales-y-politicos-para/congress-papers/1389/

Li, M., Turki, N., Izaguirre, C. R., DeMahy, C., Thibodeaux, B. L., & Gage, T. (2021). Twitter as a tool for social movement: An analysis of feminist activism on social media communities. *Journal of Community Psychology, 49*(3), 854–868. https://doi.org/10.1002/JCOP.22324

Ljubljana-University. (2021). *Orange data mining—Data mining*. https://orangedatamining.com/

Marín-Dueñas, P. P., Simancas-González, E., & Berzosa-Moreno, A. (2019). Twitter and political communication: the case of the Partido Popular and Podemos in the 2016 general elections. *Cuadernos. Info*, *45*, 129–144. https://doi.org/10.7764/cdi.45.1595

Mazzoleni, G. (2010). *La comunicación política*. Alianza Editorial.

Medina, I. G., Miquel-Segarra, S., & Navarro-Beltrá, M. (2018). El uso de Twitter en las marcas de moda. Marcas de lujo frente a marcas low-cost. *Cuadernos. Info, 42*, 55–70. https://doi.org/10.7764/CDI.42.1349

Microsoft. (2021). *About Power Query in Excel—Excel*. https://support.microsoft.com/en-us/office/about-power-query-in-excel-7104fbee-9e62-4cb9-a02e-5bfb1a6c536a

Morales-i-Gras, J. (2020). *Datos masivos y minería de datos sociales: conceptos y herramientas básicas* (p. 32). Universitat Oberta de Catalunya.

Mrvar, A., & Batagelj, V. (2021). *Programs for Analysis and Visualization of Very Large Networks Reference Manual*. http://mrvar.fdv.uni-lj.si/pajek/pajekman.pdf

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me? Message Content in social awareness streams. *Proceedings of the 2020 ACM Conference on Computer Supported Cooperative Work—CSCW'10*.

Navas, A. (2018). *Modelo de variables de desempeño e impacto en Twitter*. Universidad de Navarra.

OpenRefine. (2021). *A free, open source, powerful tool for working with messy data*. https://openrefine.org/

Orbegozo-Terradillos, J., Larrondo-Ureta, A., & Morales-i-Gras, J. (2020). Influencia del género en los debates electorales en España: análisis de la audiencia social en #ElDebateDecisivo y #L6Neldebate. *Profesional De La Informacion*. https://doi.org/10.3145/epi.2020.mar.09

Otero, P., Gago, J., & Quintas, P. (2021). Twitter data analysis to assess the interest of citizens on the impact of marine plastic pollution. *Marine Pollution Bulletin*. https://doi.org/10.1016/J.MARPOLBUL.2021.112620

Pang, N., & Law, P. W. (2017). Retweeting #WorldEnvironmentDay: A study of content features and visual rhetoric in an environmental movement. *Computers in Human Behavior, 69*, 54–61. https://doi.org/10.1016/j.chb.2016.12.003

Pear-Analytics. (2009). *Twitter study*. http://www.slideshare.net/stephendann/twitter-analytics

Peplow, A., Thomas, J., & AlShehhi, A. (2021). Noise Annoyance in the UAE: A Twitter Case Study via a Data-Mining Approach. *International Journal of Environmental Research and Public Health, 18*(4), 1–10. https://doi.org/10.3390/IJERPH18042198

ProQuest. (2021). *World Environment Day (05-06-2021 / 06-06-2021)*. https://www.proquest.com/results/9542C419B6CB45C0PQ/1?accountid=17248

Reyes-Menendez, A., Saura, J. R., & Alvarez-Alonso, C. (2018). Understanding #worldenvironmentday user opinions in twitter: A topic-based sentiment analysis approach. *International Journal of Environmental Research and Public Health, 15*, 1–18. https://doi.org/10.3390/ijerph15112537

Ruiz-Soler, J. (2017). Twitter research for social scientists: A brief introduction to the benefits, limitations and tools for analysing Twitter data. *Revista Dígitos, 1*(3), 17–32. https://doi.org/10.7203/rd.v1i3.87

Statista. (2021). *Global penetration social media 2020 | Statista*. https://www.statista.com/statistics/274773/global-penetration-of-selected-social-media-sites/

The Guardian. (2009). *China blocks Twitter, Flickr and Hotmail ahead of Tiananmen anniversary* . https://www.theguardian.com/technology/2009/jun/02/twitter-china

UN. (2021). *World Environment Day: Millions rally behind movement to restore the earth*. https://www.unep.org/news-and-stories/story/world-environment-day-millions-rally-behind-movement-restore-earth

UNEP. (2021). *Private Sector Engagement*. https://www.unep.org/about-un-environment/private-sector-engagement

Wasserman, S., & Faust, K. (1994). Social Network analysis: Methods and applications. *Cambridge University Press*. https://doi.org/10.1017/CBO9780511815478

WED. (2021). *World Environment Day*. https://www.worldenvironmentday.global/

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on Twitter. In *International World Wide Web Conference Committee (IW3C2)*, 705–714.