

# Green energy: identifying development trends in society using *Twitter* data mining to make strategic decisions

**Enara Zarrabeitia-Bilbao; Jordi Morales-i-Gras; Rosa-María Rio-Belver; Gaizka Garechana-Anacabe**

**Nota:** Este artículo se puede leer en español en:  
<https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86560>

How to cite this article:

**Zarrabeitia-Bilbao, Enara; Morales-i-Gras, Jordi; Rio-Belver, Rosa-María; Garechana-Anacabe, Gaizka (2022).** “Green energy: identifying development trends in society using *Twitter* data mining to make strategic decisions”. *Profesional de la información*, v. 31, n. 1, e310114.  
<https://doi.org/10.3145/eipi.2022.ene.14>

Article received on June 10<sup>th</sup> 2021  
Approved on October 7<sup>th</sup> 2021



**Enara Zarrabeitia-Bilbao** ✉  
<https://orcid.org/0000-0002-2347-3885>  
Universidad del País Vasco/Euskal Herriko Unibersitatea  
Escuela de Ingeniería de Bilbao  
Plaza Ingeniero Torres Quevedo, 1  
48013 Bilbao, Spain  
[enara.zarrabeitia@ehu.eus](mailto:enara.zarrabeitia@ehu.eus)



**Jordi Morales-i-Gras**  
<https://orcid.org/0000-0003-4173-3609>  
Universidad del País Vasco/Euskal Herriko Unibersitatea  
Facultad de CC.SS. y de la Comunicación  
Campus Leioa, Barrio Sarriena, s/n  
48940 Leioa (Bizkaia), Spain  
[morales.jordi@gmail.com](mailto:morales.jordi@gmail.com)



**Rosa-María Rio-Belver**  
<https://orcid.org/0000-0002-4244-9098>  
Universidad del País Vasco/Euskal Herriko Unibersitatea  
Escuela de Ingeniería de Vitoria-Gasteiz  
C/ Nieves Cano, 12  
01007 Vitoria, Spain  
[rosamaria.rio@ehu.eus](mailto:rosamaria.rio@ehu.eus)



**Gaizka Garechana-Anacabe**  
<https://orcid.org/0000-0002-1913-3239>  
Universidad del País Vasco/Euskal Herriko Unibersitatea  
Facultad de Economía y Empresa (Elcano)  
C/ Elcano, 21  
48008 Bilbao, Spain  
[gaizka.garechana@ehu.eus](mailto:gaizka.garechana@ehu.eus)

## Abstract

This study analyzes *Twitter*'s contribution to green energy. More than 200,000 global tweets sent during 2020 containing the terms “green energy” OR “greenenergy” were analyzed. The tweets were captured by web scraping and processed using algorithms and techniques for the analysis of massive datasets from social networks. In particular, relationships between users (through mentions) were determined according to the Louvain multilevel algorithm to identify communities and analyze global (density and centralization) and node-level (centrality) metrics. Subsequently, the content of the conversation was subject to semantic analysis (co-occurrence of the most relevant words), hashtag analysis (frequency analysis), and sentiment analysis (using the VADER model). The results reveal nine main communities and their leaders, as well as three main topics of conversation and the emotional state of the digital discussion. The main communities revolve around politics, socioeconomic issues, and environmental activism, while the conversations, which have developed mostly in positive terms, focus on green energy sources and storage, being aligned with the main communities identified, i.e., on political, socioeconomic, and climate change issues. Although most of the conversations have been about socioeconomic issues, the presence of leading company accounts was minor. The main aim of this work is to take the first steps toward an innovative competitive intelligence methodology to study and determine trends within different scientific fields or technologies in society that will enable strategic decisions to be made.

## Keywords

Green energy; *Twitter*; Social network analysis; Semantic analysis; Sentiment analysis; Big data; Business intelligence; Data analytics; Text analytics; Social analytics; Social networks; Social media; Environment; Renewable energy.

## 1. Introduction

The emergence of the Internet, its massive expansion since the 1990s around the world, and its subsequent establishment in everyday life through myriad connected devices have facilitated the materialization of what it is known as big data:

“Collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications” (Oguntimilehin; Ademola, 2014; Morales-i-Gras, 2020).

Regarding this new communicative paradigm, it is clear that, by being “big”, the adjective that precedes “data”, the importance is given to volume. However, it should be noted that there are also other issues that are equally important when it comes to characterizing this communicative paradigm; in other words, size matters, but it is not everything (Morales-i-Gras, 2020). Other concepts that should also be considered are, in addition to volume: variety, velocity, value, veracity, validity, virtuality, visibility, variability, and complexity (Laney, 2001; Khan; Uddin; Gupta, 2014; Oguntimilehin; Ademola, 2014; Patgiri; Ahmed, 2016; Morales-i-Gras, 2020).

Thus, one of the main features of what we call big data is the variety in the format and sources of data. The big data paradigm is simultaneously fed by data coming from financial services, commerce, industry, healthcare, etc., and to a high degree by social networks or social media (Morales-i-Gras, 2020), with *Twitter* being one of its main representatives.

With the development of Web 2.0 and microblogging, *Twitter* plays an important role as a local, national, and international conversation space, and the different discussions created on the digital platform have a great impact on various sectors of real life (Cossu; Dugué; Labatut, 2015; Li et al., 2016). Therefore, *Twitter*, among others, contributes to the public’s understanding of science (Veltri, 2012).

Likewise, more and more companies are beginning to realize the importance that social media in general, and *Twitter* in particular, should have in their business strategies. By gathering and analyzing social media data along with other corporate business intelligence, companies can obtain a greater understanding of who their competitors and consumers are, leading them to make better decisions (Soussan; Trovati, 2020). Companies are therefore working on tools that allow them to filter the information they collect on *Twitter*, to convert this information into valuable data and subsequently, using business intelligence, into knowledge. Accordingly, there are various ways to obtain competitive business intelligence through *Twitter*, such as analyzing the tweets of customers and competitors, identifying followers of the competitors, or monitoring a strategic topic, as applied herein.

Social media in general, and *Twitter* in particular, have transformed public conversation on a variety of topics. These open discussions play a key role in shaping public opinion on different issues such as climate change, global warming, and environmental activism (Holmberg; Hellsten, 2015; Veltri; Atanasova, 2017; Reyes-Menéndez; Saura; Álvarez-Alonso, 2018; Dahal; Kumar; Li, 2019).

In this study we focus on tweets regarding green energy, a subset of renewable energy that represents those renewable energy resources and technologies that provide the greatest environmental benefit. Those main sources are solar, wind, geothermal, biogas, eligible biomass, and low-impact small hydroelectric (Figure 1) (EPA, 2019).

Conventional power	Renewable energy		
Coal	Large hydropower	Green power	
Nuclear		Wind	Solar
Oil	Municipal solid waste	Geothermal	Biogas
Natural gas		Low-impact hydropower	
<b>Less beneficial</b>	<b>More beneficial</b>	<b>The most beneficial</b>	

Figure 1. Green power based on its relative environmental benefits (EPA, 2019).

According to the *World Meteorological Organization (WMO)*, 2019 was the second warmest year on record, after 2016. The development of environmentally friendly energy is essential to combat climate change and limit its most devastating effects. In this regard, conventional energy sources are a major source of pollution and greenhouse gas emissions. By 2030, the European Union wants 32% of the energy that we consume to come from renewable sources and aims to lead the transition to clean energy and the fulfillment of the objectives established by the 2016 Paris Agreement (*European Parliament*, 2018).

Hence, switching to green energy could help improve the environment, as well as support renewable energy development, reduce the carbon footprint associated with purchased electricity, and protect against future electricity price increases and volatility, among other effects (EPA, 2019).

This study aims to use *Twitter* to explore societal trends in green energy. For this purpose, the main communities that form part of the digital conversation and associated dialogs were analyzed. The main idea is to establish the first steps of an innovative methodology of competitive intelligence to study and determine trends within different scientific fields or technologies in society that will allow us to make strategic decisions.

“*Twitter* contributes to the public’s understanding of science”

## 2. *Twitter* data mining and environmental issues

The process by which data are exploited as raw material and value generated is known as data mining, an operation essentially consisting of capturing a series of information records and interpreting them to create a pattern that provides actionable insights (Morales-i-Gras, 2020).

Like many other social networks, *Twitter* is a great source of data. For this reason, numerous data-mining studies have focused on this social platform in general and on environmental issues in particular. Thus, environmental issues such as sustainability, climate change, global warming, pollution, energy, and specifically, renewable energy, among other issues, have been studied using data provided by *Twitter*.

Sustainability, the most generic concept, is widely used to refer to issues that are good and desirable, usually considered in the context of environmental issues (Pilär *et al.*, 2019). Analyses of the main communities and topics related to the hashtag #sustainability on *Twitter* have concluded that environment, climate change, and green energy, among others, are concepts related to sustainability (Adi, 2018; Pilär *et al.*, 2019; Shen; Luong; Pham, 2021). Likewise, sociopolitical dimensions of energy sustainability transitions have been analyzed using *Twitter* data (Labonte; Rowlands, 2021).

Nonetheless, it should be noted that, since the end of 2016 (the hottest year in history, according to the *WMO*), global warming and thus climate change have become hot topics among other environmental issues (Mooney; Winstanley; Corcoran, 2009), which has also been reflected in several studies based on *Twitter* data. Some studies have applied techniques such as sentiment analysis to a dataset of climate-change-related tweets, revealing how sentiment has changed over time and location (Cody *et al.*, 2015; Dahal; Kumar; Li, 2019; Prabhakar, 2019), or topic modeling to identify discussion topics on climate change (Dahal; Kumar; Li, 2019; Prabhakar, 2019). Differences in *Twitter* communication on the subject have even been studied as a function of gender (male or female) (Holmberg; Hellsten, 2015).

Pollution is another topic that has been studied recently using data mining on *Twitter*. An approach for automatically classifying tweets on pollution and traffic has been constructed using supervised machine learning algorithms for text classification (Chamorro *et al.*, 2020), as well as user and content analysis of PM 2.5 (particulate matter with a diameter less than 2.5  $\mu\text{m}$ ) as an indicator of air pollution (Chen; Tu; Zheng, 2017).

For energy, studies have been carried out on both conventional and renewable or green energies. As stated above, various global problems have made renewable energies an option not for the future but for the present, while traditional fossil fuels are losing their central role in the most industrialized countries (Fernández-Arias, 2017). Similarly, nuclear energy has been displaced from its preferential position in the energy system of several countries; this is due to political, economic, and even technical factors, but is mostly due to the great social controversy that has existed and still exists around its development and exploitation (Fernández-Arias; Cuevas; Vergara, 2021). Consequently, nuclear energy has been studied repeatedly on *Twitter*, and several studies have analyzed the sentiments and opinion of *Twitter* users on nuclear energy (Kim *et al.*, 2016; Gupta; Ripberger; Wehde, 2018; Khatua *et al.*, 2020).

In the case of renewable energies, although they are an indisputable reality, public acceptance and support are factors relevant to both renewable energy policies and market conditions (Kim *et al.*, 2020). Accordingly, studies have analyzed the feelings of *Twitter* users in relation to renewable energies in general (Jain; Jain, 2019) as well as specific types such as solar energy (Reboredo; Ugolini, 2018; Li *et al.*, 2019; Kim *et al.*, 2020).

In general, environmental issues debated on *Twitter* are analyzed through discussions on the main topics, the resulting communities, and above all, sentiment analysis.

The aim of the current study is to use *Twitter* to identify the actors, topics, and opinions and thereby produce a decision-making tool. This is done herein for green energies, a subject for which such analysis is lacking to date. To achieve this purpose, a combination of data mining and big data techniques and tools are applied.

## 3. Research methodology

The primary goal of this study is to identify green energy development trends in society using data mining on *Twitter*. The procedure and tools used in this study are summarized in Figure 2.

The first stage of the analysis was to specify green energy as the subject. Then, an appropriate search query was built and adapted to *Twitter*: “green energy” OR “greenenergy.” The first term was chosen because it is the exact term of study, while the second one was chosen to collect tweets containing, among others, the hashtag #greenenergy.

Data were collected globally from 1 January to 31 December 2020 through web scraping, which relies on a set of auto-

mated techniques to extract information systematically from websites (Morales-i-Gras, 2020). A company specializing in data capture provided the tool to scrape data from *Twitter* and the obtained tweets.

*OpenRefine* [an open-source Java-based tool that allows data to be cleaned and organized (OpenRefine, 2021)] and *Orange Data Mining* [an open-source tool for data mining and predictive analytics (Ljubljana University, 2021)] software packages were used to clean and refine the data.

For the analysis, the empirical approach was carried out in two stages (Figure 3).

First, data were processed and studied using social network analysis (SNA) techniques (Larrondo-Ureta; Morales-i-Gras; Orbegozo-Terradillos, 2019; Orbegozo-Terradillos; Larrondo-Ureta; Morales-i-Gras, 2020). This first stage analyzed the relationships established between users in the digital conversation through *Twitter* mentions; i.e., once the mentions had been extracted, a network was synthesized based on which users mentioned other users in the conversation itself.

Analysis of the communities (clusters) involved was carried out with the Louvain multilevel algorithm (Blondel et al., 2008), using the appropriate algorithm for such a case as presented in the work of Morales-i-Gras (2017). This algorithm enables the generation of communities of densely interconnected nodes (users or profiles), i.e., communities that engage with different conversations identified through the data (mentions). The algorithm was applied randomly and considers that the network is weighted. The nodes of the network were grouped sequentially, and the gain or loss in the modularity statistic (Newman; Girvan, 2004) was evaluated permanently to estimate the descriptive quality of a community partition by measuring the portion of connections between actors that fall within those modules. The resulting community partition can also be evaluated using this statistic, where a modularity greater than 0.3 is considered to be statistically significant (Orbegozo-Terradillos; Morales-i-Gras; Larrondo-Ureta, 2019). Different metrics were generated using *Pajek* [an open-source software for analysis and visualization of large networks (Mrvar; Batagelj, 2021)], and graphs were visualized using *Gephi* [an open-source Java-based tool for network analysis and visualization (Bastian; Heimann; Jacomy, 2009)]. *Power Query* [a data preparation and data transformation tool in *Excel* (Microsoft, 2021)] was used to produce a more detailed analysis of the communities.

Regarding the specific metrics used to analyze the relationship between users, note that, within SNA, two fundamental levels of analysis can be recognized, corresponding to complete networks and individual actors or nodes (Aguilar-Gallegos; Martínez-González; Aguilar-Ávila, 2017).

Since a social network is defined by the links formed by the actors within it, it is first important to describe the whole network. To this end, two of the most common metrics used in SNA that help characterize the entire network were analyzed

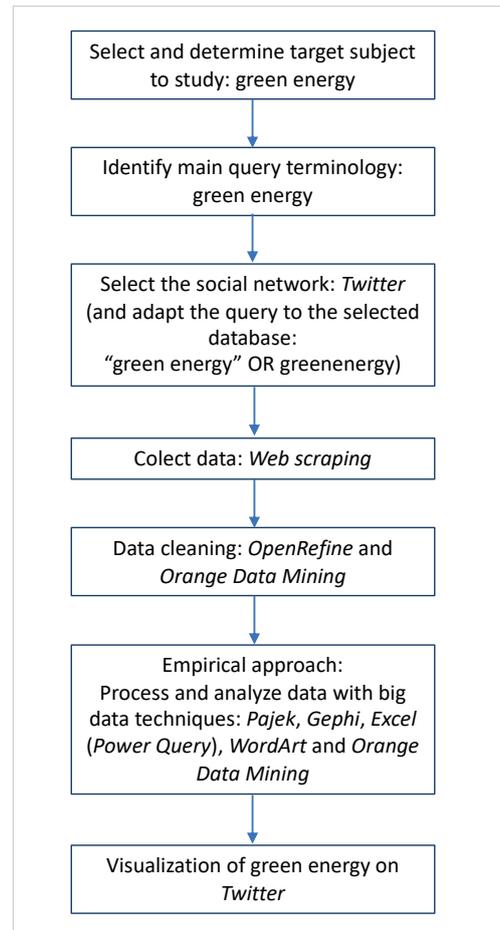


Figure 2. Workflow of procedure and tools used

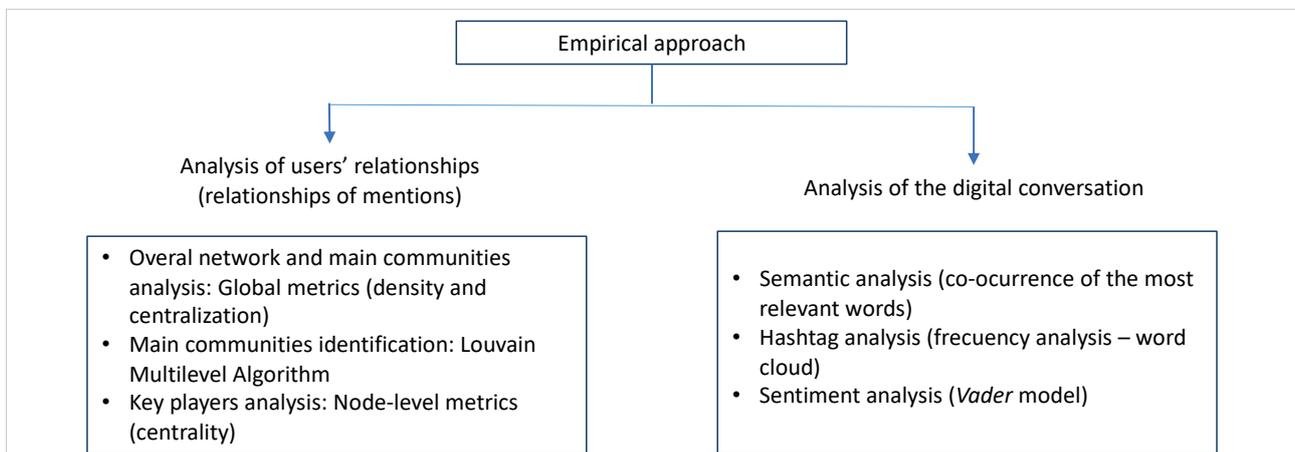


Figure 3. Empirical approach of the research

in this study, namely density and centralization (Aguilar-Gallegos; Martínez-González; Aguilar-Ávila, 2017). On the one hand, density is a measure of the cohesion among the actors in the network (Borgatti; Everett; Johnson, 2013), concretely, the number of existing links in the network, presented as a proportion of the number of possible links. On the other hand, the level of centralization of the network is an indicator that measures the difference between the levels of centrality of the actors, providing information about the existence of dominant nodes (Freeman, 1978). Specifically, the input degree centralization, output degree centralization, and betweenness centralization indicate how close the network is to behaving like a star network, with one actor playing a central role and controlling the entire network (for values close to 1) or how distant it is from such behavior (for values close to 0) (Velázquez-Álvarez; Aguilar-Gallegos, 2005). The input degree centralization reveals whether the receipt of mentions is centralized on specific users, while the output degree centralization depicts whether the output of mentions is centralized on specific users. Meanwhile, the betweenness centralization can show whether the intermediation in the network is distributed horizontally or concentrated in the hands of a few.

Regarding the node-level metrics, the centrality degree (input and output) and betweenness centrality are considered to be first-order indicators (that directly link two nodes through a single step) (Aguilar-Gallegos *et al.*, 2016) as traditionally used in SNA. The input degree refers to the number of links that an actor receives from others, whereas the output degree is the number of links that an actor sends to others (De-Nooy; Mrvar; Batagelj, 2018). It is thus possible to analyze which users receive the most mentions (i.e., the discussion leaders) and those who emit the most mentions (i.e., discussion drivers). The betweenness is based on the frequency with which a node is located among the shortest (geodesic) paths connecting pairs of other nodes in the network (Freeman, 1978) (thus being influential in the transmission of information).

In the second stage of the empirical approach, a semantic analysis of the overall digital conversation was carried out. A network was generated between the tweets and the most relevant words contained in the tweets (the top 150 that appeared most frequently, to focus on the most relevant discourses in the network). Subsequently, after eliminating the weakest connections, a graph of relevant words (nodes) in the conversation was generated on the basis of how many tweets cooccurred (edges). To achieve this, *Pajek* was used to transpose the two-mode network into a single-mode network, and *Gephi* was used for the visualization and study of the relationships between words (communities or topics of the conversation generated).

In addition, to complete the contextualization of the conversations, the most used hashtags were studied through their frequency of occurrence and word cloud visualization (Ahuja; Shakeel, 2017). *WordArt* [an online word cloud art creator (WordArt, 2021)] was applied for this visualization.

In parallel, the emotional state of the digital conversation was studied using sentiment analysis with heuristics rules. The *Valence Aware Dictionary and sEntiment Reasoner (Vader)* model was applied because it can effectively be used to investigate the sentiment of *Twitter* users (Shen; Luong; Pham, 2021). This *Python*-based library is a *lexicon*-enhanced, *rule-based* sentiment analysis model for social media texts (Hutto; Gilbert, 2014). The compound output variable represents the total sentiment of a tweet, with -1 representing the most negative sentiment (strongly negative) and 1 the most positive sentiment (strongly positive) (Ljubljana University, 2021). Hence, a numerical score is given to each tweet, where positive values from 0 to 1 correspond to positive emotional states toward a subject and negative values from 0 to -1 correspond to a negative emotional state toward a subject, with values close to zero corresponding to a neutral emotional state toward a subject. The *Orange Data Mining* tool was used for all this analysis.

“The reciprocal influence of *Twitter* and television is demonstrated even for more scientific or technical topics such as green energy”

#### 4. Results and discussion

The query retrieved 236,233 tweets about green energy during 2020 from 107,873 unique users. The temporal distribution of the number of tweets per day is shown in Figure 4.

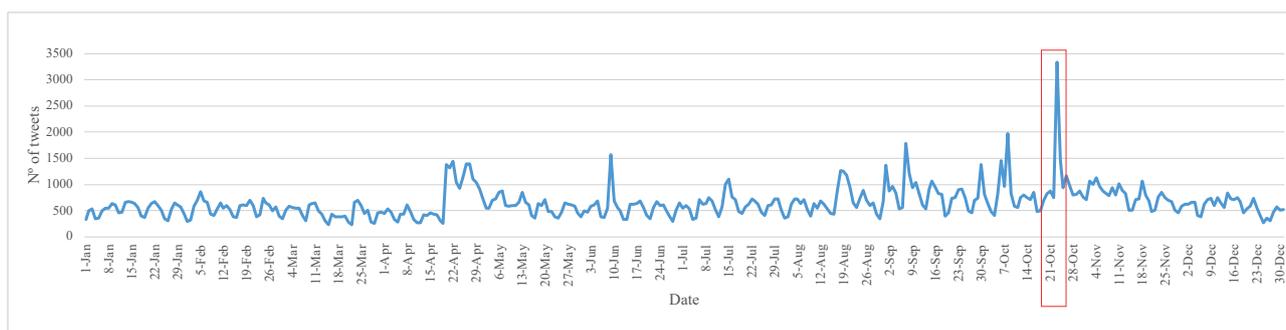


Figure 4. General trend of tweets throughout 2020

As expected, the temporal distribution of the tweets was not constant, with different events throughout the year generating more or less activity on the network.

Unusually intense activity was observed on 23 October<sup>1</sup> (3,337 tweets compared with the average of 645 per day), influenced by the televised US election debate, where one of the topics discussed was climate change. For the first time since 2008, climate change appeared as an issue in the televised debate between the two candidates in the US presidential election. The roughly 10-min dialog (*YouTube*, 2020) generated enormous activity on the digital network for green energies, among others. Hence, the reciprocal influence of *Twitter* and television (**Macmillan**, 2015) is demonstrated, even for more scientific or technical topics such as green energy.

Likewise, and also as expected, almost all the tweets were in English (since an English term was used for the search), with the second (Italian, 0.87%) and third (Spanish, 0.81%) most widely used languages lagging far behind.

Finally, to provide an overview of the tweets published, note that 63.6% were original tweets, 36.3% were replies, and 0.1% were quotes.

**4.1. Relationships established between users**

Regarding the captured conversation, Table 1 presents the values obtained based on the relationships established between the users, through *Twitter* mentions, in the overall digital conversation:

Table 1. Conversation metrics

Density	0.000009
Input degree centralization	0.021358
Output degree centralization	0.002479
Betweenness centralization	0.000507
Number of clusters	13,262
Modularity	0.860586

The density data indicate that only 0.0009% of the possible connections between actors materialized, suggesting that there are still many pending strategic interpellations to be explored for green energy. The input degree centralization was very low (2.14%), thus small groups of users do not monopolize the reception of the mentions on the network. Moreover, the output degree centralization was also very low (0.25%), so there is no single group from which the majority of mentions are being issued on the network and the mentions are fairly well distributed. According to the betweenness centralization (0.05%), the intermediation in the network is distributed horizontally instead of being capitalized by a few users.

Likewise, 13,262 different communities were identified in the interaction graph, with a modularity figure of 0.86, indicating a community structure with high mathematical significance. Of these communities, only nine accounted for

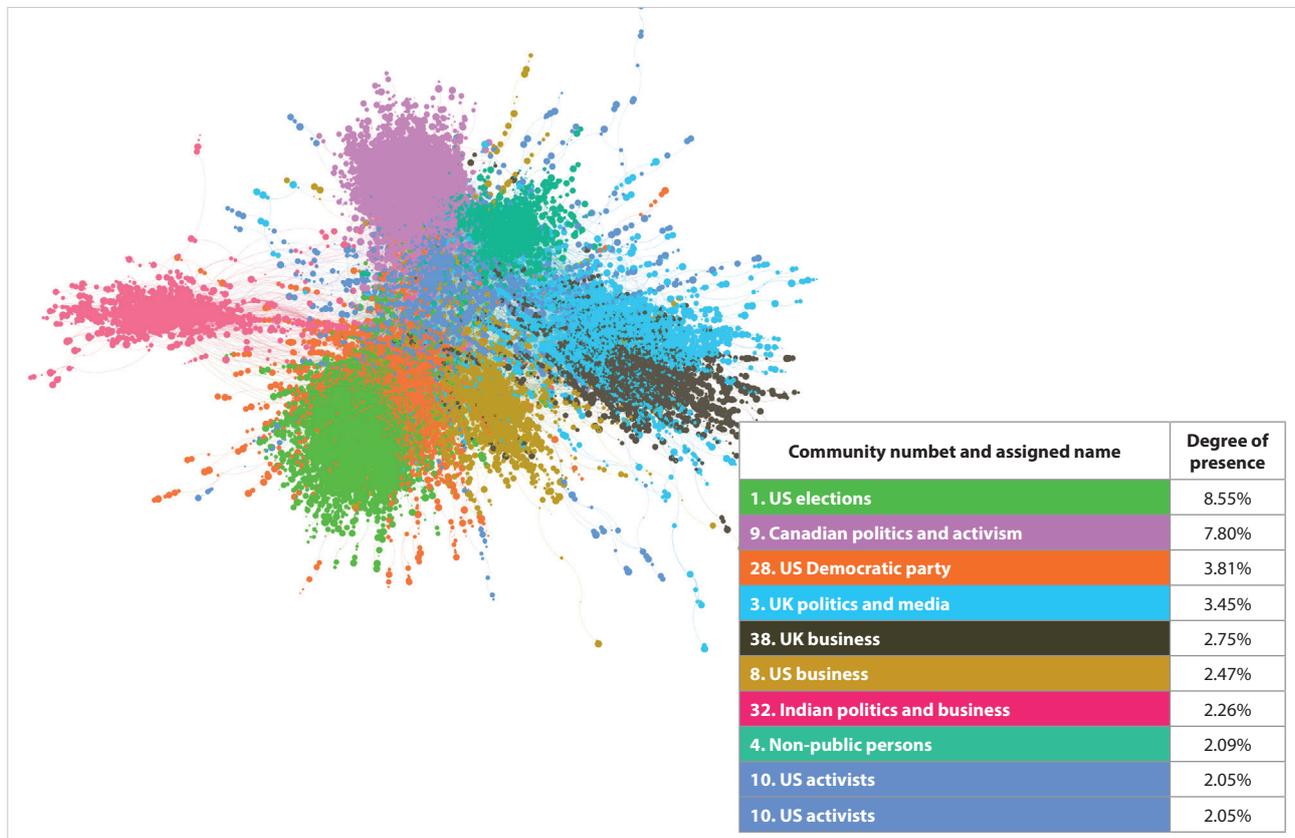


Figure 5. Network after filtering the nine most important communities

more than 2% of the actors in the network, collectively accounting for 35.25% of the actors in the network. Therefore, this represents an extremely fragmented conversation with no absolute leadership.

“The digital sphere contains topics of conversation around green energy, such as more technical, socioeconomic and political, or environmental ones”

Figure 5 shows the position that each community holds in the overall network, and their connections with the other communities. The first column (in Figure 5) corresponds to “community number (assigned name),” where the numbers are randomly assigned by the (Louvain multilevel) algorithm to all communities and should not be interpreted as an ordering. The communities are named according to their characteristics and leaders. The second column corresponds to the node (user) percentage that the community represents in the total network, i.e., the “degree of presence” that the community has in the whole network. Therefore, the main communities are not ranked by community number but by their degree of presence on the network.

Likewise, Table 2 presents the different metrics for each of the communities, i.e., showing how equally or unequally the possible connections between actors in each community, the reception and output of mentions, as well as the behavior of intermediation (Table 2) are distributed.

Table 2. Conversation metrics of the main communities

Community	Density	Input degree centralization	Output degree centralization	Betweenness centralization
1	0.000110	0.202083	0.006532	0.000003
9	0.000187	0.052864	0.013859	0.016102
28	0.000218	0.187076	0.006533	0.000012
3	0.000245	0.074105	0.021161	0.000689
38	0.000319	0.053904	0.026552	0.012724
8	0.000350	0.188600	0.018331	0.000076
32	0.000444	0.084158	0.022975	0.000416
4	0.001489	0.040848	0.111306	0.022971
10	0.000407	0.106413	0.015036	0.000110

As expected, the density of the main communities, i.e., the coexistence of the most prominent communities, was higher than the density of the overall network. However, the densities remain low. This may suggest that some of the nodes in the various communities may be capitalizing on many of the links.

Nevertheless, although the input degree centralization of the highlighted communities is higher than that of the general network, that metric of these communities is not high. There are communities with diverse protagonists and there is a tendency to interact with more than a single actor, i.e., communities where many actors mention or interact with many other actors (Larrondo-Ureta; Morales-i-Gras; Orbegozo-Terradillos, 2019). Even so, it is worth noting that in communities 1, 28, 8, and 10 the reception of mentions is more monopolized than in the rest.

In the case of output degree centralization, although community 4 stands out from the other communities, this metric is also low for all communities. This suggests that there are no bots (accounts with automated behavior), although this must be verified by other methods.

In the case of the most important communities, the networks continue to be distributed in a very horizontal way instead of being concentrated in the hands of a few.

Once the main communities were isolated, it was possible to observe their respective leaderships (top five of each community) according to the input degree (Table 3).

“A triple typology of actors participating in the interactions of the conversation can be observed, namely figures from the political system, actors from the business world, and agents from the activist sphere”

Table 3. Brief description of the main communities' leaders

	Leaders (input degree)	Location and profile of leaders
1	@realdonaldtrump, @joebiden, @gop, @kamalaharris, @speakerpelosi	United States Leading politicians and parties in the elections.
9	@justintrudeau, @elizabethmay, @jkenney, @gmbutts, @fordnation	Canada Senior politicians and activists.
28	@aoc, @berniesanders, @sensanders, @andrewyang, @ewarren	United States Senior politicians in the democratic party.
3	@borisjohnson, @bbcnews, @rishisunak, @greenpeace, @googlenews	United Kingdom/international Senior politicians; an environmental NGO; media.
38	@octopus_energy, @guardian, @bulbuk, @ecotricity, @green_energy_uk	United Kingdom British sustainable energy companies and one media outlet.
8	@elonmusk, @tesla, @rbreich, @jimcramer, @gerberkawasaki	United States A business magnate; a big company; an economist; a television personality (related to finance); a financier.
32	@narendramodi, @adanionline, @pmoindia, @gautam_adani, @mnreindia	India The Prime Minister and a government ministry; a big group company and the chairman of the group.
4	@geraldkutney, @friendsoscience, @dawntj90, @jwspry, @roypentland	Canada/Australia/undetermined Non-public personas and groups with high <i>Twitter</i> activity related to climate change, among other issues.
10	@mmflint, @billmckibben, @joshfoxfilm, @jeffgibbstc, @algore	United States Environmental activists from different disciplines (filmmakers, writers, and politicians).

Analyzing the top five leaders in each of the communities, a triple typology of actors participating in the interactions of the conversation can be observed, namely figures from the political system (clusters 1, 9, 28, 3, and 32), actors from the business world (clusters 38, and 8), and agents from the activist sphere (clusters 4 and 10). In any case, the political field remains predominant.

Cluster 10 (the smallest) is most connected with the other communities, as shown in Figure 5. These are globally known agents whose influence stems from the field of entertainment rather than political or business matters; they are involved in the production of content consumed worldwide.

Although more in-depth analysis is required for confirmation, it seems that the borders between territories also mark the relationships within the network. Leaders tend to be grouped according to country: in other words, the “green energy” topic does not necessarily generate interterritorial conversations or unite cross-border public debates. This is therefore not perceived to be an international issue or one with a common global strategy agenda. This may be due to each country’s particular legislation, which does not comply with the environmentalist maxim: think globally, act locally.

Likewise, if one analyzes the centrality metrics of the general network (Table 4), the main interpellations favor political agents (with the exception of @elonmusk) (input degree), but the emission of mentions was headed, as expected, by nonpublic personas and groups with high *Twitter* activity (output degree). Finally, although there is not much intermediary power in the system, it is worth noting that it is a company’s account that has the most power (betweenness centrality).

Table 4. Centrality metrics (top 10 accounts)

Label/community	Input centrality degree	Label/community	Output centrality degree	Label/community	Betweenness centrality			
@realdonaldtrump	1	3,237	@kctaz	4	377	@octopus_energy	38	0.000507
@joebiden	1	2,439	@0sundance	4	361	@jwspry	4	0.000503
@aoc	28	1,528	@paprkalady	4	290	@stanleedia	3	0.000430
@elonmusk	8	883	@mal1976retaf	2	285	@jorgenfaxholm	3	0.000428
@berniesanders	28	796	@dawnlou59339714	357	257	@drolkrad_eht	4	0.000417
@youtube	50	777	@barbarajdurkin	4	243	@dawntj90	4	0.000385
@justintrudeau	9	690	@medbennett	4	243	@geraldkutney	4	0.000371
@gop	1	625	@roypentland	4	233	@joaquinblume	3	0.000355
@elizabethmay	9	624	@thedemorats	86	220	@glenskurka	9	0.000282
@jkenney	9	558	@dawntj90	4	214	@quea_ali	9	0.000256



Although Figure 6 shows that the digital sphere contains different topics of conversation around green energy, such as more technical topics (energy sources, storage systems), socioeconomic (welfare and work) and political topics (US elections), or environmental topics (environmental activism), analysis of the hashtags labeling the tweets studied shows that the vast majority are grouped around concepts related to renewable energy and environmental sustainability. The top 10 most used hashtags include #renewableenergy, #renewables, #solar, #solarenergy, and #solarpower (related to renewable energy) and #cleantech, #green, #climatechange, and #sustainability (related to environmental sustainability), as highlighted in Table 5.

Table 6. Occurrence frequency of hashtags related to science, technology, and innovation

Position	Hashtag	Frequency
13	#technology	2,807
16	#innovation	2,130
21	#science	1,636
22	#tech	1,485
34	#greentech	1,259
40	#cleantech	1,118

In addition, it is worth mentioning the hashtags related to science (#science), technology (#technology, #tech, #greentech, #cleantech), and innovation (#innovation) (Table 6), which are undisputed elements of sustainable growth (Unesco, 2021).

The emotional state expressed in the digital conversation about green energy may be relevant because emotions can be important for, among other things, information diffusion and solidarity building (Kramer; Guillory; Hancock, 2014; Margolin; Liao, 2018). In the sentiment analysis, after removing all non-English tweets, 223,237 tweets were analyzed. Figure 8 shows the average tweet sentiment per day (summing the sentiment scores of all tweets on each day and dividing by the number of tweets on that day).

The mean for all days is greater than 0. There are no major jumps over different days, but the largest peak in negative sentiment occurred on April 27 and the largest peak in positive sentiment on June 9.

On 21 April (the eve of the 50th anniversary of the first Earth Day), Michael Moore “@mmflint” (the executive producer) posted the documentary “Planet of the Humans” (written, directed, and produced by Jeff Gibbs “@jeffgibbstc”) on YouTube for free viewing (Gibbs; Moore, 2019). The main thesis of the documentary is that renewable and green energies are not going to prevent the ecocide of the planet as their manufacture, maintenance, and recycling also depend on fossil fuels. The film argues that green energy sources are in fact neither renewable nor sustainable, and that the only way to save the world is to reduce the consumption and waste of resources and energy. The digital conversation around the documentary and its main thesis reached its most negative emotional state peak on 27 April, with tweets such as

“The borders between territories also mark the relationships within the network”

“The film is all about ‘green’ energy, totally reliant on FFs. Scams, frauds and lies all uncovered. Your weak attempt to deflect from this truth won’t work. Your damage limitation is embarrassing you”  
<https://twitter.com/redirect/status/1254891772825796614>

“Hmmm who should he apologize to 1) the Billionaires ripping off the world, 2) the lying, corrupt, green energy pigs ripping off their fellow citizens, or 3) the moronic fools touting the fake, fake, fake mantra of anthropologic climate change - you tell me?”  
<https://twitter.com/redirect/status/1254611718447099904>

On 9 June, Adani Green Energy Limited company won the first-of-its-kind manufacturing-linked solar agreement from the Solar Energy Corporation of India (SECI). As part of this award, the company will develop 8 GW worth of solar projects, along with a commitment that will see Adani Solar establish 2 GW of additional solar cell and module manufacturing capacity. This award, the largest ever of this type in the world, involves a single investment of US \$6 billion (Adani, 2020). The digital conversation around the award peaked at its most positive emotional state on the same day, with tweets such as

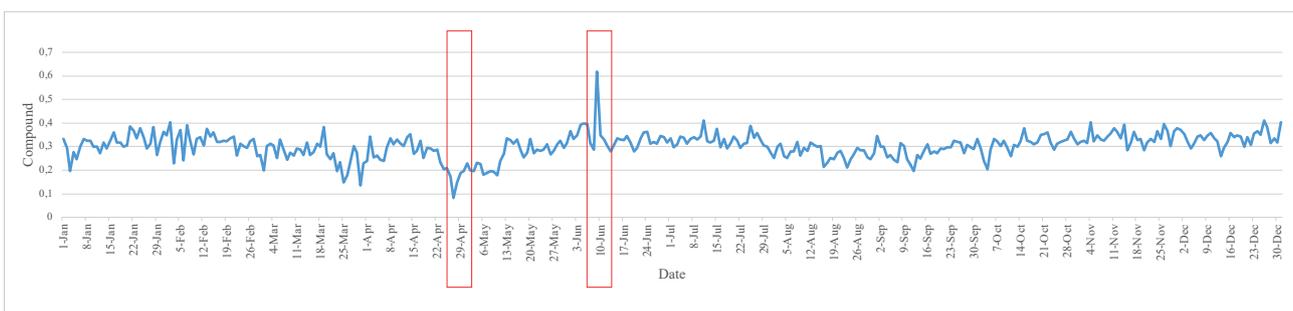


Figure 8. Average tweet sentiment per day

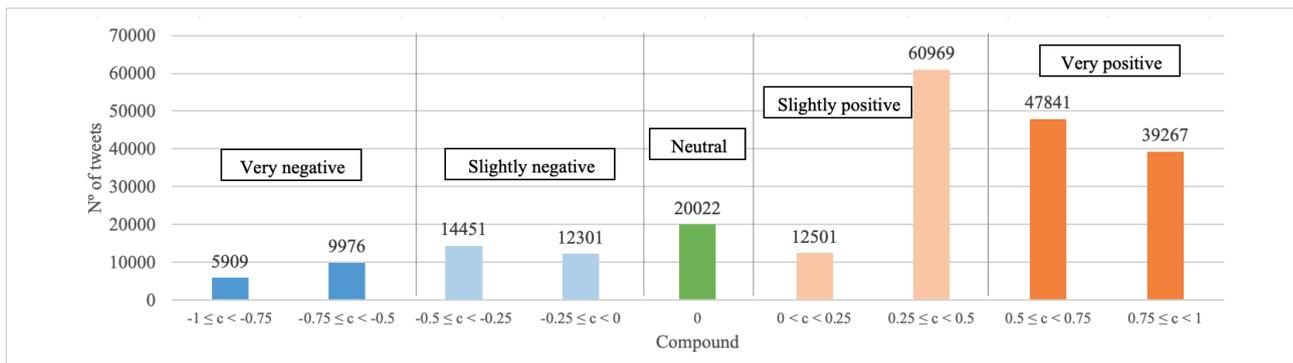


Figure 9. Positive and negative sentiment score of the digital conversation

“Adani Green Energy wins the world’s largest solar award, a proud moment for Adani Group. Really this is a magnificent news for India’s green energy landscape; Adani Green Energy wins SECI award. Kudos!”

<https://twitter.com/redirect/status/1270318548456480768>

“It is reported that Adani Green Energy wins the latest solar award from SECI and Adani Solar will be featured amongst the top 10 solar players by capacity in the BNEF Tier-1 list. This is exemplary. It is superb to know that Adani Green Energy, the latest SECI award winner”

<https://twitter.com/redirect/status/1270320664663093248>

Figure 9 confirms that the overall discussion in the digital sphere was positive. According to the established classification, 72% of the tweets were positive; moreover, 39% of the tweets were categorized as very positive. It seems that, when different actors discuss green energy issues, they do so in positive terms. This seems hopeful, since the cause of environmental problems, among others, and opportunities for addressing them, depend on the perceptions, attitudes, and behavior of society (Carvalho, 2009), and positive messages make people feel more identified, motivated, and in many cases, hopeful.

## 5. Limitations and other considerations

One of the most important and socially transcendent challenges facing the big data paradigm is related to the analysis and interpretation of data. Some authors from the artificial intelligence field (Campolo *et al.*, 2017) have already asked the scientific community to abandon the use of “black box” algorithms and to design study methods based on simpler, more transparent algorithms that facilitate analytical and interpretative work. This shift, which is taking place among researchers in big data, artificial intelligence, and data mining in general (Edizel *et al.*, 2020), represents a great opportunity for scientists in different fields, but especially for those in the social and communication sciences. In this regard, the methodology and tools described in this study could be applied in various investigations to approach the study from different perspectives, depending on the research objectives, although such an analytical strategy will always depend on the interpretative skills of the researcher.

“Twitter has become a reflection of the opinions and concerns of society in general, and of specific communities such as scientists, companies, political parties, and social movements, on specific subjects, such as green energy”

Likewise, in the case of the data source used, it is worth mentioning the methodological limitations of research using *Twitter* because of the bias of the data collected and bias of representation when making general assumptions, as well as other problems, for example, the language used by the users (Ruiz-Soler, 2017).

Finally, it should be emphasized that the success of such research is largely determined by the appropriate selection of the terms used in the data search strategy. In this sense, as *Twitter* is a platform used by a wide spectrum of society, different accounts may not be using the correct term to refer to a particular topic (most likely when the topics are technical), or the term may be used incorrectly.

## 6. Conclusions and future research

*Twitter* offers a unique research perspective to the scientific community interested in, among other things, social movements, public debate or contemporary public controversies, and deliberative processes. Through specific methodological processes (big data techniques), digital conversations can be described using semantic approaches, and social, political, and/or business processes can also be explained and interpreted. This not only reveals what happens, but how and why, and with what effects (Casero-Ripollés, 2018).

Since the introduction of *Twitter* in October 2006, this microblogging platform has attracted increasing interest from various fields such as academia, politics, and business. With 353 million active users in 2020, *Twitter* has become a reflec-

tion of the opinions and concerns of society in general, and of specific communities such as scientists, companies, political parties, and social movements, on specific subjects, such as green energy.

In this study, data obtained from *Twitter* were used and processed to provide information that will help society in general, and the scientific community in particular, by putting the social phenomena that takes place around green energies into context, with the aim of better understanding the dynamics and changes in society around green energy and for making strategic decisions.

In the case of green energy, the digital conversation generated is a discussion with generally low centralization, and more specifically in the case of the main communities. The tendency is for many actors to interact and organize themselves based on their links (mentions) to many other actors, without the presence of an absolute leader in charge of dynamizing user participation.

The main communities that form part of the digital conversation are the political, business, and activist communities. Regarding the conversations generated on the network, these were carried out in positive terms and were centered around green energy sources and storage, in general, and more specifically, in line with the communities observed, around socioeconomic and political aspects, and the impact of climate change.

Likewise, although most of the conversations were about socioeconomic aspects, the presence of leading company accounts is low. Social networks are an essential tool for companies, particularly small and medium-sized enterprises (SMEs) that, with a good strategy, can meet many of their marketing or business generation objectives via social networks, even with a limited budget for large campaigns. However, in this case, the use of *Twitter* is not very high compared with other communities.

Nonetheless, although company participation is low, equipment manufacturers or energy suppliers should be interested in what is said on *Twitter* concerning the topic of their business activity, as well as the communities that lead such discussions. Hence, this analysis may be of great interest as it allows public opinion trends or streams to be incorporated into strategic business plans, becoming a tool for “listening” to the social environment.

Future work related to this study could take many forms. In-depth analysis of the conversation in each of the most relevant communities would be interesting, extending the study to identify the development trends in science using data from *WoS* and *Scopus*, using the available information (selected words) to categorize future tweets according to the topic of conversation through supervised machine learning algorithms, or carrying out sentiment analysis based on machine learning rules instead of heuristic rules. In addition, concerning the topic of *Twitter* data mining and environmental issues, it would be interesting to compare the impact of different energy topics or subjects on society.

“ Social networks are an essential tool for companies, particularly small and medium-sized enterprises (SMEs) that, with a good strategy, can meet many of their marketing or business generation objectives via social networks ”

“ Although company participation is low, equipment manufacturers or energy suppliers should be interested in what is said on *Twitter* concerning the topic of their business activity, as well as the communities that lead such discussions ”

## 7. Note

1. The US election debate was held on 22 October in the USA; however, the UTC time at which the debate began was 02:00 on 23 October in Spain.

## 8. References

Adani (2020). “Adani green energy wins the world’s largest solar award”. *Newsroom of Adani Green Energy Limited (AGEL)*, Jun 9.

<https://www.adanigreenenergy.com/newsroom/media-releases>

Adi, Ana (2018). “#Sustainability on *Twitter*: loose ties and green-washing CSR”. In: Grigore, Georgiana; Stancu, Alin; McQueen, David (eds.). *Corporate responsibility and digital communities. An international perspective towards sustainability*. Palgrave studies in governance, leadership and responsibility. Palgrave Macmillan, pp. 99-122. ISBN: 978 3 319 63480 7

[https://doi.org/10.1007/978-3-319-63480-7\\_6](https://doi.org/10.1007/978-3-319-63480-7_6)

Aguilar-Gallegos, Norman; Martínez-González, Enrique-Genaro; Aguilar-Ávila, Jorge (2017). *Análisis de redes sociales: conceptos clave y cálculo de indicadores*. Chapingo, México: Universidad Autónoma Chapingo (UACH). Centro de investigaciones económicas, sociales y tecnológicas de la agroindustria y la agricultura mundial (Ciestaam). ISBN: 978 607 12 0487 5  
<https://www.redinnovagro.in/pdfs/indicadores.pdf>

- Aguilar-Gallegos, Norman; Martínez-González, Enrique-Genaro; Aguilar-Ávila, Jorge; Santoyo-Cortés, Horacio; Muñoz-Rodríguez, Manrubio; García-Sánchez, Edgar-Iván** (2016). "Social network analysis for catalysing agricultural innovation: from direct ties to integration and radiality". *Estudios gerenciales*, v. 32, n. 140, pp. 197-207.  
<https://doi.org/10.1016/J.ESTGER.2016.06.006>
- Ahuja, Vandana; Shakeel, Moonis** (2017). "Twitter presence of jet airways-deriving customer insights using netnography and wordclouds". *Procedia computer science*, v. 122, pp. 17-24.  
<https://doi.org/10.1016/j.procs.2017.11.336>
- Bastian, Mathieu; Heymann, Sebastien; Jacomy, Mathieu** (2009). "Gephi: an open source software for exploring and manipulating networks". In: *International AAAI conference on weblogs and social media*.  
<https://gephi.org/publications/gephi-bastian-feb09.pdf>
- Blondel, Vincent D.; Guillaume, Jean-Loup; Lambiotte, Renaud; Lefebvre, Etienne** (2008). "Fast unfolding of communities in large networks". *Journal of statistical mechanics: theory and experiment*, v. 8, n. 10, pp. 1-12.  
<https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Borgatti, Stephen P.; Everett, Martin G.; Johnson, Jeffrey C.** (2013). *Analyzing social networks*. London: SAGE Publications Limited. ISBN: 978 1 52 64 0410 7
- Campolo, Alex; Sanfilippo, Madelyn; Whittaker, Meredith; Crawford, Kate** (2017). *AI now 2017 report*.  
[https://ainowinstitute.org/AI\\_Now\\_2017\\_Report.pdf](https://ainowinstitute.org/AI_Now_2017_Report.pdf)
- Carvalho, Anabela** (2009). "Communication for sustainable policy: connecting science, society and government". *Science for environment policy. Environmental communication*, n. 17.  
[https://ec.europa.eu/environment/integration/research/newsalert/pdf/17si\\_en.pdf](https://ec.europa.eu/environment/integration/research/newsalert/pdf/17si_en.pdf)
- Casero-Ripollés, Andreu** (2018). "Research on political information and social media: Key points and challenges for the future". *El profesional de la información*, v. 27, n. 5, pp. 964-974.  
<https://doi.org/10.3145/epi.2018.sep.01>
- Chamorro, Verónica; Rivera, Richard; Varela-Aldás, José; Castillo-Salazar, David; Borja-Galeas, Carlos; Guevara, César; Arias-Flores, Hugo; Fierro-Saltos, Washington; Hidalgo-Guijarro, Jairo; Yandún-Velasteguí, Marco** (2020). "Twitter mining for multiclass classification events of traffic and pollution". In: *IHSED 2019. International conference on human systems engineering and design: Future trends and applications*, pp. 1030-1036.  
[https://doi.org/10.1007/978-3-030-27928-8\\_153](https://doi.org/10.1007/978-3-030-27928-8_153)
- Chen, Wenhong; Tu, Fangjing; Zheng, Pei** (2017). "A transnational networked public sphere of air pollution: analysis of a Twitter network of PM2.5 from the risk society perspective". *Information, communication & society*, v. 20, n. 7, pp. 1005-1023.  
<https://doi.org/10.1080/1369118X.2017.1303076>
- Cody, Emily M.; Reagan, Andrew J.; Mitchell, Lewis; Dodds, Peter-Sheridan; Danforth, Christopher M.** (2015). "Climate change sentiment on Twitter: an unsolicited public opinion poll". *PloS one*, v. 10, n. 8.  
<https://doi.org/10.1371/journal.pone.0136092>
- Cossu, Jean-Valère; Dugué, Nicolas; Labatut, Vincent** (2015). "Detecting real-world influence through Twitter". In: *2<sup>nd</sup> European network intelligence conference*.  
<https://doi.org/10.1109/ENIC.2015.20>
- Dahal, Biraj; Kumar, Sathish A. P.; Li, Zhenlong** (2019). "Topic modeling and sentiment analysis of global climate change tweets". *Social network analysis and mining*, v. 9, art. 24.  
<https://doi.org/10.1007/s13278-019-0568-8>
- De-Nooy, Wouter; Mrvar, Andrej; Batagelj, Vladimir** (2018). *Exploratory social network analysis with Pajek*. Cambridge University Press. ISBN: 978 1 108 47414 6
- Edizel, Bora; Bonchi, Francesco; Hajian, Sara; Panisson, André; Tassa, Tamir** (2020). "FaiRecSys: mitigating algorithmic bias in recommender systems". *International journal of data science and analytics*, v. 9, pp. 197-213.  
<https://doi.org/10.1007/S41060-019-00181-5>
- EPA** (2019). *What is green power?*. United States environmental protection Agency.  
<https://www.epa.gov/greenpower>
- European Parliament** (2018). "Directive (EU) 2018/2001 of the European parliament and of the council on the promotion of the use of energy from renewable sources". *Official journal of the European Union*, n. L 328, 21/12/2018.  
<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L2001>
- Fernández-Arias, Pablo** (2017). *Análisis de los factores que influyeron en la evolución y desarrollo del reactor nuclear PWR*. Tesis doctoral. Universidad de Salamanca.

- Fernández-Arias, Pablo; Cuevas, Ana; Vergara, Diego** (2021). "Controversia nuclear en España: la central de Lemóniz". *Revista CTS*, v. 16, n. 46, pp. 199-218.  
<http://ojs.revistacts.net/index.php/CTS/article/view/218>
- Freeman, Linton C.** (1978). "Centrality in social networks: conceptual clarification". *Social networks*, v. 1, n. 3, pp. 215-239.  
[https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Gibbs, Jeff; Moore, Michael** (2019). *Planet of the humans*.  
<https://www.youtube.com/watch?v=Zk11vl-7czE>
- González-Fernández-Villavicencio, Nieves** (2014). "El #hashtag ya tiene historia". *Anuario ThinkEPI*, v. 8, pp. 326-330.  
<https://thinkepi.profesionaldelainformacion.com/index.php/ThinkEPI/article/view/29601>
- Gupta, Kuhika; Ripberger, Joseph; Wehde, Wesley** (2018). "Advocacy group messaging on social media: using the narrative policy framework to study *Twitter* messages about nuclear energy policy in the United States". *Policy studies journal*, v. 46, n. 1, pp. 119-136.  
<https://doi.org/10.1111/psj.12176>
- Hanneman, Robert 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*.  
<http://revista-redes.rediris.es/webredes/textos/cap6.pdf>
- Holmberg, Kim; Hellsten, Iina** (2015). "Gender differences in the climate change communication on *Twitter*". *Internet research*, v. 25, n. 5, pp. 811-828.  
<https://doi.org/10.1108/IntR-07-2014-0179>
- Hutto, Clayton J.; Gilbert, Eric** (2014). "Vader: A parsimonious rule-based model for sentiment analysis of social media text". In: *Proceedings of the 8th international AAAI conference on weblogs and social media*, pp. 216-225.  
<https://ojs.aaai.org/index.php/ICWSM/article/view/14550/14399>
- Jain, Achin; Jain, Vanita** (2019). "Sentiment classification of *Twitter* data belonging to renewable energy using machine learning". *Journal of information and optimization sciences*, v. 40, n. 2, pp. 521-533.  
<https://doi.org/10.1080/02522667.2019.1582873>
- Khan, M. Ali-ud-din; Uddin, Muhammad-Fahim; Gupta, Navarun** (2014). "Seven V's of big data understanding big data to extract value". In: *Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education*, pp.1-5.  
<https://doi.org/10.1109/ASEEZone1.2014.6820689>
- Khatua, Aparup; Cambria, Erik; Ho, Shirley S.; Na, Jin-Cheon** (2020). "Deciphering public opinion of nuclear energy on *Twitter*". In: *2020 International joint conference on neural networks (IJCNN)*.  
<https://doi.org/10.1109/IJCNN48605.2020.9206903>
- Kim, Jiyoun; Brossard, Dominique; Scheufele, Dietram A.; Xenos, Michael** (2016). "Shared" information in the age of big data: exploring sentiment expression related to nuclear energy on *Twitter*". *Journalism and mass communication quarterly*, v. 93, n. 2, pp. 430-445.  
<https://doi.org/10.1177/1077699016640715>
- Kim, Serena Y.; Ganesan, Koushik; Dickens, Princess; Panda, Soumya** (2020). "Public sentiment toward solar energy: opinion mining of *Twitter* using a transformer-based language model". *Sustainability*, v. 13, n. 5, 2673.  
<https://doi.org/10.3390/su13052673>
- Kramer, Adam D. I.; Guillory, Jamie E.; Hancock, Jeffrey T.** (2014). "Experimental evidence of massive-scale emotional contagion through social networks". In: *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 24, pp. 8788-8790.  
<https://doi.org/10.1073/pnas.1320040111>
- Labonte, Dane; Rowlands, Ian H.** (2021). "Tweets and transitions: exploring *Twitter*-based political discourse regarding energy and electricity in Ontario, Canada". *Energy research and social science*, v. 72, p. 101870.  
<https://doi.org/10.1016/j.erss.2020.101870>
- Laney, Doug** (2001). *3D data management: controlling data volume, velocity, and variety*. META Group Inc.  
[https://www.academia.edu/37216189/Application\\_Delivery\\_Strategies](https://www.academia.edu/37216189/Application_Delivery_Strategies)
- Larrondo-Ureta, Ainara; Morales-i-Gras, Jordi; Orbegozo-Terradillos, Julen** (2019). "Feminist hashtag activism in Spain: Measuring the degree of politicisation of online discourse on #yosítecreo, #hermanayositecreo, #cuéntalo y #noestáso-la". *Communication & society*, v. 32, n. 4, pp. 207-221.  
<https://doi.org/10.15581/003.32.4.207-221>

- Li, Qiudan; Jin, Zhipeng; Wang, Can; Zeng, Daniel-Dajun** (2016). "Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems". *Knowledge-based systems*, v. 107, pp. 289-300.  
<https://doi.org/10.1016/j.knosys.2016.06.017>
- Li, Ruopu; Crowe, Jessica; Leifer, David; Zou, Lei; Schoof, Justin** (2019). "Beyond big data: social media challenges and opportunities for understanding social perception of energy". *Energy research and social science*, v. 56, p. 101217.  
<https://doi.org/10.1016/j.erss.2019.101217>
- Ljubljana University (2021). *Orange data mining*.  
<https://orangedatamining.com>
- Macmillan, Gordon** (2015). "8 reasons why *Twitter* is the home of TV conversation". *Twitter blog*, 29 September.  
[https://blog.twitter.com/en\\_gb/a/en-gb/2015/8-reasons-why-twitter-is-the-home-of-tv-conversation.html](https://blog.twitter.com/en_gb/a/en-gb/2015/8-reasons-why-twitter-is-the-home-of-tv-conversation.html)
- Margolin, Drew; Liao, Wang** (2018). "The emotional antecedents of solidarity in social media crowds". *New media & society*, v. 20, n. 10, pp. 3700-3719.  
<https://doi.org/10.1177/1461444818758702>
- 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>
- Mooney, Peter; Winstanley, Adam; Corcoran, Pdraig** (2009). "Evaluating *Twitter* for use in environmental awareness campaigns". *Proceedings of the China-Ireland information and communications technologies conference (CICT 2009)*. Maynooth: Department of Computer science, NUI Maynooth ER, pp. 83-86.  
<https://core.ac.uk/download/pdf/297012769.pdf>
- Morales-i-Gras, Jordi** (2017). *Soberanías enredadas: una perspectiva reticular, constructural y agéntica hacia los relatos soberanistas vasco y catalán contemporáneos en Twitter*. Tesis doctoral. Universidad del País Vasco / Euskal Herriko Unibertsitatea.  
<https://addi.ehu.es/handle/10810/22686>
- Morales-i-Gras, Jordi** (2020). *Datos masivos y minería de datos sociales: conceptos y herramientas básicas*.  
<https://jordimorales.com/es/recursos>
- Mrvar, Andrej; Batagelj, Vladimir** (2021). *Programs for analysis and visualization of very large networks. Reference manual*.  
<http://mrvar.fdv.uni-lj.si/pajek/pajekman.pdf>
- Newman, Mark E. J.; Girvan, Michelle** (2004). "Finding and evaluating community structure in networks". *Physical review E*, v. 69, n. 2, 026113.  
<https://doi.org/10.1103/PhysRevE.69.026113>
- Oguntimilehin, Abiodun; Ademola, Emmanuel-Ojo** (2014). "A review of big data management, benefits and challenges". *Journal of emerging trends in computing and information sciences*, v. 5, n. 6, pp. 433-438.  
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.677.8629&rep=rep1&type=pdf>
- OpenRefine (2021). *A free, open source, powerful tool for working with messy data*.  
<https://openrefine.org>
- Orbegozo-Terradillos, Julen; Larrondo-Ureta, Ainara; Morales-i-Gras, Jordi** (2020). "Influencia del género en los debates electorales en España: análisis de la audiencia social en #ElDebateDecisivo y #L6Neldebate". *El profesional de la información*, v. 29, n. 2.  
<https://doi.org/10.3145/epi.2020.mar.09>
- Orbegozo-Terradillos, Julen; Morales-i-Gras, Jordi; Larrondo-Ureta, Ainara** (2019). "Feminismos indignados ante la justicia: la conversación digital en el caso de La Manada". *IC revista científica de información y comunicación*, n. 16, pp. 211-247.
- Patgiri, Ripon; Ahmed, Arif** (2016). "Big data : The V's of the game changer paradigm". In: *2016 IEEE 18<sup>th</sup> international conference on high performance computing and communications; IEEE 14<sup>th</sup> international conference on smart city; IEEE 2<sup>nd</sup> international conference on data science and systems (HPCC/SmartCity/DSS)*, pp. 17-24.  
<https://doi.org/10.1109/HPCC-SmartCity-DSS.2016.0014>
- Pilař, Ladislav; Kvasničková-Stanislovská, Lucie; Pitrová, Jana; Krejčí, Igor; Tichá, Ivana; Chalupová, Martina** (2019). "Twitter analysis of global communication in the field of sustainability". *Sustainability*, v. 11, n. 24, 6958.  
<https://doi.org/10.3390/su11246958>
- Prabhakar, Kaila-Rajesh** (2019). "Climate change and *Twitter*. An empirical analysis of environmental awareness and engagement". *Disaster advances*, v. 12, n. 9, pp. 10-15.
- Reboredo, Juan C.; Ugolini, Andrea** (2018). "The impact of *Twitter* sentiment on renewable energy stocks". *Energy economics*, v. 76, pp. 153-169.  
<https://doi.org/10.1016/j.eneco.2018.10.014>

**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>

**Ruiz-Soler, Javier** (2017). "Twitter research for social scientists: a brief introduction to the benefits, limitations and tools for analysing Twitter data". *Dígitos*, v. 1, n. 3, pp. 17-32.

<https://revistadigitos.com/index.php/digitos/article/view/87>

**Shen, Chien-Wen; Luong, Thai-Ha; Pham, Tuan** (2021). "Exploration of social media opinions on innovation for sustainable development goals by topic modeling and sentiment analysis". In: *Research and innovation forum 2020. RiiForum 2020*, pp. 459-471.

[https://doi.org/10.1007/978-3-030-62066-0\\_35](https://doi.org/10.1007/978-3-030-62066-0_35)

**Soussan, Tariq; Trovati, Marcello** (2020). "Twitter analysis for business intelligence". In: Barolli Leonard; Nishino, Hsing-Chung; Miwa, Hiroyoshi. *Advances in intelligent networking and collaborative systems. INCoS 2019. Advances in intelligent systems and computing*, v. 1035, pp. 473-480. ISBN: 978 3 030 29035 1

[https://doi.org/10.1007/978-3-030-29035-1\\_46](https://doi.org/10.1007/978-3-030-29035-1_46)

Unesco (2021). *Invertir en ciencia, tecnología e innovación*.

<https://es.unesco.org/themes/invertir-ciencia-tecnologia-e-innovacion>

**Velázquez-Álvarez, O. Alejandro; Aguilar-Gallegos, Norman** (2005). *Manual introductorio al análisis de redes sociales (medidas de centralidad)*.

[http://revista-redes.rediris.es/webredes/talleres/Manual\\_ARC.pdf](http://revista-redes.rediris.es/webredes/talleres/Manual_ARC.pdf)

**Veltri, Giuseppe A.** (2012). "Microblogging and nanotweets: nanotechnology on Twitter". *Public understanding of science*, v. 22, n. 7, pp. 832-849.

<https://doi.org/10.1177/0963662512463510>

**Veltri, Giuseppe A.; Atanasova, Dimitrinka** (2017). "Climate change on Twitter: Content, media ecology and information sharing behavior". *Public understanding of science*, v. 26, n. 6, pp. 721-737.

<https://doi.org/10.1177/0963662515613702>

WordArt (2021). *Word cloud art creator*.

<https://wordart.com>

YouTube (2020). *Donald Trump vs Joe Biden: Full presidential debate | US Election 2020*.

<https://www.youtube.com/watch?v=GBLCSdx6lk>

**Zeifer, Bárbara** (2020). "El hashtag contestatario: cuando los hashtags tienen efectos políticos". *Dígitos. Revista de comunicación digital*, v. 6, pp. 101-118.

<https://doi.org/10.7203/rd.v1i6.178>



<http://www.profesionaldelainformacion.com>

**Bienvenido a EPI**  
Revista científica internacional

e-ISSN: 1699-2407  
<https://doi.org/10.3145/EPI>

Revista internacional de  
**Información y Comunicación**  
indexada por WoS Social Sciences Citation Index (Q3),  
Scopus (Q1) y otras bases de datos

Factor de impacto JCR:  
**JIF 2020=2,253**

Scopus/SCImago Journal Rank:  
**SJR 2020=0,698**