



From Sustainable Development Goals to sustainable industry, innovation and infrastructure: insights from the digital sphere

Enara Zarrabeitia-Bilbao¹ · Maite Jaca-Madariaga¹ · Rosa-María Rio-Belver² · Izaskun Alvarez-Meaza¹

Received: 28 September 2023 / Accepted: 27 November 2023
© The Author(s) 2023

Abstract

The aim of this study is to analyze and understand the digital discussion on Twitter (from 2016 to 2022) of the SDGs in general and SDG 9 in particular, based on a comparative approach and with a methodology using Python libraries for advanced data analysis, social network analysis (SNA) methods and artificial neural networks (ANN) models. To this end, 6,323,139 tweets about SDGs in general and 2,892,922 about specific SDGs were retrieved for further analysis. The results obtained show that SDG 9 generated less interest and a lower presence of women in the social discussion than other SDGs over the seven years studied; furthermore, the number of tweets about SDG 9 has decreased. However, the digital conversation among different actors does develop in a cohesive manner, sharing leadership and space. This study shows that there are exceptional peaks in the digital activity and the SDG topic goes from “sidestream” to “mainstream” in terms of the digital public debate when certain celebrities (specifically, the Korean music group BTS) interact with the initiative. Finally, SDGs do not generate controversy and there is no substantial difference in the distribution of sentiment and emotions between different periods and different SDGs.

Keywords Sustainable Development Goals · SDGs · SDG 9 · Social networks · Social network analysis · Artificial neural networks · Twitter

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

Maite Jaca-Madariaga
maite.jacamadariaga@ehu.eus

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

Izaskun Alvarez-Meaza
izaskun.alvarez@ehu.eus

¹ Industrial Organization and Management Engineering Department, Faculty of Engineering, University of the Basque Country, Paseo Rafael Moreno “Pitxitxi” n. 2/3, 48013 Bilbao, Spain

² Industrial Organization and Management Engineering Department, Faculty of Engineering, University of the Basque Country, Nieves Cano 12, 01006 Vitoria-Gasteiz, Spain

1 Introduction

The Sustainable Development Goals (SDGs) were adopted by the United Nations, on September 25, 2015; the aim being to end poverty, preserve the planet and ensure prosperity as part of a new sustainable development agenda: the 2030 Agenda (UN Press, 2015; United Nations, 2023b). Hence, the 2030 Agenda and the SDGs have become the new moral commitment and the map that sets the course for governments, cities, businesses, organizations and citizens over a 15-year time span (Herranz de la Casa & García Caballero, 2021).

The Agenda has 17 interconnected SDGs with 169 associated targets and more than 232 unique indicators for transforming the world (Roser & Ortiz-Ospina, 2018; United Nations, 2015). Within these 17 goals, number 9 (SDG 9) is about "building resilient infrastructures, promoting inclusive and sustainable industrialization and fostering innovation" (United Nations, 2023a). A sound economy and society are the basis for sustainable development. Therefore, given the unstable economic landscape and social inequalities, sustained growth has to include an industrialization that provides equal opportunities for all, supported by innovation and resilient infrastructures (United Nations, 2016). In addition, in the post-pandemic era, SDG 9 is considered a priority for low and lower-middle income countries (Yuan et al., 2023).

In order to implement and achieve the SDGs in general and the SDG 9 in particular, the involvement of all kinds of agents is necessary. Therefore, a formal or traditional communication strategy may not be the best option to raise awareness of the SDGs. Also, traditional data sources, while valuable and necessary, are not sufficient to measure the development of the SDGs and new, non-traditional, data sources are needed (Fritz et al., 2019; Ghermandi et al., 2023; Grover et al., 2021). Traditional polling has been criticized for its recent errors and despite the fact that the method has been adapted and improved, there are many who question the reliability of polls and believe that they should be complemented with other methodologies that allow observation from other angles and approaches (Gutiérrez-Rubí, 2021).

In this sense, social media defined as a set of tools that enable and promote social interactions between users in the digital sphere (Hansen et al., 2011) have been used in lots of fields to understand many different phenomena (Roldán-Álvarez et al., 2021). Among all the existing social media, Twitter¹ is a social network that allows users to easily interact with each other and to exchange information (Roldán-Álvarez et al., 2021).

Tweets are considered an important resource for studying public awareness. The opinions and comments published via tweets can be used to raise public awareness and to help governments and enterprises to understand citizens' opinions (Li et al., 2019). Twitter's great capacity for dissemination allows users themselves (whether public institutions, businesses or citizens) to influence public opinion. The impact of Twitter has been demonstrated in various fields, such as politics, social movements, communication, business, healthcare, crisis detection and management (Marzouki et al., 2021; Zarrabeitia-Bilbao et al., 2022).

Motivated by the above statements, this research paper analyzes the discussions about SDGs in general and SDG 9 in particular, which took place over a seven-year period (from 2016 to 2022) on the Twitter platform. Thus, the main aim of this research is to understand

¹ Twitter officially rebranded to "X" after owner Elon Musk changed its iconic bird logo on July 24, 2023 (<https://twitter.com/elonmusk/status/1683378289031761920>). However, during the time span of this research, the social network was Twitter; therefore, it was decided to use this term throughout the work.

how SDGs in general, and specifically SDG 9, are discussed on Twitter and to draw valuable conclusions from these discussions in the digital sphere. More specifically, the results of this research will serve to better understand the social reality of a key social issue that has been agreed upon within the institutional sphere, and is a wake-up call for politicians, nongovernmental organization (NGOs) and business organizations across the world.

2 Social networks data mining and Sustainable Development Goals

Monitoring and identifying social trends is essential for government and enterprises strategic planning and practices (Li et al., 2019). In this regard, social media are communication tools that provide knowledge about trends and the sensitization of citizens to social issues. Thus, through social media, different actors can obtain useful information and make decisions in order to raise awareness and improve the achievement of the SDGs (Galiano-Coronil & Ortega-Gil, 2019). It has been proven that social media can do a lot for the attainment of the SDGs (Carrasco-Polaino & Piqueiras Conlledo, 2023; Porcelli, 2020) and, as will be seen below, several studies have analyzed the trend or social impact of the SDGs through data obtained from social media.

Among the various scientific studies that analyze the presence of the SDGs in the digital sphere, Twitter is the social network par excellence used to obtain data (Bautista-Puig & Dudek, 2019; Carrasco-Polaino & Piqueiras Conlledo, 2023; Franco-Riquelme & Rubalcaba, 2021; Galiano-Coronil et al., 2023; Grover et al., 2021; Marzouki et al., 2021; Patuelli & Saracco, 2023; Roldán-Álvarez et al., 2021; Singh et al., 2023). Unlike other social networks such as Facebook and Instagram, which are more related to personal life or entertainment, or LinkedIn which is based more on work or professional relationships, Twitter is a social network used to comment and discuss social topics such as SDGs. Twitter is considered a virtual agora that promotes interactions and shares thoughts, opinions and beliefs, which makes it an adequate source of scientific research data for this kind of research field (Burnap & Williams, 2016; Galiano-Coronil et al., 2023).

The Organisation for Economic Co-operation and Development (OECD) reports that the countries and their governments are primarily responsible for responding to the 2030 Agenda (OECD, 2016); however, no single government can meet a challenge as big as the 2030 Agenda alone. In this sense, the initiative OpenODS states that without the actions of other agents (public and private sector, or citizens), no progress can be made in achieving the 17 SDGs (OpenODS, 2023). Therefore, different researchers have focused their studies on monitoring the discussion on Twitter surrounding SDGs, in order to analyze the contribution or influence of different actors in achieving them. In this sense, the following have been analyzed in previous research: the public sector, through the contribution of public administration (Carrasco-Polaino & Piqueiras Conlledo, 2023) and political leaders (Grover et al., 2021); the private sector, through the contribution of different companies (Franco-Riquelme & Rubalcaba, 2021; Patuelli & Saracco, 2023); and the citizenry (Roldán-Álvarez et al., 2021).

Furthermore, although most research has focused on the study of SDGs in general, specific SDGs have also been exclusively monitored, such as SDG11 (sustainable cities and communities) (Marzouki et al., 2021) or SDG13 (climate action) (Singh et al., 2023).

Twitter's data collection strategy regarding SDGs is based either on the tracking of individual Twitter accounts of public institutions (Carrasco-Polaino & Piqueiras Conlledo, 2023; Galiano-Coronil et al., 2023), public officials (Carrasco-Polaino &

Piqueiras Conlledo, 2023; Grover et al., 2021) or companies (Patuelli & Saracco, 2023), or on the use of keywords as a search query (Bautista-Puig & Dudek, 2019; Marzouki et al., 2021; Roldán-Álvarez et al., 2021; Singh et al., 2023). There are also studies that combine both strategies (Franco-Riquelme & Rubalcaba, 2021).

In the case of studies that use an individual account or a limited number of Twitter accounts as their reference, it is relatively easy to detect the gender of the users. In this sense, there are studies that analyze the behavior of users in relation to SDGs according to their gender (Grover et al., 2021), but not in the case of more generic and massive studies that use keywords as a search query, where a large number of Internet users interact.

In all these scientific studies, different social media analytics (SMA) methodologies, techniques or tools were used. One of the techniques widely used is social network analysis (SNA) (Carrasco-Polaino & Piqueiras Conlledo, 2023; Patuelli & Saracco, 2023; Roldán-Álvarez et al., 2021), an empirical exploration technique that is very useful for analyzing social media activity due to its interconnected nature and its own algorithms and metrics for evaluating the properties of each node, as well as the network as a whole (Morales-i-Gras, 2020). Moreover, all of them use some classic SMA techniques such as sentiment, engagement, polarity, objectivity or topic analysis. However, it is worth noting that more powerful, sophisticated and versatile tools like artificial neural networks (ANN) or deep learning algorithms—used to provide more precise answers for more complex realities—are not used to improve the techniques previously mentioned.

3 Research questions and methodology

The aim of this study is to analyze and understand the digital conversation of SDGs in general and SDG 9 in particular, based on a comparative approach and using Big Data processing methods and tools related to the digital footprint on Twitter. In other words, the aim is to generate social knowledge concerning the effects on digital public opinion of a particular event, based on the information obtained from the digital interactions of a given social network.

Moreover, the research process remains in the hands of the scientific community, so that it can be replicated in scientific research into other social phenomena.

3.1 Research questions

Due to the wide range of information that digital Big Data can provide, it is essential in this type of study to pose certain questions to guide the research process. Hence, this scientific study formulates the following three research questions:

- *RQ1*. What has the social presence of SDGs and SDG 9 been like on Twitter?
- *RQ2*. What are the main characteristics of the overall discussion generated and who is talking about SDGs and SDG9 on Twitter?
- *RQ3*. What sentiments and emotions are communicated in discussions surrounding SDGs and SDG 9 on Twitter?

3.2 Research methodology design

To accomplish the research aim we divided the research methodology into two phases: data preparation and data analysis.

3.2.1 Data preparation

The social network Twitter was used to obtain the data under study. More specifically, the tweets were obtained through Twitter API for Academic Research and Twarc2 (a Python library) (Twarc, 2023; Twitter, 2023).

For the retrieval of tweets related to SDGs in general, the query used was: “#SDG” OR “#SDGs”; and, for the retrieval of tweets related to specific SDGs 17 different searches were carried out, one for each SDG, using the hashtags corresponding to each one (i.e., from #SDG1 to #SDG17). The data were collected retroactively from January 1, 2016, to December 31, 2022, in order to include complete years since the adoption of the SDGs on September 25, 2015.

Once the tweets were obtained, it was necessary to transform and prepare the data for further data analysis. The data obtained, in an unstructured format (json format), were transformed into structured data (csv format) for further analysis (Twarc, 2023). Once the data were in a structured format, they were pre-processed (data cleaning) to prepare them for information extraction (OpenRefine, 2023; VantagePoint, 2023).

One of the decisions made was to eliminate tweets that were not in English from the initial corpus to give coherence to the study as a whole.

3.2.2 Data analysis

After collecting and pre-processing data, the empirical approach of the research was carried out via Python libraries for advanced data analysis, social network analysis (SNA) methods and artificial neural networks (ANN) models.

In order to answer the first research question, an analysis of the temporal distribution of tweets related to SDGs in general, and of tweets related to specific SDGs in particular (with special emphasis on SDG 9), was carried out. Moreover, for the purpose of analyzing women's presence in the digital debate the gender-guesser Python library (Saeta-Pérez, 2016) was implemented to estimate the gender of the users who posted the collected tweets.

Regarding the second research question, first of all, two networks (one for SDGs in general and another for SDG 9 in particular) were synthesized, based on which Twitter users mentioned other users in the conversation itself (Bastian et al., 2009; Mrvar & Batagelj, 2016). Then, in order to analyze the relationships established between the users of the digital conversation, different global metrics were calculated. It was thus possible to analyze the morphology of the networks obtained. Subsequently, an analysis of communities generated in the digital sphere discussions was carried out via the Louvain Multilevel Algorithm, an algorithm for performing community or clustering detection in very large networks by maximizing a modularity function within short computing times (Blondel et al., 2008). Likewise, the community leaders were identified via input degree centrality metric, a metric that measures the number of mentions received by a user on the network, either by direct mentions, replies or retweets (Zarrabeitia-Bilbao et al., 2022).

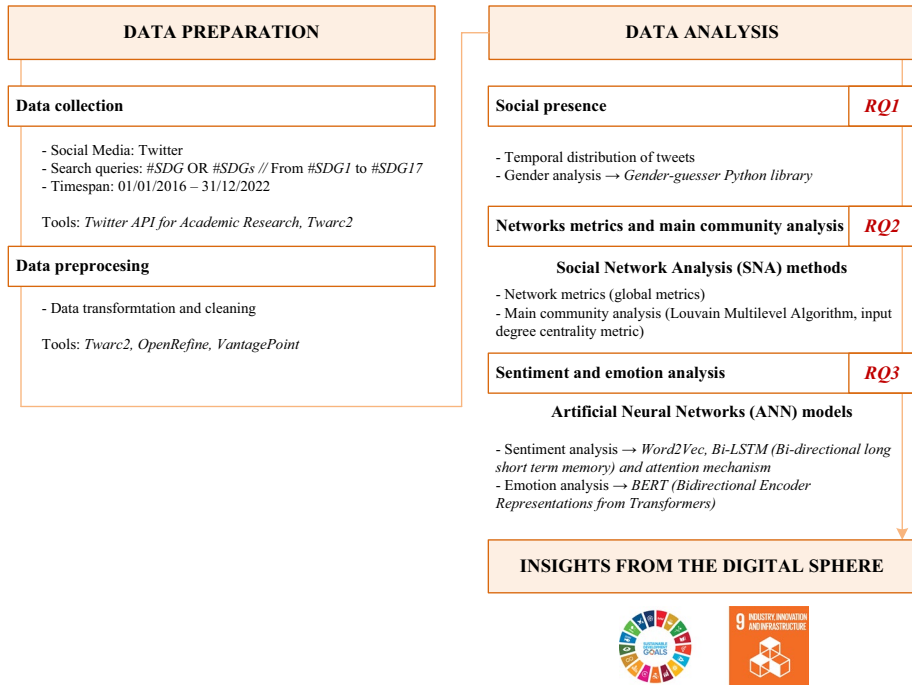


Fig. 1 Overview of the methodology

Finally, to answer the last research question, in the third stage of the empirical approach, artificial neural networks were used to process the tweets and analyze the sentiments and emotions generated in the discussions. Regarding the sentiment analysis of the digital conversation, the SDG tweets were run through an artificial neural network model that includes Word2Vec, Bi-LSTM (Bi-directional long short-term memory) and attention mechanism (accuracy of the model: 84.12%) in order to obtain the sentiment score, to measure its positivity or negativity (Jaca-Madariaga et al., 2023; Zarrabeitia-Bilbao et al., 2023). In addition to this, to further complete the sentiment mining and analyze the emotions generated about SDGs, a multi-class emotion classification was carried out. For this purpose, an artificial neural network model highly used in NLP (Natural Language Processing) was used: BERT (Bidirectional Encoder Representations from Transformers), a large-scale model with transformers inside it (Devlin et al., 2018). Specifically, BertBase Uncase Emotion model (accuracy of the model: 94.05%) was used to detect the emotions of the tweets: sadness, joy, love, anger, fear and/or surprise (Savani, 2023).

Figure 1 provides an overview of the methodological framework of the scientific work.

4 Results and discussion

Once the proposed methodology had been implemented, the results were obtained and are shown below along with the corresponding discussion.

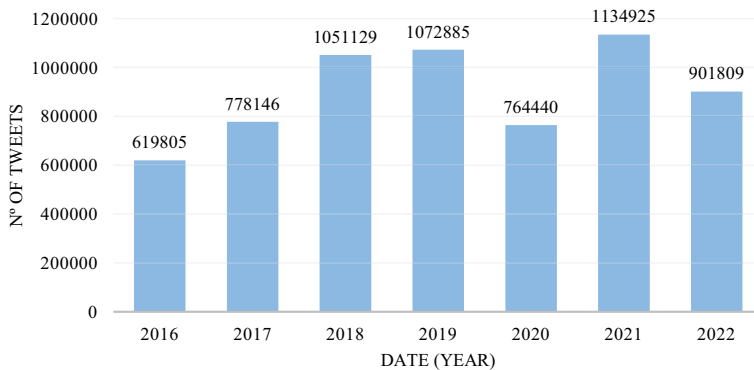


Fig. 2 Number of tweets (in English) per year (from 2016 to 2022) related to SDGs

4.1 Social presence of SDGs and SDG 9 (RQ1)

In total, 9,352,738 tweets about SDGs in general from 2016 to 2022 were retrieved (using the search query: “#SDG” OR “#SDGs”), of which 6,323,139 were in English, i.e., those used henceforth for the data analysis. It can be seen that the majority of the tweets (67.61%) were posted in English, which is to be expected since the search terms used were in English and nowadays it is considered the institutional, business and scientific language (Roldán-Álvarez et al., 2021), with the second (Japanese, 25.25%) and third (Spanish, 0.94%) most widely used languages, lagging far behind.

Figure 2 shows that Twitter user interest in the subject increased from 2016 to 2019. In 2020, a significant decline in terms of number of tweets can be observed and we can infer that the COVID’19 pandemic, in addition to negatively impacting progress on the SDGs (Madurai-Elavarasan et al., 2022; Yuan et al., 2023), also negatively impacted user engagement in terms of SDGs (Marzouki et al., 2021).

As for the number of tweets about specific SDGs in the same period (from 2016 to 2022), in total, 3,198,839 tweets about the different SDGs were retrieved (using 17 different searches queries: from #SDG1 to #SDG17), of which 2,892,922 are in English; i.e., those used for the data analysis from here on in. It can be seen that in this case also, the majority of the tweets (90.44%) were posted in English, as expected, with the second (Japanese, 1.13%) and third (Spanish, 0.93%) most widely used languages, lagging far behind.

Figure 3 shows that social network user interest varies in terms of the different SDGs. SDG 9 in particular generated the least interest among all of them. Moreover, in addition to having generated less interest than the rest of the SDGs, this has decreased over the last five years (Fig. 4).

The temporal distribution of the number of daily tweets related to the SDGs (Fig. 5) shows that, although the distribution of tweets is relatively homogeneous over the seven years under study, different topics of discussion or events led to an increase in interactions in the digital sphere conversation. While on average 2,473 tweets were posted per day, there was a substantial increase in posts (164,746 tweets) on September 20, 2021. This increase in interaction on the social network was due to the participation of the members of

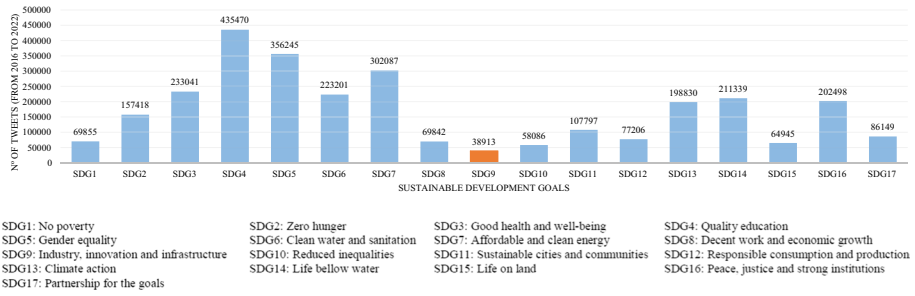


Fig. 3 Number of tweets (in English) related to the different SDGs over the period 2016–2022

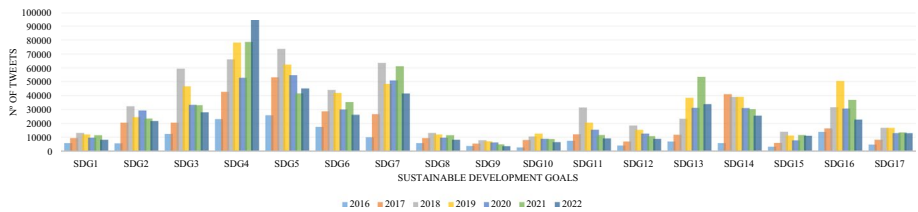


Fig. 4 Number of tweets per year (from 2016 to 2022) related to the different SDGs

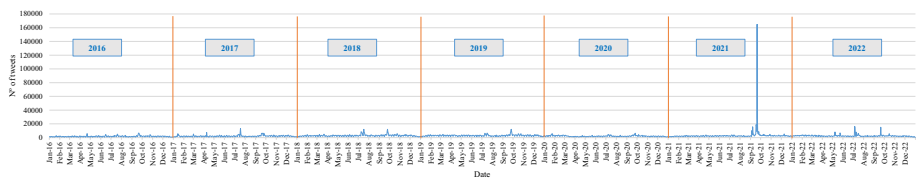


Fig. 5 General daily trend of tweets related to SDGs over the period 2016–2022

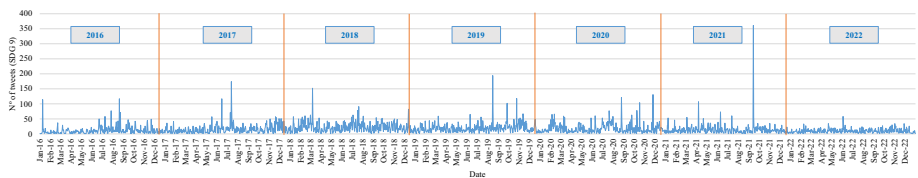


Fig. 6 General daily trend of tweets related to SDG 9 over the period 2016–2022

the well-known South Korean music group BTS² in “the second SDG Moment of the Decade of Action” (SDG Moment 2021) (United Nations, 2021). Fans of the band followed the musicians’ performance through the UN’s YouTube channel, flooding a live chat with

² For more information: <https://ibighit.com/bts/eng/>

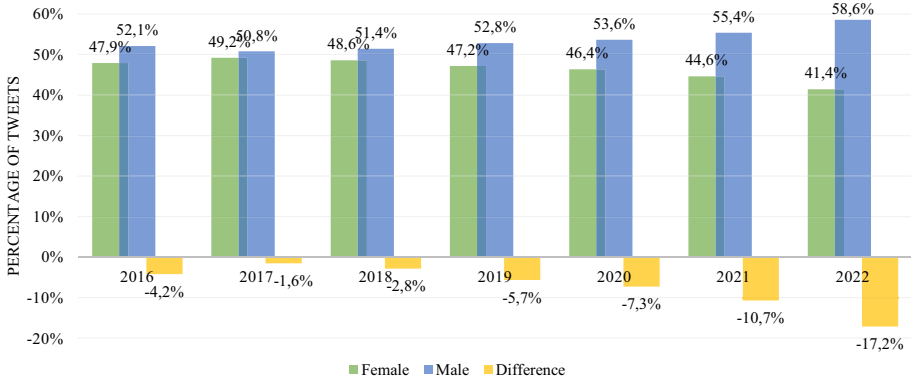


Fig. 7 Percentage presence of female and male in tweets per year (from 2016 to 2022) related to SDGs

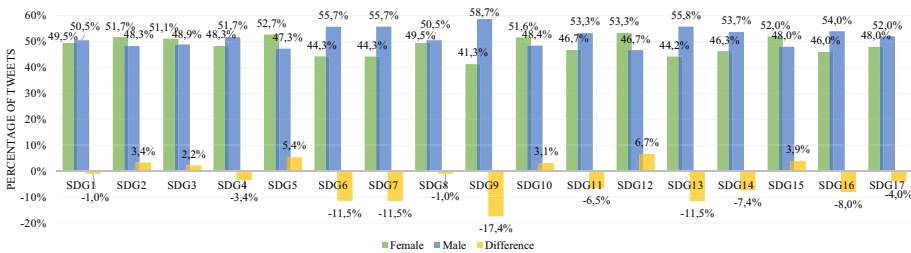


Fig. 8 Percentage presence of female and male in tweets related to the different SDGs over the period 2016–2022

messages (Vigdor, 2021), which was also reflected in the viralization of tweets mentioning the event.³

The daily distribution of tweets related specifically to SDG 9 was relatively homogeneous in the seven years under study (Fig. 6). Nevertheless, in this case also, specific topics or events led to increased interaction among the users of this social network. While the average number of tweets posted per day was 15, there was a substantial increase in posts (361 tweets) on September 14, 2021. This increase in tweets was due to a series of messages posted by the user @FoodSystems (UN Food Systems Coordination Hub) about how the food system can contribute to SDG 9.⁴

As far as gender distribution is concerned, it should be noted that the gender of the users was detected in 38% of those who tweeted about SDGs in general, and in 37% of those who tweeted about the specific SDGs. In the rest of cases, the users were non-individual users (different types of organizations, brands, etc.) or the variable "author name" does not give enough information for gender detection.

Figure 6 shows the gender distribution over the years analyzed in relation to SDGs in general. Considering that the percentage distribution of Twitter users worldwide (as of

³ For more information: <https://twitter.com/aminajmohammed/status/1439940910855475203?lang=zh-Hant>.

⁴ For more information: <https://twitter.com/i/events/1443857307423682561?lang=es>

Table 1 Metrics extracted over the period 2016–2022 from the analyzed conversations of SDGs and SDG 9 (Zarrabeitia-Bilbao et al., 2022)

Network metrics	Overall networks	
	SDGs	SDG 9
Total impacts (tweets or retweets)	6,323,139	38,913
Users (nodes)	1,308,675	22,139
Average impacts (per user)	6.64	2.50
Arcs (interactions)	8,690,864	55,311
Density	0.00000507	0.00011285
Average degree	13.28192867	4.99670265
Input degree centralization	0.17501267	0.08287038
Output degree centralization	0.04175326	0.04849355
Number of clusters	6390 (8 > 4% of nodes)	735 (8 > 3% of nodes)
Modularity (Louvain multi-level algorithm)	0.621027	0.830560

January 2023) is 37% female and 63% male (DataReportal, 2023), it is worth noting the high presence of women in the digital sphere conversation about SDGs.

From a gender perspective, this conversational phenomenon and this space or forum could be considered as approaching equality in terms of the number of male and female participants. However, it has also been observed that this presence has decreased in recent years (Fig. 7).

Regarding the presence of male and female profiles in the conversations established in relation to specific SDGs, differences were observed depending on the different SDGs (Fig. 8). There are specific SDGs (SDG 2, 3, 5, 10, 12 and 15) in which the presence of female users surpasses that of male users; however, there are more specific SDGs in which the presence of men is greater than that of women, with SDG 9 standing out among all of them.

4.2 Networks metrics and main community analysis of SDGs and SDG 9 (RQ2)

The digital conversations about SDGs in general and about SDG 9 in particular were captured and transformed into two weighted directed network graphs. Nodes represent a Twitter user and arcs represent Twitter mentions (from one user to another user) either in the form of a retweet, a direct reply or a mention within a single tweet. Table 1 provides information about the general morphology of both generated digital conversations.

The first of the networks contains a total of 1,308,675 actors (nodes) and 8,690,864 connections (arcs), summarizing an average of 6.64 impacts per user. Each actor in the network is connected on average to 13.28 other actors and the density variable indicates that only 0.0005% of the possible connections between actors have materialized, which suggests that there are still many strategic questions to be explored by SDGs as a whole. In the case of the second network, it contains a total of 22,139 actors and 55,311 connections, summarizing an average of 2.5 impacts per user. Each actor in the network is connected on average to five other actors, and the density indicates that 0.01% of the possible connections between actors have materialized, which, although a low figure, is higher than that of the previous network. The density figures obtained suggest that there are numerous discussions in which

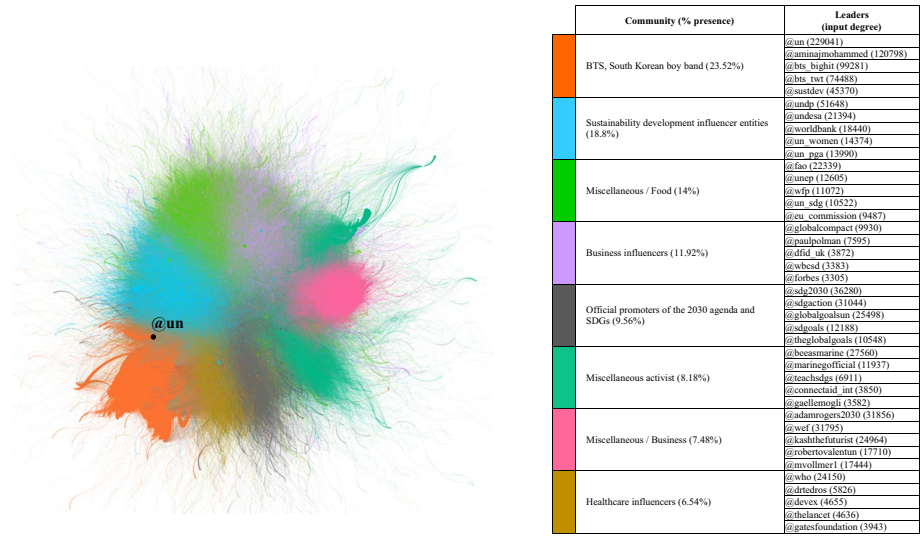


Fig. 9 Network of most important communities over the period 2016–2022 for SDGs

Twitter users tend to establish links with a really small number of users (Rovira-Sancho & Morales-i-Gras, 2022).

Neither the input degree centralization (17.5% and 8.3%) nor the output degree centralization (4.2% and 4.8%) is high, so it can be inferred that there were no small groups of nodes that monopolized the reception of mentions, nor that sent out most of the mentions.

After applying the Louvain Multilevel Algorithm, 6390 communities were identified for the first network and 735 communities for the second network. These figures show how fragmented the analyzed conversations are, which is usual in massive data analytical strategies (Orbegozo-Terradillos et al., 2019). The modularity is 0.62 and 0.83, respectively, indicating a very good partitioning quality and reinforcing the aforementioned idea that the groups of nodes (users) are very diverse and tend to relate only to nodes (users) belonging to their clusters or communities.

From the debate about SDGs, the following were isolated for further analysis: the eight main communities that gather more than 4% of the total nodes (64.15% of the nodes of the overall network in total); and, from the debate on SDG 9 eight communities that gather more than 3% of the total nodes (34.86% of the nodes of the overall network in total).

Figure 9, on the one hand, shows the most important communities of the network that has been synthesized over the period 2016–2022 for SDGs in general. The position that each main community holds in the overall network and their connections with the other main communities show that all main communities are interconnected, i.e., the network formed by the main communities is not polarized.

An overview of the graph, taking into account its structure, provides some information about some indications that lead us to affirm that it is a cohesive, proportionate and balanced conversation. In this respect, the communities generated are similar in size, are located at similar distances from the center of the graph and connect with a similar number of groups. In that sense, it could be considered a perfect graph in "morphological" terms.

Regarding the actors present in the conversation, on the one hand, the diversity of social references (activists, influencers, etc.) stands out, as does, on the other hand, the absence of

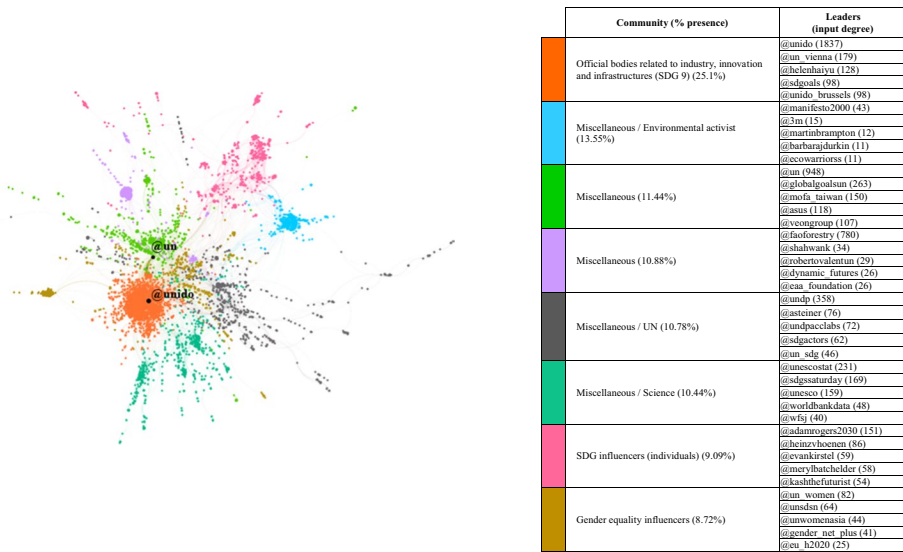


Fig. 10 Network of most important communities over the period 2016–2022 for SDG 9

two fundamental actors in the public debate: the media and political or institutional agents, beyond the very promoters of the phenomenon, the United Nations (UN).

As for the latter entity, even though it is the leader of the general conversation (the most mentioned account), its relatively remote or isolated location in the graph (bottom left) is significant and can be interpreted as its inability to cohere and centralize the process. In addition, the UN (@un) is located in one of the worst placed communities (orange color) in terms of connectivity: It only "touches" or connects directly and without intermediaries with two clusters (blue and light brown colors), which denotes a certain isolation in communicative terms. This could derive from the fact that the impulse and presence of this institution is marked by the performance of the K-Pop band BTS in a specific event and by the large increase in tweets produced by that act.

Figure 10, on the other hand, shows the graph referring to the digital conversation around SDG 9. From a general and comparative view, there are some noteworthy details, such as, for example, the role played in this case by the main institutional accounts of the organizers (@unido and @un). Spatially, these accounts are, in this case, in much more central positions and better connected to the rest of the conversation communities. This indicates that, in this specific SDG, the debate or discussion is supported and depends to a greater extent on the official and institutional activity of UN-related accounts.

Likewise, it can be observed that civil society also takes part in the debate. However, the private sector does not play a key role in any of the main communities.

4.3 Sentiment and emotion analysis of SDGs and SDG 9 (RQ3)

Sentiment and emotion analysis of Twitter data aims to describe people's attitudes toward the SDGs in general and SDG 9 in particular.

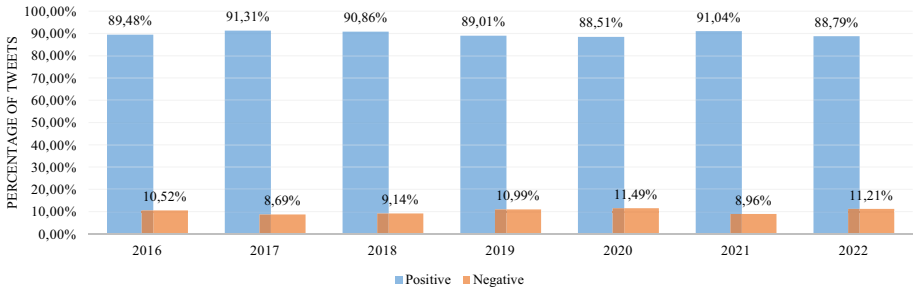


Fig. 11 Positive and negative tweets per year (from 2016 to 2022) related to SDGs

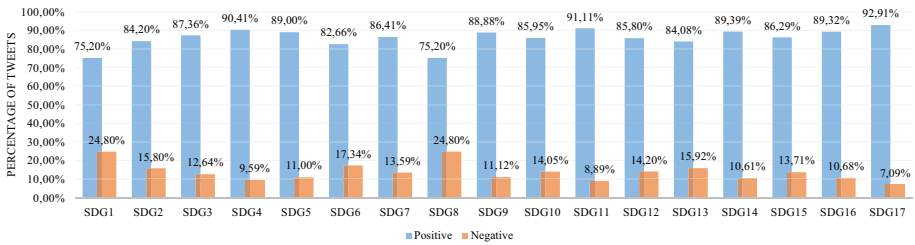


Fig. 12 Positive and negative tweets related to the different SDGs over the period 2016–2022

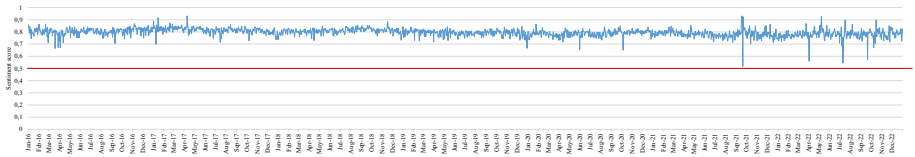


Fig. 13 Average sentiment per day of tweet related to SDGs over the period 2016–2022

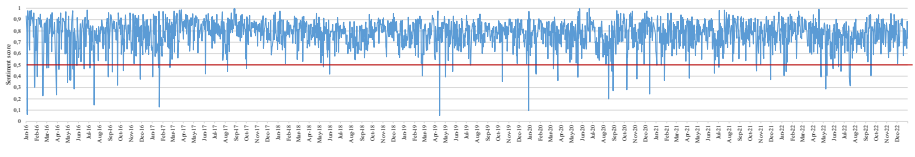


Fig. 14 Average sentiment per day of tweets related to SDG 9 over the period 2016–2022

Using artificial neural networks (Word2Vec, Bi-LSTM and attention mechanism) the sentiment classification model applied detected whether each tweet showed a positive or a negative sentiment. Figure 11 indicates how the proportion of positive and negative tweets related to the SDGs in general remains practically constant in each of the years analyzed. In addition, it seems that when different actors talk about SDGs, they tend to do so in positive terms. As far as the different SDGs are concerned, communication remains mostly positive (Fig. 12), but there are differences between the specific SDGs. In the case of SDG 1 (no poverty) and SDG 8 (decent work and economic growth), the

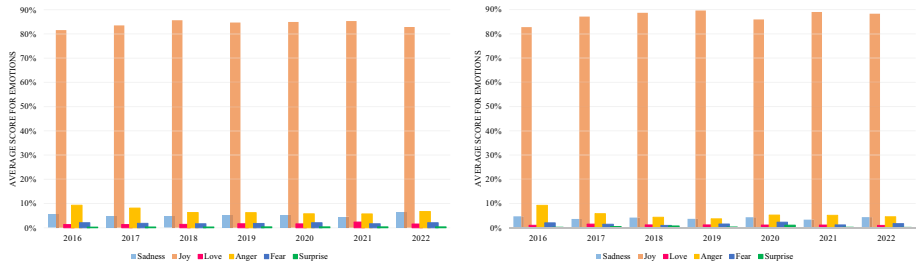


Fig. 15 Emotions in tweets per year (from 2016 to 2022) related to SDGs (left graph) and SDG 9 (right graph)



Fig. 16 Emotions in tweets related to the different SDGs over the period 2016–2022

two SDGs that were most affected by the COVID-19 pandemic (Madurai-Elavarasan et al., 2022), the percentage of negatives is almost 25%.

As for sentiment evolution detected through the neural network sentiment model, Figs. 13 and 14 show the sentiment score for the SDGs and SDG 9 tweets over the days for each of the seven years under study. Taking into account that a sentiment score above 0.5 is considered a positive sentiment, while a lower value is a negative sentiment, it can be observed that in the case of the SDGs in general, the average sentiment per day was relatively homogeneous and always positive. In the case of SDG 9, although the average sentiment per day was practically always positive, this average was not homogeneous over time and a more detailed analysis of the negative peaks that appeared would be desirable. In this sense, an analysis of the most negative peaks (values below 0.1) shows that these are activist tweets focused on criticizing or pointing out some social need related to the subject, and that directly or indirectly call for an eventual institutional intervention.

Regarding emotion analysis, using BERT Artificial Neural Network, an emotion multi-classification was carried out on six emotions: sadness, joy, love, anger, fear and/or surprise. In this way, a score was calculated for each of the tweets pertaining to each of the labels that indicate the probability of showing these emotions. Figure 15 shows how the proportion of different emotions related to the SDGs in general and SDG 9 in particular remains practically constant in each of the years analyzed, with “Joy” being the predominant emotion. In terms of the different SDGs, although the predominant emotion continues to be “Joy” (positive sentiment) in all cases, a “negative sentiment” increase can also be observed through the “sadness” emotion for SDG 1 (no poverty) and SDG 8 (decent work and economic growth) (Fig. 16).

5 Conclusions, limitations and future works

In recent years, there has been a common topic of conversation all over the world that has generated numerous interactions around it: the 2030 Agenda and its SDGs. This study analyzes and understands the digital conversation on Twitter around this phenomenon, with the aim of generating social knowledge extracted from the digital footprint. In this regard, the main findings of this research, which focuses on comparing the digital discussion of one of the SDGs (SDG 9: Industry, Innovation and Infrastructure) with the digital discussion around the SDGs as a whole, are set out below.

Our results show that the social interest in the digital sphere generated by SDG 9, related to industry, innovation and infrastructure, is lower and has a more reduced female presence than that generated by the other SDGs, which cover many social, economic and political areas. It is significant that, for example, discussion or interaction in relation to SDG 4 (quality education) is ten times that of SDG 9. In other words, although all the SDGs are intended to occupy the media, political and social agenda on equal terms, it has been demonstrated that in the digital public sphere there are notable differences in terms of importance and social relevance. Specifically, SDG 9 is the Sustainable Development Goal that has the least impact on the digital debate and the number of tweets about it has decreased in recent years. However, the digital conversation among the few actors involved does develop in a cohesive manner, sharing leadership and space.

These data, already with great scientific and social potential, could be a clear indication of what citizens' main priorities or concerns about SDGs were during the period analyzed (from 2016 to 2022). If the institutions driving this initiative had this information, they would have a better understanding of the evolution of their initiative and, in a short space of time, could make strategic decisions to influence the digital social conversation itself. In fact, it is not a matter of the SDGs competing to occupy more space on the social, media and institutional agendas, but rather of confirming a reality and balancing the space occupied by the different topics according to certain interests, so as not to definitively corner certain social issues.

In this sense, it is essential to obtain this type of information based on the digital footprint of interactions on social networks such as Twitter, because in this way ad hoc communication strategies could be designed for different institutional actions or, in this case, for the impulse and promotion of each SDGs.

To reinforce the aforementioned idea, this research has found other relevant data, such as the number of actors mobilized or involved in SDG 9—just 22,000 during the entire period analyzed. Moreover, the mobilized citizens belong to a very specialized and informed sector, a specific niche that runs the risk of being hidden in the digital public sphere, without transcendence in global or mass terms.

In this respect, the promoters of the initiative are faced with a relevant fact about the type of conversation that their initiative generates, involving only a specialized public with a strong interest in the subject. It is therefore up to them to reflect on possible measures to ensure that the topic has greater coverage, so that a larger part of the population feels involved, and thus reaching a wider audience.

How to overcome this eventuality if the UN's goal is to raise awareness among as much of society as possible? The study shows that there are exceptional peaks of digital activity where the initiative goes from "sidestream" to "mainstream" in terms of digital public debate, when certain celebrities (in this case, the Korean music group BTS)

formally interact with the initiative. In this regard, it is worth noting the importance of involving key players from the social, cultural or sports world to mobilize various social sectors.

On the other hand, the application of sentiment and emotion analysis models in digital conversations has great potential for understanding contemporary public debates. The results show that the SDGs do not generate controversy and there is no substantial difference in the distribution of sentiment and emotions between different periods and different SDGs.

Like any other research study, this one is also not without limitations. Like other social networks, Twitter is associated with a tendency toward digital endogamy and the creation of digital niches, known as echo chambers (González, 2011). Likewise, Twitter shows other limitations that are due to the bias of the data gathered, the representation bias when making general assumptions and the language employed by digital users (Zarrabeitia-Bilbao et al., 2023). However, in this case, through a relevant social media tool for research and understanding the public, it has been possible to achieve the aim of social listening and forecasting analysis of SDGs.

It would be interesting for future lines of research to go deeper into understanding of the social impact of SDGs in general and SDG 9 in particular, through, for example, the use of further social network analysis metrics (global and node level metrics) or other artificial neural networks models for understanding topics of discussion, among others.

Likewise, noting the positive effect, in terms of social impact, of involving relevant actors in society such as celebrities or influencers, future studies could focus on analyzing the positive and/or negative consequences of including this type of actors in this type of actions. In this respect, questions related to the viralization patterns of a digital conversation, the temporal length of the digital "boom" that these actors bring about, the residual engagement or secondary impacts of new actors in the conversation, among others, could be answered.

In addition, it would also be very interesting to integrate the analysis of how, who and what is being discussed about SDGs in languages other than English, studying interaction patterns according to the language used to intervene in the digital conversation.

Finally, it would also be worthwhile to carry out a deeper comparative study for different SDGs from a gender perspective.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare none.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- 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*.
- Bautista-Puig, N., & Dudek, J. (2019). Can Twitter hashtags be used for field delineation? The case of sustainable development goals (SDGs). In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *17th international conference on scientometrics and informetrics, ISSI 2019 - proceedings* (Vol. 2, pp. 2774–2775).
- 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>
- Burnap, P., & Williams, M. L. (2016). Us and them: Identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science*, 5(11), 1–15. <https://doi.org/10.1140/epjds/s13688-016-0072-6>
- Carrasco-Polaino, R., & Piqueiras Conlledo, P. (2023). Communicating the SDGs to create citizen awareness: an analysis of the activity of the public administration on Twitter. *Revista Mediterránea De Comunicación/mediterranean Journal of Communication*, 14(2), 69–81. <https://doi.org/10.14198/MEDCOM.24107>
- DataReportal. (2023). *Twitter statistics and trends*. <https://datareportal.com/essential-twitter-stats>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL HLT 2019 - 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies - proceedings of the conference*, 4171–4186. <https://doi.org/10.48550/arxiv.1810.04805>
- Franco-Riquelme, J. N., & Rubalcaba, L. (2021). Innovation and sdgs through social media analysis: Messages from fintech firms. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(165), 1–18. <https://doi.org/10.3390/joitmc7030165>
- Fritz, S., See, L., Carlson, T., Haklay, M., Oliver, J. L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L. A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., & West, S. (2019). Citizen science and the United Nations sustainable development goals. *Nature Sustainability*, 2, 922–930. <https://doi.org/10.1038/s41893-019-0390-3>
- Galiano-Coronil, A., Gil, M. O., Macías-Varela, B., & Ravina-Ripoll, R. (2023). An approach for analysing and segmenting messages about the SDGs on Twitter from the perspective of social marketing. *International Review on Public and Nonprofit Marketing*. <https://doi.org/10.1007/s12208-023-00369-3>
- Galiano-Coronil, A., & Ortega-Gil, M. (2019). Marketing social, indicadores de bienestar y ODS. Análisis de la cuenta oficial del gobierno de España @desdelamoncloa. *Retos*, 9(18), 219–238. <https://doi.org/10.17163/ret.n18.2019.03>
- Ghermandi, A., Langemeyer, J., Van Berkel, D., Calcagni, F., Depietri, Y., Egarter Vigl, L., Fox, N., Havinga, I., Jäger, H., Kaiser, N., Karasov, O., McPhearson, T., Podschun, S., Ruiz-Frau, A., Sinclair, M., Venohr, M., & Wood, S. A. (2023). Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use. *One Earth*, 6(3), 236–250. <https://doi.org/10.1016/j.oneear.2023.02.008>
- González, V. (2011). Educación para la ciudadanía democrática en la cultura digital. *Comunicar*, 18(36), 131–138. <https://doi.org/10.3916/C36-2011-03-04>
- Grover, P., Kumar Kar, A., & Gupta, S. (2021). Influence of political leaders on sustainable development goals-insights from twitter. *Journal of Enterprise Information Management*, 34(6), 1893–1916. <https://doi.org/10.1108/JEIM-07-2020-0304>
- Gutiérrez-Rubí, A. (2021, May 5). Nueva demoscopia para comprender lo complejo. *CincoDías*. https://cincodias.elpais.com/cincodias/2021/05/04/opinion/1620131187_017335.html
- Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). Social media: New technologies of collaboration. In D. L. Hansen, B. Shneiderman, & M. A. Smith (Eds.), *Analyzing social media networks with NodeXL. Insights from a connected world* (pp. 11–29). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-382229-1.00002-3>
- Herranz de la Casa, J. M., & García Caballero, S. (2021). Communication of sustainable development goals in social economy organizations. *CIRIEC-España, Revista De Economía Pública, Social y Cooperativa*, 101, 165–191. <https://doi.org/10.7203/CIRIEC-E.101.18393>
- Jaca-Madariaga, M., Zarrabeitia-Bilbao, E., Río-Belver, R. M., & Moens, M. F. (2023). Sentiment analysis model using Word2vec, Bi-LSTM and attention mechanism. In F. P. García Márquez, I. Segovia-Ramírez, P. J. Bernalte-Sánchez, & A. Muñoz-del-Río (Eds.), *Lecture notes on data*

- engineering and communications technologies* (Vol. 160, pp. 239–244). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-27915-7_43/COVER
- Li, X., Xie, Q., Jiang, J., Zhou, Y., & Huang, L. (2019). Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: The case of perovskite solar cell technology. *Technological Forecasting and Social Change*, 146, 687–705. <https://doi.org/10.1016/J.TECHFORE.2018.06.004>
- Madurai-Elavarasan, R., Pugazhendhi, R., Shafiqullah, G. M., Manoj-Kumar, N., Taufiqul-Arif, M., Jamal, T., Singh-Chopra, S., & Dyduch, J. (2022). Impacts of COVID-19 on sustainable development goals and effective approaches to maneuver them in the post-pandemic environment. *Environmental Science and Pollution Research*, 29, 33957–33987. <https://doi.org/10.1007/s11356-021-17793-9>
- Marzouki, A., Chouikh, A., Mellouli, S., & Haddad, R. (2021). From sustainable development goals to sustainable cities: A social media analysis for policy-making decision. *Sustainability*, 13(15), 8136. <https://doi.org/10.3390/SU13158136>
- Morales-i-Gras, J. (2020). *Datos masivos y minería de datos sociales: conceptos y herramientas básicas* (pp. 1–32). Universitat Oberta de Catalunya. <https://jordimorales.com/wp-content/uploads/2021/07/Datos-masivos.pdf>
- Mrvar, A., & Batagelj, V. (2016). Analysis and visualization of large networks with program package Pajek. *Complex Adaptive Systems Modeling*. <https://doi.org/10.1186/s40294-016-0017-8>
- OECD. (2016). *Better Policies for 2030. An OECD action plan on the sustainable development goals*. <https://www.oecd.org/dac/BetterPoliciesfor%202030.pdf>
- OpenODS. (2023). *¿Quién tiene la responsabilidad de cumplir la Agenda 2030? – OpenODS*. <https://openods.es/es/quien-tiene-la-responsabilidad-de-cumplir-la-agenda-2030/>
- OpenRefine. (2023). *A free, open source, powerful tool for working with messy data*. <https://openrefine.org/>
- Orbegozo-Terradillos, J., Morales-i-Gras, J., & Larrondo-Ureta, A. (2019). Feminisms outraged at justice: the online conversation on “La Manada” case. *Ic–revista Científica De Información y Comunicación*. <https://doi.org/10.12795/IC.2019.i01.08>
- Patuelli, A., & Saracco, F. (2023). Sustainable development goals as unifying narratives in large UK firms’ Twitter discussions. *Scientific Reports*, 13(7017), 1–17. <https://doi.org/10.1038/s41598-023-34024-y>
- Porcelli, A. M. (2020). The value of social networks in the progress of sustainable development goals. *Lex Social*, 10(1), 595–630. <https://doi.org/10.46661/lexsocial.5078>
- UN Press. (2015, September 25). *Unanimously adopting historic sustainable development goals, general assembly shapes global outlook for prosperity, peace*. <https://press.un.org/en/2015/ga11688.doc.htm>
- Roldán-Álvarez, D., Martínez-Martínez, F., Martín, E., & Haya, P. A. (2021). Understanding discussions of citizen science around sustainable development goals in Twitter. *IEEE Access*, 9, 144106–144120. <https://doi.org/10.1109/ACCESS.2021.3122086>
- Roser, R., & Ortiz-Ospina, M. (2018). *Measuring progress towards the Sustainable Development Goals*. SDG-Tracker. <https://sdg-tracker.org/>
- Rovira-Sancho, G., & Morales-i-Gras, J. (2022). Femitags for feminist connected crowds in Latin America and Spain. *Acta Psychologica*, 230, 103756. <https://doi.org/10.1016/J.ACTPSY.2022.103756>
- Saeta-Pérez, I. (2016). *gender-guesser - PyPI*. Python package index. <https://pypi.org/project/gender-guesser/>
- Savani, B. (2023). *Bert base uncased emotion - hugging face*. <https://huggingface.co/bhadresh-savani/bert-base-uncased-emotion>. Accessed 17 Aug 2023.
- Singh, P., Kaur, S., Baabdullah, A. M., Dwivedi, Y. K., Sharma, S., Sawhney, R. S., & Das, R. (2023). Is #SDG13 trending online? Insights from climate change discussions on Twitter. *Information Systems Frontiers*, 25, 1–21. <https://doi.org/10.1007/s10796-022-10348-4>
- Twarc. (2023). *Twarc2*. https://twarc-project.readthedocs.io/en/latest/twarc2_en_us/
- Twitter. (2023). *Twitter API for academic research*. <https://developer.twitter.com/en/products/twitter-api/academic-research>
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. <https://sdgs.un.org/2030agenda>
- United Nations. (2016). *Industria, innovación e infraestructura: Por qué es importante*. https://www.un.org/sustainabledevelopment/en/wp-content/uploads/sites/3/2016/10/9_Spanish_Why_it_Matters.pdf
- United Nations. (2021). *SDG moment 2021*. <https://www.un.org/sustainabledevelopment/sdg-moment-2021/>
- United Nations. (2023a). *SDG 9*. <https://sdgs.un.org/goals/goal9>

- United Nations. (2023b). *The 17 goals*. <https://sdgs.un.org/goals>
- VantagePoint. (2023). *The vantagepoint*. <https://www.thevantagepoint.com/>
- Vigdor, N. (2021). BTS took center stage at the U.N. Over one million fans watched live. *The new york times*. <https://www.nytimes.com/2021/09/20/world/asia/bts-un-performance.html>
- Yuan, H., Wang, X., Gao, L., Wang, T., Liu, B., Fang, D., & Gao, Y. (2023). Progress towards the sustainable development goals has been slowed by indirect effects of the COVID-19 pandemic. *Communications Earth & Environment*, 4(184), 1–13. <https://doi.org/10.1038/s43247-023-00846-x>
- Zarrabeitia-Bilbao, E., Jaca-Madariaga, M., Rio-Belver, R. M., & Álvarez-Meaza, I. (2023). Nuclear energy: Twitter data mining for social listening analysis. *Social Network Analysis and Mining*, 13, 1–17. <https://doi.org/10.1007/s13278-023-01033-8>
- Zarrabeitia-Bilbao, E., Rio-Belver, R.-M., Alvarez-Meaza, I., & Martínez de Alegría-Mancisidor, I. (2022). World Environment day: Understanding environmental programs impact on society using twitter data mining. *Social Indicators Research*, 164, 263–284. <https://doi.org/10.1007/s11205-022-02957-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.