The impact of science communication on Twitter: The case of Neil deGrasse Tyson

El impacto del discurso científico en Twitter: El caso de Neil deGrasse Tyson

Elena Denia, Predoctoral Fellow in Science Communication, Institute of Innovation and Knowledge Management (INGENIO), CSIC/Polytechnic University of Valencia (Spain) (elenadeniacc@gmail.com) (https://orcid.org/0000-0002-9187-9051)

ABSTRACT

Public perceptions of science have been studied extensively since the mid-twentieth century. The aim of this project is to explore the interaction between science and the public in the digital world as a complement to traditional studies on the societal impact of science, particularly on the social network Twitter. It thus proposes a low-cost, easily reproducible methodology involving the design of an algorithm operating on representative sets of tweets to analyse their content by using computational techniques of data mining and natural language processing. To test this methodology, I analyse the communications of the popular science communicator Neil DeGrasse Tyson. The impact of the information is calculated in terms of 1) likes and retweets; 2) suggested formulas for measuring the popularity and controversial nature of the content; and 3) the semantic network. Relevant elements of the communications are then identified and classified according to the categories of "science", "culture", "political-social", "beliefs", "media" and "emotional". The results reveal that content with an emotional charge in the communicator's message triggers a substantially more profound response from the public, as do references to socio-political issues. Moreover, numerous concepts peripheral to the scientific discussion arouse more interest than the concepts central to the communication. Both these results suggest that science is more interesting when it is linked to other issues.

RESUMEN

La percepción social de la ciencia se ha estudiado ampliamente desde mediados del siglo XX. El presente proyecto pretende abordar la interacción ciencia-público en el marco de la vida digital para complementar los estudios clásicos sobre impacto social de la ciencia, en particular en la red social Twitter. Se presenta así una propuesta metodológica con el diseño de un algoritmo que opera sobre conjuntos representativos de tweets para analizar su contenido utilizando técnicas computacionales de minería de datos y procesamiento del lenguaje natural, fácilmente reproducible por otros investigadores y de bajo coste. Para probar la herramienta, se analiza el discurso del popular divulgador Neil DeGrasse Tyson. El impacto de la información se calcula en términos de: 1) likes y retuit; 2) medidas sugeridas para la popularidad y el grado de contenido polémico; y 3) la red semántica. Tras identificar y clasificar los elementos relevantes del discurso por las categorías «ciencia», «cultura», «político-social», «creencias», «medios» y «emocional», los resultados revelan que una transmisión con carga emocional en el mensaje del divulgador despierta una respuesta sustancialmente más profunda en el público, así como la alusión a cuestiones socio-políticas. Además, numerosos conceptos periféricos a la discusión científica suscitan mayor interés que los propios centrales en el discurso. Ambos resultados sugieren que la ciencia interesa en mayor medida cuando va ligada a otros aspectos.

KEYWORDS | PALABRAS CLAVE

Twitter, communication, science, dissemination, impact, public, participation, computational analysis.
Twitter, comunicación, ciencia, divulgación, impacto, público, participación, análisis computacional.

1. Introduction

Public perceptions of science have been studied extensively since the mid-twentieth century by means of public opinion polls like the Eurobarometers in Europe, the National Science Foundation (NSF) surveys in the United States, or the reports of the Spanish Foundation for Science and Technology (FECYT) in Spain, among others. Ever since they were introduced, these surveys have sought to measure public interest in, knowledge of, and attitudes towards science (Davis, 1958), although they have not been immune to criticism (Bauer et al., 2007; Pardo, 2001). Science communication in particular has often been presented as an essential strategy for fostering public engagement with science (European Commission, 2008).

Given that the current trend in science communication is to involve the public in scientific dialogue (Nisbet & Scheufele, 2009), the performance of science communicators on social networks, which have massive numbers of user accounts and high levels of participation, has become a question of particular interest. Academics have highlighted the need for a better understanding of how these new virtual environments affect science communication (Brossard & Scheufele, 2013). At the same time, social networks themselves offer the opportunity for close-range investigation of public debates about scientific issues, with special attention to new voices and different contexts (Kapoor et al., 2018; Shan et al., 2014), on media platforms where content is often accessed without mediators. It is for this reason that studies of social networks can constitute a useful complement to traditional surveys of public perceptions of science (Li et al., 2019). With respect to public participation, researchers have highlighted the need to assess public involvement in open discussions about science in these environments, considering aspects like ease of access to content, the type of information disseminated and even the type of audience, among other factors (López-Pérez & Olvera-Lobo, 2019). It has been suggested, however, that audiences that follow science accounts do not generally interact with them, as in most cases they use them merely to keep updated (Álvarez-Bornstein & Montesi, 2019). It is worth noting that the content generated every day on Twitter is intimately linked to current scientific developments (Veltri, 2013; Wilkinson & Thelwall, 2012; Zhao et al., 2011), a quality that makes it especially attractive for research on science communication (Büchi, 2016), and that constitutes one of the most powerful reasons for its selection for this study, along with the open access it offers to bulk data. On the other hand, it has been suggested that the best strategy for communication on Twitter is actively focusing on increasing followers rather than relying on keyword searches to make scientific content visible (Mohammadi et al., 2018). With this in mind, to test the methodology proposed in this article I have chosen to use the Twitter account of the popular science communicator Neil de Grasse Tyson, who has gathered a very large number of followers (more than 13 million in 2019).

The objective here is to present the possibilities offered by a methodology designed to analyse sets of tweets and assess the impact of communications on Twitter about scientific issues, with the aim of revealing unexplored dimensions of the public interest in science and attempting to answer a question that has been inspired by the results of traditional surveys: What triggers an interest in science? Scientific advances and discoveries or aspects associated with everyday human life, like cultural, political, or even emotional factors? Attempting to answer this question will help reveal trends that indicate the most effective ways of communicating science. My starting hypothesis is that the social network Twitter, understood as a space for public participation, could prove useful for learning about these issues. To address the questions outlined above, this study includes the following stages: (1) development of an algorithm in Swift programming language that can analyse large sets of data and measure the degree of interest and impact of information released on Twitter based on conversations about science; (2) application of the algorithm to a set of publicly available data extracted from the account of the astrophysicist Neil de Grasse Tyson; (3) graphic representation and interpretation of results; and (4) assessment of the scope of the study.

It should be noted first of all that there is very little consensus about what methods are reliable for research on Twitter and what information it can reveal to us (Veltri & Atasanova, 2015), since rigorous methodologies that would permit reliable systematic analysis have yet to be developed (Kahle et al., 2016). However, there are various exploratory studies of science communication on Twitter in specific areas; for example, the studies that have aroused the most academic attention are those related to public

perceptions of risk, like research on the climate debate (Pearce et al., 2014) or studies of health issues aimed at understanding the emotional stance of the public (Becker et al., 2016). To study the content of tweets, computational techniques are used in order to systematically analyse large volumes of data, using text mining, natural language processes, network analysis, etc., and qualitative approaches with the participation of a human encoder who is aware of the conceptual context of the communication being researched and is therefore able to draw subtle information out of the tweets (Uren & Dadzie, 2015). This study therefore combines a mixed methodology to leverage the potential of both types of analysis.

It is worth noting that due to the format of Twitter, users express themselves in very brief terms, carefully choosing relevant words to reflect their ideas, which in principle makes it easier to explore key elements based on frequently used words. In this sense, the content of tweets allows for a semantic representation (Narr et al., 2011) that facilitates the analysis both of elements central to the communication, and of peripheral elements in simple terms, assigning them levels of relevance. Another common approach is what is known as "sentiment analysis" applied to tweet content, and in this study I also attempt to measure the emotional charge of the communications, given that emotional messages on Twitter are more likely to be retweeted (Stieglitz & Dang-Xuan, 2013; Veltri & Atanasova, 2015) and reflect the emotional perceptions of users expressed in natural language (Dehkharghani et al., 2014), generally identified based on predetermined lists of words.

Finally, it is also important to acknowledge the studies that suggest users are more inclined to tweet about their personal daily activities than about informational publications (80% compared to 20%) (Dann, 2010; Naaman et al., 2010). Twitter is, after all, both an information network and a social network (Myers et al., 2014), and its use as a research tool can take a wide range of approaches, from studying the potential to engage audiences by retweets (Kwak et al., 2010) to the activity of scientific journalists on the platform (Arrabal & De-Aguilera, 2016), or even the role of teachers in motivating and engaging their students (Santoveña & Bernal, 2019). There is also a widely accepted idea that personal profiles work better than institutional accounts in terms of interactions with the public (Pérez-Rodríguez et al., 2018).

2. Material and methods

2.1. Methodological proposal

For this study I propose a methodology aimed at identifying trends among audiences of science communication through an analysis of communications that are openly available on Twitter. The set of tweets to be studied can be collected easily through the application programming interface (API) offered by the platform itself (Twitter, 2019), for example, using a simple algorithm programmed in R language. In this case, it is recommended to clean the text of the messages with the tidytext R package (Silge & Robinson, 2016), which processes the text and prepares it for analysis, eliminating words with no semantic value like connectors or stopwords. Then, to analyse the content of representative sets of tweets, an algorithm in Swift language has been designed. The functions of this algorithm are described in this section. Quantitative techniques for performing systematic analysis are combined with a qualitative approach for classification of terms into categories.

First of all, to determine what the content in the sample is about and how it can be quantified in order to estimate its impact, the basic, indivisible unit of analysis used in this methodology is the word or term (Blei et al., 2003). The measurements to estimate impact are based on its frequency of use and on the number of retweets and likes associated with it, appropriately standardised based on the number of times the word appears in the sample. In addition, to reveal the extent to which a particular type of information is more interesting to the receiving audience than another, two other coefficients are proposed: "popularity" and "polemicity". The idea underlying the popularity indicator is that the more a word is retweeted (i.e., the more retweets it accumulates) the more popular it is considered, also taking into account how often it appears in the sample. For example, an unusual word with a lot of retweets would be considered especially popular. To define this, the variable retweetRate (retweet ratio) has been used, providing a measurement of the interest aroused by the term, i.e., the extent to which it is shared. It is important to note that popularity may be either positive or negative, as content that a given user disapproves of may be retweeted and thus gain visibility. On the other hand, polemicity, understood as the degree of

controversy triggered by the information, is defined as the ratio of the retweetRate to the favoriteRate (like ratio), which will be higher if the content of the tweet is retweeted widely but receives fewer likes. The variables described above and the popularity and polemicity indicators for words and categories are presented in Table 1, being comparable by definition between different datasets.

Table 1. Variables integrated into the algorithm		
Variable	**Definition**	**Equation**
Frequency (word)	Number of appearances of a word divided by the total number of classified words present in the user's communications	$\text{freq}(w_i) = \frac{\text{wordCount}(w_i)}{\sum_j^k \text{wordCount}(w_j)}$
favoriteRate	Number of likes received by a particular word divided by the number of tweets that word contains.	$\text{favoriteRate}(w_i) = \frac{\text{favoriteCount}(w_i)}{\#\text{tweetsContaining}(w_i)}$
retweetRate	Number of retweets received by a particular word divided by the number of tweets containing that word.	$\text{retweetRate}(w_i) = \frac{\text{retweetCount}(w_i)}{\#\text{tweetsContaining}(w_i)}$
Popularity (word)	Number of retweets (weighted) divided by the frequency of a particular word.	$\text{popularity}(w_i) = \frac{\text{retweetRate}(w_i)}{\text{frequency}(w_i)}$
Polemicity (word)	Number of retweets (weighted) divided by the number of favourites (weighted) of a particular word.	$\text{polemicity}(w_i) = \frac{\text{retweetRate}(w_i)}{\text{favRate}(w_i)}$
Popularity (category)	Number of retweets (weighted) divided by the frequency of a particular category.	$\text{pop}_{cat} = \sum_{cat} \frac{\text{wordCount}(w_i) * \text{pop}(w_i)}{\text{wordCount}(cat)}$
Polemicity (category)	Number of retweets (weighted) divided by the number of favourites (weighted) of a particular category.	$\text{pol}_{cat} = \sum_{cat} \frac{\text{wordCount}(w_i) * \text{pol}(w_i)}{\text{wordCount}(cat)}$

Note. Where w_i is a word or term in the sample and cat is a category.

On the other hand, to determine the level of interest triggered by scientific content either on its own or correlated with other types of information in the tweet, a classification of words by category is proposed. This method has been used before in other studies investigating the relative interest in science compared to other topics on Twitter (Zhao et al., 2011). For this study, the categories proposed (which are described in Table 2) are the following: "science", "culture", "political-social", "beliefs", "media", and "emotional". These categories have been defined using my own criteria inspired by previous studies. Specifically, the category "culture" is drawn from proposals that explore cultural factors in studies of public perceptions of science (Bauer et al., 2012; Pardo, 2001); "political-social" is based on Twitter studies that highlight concerns of this type in relation to scientific controversies, usually in the area of climate change (Pearce et al., 2014); "beliefs" refers to the frequent interactions between science and religion and to the growing interest in studies of pseudoscience (Moreno-Castro et al., 2019); "media" is included due to its relevance to communication studies; and "emotional" is derived from studies that apply sentiment analysis to tweets, along with the observation that "an emotional connection [...] can be a powerful 'way in' to a science experience for non-experts, capturing initial attention and increasing feelings of bonding with the communicator or other participants" (Kaiser, 2014: 28).

Table 2. Description of categories for analysis of communications on Twitter	
Category	**Description**
Science	Words that refer to scientific issues, including: scientific concepts ("planet", "physics"), methodological terms ("deduction"), and scientists (both historical, like "Galileo", and contemporary, like "bgreene").
Emotion	Words containing an emotional charge, expressing feelings, including: emotions ("happy", "fear", "awesome"), expressions ("yup", "yeah"), and evaluations ("lovely").
Culture	Words referring to culture, including: the cultural industry ("film", "song"), popular figures ("ladygaga") and cultural activities ("football").
Political-social	Words referring to political content and sociological aspects ("electorate", "president", "Americans") as well as social indicators ("poverty").
Media	Words referring to different types of media, including: social networks ("Facebook") and mass media ("audience", "news").
Beliefs	Words referring to beliefs, including: religion ("faith"), pseudoscience, esotericism and ufology ("homoeopathic", "astrology").

The classification of words into categories is done manually based on their meaning. Computer-assisted research methods are not suited to this type of content analysis as such methods assume that the terms have the same meaning in any context (Matthes & Kohring, 2008), while the use of a human

encoder can ensure a better interpretation of the context of the discussion. Thanks to the algorithm design, this categorisation is relatively simple when the words are organised in order of relevance in a list based on frequency of use and cumulative likes and retweets. At the same time, the algorithm also generates two files with nodes and edges organised in a visual representation in the form of a semantic network (for example, using the popular software Gephi) that shows the relationships between the metrics proposed and the content of the tweets collected, such as depicting the weight of each category in comparison with the others, or showing whether the scientific concepts present in the data sample are central to the communication or merely peripheral. The tasks required for the analysis are detailed in Table 3.

Table 3. Tasks proposed for analysis of a sample of tweets		
Tool	Task	Description
Algorithm in Swift	Statistics on terms	Various statistical parameters are defined based on frequency of use, retweets and likes: frequency divided by the total number of words; retweet ratios and like ratios (retweetRate and favoriteRate, respectively); and popularity and polemicity indicators.
	List of words	A CSV file is generated containing a list of words in order of relevance in the set of tweets using an approximate measurement calculated based on weighted variables: frequency, retweetRate and favoriteRate.
	Word associations	Word matches in the tweets are identified with an automated system.
	Depiction of semantic network	The data is prepared for a visual depiction of the matches between words in the tweets to assess the degree of centrality of the concepts that proved the most appealing.
Human encoder	Categorisation of words	Terms are classified manually into categories ("science", "culture", political-social", "beliefs", "media", and "emotional") in the CSV file with the word list.

It is important to note that the list of categorised words is cumulative, which means that it will always grow in size and thus favour subsequent analysis applied to new sets of tweets, for which much of the work will already be done, apart from the classification of the most relevant words in the new communication by executing the algorithm. This will provide increasingly refined analyses of science communications, easily reproducible by other researchers and at a low cost in terms of intellectual and financial resources compared to public opinion polls, which require extensive work and substantial funding, and which have static results. Of course, this algorithm design is not restricted to the area of science, as its variables are compatible with other conceptual spaces. This tool is available to any researchers who may need it, simply by sending me a request via email.

2.2. Reliability of categorisation

To validate the categorisation of words, six members of the ScienceFlows research group (Science-Flows, 2019) worked independently to classify the set of the first 50 relevant words identified by the algorithm in the data sample (see Section 3.1) into the categories created. For each word, the degree of accuracy was calculated based on my classification, with any score above the minimum threshold of 75% considered valid. A reliability level of 82% was estimated for the manual classification, although it serves merely as an indicative estimate for exploring certain trends in the communications under study.

2.3. Limitations of Twitter for data extraction

Although Twitter's API is free and facilitates access to millions of tweets including metadata, each search returns around 3,000 tweets, which constitute random samples from larger datasets, and thus there are limitations on the information that researchers can collect. For keyword searches, it needs to be borne in mind that the sample returned by the API is from the last nine days, which means the performance of cross-sectional studies on this platform requires the implementation of a mechanism for systematic data extraction in real time to obtain samples covering longer periods. On the other hand, when tweets by specific users are collected, the random sample includes tweets dating back to when the profile was first created.

Another obvious weakness in the use of Twitter is the superficial distinction between countries offered by the tool, resulting in insufficient demographic data and a homogeneous audience description. In addition, some authors point out that social network users are only one subset of the general public and

should therefore not be taken as representative (Murphy et al., 2014). Nevertheless, it is clear that as internet access increases around the world, so too grows the number of people participating in social networks. In any event, as the objective of this study is to explore trends, this limitation is not a serious problem in this case.

### 3. Analysis and results: The impact of Neil deGrasse Tyson's public communications
3.1. Data

To test the tool, a set of tweets was extracted using the API for the account of a specific user, Neil deGrasse Tyson ("@neiltyson"). The resulting file contains the full text of the messages together with a series of properties stored in columns with their associated values, of which the number of likes and the number of retweets are of special relevance to the analysis of the impact of the information. The resulting sample, which only includes tweets written by the user (no retweets), contains 3,005 tweets posted between 2012-10-05 and 2019-06-19. This download represents 49.5% of tweets posted by Tyson on the dates of the search, out of a total of 6,974 tweets. After cleaning the text, a file with 24,484 relevant words and their associated statistics was obtained.

3.2. Analysis

After identifying the most relevant words with the algorithm, the top 1,250 terms were classified manually. These offer an idea of the thematic preferences of the user in question. Among the most frequently used words are scientific concepts like "Mars", "space", and "physics". This result is to be expected given that the account holder is a science communicator who works in the field of astrophysics. Also appearing frequently are concepts unrelated to science itself, such as the word "film", making reference to the cultural industry of cinema, and the word "happy", referring to an emotional state.

Figure 1 contains a series of pie charts graphically depicting the relative importance of each category in Tyson's communications, based on: 1) his own thematic preferences; 2) cumulative likes by category; and 3) cumulative retweets by category.

Figure 1. Communication preferences, cumulative retweets and likes by category, popularity and polemicity in communications by @neiltyson
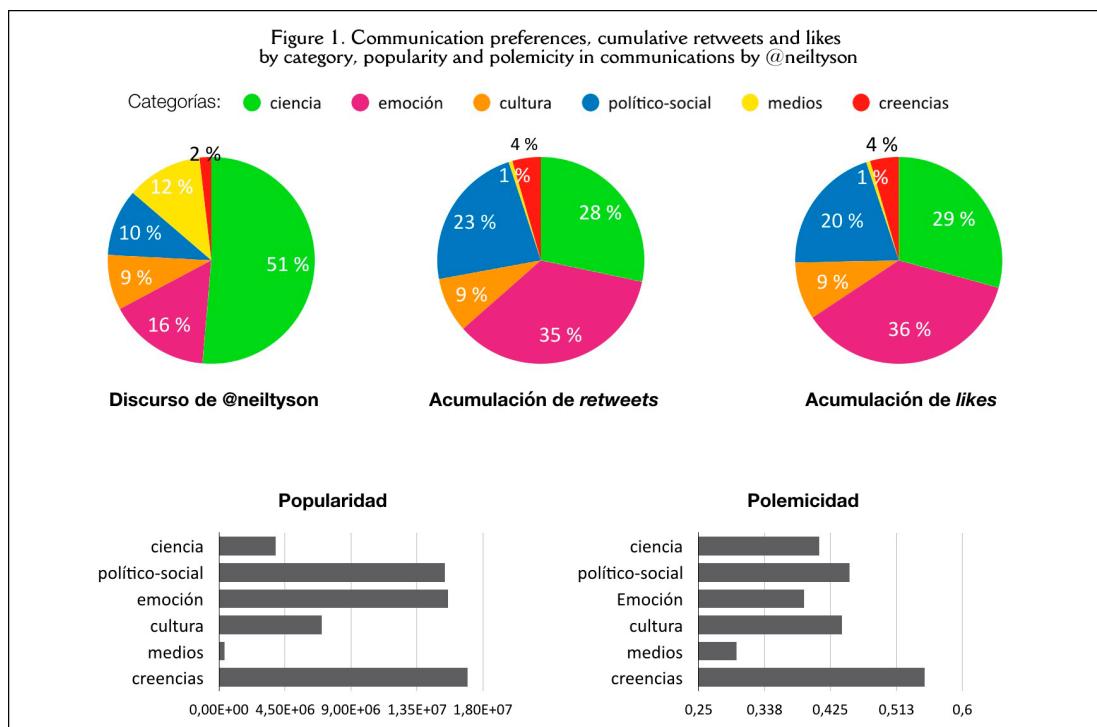
In the first chart, it is clear that the frequency of words with scientific content is high in Tyson's communications, making up more than half of the sample categorised (51%). Another 16% of his communications contain emotional expressions, 12% are words referring to the media (for example, citing his own radio program "StarTalk"), 8% are words with political and/or social content, 9% are words related to cultural issues, and 2% are related to beliefs.

To estimate the impact of his communications, the other two pie charts show how these thematic preferences trigger different quantities of likes and retweets for each category. Here we find that the impact of words containing an emotional charge (e.g. "joy", "shit", "hostile" or "cry") is markedly bigger than their representation in Tyson's communications would suggest. Words referring to social or political issues also have a much bigger impact, while the impact of scientific terms is smaller. On the other hand, words containing cultural information and those referring to beliefs have an impact in proportion with their relative presence in his communications. The percentages for cumulative likes are: "science": 28%; "emotion": 35%' "culture": 9%; "political-social": 23%; "media": 1%; "beliefs": 4%; while the percentages for cumulative retweets are: "science": 29%; "emotion": 36%' "culture": 9%; "political-social": 20%; "media": 1%; "beliefs": 4%.

Finally, in order to gauge how popular or controversial the content of Tyson's communications is, the popularity and polemicity coefficients were calculated for the different categories. These are also represented in Figure 1 with bar graphs (using comparative measurements whose values are not meaningful in themselves). The most popular communications are those related to beliefs, political or social issues and emotions, while the most unpopular are those related to media. Similarly, the level of polemical content in the communications is clearly higher in the category of beliefs than in any other category.

The final step was to find out what happens when scientific information is combined with other types of information, which was explored by representing the semantic network of Tyson's communications, also classified into categories, showing the relationships and connections between words. This depiction can be found in Figure 2, where the size of the nodes represents: 1) the communication preferences of the communicator (a node is bigger when a particular term appears more often in the sample); 2) the terms that received the most retweets; and 3) the terms that received the most likes. As can be seen, Tyson constructs his communication through central scientific concepts (network on the left), but it is evident in the other two conceptual network depictions that the impact of information on users is bigger for terms that are peripheral to the communication, especially for the categories of "emotion" and "political-social".

Figure 2. Semantic network of communications by @neiltyson classified by categories
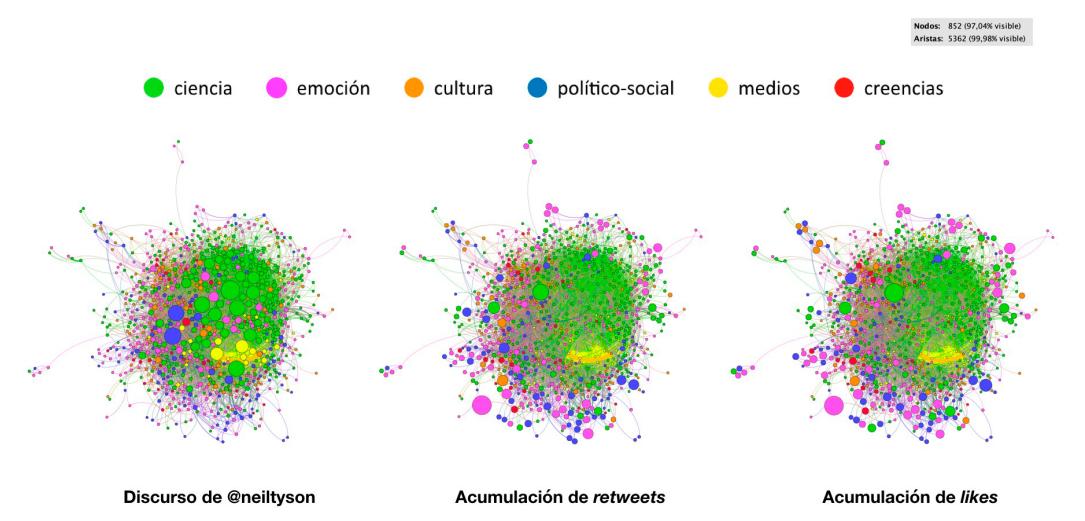
Note. For this representation the Gephi program has been used, applying the Force Atlas algorithm to a sample of 852 nodes and 5,362 edges. In the network, the nodes repel each other and the edges attract each other.

4. Discussion and conclusions

The results discussed in the previous section have two dimensions: the strictly methodological dimension and the dimension related to the case chosen for this study. Based on the analysis of the case study with the tool proposed, the most striking result is that scientific content gives way to other types of information in the collected set of tweets by the science communicator Neil deGrasse Tyson. Although Tyson's communications are dominated by words associated with scientific content, the tweets with an emotional charge are the ones that receive the most attention from users, in terms both of likes and of retweets, followed by words referring to social and political content. Conversely, the media category is clearly the one with the weakest result in terms of the interest it arouses, and it also has the lowest popularity rate, despite the fact that its presence in Tyson's communications is quite high. This phenomenon is left open to interpretation.

In light of these results, this study constitutes a data-based approach that suggests that science on its own is not as interesting to the public as other subjects. Although it only studies one specific case, it is worth highlighting that Tyson is a science communicator with a huge influence and a large number of followers, and the sample used for this study contains half of his tweets. It should also be noted that messages posted by Tyson go first to his followers, an audience supposedly interested in science (although they are also subsequently disseminated via retweets).

The values generated for the two proposed indicators of popularity and polemicity reveal some clear differences between the two in the individual categories that may constitute points of interest. For example, the "science" category, despite not being particularly popular with receiving users, exhibits a certain level of polemical content, perhaps because of Tyson's social activism on issues like the climate crisis.

While the idea of polemicity is based on identifying which topics generate the most debate or controversy, popularity is a measurement of the attention they receive, regardless of whether that attention is positive or negative. The aspects revealed by these calculations might otherwise go unnoticed if there is a low incidence of words used in the sample in a particular category; this is the case, for example, of "beliefs", which proves an extremely popular and controversial category, despite its minimal presence in the set studied. Presumably due to their sensitive nature, these are issues that do not predominate in the communications of the user studied but that have a big impact in the Twitter ecosystem. It should also be noted that while the results based on likes and retweets are quite somewhat similar, an interpretation for this that seems plausible is that the action of retweeting gives the message greater visibility, thereby increasing the potential number of likes.

With respect to word matches in the semantic network, represented to evaluate the levels of centrality of terms with differing degrees of appeal, it is clear for the sample examined that concepts peripheral to scientific discussion, i.e., referring to adjacent issues, are of more interest than central concepts, and are for the most part non-scientific. One possible interpretation for this is that there are particular subjects that suddenly attract attention, but they are not subjects that normally appear in Tyson's communications. This question requires more in-depth analysis to identify the causes. On the other hand, the similarity between the two semantic networks depicting the cumulative number of likes and retweets in nodes may again be related to the fact that the spreading of a tweet on Twitter favours the accumulation of likes by giving greater visibility to it and exposing it to other users.

This case study thus hints at specific strategies for strengthening science communication: linking scientific information to socio-political issues and/or expressing it in emotional terms. Of course, as this study is limited to the particular case of Tyson's tweets, the results obtained cannot be extrapolated to the whole study universe of science communication, and my intention is certainly not to make such a generalisation, but merely to point to a trend that should be researched in greater depth.

Given that the results of applying the tool proposed here to the particular case of a famous science communicator are reasonably consistent, it is worth asking: Would studies of the communications of other science communicators offer similar results? What differences would be identified in the case of institutional accounts, or if current scientific issues are the focus instead of users? It would thus be worthwhile to apply this tool to other specific Twitter profiles and to general discussions that receive significant media attention. Another incentive for further research is the potential of the tool to provide

comparable periodic assessments and to support governments in the preparation of specific scientific communication plans or similar initiatives (to offer one example).

In short, due to the limitations of surveys to gauge public perceptions of science (Bauer et al., 2012; Pardo, 2001), this study has effectively considered a number of unexplored areas, confirming the potential of research on social networks to complement such surveys (Li et al., 2013). In particular, the confirmation with empirical data of the effect of emotional content in scientific communication is especially noteworthy (Kaiser, 2014).

Funding Agency

References
Alvarez-Bornstein, B., & Montesi, M. (2019). Who is interacting with researchers on Twitter? A survey in the field of Information Science. *JLIS*, *10*(2), 87-106. https://doi.org/10.4403/jlis.it-12530

Arrabal, G., & De-Aguilera, M. (2016). Comunicar en 140 caracteres. Cómo usan Twitter los comunicadores en España. [Communicating in 140 characters. How journalists in Spain use Twitter]. *Comunicar*, *46*, 9-17. https://doi.org/10.3916/C46-2016-01

Bauer, M.W., Allum, N., & Miller, S. (2007). What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understanding of Science*, *16*(1), 79-95. https://doi.org/10.1177/0963662506071287

Becker, B.F., Larson, H.J., Bonhoeffer, J., van Mulligen, E.M., Kors, J.A., & Sturkenboom, M.C. (2016). Evaluation of a multinational, multilingual vaccine debate on Twitter. *Vaccine*, *34*(50), 6166-6171. https://doi.org/10.1016/j.vaccine.2016.11.007

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022. https://bit.ly/2wQLaGj

Brossard, D., & Scheufele, D.A. (2013). Science, New Media, and the Public. *Science*, *339*(6115), 40-41. https://doi.org/10.1126/science.1232329, https://dx.doi.org/10.1126/science.1232329

Büchi, M. (2017). Microblogging as an extension of science reporting. *Public Understanding of Science*, *26*(8), 953-968. https://doi.org/10.1177/0963662516657794

Dann, S. (2010). Twitter content classification. *First Monday*, (12), 15-15. https://doi.org/10.5210/fm.v15i12.2745

Davis, R.C. (1958). *The public impact of science in the mass media*. Institute for Social Research, University of Michigan. https://stanford.io/2w9teGk

Dehkharghani, R., Mercan, H., Javeed, A., & Saygin, Y. (2014). Sentimental causal rule discovery from Twitter. *Expert Systems with Applications*, *41*(10), 4950-4958. https://doi.org/10.1016/j.eswa.2014.02.024

European Commission (Ed.) (2008). Public engagement in science. Publications Office of the European Union. https://bit.ly/2uB98Vg

Kahle, K., Sharon, A.J., & Baram-Tsabari, A. (2016). Footprints of fascination: Digital traces of public engagement with particle physics on CERN's social media platforms. *PLoS One*, *11*(5), e0156409. https://doi.org/10.1371/journal.pone.0156409

Kaiser, D., Durant, J., Levenson, T., Wiehe, B., & Linett, P. (2014). *The evolving culture of science engagement: an exploratory initiative*. MIT & Culture Kettle. https://bit.ly/2Wjy5hG

Kapoor, K.K., Tamilmani, K., Rana, N.P., Patil, P., Dwivedi, Y.K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers*, *20*(3), 531-558. https://doi.org/10.1007/s10796-017-9810-y

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? . In Rappa, M., & Jones, P. (Eds.), *Proceedings of the 19th International Conference on World Wide Web* (pp. 591-600). ACM. https://doi.org/10.1145/1772690.1772751

Li, R., Crowe, J., Leifer, D., Zou, L., & Schoof, J. (2019). Beyond big data: Social media challenges and opportunities for understanding social perception of energy. *Energy Research & Social Science*, *56*, 101217. https://doi.org/10.1016/j.erss.2019.101217

López-Pérez, L., & Olvera-Lobo, M.D. (2019). Participación digital del público en la ciencia de excelencia española: Análisis de los proyectos financiados por el European Research Council. *El Profesional de la Información*, *28*, 1-10. https://doi.org/10.3145/epi.2019.ene.06

Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, *58*(2), 258-279. https://doi.org/10.1111/j.1460-2466.2008.00384.x

Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K.L. (2018). Academic information on Twitter: A user survey. *PLoS One*, *13*(5), e0197265. https://doi.org/10.1371/journal.pone.0197265

Moreno-Castro, C., Corell-Doménech, M., & Camano-Puig, R. (2019). Which has more influence on perception of pseudo-therapies: The media's information, friends or acquaintances opinion. *Communication & Society*, *32*, 35-49. https://doi.org/10.15581/003.32.3.35-48

Murphy, J., Hill, C., & Dean, E. (2013). *Social media, sociality, and survey research*. John Wiley & Sons. https://doi.org/10.1002/9781118751534.ch1

Murphy, J., Link, M.W., Childs, J.H., Tesfaye, C.L., Dean, E., ... Harwood, P. (2014). Social media in public opinion research: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly, 78*(4), 788-794. https://doi.org/10.1093/poq/nfu053

Myers, S.A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network? the structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 493-498). ACM. https://doi.org/10.1145/2567948.2576939

Naaman, M., Boase, J., & Lai, C.H. (2010). Is it really about me? message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 189-192). ACM. https://doi.org/10.1145/1718918.1718953

Narr, S., Luca, E.W.D., & Albayrak, S. (2011). Extracting semantic annotations from twitter. In *Proceedings of the fourth workshop on Exploiting semantic annotations in information retrieval* (pp. 15-16). ACM. https://doi.org/10.1145/2064713.2064723

Nisbet, M.C., & Scheufele, D.A. (2009). What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany, 96*(10), 1767-1778. https://doi.org/10.3732/ajb.0900041

Pardo, R. (2001). La cultura científico-tecnológica de las sociedades de la modernidad tardía. *Treballs de la Societat Catalana de Biologia, 51*, 35-63. https://bit.ly/2T0n8B5

Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate Change on Twitter: Topics, communities and conversations about the 2013 IPCC working group 1 report. *PLoS One, 9*(4). https://doi.org/10.1371/journal.pone.0094785

Pérez-Rodríguez, A.V., González-Pedraz, C., & Berrocal, J.L.A. (2018). Twitter como herramienta de comunicación científica en España. Principales agentes y redes de comunicación. *Communication Papers, 7*(13), 95-95. https://doi.org/10.33115/udg_bib/cp.v7i13.21986

Santoveña, S., & Bernal, C. (2019). Explorando la influencia del docente: Participación social en Twitter y percepción Académica. [Exploring the influence of the teacher: Social participation on Twitter and academic perception]. *Comunicar, 58*, 75-84. https://doi.org/10.3916/C58-2019-07

ScienceFlows (Ed.) (2019). Scienceflows. https://bit.ly/2wGZ8dB

Shan, L., Áine Regan, Brún, A.D., Barnett, J., van der Sanden, M.C.A., ... Áine McConnon (2014). Food crisis coverage by social and traditional media: A case study of the 2008 Irish dioxin crisis. *Public Understanding of Science, 23*(8), 911-928. https://doi.org/10.1177/0963662512472315

Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R. *The Journal of Open Source Software, 1*(3), 37-37. https://doi.org/10.21105/joss.00037

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media: Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217-248. https://doi.org/10.2753/mis0742-1222290408

Twitter (Ed.) (2019). Application programming interface. https://developer.twitter.com

Uren, V., & Dadzie, A.S. (2015). Public science communication on Twitter: A visual analytic approach. *Aslib Journal of Information Management, 67*(3), 337-355. https://doi.org/10.1108/ajim-10-2014-0137

Veltri, G. (2013). Microblogging and nanotweets: Nanotechnology on Twitter. *Public Understanding of Science, 22*(7), 832-849. https://doi.org/10.1177/0963662512463510

Veltri, G.A., & Atanasova, D. (2017). Climate change on Twitter: Content, media ecology and information sharing behaviour. *Public Understanding of Science, 26*(6), 721-737. https://doi.org/10.1177/0963662515613702

Wilkinson, D., & Thelwall, M. (2012). Trending Twitter topics in English: An international comparison. *Journal of the American Society for Information Science and Technology, 63*(8), 1631-1646. https://doi.org/10.1002/asi.22713

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., ... Li, X. (2011). Comparing Twitter and traditional media using topic models. In *Lecture Notes in Computer Science: Vol 6611. Advances in Information Retrieval*, volume 6611 (pp. 338-349). Springer. https://doi.org/10.1007/978-3-642-20161-5_34