Social Media Diffusion by Spanish Public Research Institutions

Ángel Hernández-García
Departamento de Ingeniería de Organización, Administración de Empresas y Estadística. ETSI de Telecomunicación. Universidad Politécnica de Madrid.
Av. Complutense 30, 28040 Madrid
Spain
angel.hernandez@upm.es

Emiliano Acquila-Natale
Departamento de Ingeniería de Organización, Administración de Empresas y Estadística. ETSI de Telecomunicación. Universidad Politécnica de Madrid.
Av. Complutense 30, 28040 Madrid
Spain
emiliano.acquila@upm.es

Julián Chaparro-Peláez
Departamento de Ingeniería de Organización, Administración de Empresas y Estadística. ETSI de Telecomunicación. Universidad Politécnica de Madrid.
Av. Complutense 30, 28040 Madrid
Spain
julian.chaparro@upm.es

Santiago Iglesias-Pradas
Departamento de Ingeniería de Organización, Administración de Empresas y Estadística. ETSI de Telecomunicación. Universidad Politécnica de Madrid.
Av. Complutense 30, 28040 Madrid
Spain
s.iglesias@upm.es

ABSTRACT

Public Research Institutions form, together with Universities, the core of scientific research in Spain. These organizations often publish their research in paper format or in their websites, but are generally falling behind in their adaptation to social media, and thus failing to reach wider audiences – especially Millennials. This study offers the preliminary results of an ongoing project aiming to identify best practices in the diffusion of scientific research by Spanish Public Institutions in social media, and more specifically in the three most popular social networking sites: Facebook, Twitter and YouTube. To do so, the study proposes a set of metrics to assess the impact and reach of content published by the different research centers belonging to Spanish Public Research Institutions in those social media platforms, presents the tools used for data extraction, collection and visualization, and identifies the research centers that are most successful in disseminating their scientific activity in social media. Following this study, a set of interviews with community managers and directors of communication of these centers will offer complementary information that shall be helpful in providing guidelines to deploy an effective scientific diffusion strategy in social media to all Spanish Public Research Institutions.

CCS CONCEPTS

• **Information systems → Information systems applications → Collaborative and social computing systems and tools → Social networking sites**

KEYWORDS

Public Research Institutions, Spain, dissemination, scientific research, social networking sites.

ACM Reference format:
Ángel Hernández-García, Emiliano Acquila-Natale, Julián Chaparro-Peláez and Santiago Iglesias-Pradas. 2018. *Social Media Diffusion by Spanish Public Research Institutions*. White paper. Universidad Politécnica de Madrid.

1 INTRODUCTION

Spanish Public Research Institutions (PRIs) are public research organizations that, together with Universities, comprise the core of the public Spanish scientific research and technological development. One of the main objectives

of PRIs is the diffusion of scientific and technical knowledge. There are currently eight PRIs in Spain, grouping a total of 196 research centers. PRIs mainly release their findings via publications in paper and their respective official web pages. While these two channels are the most prevalent across all research centers, the scope and reach of the knowledge is most often limited to professionals within each sector and to researchers in the different fields of study of each PRI. However, there is a strong belief in European societies that researchers should report their findings and activities to the general public, that they are quite inefficient at doing so, and that states should increase their effort to attract younger audiences and women into sciences and scientific research [1]. Moreover, and despite the improving perception among Spaniards that they are well informed about science and technology [2], Spanish citizens have the lowest rank regarding objective knowledge about scientific concepts [3], which evidences the need for development and implementation of policies aiming to promote scientific culture among Spanish general public.

The development of Web 2.0 and social networking sites (SNS), and their intrinsic interactive and participative nature, offer a unique opportunity to broaden the reach and impact of scientific advances, as SNS allow immediate diffusion through interpersonal networks. Furthermore, the use of SNS makes it possible to access scientific knowledge for people outside of the common PRI spheres, especially Millennials whose consumption of information has shifted towards these kinds of platforms. In fact, the Internet has become the most cited source of information, even over television; and younger audiences perceive higher influence from SNS, blogs and specialized media when searching for science-related information [2].

A preliminary observation of the different PRIs shows that there are no unified policies in the way scientific research is communicated to the public, more so in the case of SNS. While a common practice by PRIs is to include an icon or series of icons redirecting to different SNS (primarily Twitter and Facebook) in their official web pages, not all PRIs offer this information or even have registered accounts in these platforms. Furthermore, and despite the fact that many PRIs use social media, it is difficult to assess the impact derived from their use. It is necessary then to establish mechanisms and metrics to perform that assessment and to have a deeper insight about how PRIs use social media for the diffusion of science, and the impact of this use in the Spanish society, in order to develop systematic procedures to improve the exposure of scientific knowledge.

Therefore, this research study aims to collect and analyze activity data of the different Spanish PRIs in the main three SNS (Facebook, Twitter and YouTube), in order to measure the reach and impact of their scientific diffusion, and to identify best practices so as to offer useful guidelines to all PRIs as a whole. As a by-product, the research also outlines a method and list of tools that researchers and governments may apply in case they wish to extend this study to other national and international PRIs. This objective is aligned with the goals of the Spanish State Research and Development plan, which aims to facilitate the access of Spanish citizens to science, technology and innovation.

2 METHODOLOGY OF THE STUDY

2.1 Method

The method used in this research covers four different stages. First, it is necessary to collect the information about all the different research centers and their presence in the three SNS under study (Facebook, Twitter and YouTube), in order to identify the correspondence between each center and its handle in the social media platform. Second, the different metrics to characterize and measure the impact of the scientific diffusion need to be established. Third, data collection requires an analysis and selection of the tool or tools used to collect the activity data of each research center in the different SNS. After data collection, data visualization is required in order to identify best practices. Finally, and after selection of the different research centers as examples of best practices in the different SNS, it will be necessary to conduct interviews with the Directors of Communication and/or community managers of these centers in order to confirm the findings from the analysis and to better understand and identify successful strategies for the diffusion of scientific knowledge among PRIs.

2.2 Sample and duration of the study

The sample used in this study includes the totality of the public research centers that are part of one of the eight PRIs. We identified 133 official Twitter accounts, 97 official Facebook accounts and 35 official YouTube accounts. In other words, two thirds of the research centers have presence in Twitter, around half of them have Facebook accounts and less than twenty percent have a YouTube channel, all of them with varying levels of activity. Most research centers do not have a verified account. It is worth noting that the accounts of individual researchers that are developing their research activity in any research center were not included in the analysis. The collection of this information took place during March 2017, and was revised in June 2017.

The collection of activity data of the different SNS accounts covered the period May-October 2017. After data analysis and visualization, the identification of best practices took place in December 2017 and, by January-February 2018, the authors are currently conducting the interviews with the selected research centers.

2.3 Metrics

For the selection of metrics to assess reach and impact of scientific diffusion, some considerations need to be accounted for. First, the APIs of the three different SNS have different restrictions about which data can be accessed without a commercial license; second, some of the metrics – e.g. total number of impressions of each Facebook post or Twitter tweet– can only be accessed by the owner of the account (i.e. the community manager or director of communications of each research center) and are therefore not available for external researchers. Thus, the metrics selected for this study include the following:

- Number of posts: number of unique tweets, posts or videos uploaded to the SNS. They are the main unit of analysis of this study.
- Retweets: number of times that a tweet has been shared (retweeted) in Twitter. They determine the reach of a given tweet.
- Favorites/Likes: they indicate whether the content uploaded to the SNS is pleasant to the audience.
- Regularity: it considers the time interval between posts/publications, and it adds context to the number of posts, by illustrating whether accounts with higher number of followers, retweets or likes have shorter or longer time intervals between publications, or if an account publishes content intensively but has a narrow reach. In this study, regularity is included only in Twitter and Facebook datasets because YouTube does not provide that information.
- Followers: This metric shows the number of users that receive the information posted by the different accounts directly. Thus, they are the primary source of retweets and shares.

The authors also considered additional metrics, that were discarded due to the difficulty to collect the data automatically, because of their variability or because the information is available only for the owner. Those metrics, which are not exempt from valuable information but fell outside of the scope of the research, include the following: active users, following (accounts followed by the account under study), number of impressions, sentiment analysis and opinion mining, mentions, comments (Facebook and YouTube), visualization times (YouTube), location, unfollows, access times and content type. For the latter, an overview of type of content (i.e. not on a per-post basis) was conducted.

2.3 Selection of tools for data collection

After performing an Internet search, the authors elaborated a list of different tools to extract and collect the data about content uploaded by the different accounts to the SNS under study. Apart from Facebook Analytics, Twitter Analytics and Google Analytics –the official software to perform analytics of one's own account– and commercial integrated solutions, the main tools included in the analysis are the following:

- Facebook data:
  - LikeAlyzer: free tool that returns, for any given account, ranked overall information about its frontpage, information provided, activity, response, and engagement. The calculation of the metrics is not disclosed.
  - Sociograph: requires Facebook authentication, and returns, in chronological order, the posts uploaded by the account, reactions, and shared content. It offers dataset export only as a premium service.
- Twitter data:
  - Twitonomy: free tool that shows the different accounts followed by one's account, and their tweets and number of favorites and retweets. It offers dataset export only as a premium service.
  - TweetStats: free tool that shows graphs and statistics of a given account. The interactive graphs show average tweets per day and month. It also provides the weekly and hourly frequency of publication, and most active retweeters. TweetStats does not give direct access to individual tweets.
- Multi-SNS data:
  - Netlytic: community-supported text and social networks analyzer that can automatically capture data from different social media sites (Twitter, Facebook, YouTube, Instagram, RSS Feed, and text/csv files). It offers a limited number of maximum datasets and records per dataset, with a premium service for large dataset requirements.

Table 1 below shows a summary of the different tools, indicating whether they allow analysis of Twitter, Facebook and YouTube, dataset export, have premium service, or allow analysis of multiple accounts.

Table 1: List of tools

Name	Tw	FB	YT	Dataset	Premium	Multi.
Netlytic	Y	Y	Y	Y	Y	Y
Likealyzer	N	Y	N	N	N	Y
Sociograph	N	Y	N	N	Y	N
Twitonomy	Y	N	N	N	Y	N
Tweetstats	Y	N	N	N	N	Y

After assessment of the volume of information generated by the different PRIs, and in order to extract the information needed to perform the study, Netlytic was

chosen to collect the data from Facebook and YouTube. For the case of Twitter, Netlytic allows users to export 5 datasets for Tier 2 free accounts, but it does not provide some information that is highly relevant for this study: number of retweets and favorites. Therefore, and despite its adequacy to build social graphs and interactions between different users, it was deemed necessary to develop a custom tool to extract data from Twitter.

This custom tool, named TweetExtract (https://github.com/TIGE-UPM/TweetExtract), is an executable code written in R that extracts the tweets of any account, the number of followers, accounts followed, date of publication of each tweet, tweet content, and number of favorites and retweets. TweetExtract also provides dataset export functionalities in CSV format.

3 DATA ANALYSIS

Upon data collection, the different content uploaded by all the research centers was incorporated to three different MS Excel spreadsheets that were used as input data for data visualization and exploration of individual posts, tweets and videos. Additionally, three dynamic tables were built to generate aggregate data of each research center.

The software Qlik Sense helped performing data visualization and exploration. Qlik Sense is one of the most popular tools for business intelligence, data analytics, visualization and exploration. Its associative engine facilitates interactive filtering, visualization and exploration of data, and its graphic user interface provides an easy way to create custom visualizations.

The identification of best practices required the analysis of different metrics, each one offering additional information over basic information, such as number of posts and likes/favorites:

- Likes per post/tweet/video (Fig. 1): this is the main metric used in the study. It gives an overall idea of how "successful" is the content published by the PRI.
- Ratio Likes per post/tweet/video-number of followers (Fig. 2, Fig 3): because the above metric is strongly biased towards accounts with high number of followers (for example, CSIC, which stands for Higher Center for Scientific Research and comprises 131 of the 196 research centers), this metric facilitates identification of "successful" accounts with a more limited audience.
- Combined graph favorites/likes, and tweets/Facebook posts (Fig. 4): given that PRIs's presence is more prevalent in Twitter and Facebook, this metric aims to unveil successful diffusion strategies using both SNS.
- Favorites and likes per post/tweet/video, in chronological order (Fig. 5): this metric facilitates analysis of specific content with distinctive or special impact in terms of reactions, if necessary.

Figure 1: Favorites per tweet (Twitter).

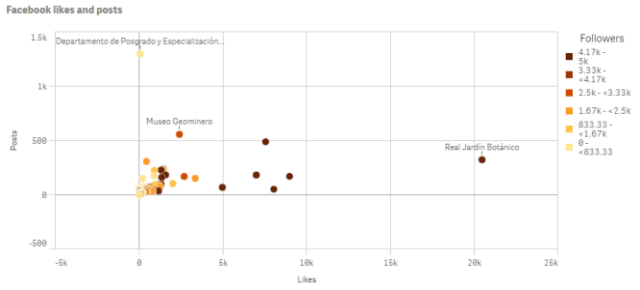

Figure 2: Facebook likes and posts (in color, the number of followers).

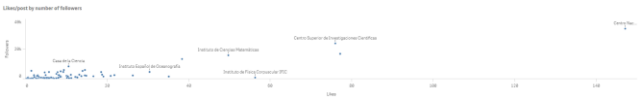

Figure 3: Facebook likes per post by number of followers.

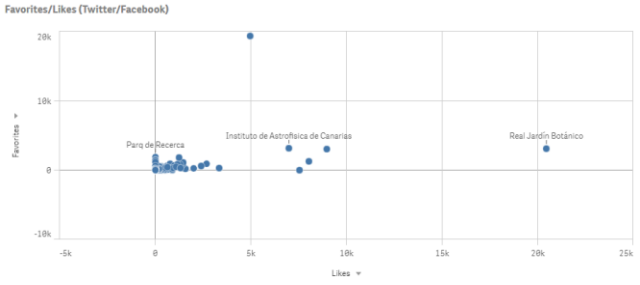

Figure 4: Twitter posts versus Facebook likes.

Figure 5: Likes per Facebook post.

4 RESULTS AND DISCUSSION

Upon exploration of the data collected, the identification of best practices consisted in the selection of the most "successful" research centers. In Fig. 1, research centers of interest should have larger area and brighter color. In Fig. 2, "successful" centers should be those at the far-right, while publication effectiveness would also require them to be placed near the horizontal axis (note that brighter yellow color would also be desirable). In Fig. 3, the best practices should gather around the far-right, nearest to the horizontal axis. Research centers positioned to the upper-right side in Fig. 4 should be considered in order to identify best practices in both Facebook and Twitter; finally, Fig. 5 will help understanding the "success" of some posts after completing the interviews with community managers and Directors of Communication of the different centers.

The preliminary analysis identifies a distinctive outlier regarding best practices in YouTube: IFT (Theoretical Physics Institute), with around 270000 followers and near 1000 likes per video uploaded. Interestingly, IFT departs from the usual publishing strategies of most research centers (generally focused on news and seminars) and offers weekly short videos (between four and ten minutes long) where their researchers try to give a comprehensive explanation, oriented to the general public, to complex topics such as particle disintegration, string theory or dark matter. Further, the success of IFT videos in YouTube is increasing after they have incorporated participatory ways in which the audience may interact with them, by posting different questions, some of which are selected to be addressed in new videos.

Regarding Facebook and Twitter, as expected, the Higher Center for Scientific Research (CSIC, the largest PRI) has the largest audience, highest activity and impact, in terms of likes and retweets. However, smaller research centers, such as the RJB (Royal Botanical Gardens) or the IFIC (Corpuscular Physics Institute), also show very high impact (retweets and likes) with medium to high, and low volume of content published, respectively. While the IFIC posts similar content –e.g. news, seminars, grants and awards– to other research centers in Facebook –the channel where they disseminate their scientific activity more successfully–, the RJB excels in showing a distinct way to communicate their research, by integrating a unique voice that is able to combine a botany-related research approach with original photographic content using examples from the Gardens, all adding a poetic vibe to each tweet and relating them to present events –e.g. the International Women's Day, or other significant dates. It is worth noting that the RJB has the double objective of disseminating their research on botany and driving visitors to the Gardens, located in the center of Madrid, the capital city of Spain. Furthermore, in the particular case of the Royal Botanic Gardens, the authors have already completed the interview with both the community manager and the director of media and communications of the research center, who state that their approach to the use of Facebook and Twitter is different, as both audiences also have different characteristics. They also state that one of the ingredients of their success is having been able to build a community that brings together general public, botany researchers and gardening lovers, which explains their regular interactions with their followers and the use of greeting messages.

The study also shows that, content aside –the authors plan to perform a content analysis after completing the interviews and further analyze the collected data–, consistency, or regularity of publication, and not volume is key to reach to a larger audience and communicate and interact with the community, both of which are essential to further improve the impact of the scientific knowledge being disseminated.

5 CONCLUSIONS

This research presents a preliminary overview of the state of dissemination of scientific knowledge by Spanish Public Research Institutions in social media networking sites. The study details the selection of adequate tools –including the development of a new software application– for data extraction and visualization of activity data of Spanish PRIs in the three most popular social networks, and identifies different research centers that excel at disseminating their research in social media by combining activity data and user-reaction data.

While the study only covers a period of six months, which may bias some of the results that are contingent on seasonal content, the authors are confident that the analyzed dataset is valid in order to identify the best practices.

Following this study, and upon a qualitative analysis of the interviews, further quantitative analysis of the dataset and content analysis of the different publications by PRIs on SNS, the authors' intention is to build a comprehensive map and guidelines for the deployment of effective strategies for dissemination of scientific knowledge by

Spanish PRIs in social media, in the hopes that it will help improve the promotion of a scientific culture in the Spanish society and widen the reach of scientific research to younger audiences.

ACKNOWLEDGMENTS

This research has received the support and collaboration of *Fundación Española para la Ciencia y la Tecnología – Ministerio de Economía y Competitividad* (FCT-2016-11112).

REFERENCES

[1] Juan I. Pérez. 2011. La función social de la cultura científica. *CIC Network* 9, 14-17.
[2] FECYT. 2012. Percepción social de la ciencia y la tecnología 2013. Fundación Española para la Ciencia y la Tecnología, FECYT.
[3] Fundación BBVA (2012). Estudio internacional de cultura científica de la fundación BBVA Comprensión de la ciencia. Departamento de Estudios Sociales y Opinión Pública.