Mapping scientific controversy in Twitter: the Maya city hoax[1]

Elena Denia*

*elenadenia@ingenio.upv.es
INGENIO (CSIC - UPV), Ciudad Politécnica de la Innovación
Camino de Vera, s/n 46022 Valencia (Spain)

ABSTRACT

The present poster reflects a study on the social diffusion of science and public attitudes toward science taking advantage of the available data of the online social network Twitter (real-time nature) and focused on a specific science new that turned out to be a hoax. I consider two lines of study of different nature: (1) on the one hand the aim is to offer insights into to what extent the structure of the network influences the information spread and serves to capture public attention, as well as identify common features of the major influencers; (2) on the other hand I carry out a deeper analysis concerning to the content of the message –tweet–, by using data mining technics, with the purpose of exploring the main elements that play a key role in terms of laypeople interest, trust and engagement, and to observe the predisposition to change opinion while dealing with a fallacious case in which there was no scientific evidence.

The particular case of study is the Maya city controversy. In May 2016, a 15-year-old schoolboy William Gadoury, from Quebec, Canada, compared maps of 22 star constellations to the ancient Maya with Google Earth images of Mexico's Yucatan peninsula. Despite there was no scientific evidence, general media spread a new saying that William found a lost Maya city. We study the reception of such a false new.

Our hypothesis is that the image of science and public opinion of scientific facts depend both on the network structure and on the content of the information. What information can we get from these two approaches of such a different nature?

INTRODUCTION

Both academia and the public authorities advocated that greater permeability between science and society today is a desirable and even essential objective. With the rise of digital social networks, that has led to virtual communities sustained in an architecture of participation, it seems reasonable to investigate novel or insufficiently studied aspects of the relationship between science and society and the public image of science. That is to consider innovative tools for measuring the social perception of science, a social aspect studied extensively over the years by *Eurobarometer* surveys in Europe, by *Fecyt* reports in Spain, among others.

The present poster reflects a study on the social diffusion of science and public attitudes toward science taking advantage of the available data of the online social network Twitter (real-time nature) and focused on a specific science new that turned out to be a hoax. I consider two lines of study of different nature: (1) on the one hand the aim is to offer insights into to what extent the structure of the network influences the information spread and serves to capture public attention, as well as identify common features of the major influencers; (2)

[1] This work was supported by Ministry of Economy and Competitiveness (Grants for predoctoral contracts for the training of PhD candidates 2014 and Proyecto EXTRA, CSO2013-48053-R).

on the other hand I propose a deeper analysis concerning to the content of the message –*tweet*–, by using data mining technics, with the purpose of exploring the main elements that play a key role in terms of laypeople interest, trust and engagement, and to observe the predisposition to change opinion while dealing with a fallacious case in which there was no scientific evidence.

The particular case of study is the Maya city controversy. In May 2016, a 15-year-old schoolboy William Gadoury, from Quebec, Canada, compared maps of 22 star constellations to the ancient Maya with Google Earth images of Mexico's Yucatan peninsula. Despite there was no scientific evidence, general media spread a new saying that William found a lost Maya city. We study the reception of such a false new.

Our hypothesis is that the image of science and public opinion of scientific facts depend both on the network structure and on the content of the information.

METHODS AND TECHNICS

The analysis software used is R –and its packages–, a powerful computer tool used widely by data miners and statisticians, which is freely available under the GNU General Public License. Also, in order to have access to the data, it is necessary to use the Twitter API (Application Programming Interface).

In our particular case, to extract the data we took the keywords 'maya city', and saved it in a file with metadata of the tweets from 2016-05-10 to 2016-05-19, resulting in a sample size of 11,495 tweets. Again, for the next time period of 9 days, to see the decreasing interest on the issue over time, we collected 1,708 tweets. The metadata includes varied information of each tweet and also about users who tweeted.

To analyze the content of the tweets we use techniques of text mining, including text cleaning, topic modelling, sentiment analysis and word associations. On the other hand, to study the structure of the network and features of the agents involved, we use social network analysis (SNA) to build the network, identify major influencers and track the message propagation.

RESULTS

We must be careful and consider different things for this kind of analysis. For example, something surprising from the data scraping is that in the Russian tweets (or linked urls) the word 'мая' occurs, meaning the month May, pronounced as 'maja' (the я is 'ja' in Russian). This is picked up by the Twitter search algorithm for our keywords. We also collected Japanese tweets, but in this case they actually refers to the case of study. If we look at one of the major influencers '@newton_science', it is the account of a science magazine in Japan.

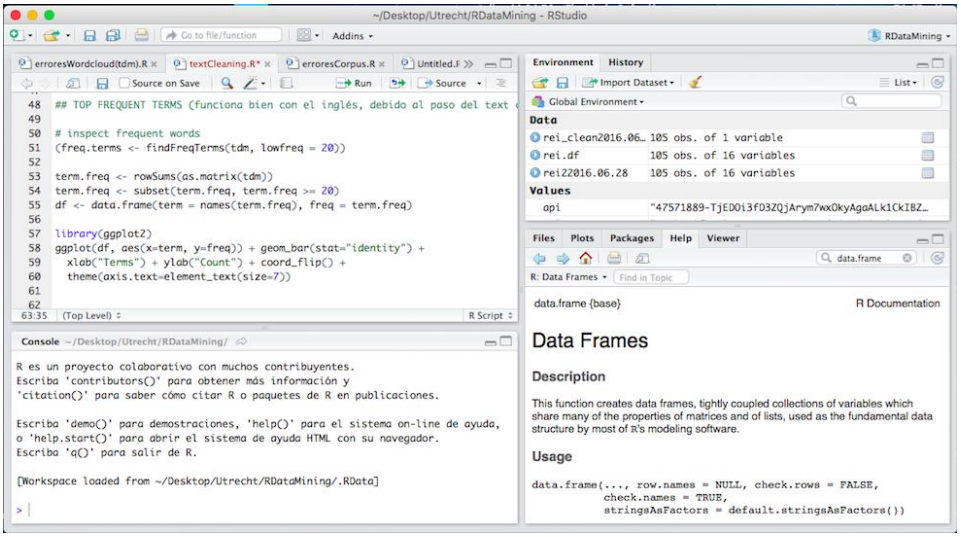
Other major influencer, those who have received more retweets, replies to tweet, or have been mentioned in conversations, etc., in the Maya city controversy is '@darrenaronovski', who is a film maker. This gives us information about agents and its social influence, further than general media.

The wordcloud revels that users pay special attention on concepts like 'teen', 'experts', 'discovers' and 'marijuana'. The last one refers to a joke, suggesting that such an image was not revealing a Maya city but a marijuana field.

CONCLUSIONS

- With the rise of digital social networks that has led to virtual communities it seems reasonable to investigate novel or insufficiently studied aspects of the relationship between science and society and the public image of science.
- The real-time nature of the study provides inexplored dimension of public opinion about science issues.
- To investigate the content of the tweet and the structure of the network we need two approaches of different nature, therefore our results are complementary but not comparable.

Figure 1: Screenshot: running the code in R.



REFERENCES

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.

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, *16*(2), 340-358.

Avellaneda, R. P. (2001). La cultura científico-tecnológica de las sociedades de la modernidad tardía. *Treballs de la Societat Catalana de Biologia*, 35-63.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. (2011, April). Comparing twitter and traditional media using topic models. In *European Conference on Information Retrieval* (pp. 338-349). Springer Berlin Heidelberg.