

Scientific Culture in the era of Twitter

Case of study: Earth-like planet in Proxima Centauri

Elena Denia

elenadenia@ingenio.upv.es

Supervisors: Pablo D'Este, Richard Woolley. INGENIO (CSIC – UPV)

Programa de Doctorado en Diseño, Fabricación y Gestión de Proyectos Industriales

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 which awakens a public debate: the discovery of a habitable planet in Proxima Centauri. 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 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.

MOTIVATION

Scientific Culture

The social perception of science is a social aspect studied extensively over the years by Eurobarometer surveys in Europe, by National Science Foundation surveys in the States and by FECYT reports in Spain, among others.

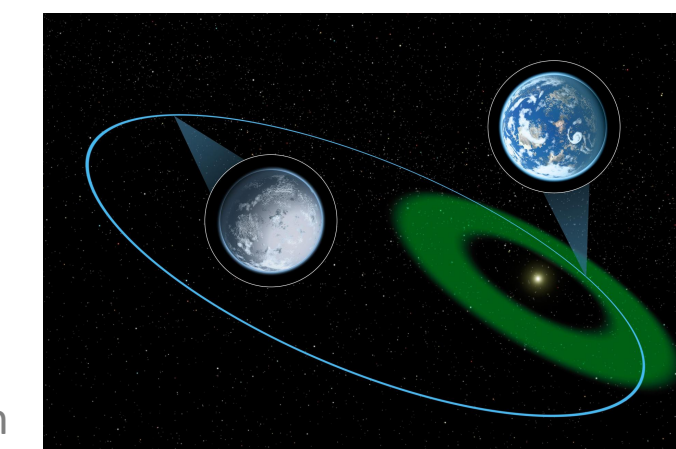


Why Twitter?

- Information exchange is public, users can participate in the public debate and express opinions.
- We can collect real-time data to analyze, including keywords and user profiles.
- Users can be profiles of people, institutions, research centres, etc.
- There is a wide collection of free tools and libraries

A habitable planet

In August 2016 The European Southern Observatory (ESO) discovered an Earth-like planet in the habitable zone in Proxima Centauri –over 4 light-years away–, the closest star system to our own. This awakened a public debate of hopes and concerns.



Credit: NASA/JPL-Caltech

METHODOLOGY

Techniques

Text mining tasks

- Extracting Tweets
- Text Cleaning
- Frequent Words and *Wordcloud*
- Word Associations
- Topic Modelling
- Sentiment Analysis

Social Network Analysis:

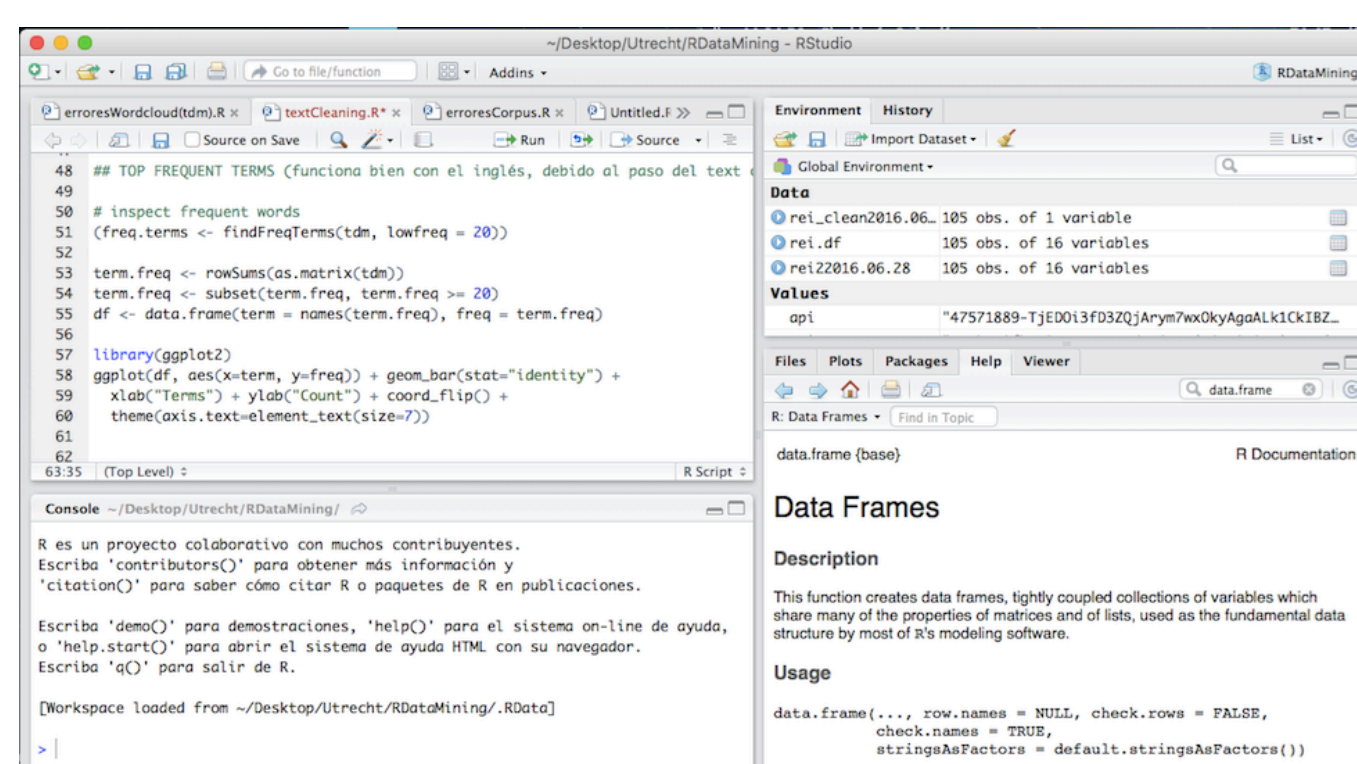
- Retrieve User Info and Followers
- Followers Map
- Active Influential Followers
- Top Retweeted Tweets
- Tracking Message Propagation

Tools

- Twitter's API (Application Programming Interface) provides programmatic access to read and write Twitter data.



- Programming language and software environment for statistical computing R.



Data

Data extracted by using keywords 'earth like planet', and save it in a file with metadata of the tweets (from 2016-08-07 to 2016-08-16), resulting in a sample size of 6,781 tweets.

Tweet	Message of 140 characters. It may contain #hashtags and @mentions, may be answered and retweeted
@mention	A <i>mention</i> is a Tweet that contains another user's @username in the body of the Tweet
#hashtag	A <i>hashtag</i> is used to index keywords or topics on Twitter, then it is easy to follow #topics
Retweet	A Tweet that you share publicly with your followers and can be shared by them.
Reply	A response to another user's Tweet that begins with the @username
Follower	Following someone on Twitter means subscribing to their Tweets as a follower, to have access to their updates.

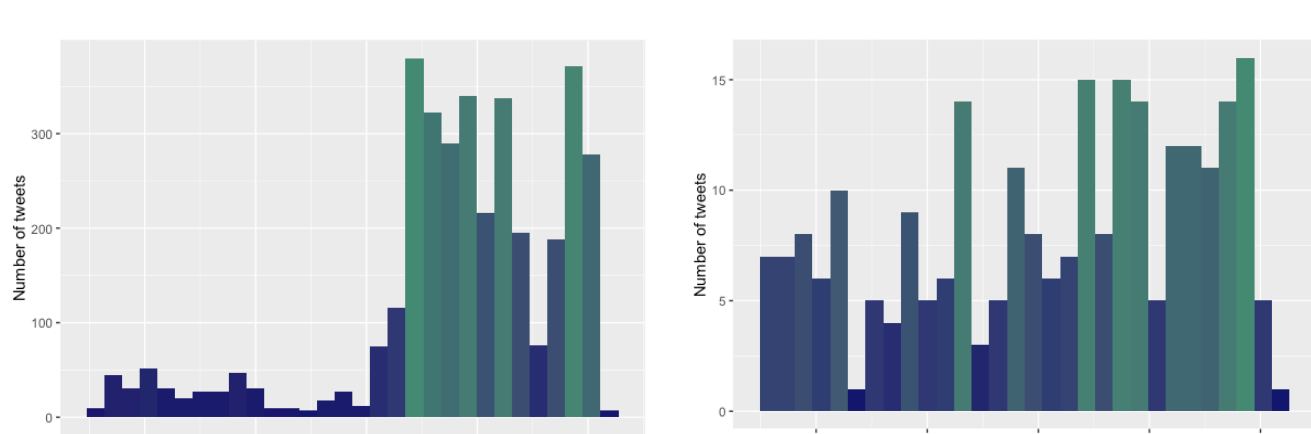
PRELIMINARY RESULTS

Two approaches of different nature: (1) **structure of the network**: information spread; public attention; common features of the major influencers (through SNA); and (2) **content of the message –tweet–**: laypeople interests, trust and engagement.

Content

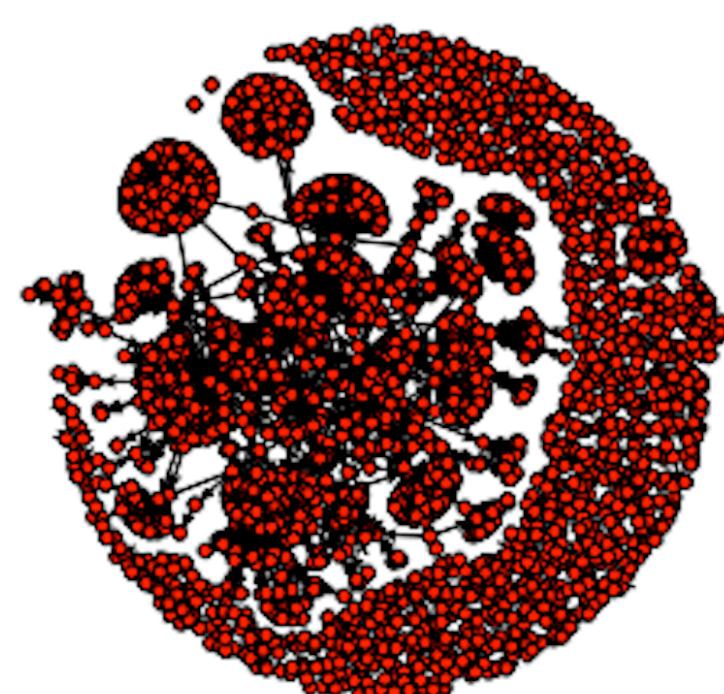


Developing more refined scripts for text cleaning and avoid noise!



Tweets by time: (a) retweets, (b) replies

Structure



Retweets network:
(directed and weighted)
2993 vertices and edges
between them.



Major influencers:

- General media
- Individuals
- Institutions

Work in progress

- Through the *Wordcloud* (or word frequency), to explore which concepts capture more attention and the conceptual context of the discussion.
- To apply sentiment analysis and identify the grade of polarization in opinions.
- What are the common features of the 'major influencers'? (profiles)
- To build the conceptual network. This is a word network to identify key topics and its relations.



CONCLUSIONS

✓ **Hypothesis**: the image of science and public opinion on scientific facts depend both on the network structure and on the content of the information.

- Social digital networks have led to virtual communities sustained in an architecture of participation.
- Applying SNA algorithms allows to detect communities formed by users (media, institutions, scientists & general public) focused on a specific science topic.
- The real-time nature of the study provides unexplored dimensions of the public opinion on science, expressed spontaneously in a public debate.
- Such a real-time data can reveal aspects that traditional surveys could be misleading.
- The conceptual context of the discussion may give clues about the key elements which create interest.