



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 (realtime 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.

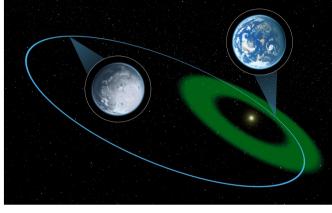


Why Twitter?

- 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 Tools Data Data extracted by using keywords 'earth like planet', and • Twitter's API (Application Programming Interface) provides **Text mining tasks** programmatic access to read and write Twitter data. save it in a file with metadata of the tweets (from Extracting Tweets 2016-08-07 to 2016-08-16), resulting in a sample size of Text Cleaning \bullet 6,781 tweets.

- 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

Programming language and software environment for statistical computing R.

• • •	~/Desktop/Utrecht/RDataMin	ning - RStudio	
ହ • (ଶ	🖹 🔹 🔝 🛛 🚔 🗋 🌧 Go to file/function 👘 🛛 🔯 🗸 🖌 Addins 🗸		🔕 RDataMinin
🖭 erre	rresWordcloud(tdm).R × 🕐 textCleaning.R* × 🕐 erroresCorpus.R × 🖗 Untitled.F ≫ 👝 🗔	Environment History	- 1
40	🖅 🔒 🖸 Source on Save 🔍 🖉 🗉 🕀 Run 🍉 🕞 Source 🔹 🗷	😅 🔒 🖙 Import Dataset 🕶 🍕	Elist •
48	## TOP FREQUENT TERMS (funciona bien con el inglés, debido al paso del text	🐴 Global Environment -	Q,
49		Data	
50	<pre># inspect frequent words</pre>	<pre>orei_clean2016.06105 obs. of 1 variable</pre>	
51	<pre>(freq.terms <- findFreqTerms(tdm, lowfreq = 20))</pre>	rei.df 105 obs. of 16 variables	
52		© rei22016.06.28 105 obs. of 16 variables	
53 54	<pre>term.freq <- rowSums(as.matrix(tdm)) term.freq <- subset(term.freq, term.freq >= 20)</pre>	Values	
55	df <- data.frame(term = names(term.freq), freq = term.freq)	api "47571889-TjED0i3fD3ZQjAr	mm7wx0kuAccAl k1CkTR7
56	an e accarrence com a name (commence), real a commencely	upt 47371003-1JE00131032QJA	ymrwaukyaguatkitkidz
57	library(ggplot2)	Files Plots Packages Help Viewer	-
58	<pre>ggplot(df, aes(x=term, y=freq)) + geom_bar(stat="identity") +</pre>		Q. data.frame
59	<pre>xlab("Terms") + ylab("Count") + coord_flip() +</pre>		G data.frame
60 61	<pre>theme(axis.text=element_text(size=7))</pre>	R: Data Frames • Find in Topic	
62		data.frame {base}	R Documentatio
	(Top Level) ¢ R Script ¢	outaine (euso)	
Conco	le ~/Desktop/Utrecht/RDataMining/ ↔	Data Frames	
Conso	e ~/Desktop/Otrecht/Rbatamining/	Data Frances	
	n proyecto colaborativo con muchos contribuyentes.	Description	
	a 'contributors()' para obtener más información y	Description	
'citat	ion()' para saber cómo citar R o paquetes de R en publicaciones.	This function creates data frames, tightly coupled collect	tions of variables which
Econik	a 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,	share many of the properties of matrices and of lists, us	
	p.start()' para abrir el sistema de ayuda HTML con su navegador.	structure by most of R's modeling software.	
	a 'a()' para salir de R.	Usage	
		Usage	
[Works	pace loaded from ~/Desktop/Utrecht/RDataMining/.RData]	data.frame(, row.names = NULL, check.r	:ows = FALSE,
		check.names = TRUE,	
>		stringsAsFactors = default.str	ingsAsFactors())

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.

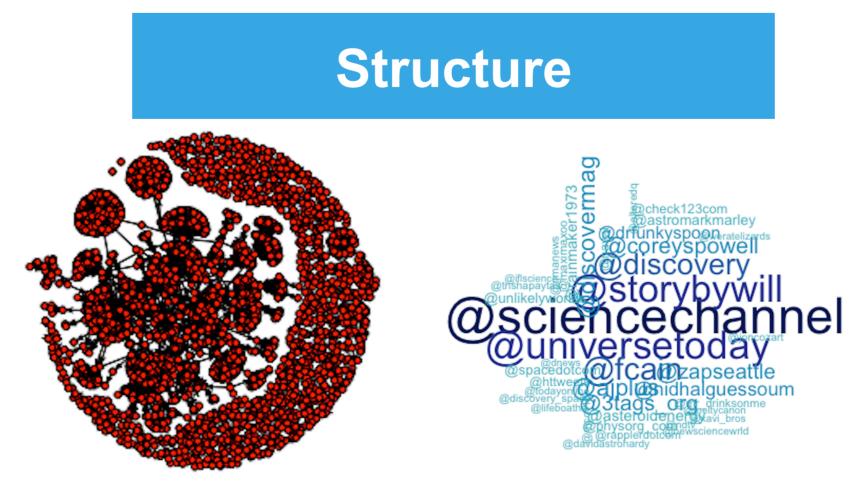
PRELIMINARY RESULTS

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.





Developing more refined scripts for text



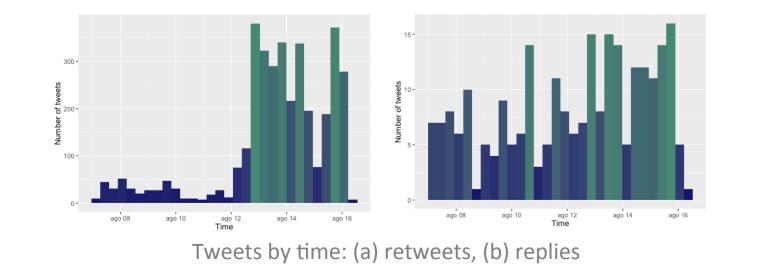
YYYYYYYYYY

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.

cleaning and avoid noise!



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.



- 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 inexplored dimensions of the public opinion on science, expressed spontaneusly 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 \bullet clues about the key elements which create interest.

IV Encuentro de Estudiantes de Doctorado de la UPV Valencia, 01 junio 2017