Collecting Twitter data with R

(and some reflexions on the limits and use of this type of data)

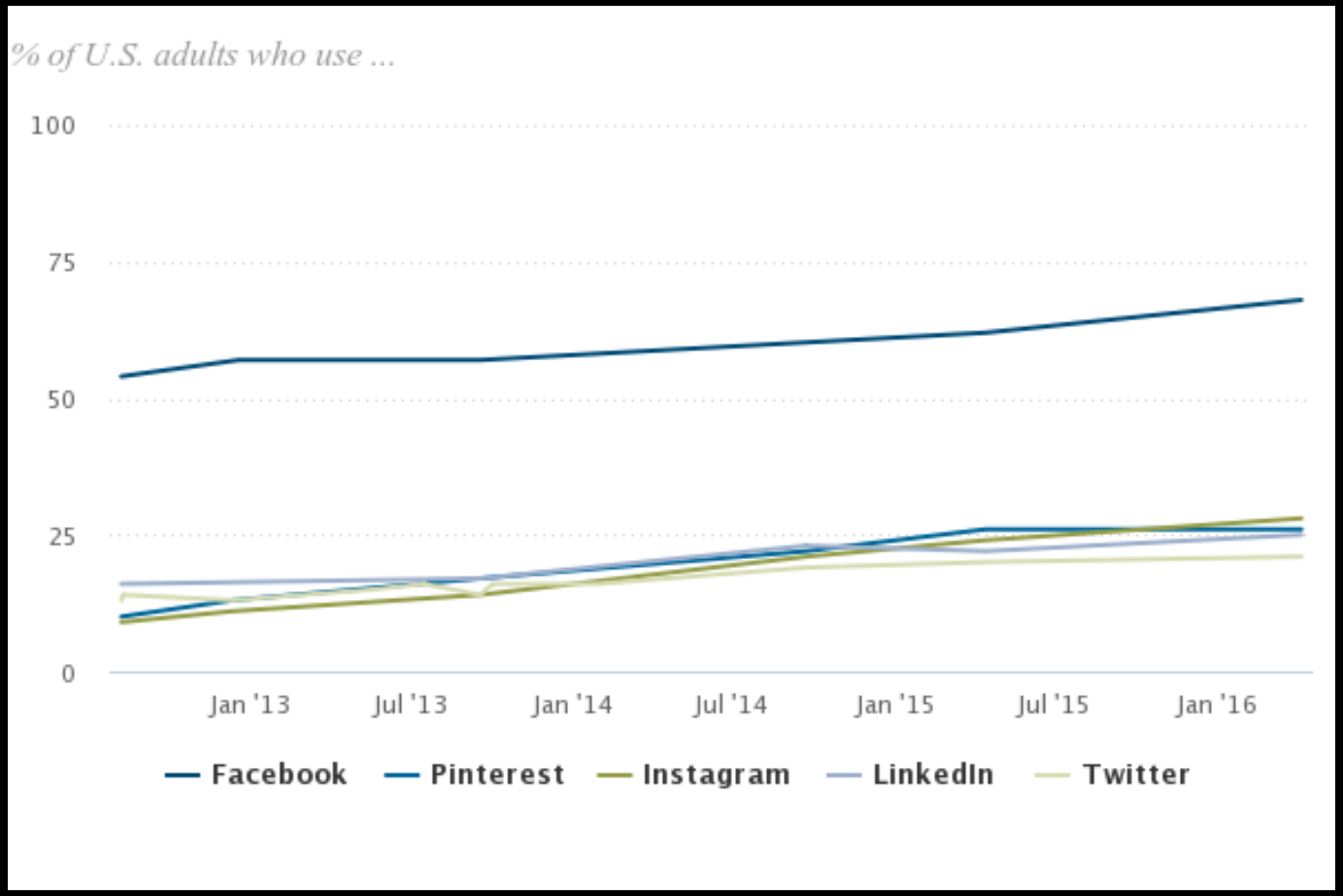
Elise Tancoigne - post-doc - Carl Vogt

R lunchs - 7 février 2017

TWITTER?

Sources: Pew Research Center &
http://www.statisticbrain.com/twitter-statistics/

- March 21, 2006
- 695,750,000 registered users
- 342,000,000 active users
- 9,100 tweets per second



Twitter Demographics

Among internet users, the % who use Twitter

	Internet users
Total	23%
Men	25
Women	21
White, Non-Hispanic	20
Black, Non-Hispanic (n=85)	28
Hispanic	28
18-29	32
30-49	29
50-64	13
65+	6
High school grad or less	19
Some college	23
College+	27
Less than $30,000/yr	21
$30,000-$49,999	19
$50,000-$74,999	25
$75,000+	26
Urban	30
Suburban	21
Rural	15

Source: Pew Research Center, March 17-April 12, 2015.

PEW RESEARCH CENTER

Biased in favor of:

- men
- in their 20s-40s
- who graduated college
- earning more than $50,000
- from urban areas

Home Notifications Messages Search Twitter Tweet

TWEETS 179 | FOLLOWING 230 | FOLLOWERS 372 | LIKES 46 | MOMENTS 0

Edit profile

Citizen Sciences

@CitizenSciences

Rethinking Science and Public Participation

citizensciences.net

Joined October 2015

40 Photos and videos

Tweets | Tweets & replies | Media

Citizen Sciences @CitizenSciences · Jan 26

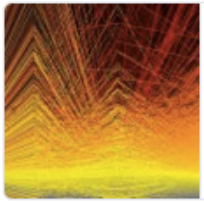
Where are you on the map? 4 things #Twitter tells us about #citizenscience on our new blog post: citizensciences.net /2017/01/26/4-t … ET

4 4

You Retweeted

Barbara Prainsack @BPrainsack · Jan 20

Come see Vanessa Heggie,Javier Lezaun, @dana_mahr1 and others @KingsCollegeLon 2 March: conference on CitizenScience



Cr...

In biomedicine, initiatives subsumed under the heading of “citizen science” are highly diverse, ranging from projects run by academic institutions to private comp...

globalhealthandsocialmedicine.com

Your Tweet activity

Your Tweets earned **3,327 impressions** over the last **28 days**

View your top Tweets

Who to follow · Refresh · View all

Jeyhan @Jeyhan Follow

Kelle Cruz @kellecruz Follow

CitizenScienceGames @CS... Follow

Find friends

Trends · Change

Home Notifications Messages Search Twitter Tweet

TWEETS 179 | FOLLOWING 230 | FOLLOWERS 372 | LIKES 46 | MOMENTS 0

Edit profile

Citizen Sciences

@CitizenSciences

Rethinking Science and Public Participation

citizensciences.net

Joined October 2015

40 Photos and videos

Your Tweet activity

Your Tweets earned 2,227 impressions

EUROPEAN CITIZEN SCIENCE ASSOCIATION

Following

ECSA

@EuCitSci FOLLOWS YOU

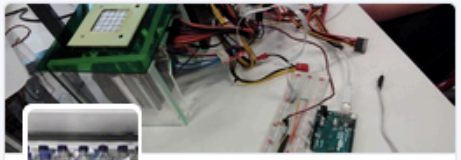
The European Citizen Science Association (ECSA) encourages the growth of the Citizen Science movement in Europe. ecsa.citizen-science.net

Following

Claire B.

@Claire_csg FOLLOWS YOU

About games, science, gamebased learning. Citizen scientist (enthusiastic newbie). meetup.com/Brighton-Citiz… @cs__games



Following

Biotop Heidelberg

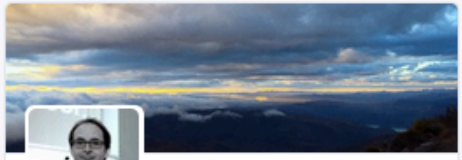
@BiotopHD FOLLOWS YOU

The DIYbio community of Heidelberg

Following

CrowdMagApp

@CrowdMagApp FOLLOWS YOU



Following

Kevin Schawinski

@kevinschawinski

Following

CC-CS Zurich

@cccs_zurich

Home Notifications Messages Search Twitter Tweet

TWEETS 179 | FOLLOWING 230 | FOLLOWERS 372 | LIKES 46 | MOMENTS 0

Edit profile

Citizen Sciences

@CitizenSciences

Rethinking Science and Public Participation

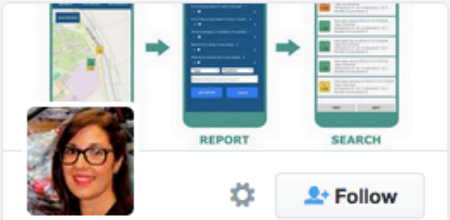
citizensciences.net

Joined October 2015

40 Photos and videos

Your Tweet activity

Your Tweets earned 2,327 impressions



Follow

Rosa Arias

@RosaAriasAlv FOLLOWS YOU

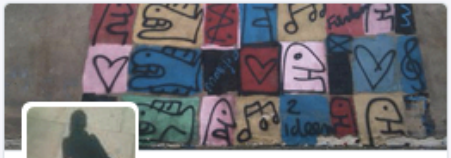
@OdourCollect R&I Consultant @finiciativas #CitizenScience #odour #H2020 #SwafS evaluator #RRI #Energy #Environment #Sustainability …

Follow

Grace Colgan

@grace_colgan FOLLOWS YOU

@ucla alum, conservationist, aspiring teacher attending @CSULBCED

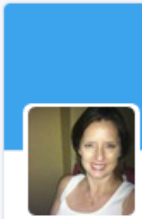


Follow

Aleks

@leksy_b FOLLOWS YOU

Art&Science enthusiast, former neuroscientist working in science and innovation policy. #DIYScience togetherscience.eu



Follow

Leaha

@LeahaRice FOLLOWS YOU

Follow

Liz MacDonald

@spaceyliz FOLLOWS YOU

Follow

CommCitSci

@UCDavisCCS FOLLOWS YOU

Tweet Activity

Citizen Sciences @CitizenSciences
Where are you on the map? 4 things **#Twitter** tells us about **#citizenscience** on our new blog post: http://citizensciences.net/2017/01/26/4-things-twitter-tells-us-about-citizen-science/ … ET

Reach a bigger audience
Get more engagements by promoting this Tweet!

Get started

Impressions	526
Total engagements	23
Link clicks	9
Retweets	4
Likes	4
Profile clicks	4
Detail expands	2

Home | Notifications | Messages | Search Twitter | Tweet

Trends · Change

#SDE2017
5,066 Tweets

#JeNeSuisPasLesEnfantsDeFillon
3,381 Tweets

#StopChasseàLHomme
2,361 Tweets

#elearning
4,449 Tweets

#opensource
@philshapiro is Tweeting about this

Denis Villeneuve
2,586 Tweets

West Bank
10.1K Tweets

#LT17uk
3,790 Tweets

#Amona
1,464 Tweets

All | Mentions — Settings

Rosa Arias and 3 others followed you — 1h

Mike Kearney @kearneymw · 23h
Four things Twitter tells us about "Citizen Science" (and 1,000 things it doesn't) citizensciences.net/2017/01/26/4-t… via @citizensciences

Jill Nugent liked a Tweet you were mentioned in — 24h
Jan 31: Four things Twitter tells us about #CitizenScience (and 1,000 things it doesn't) citizensciences.net/2017/01/26/4-t… via @citizensciences

ECSA Retweeted a Tweet you were mentioned in — Jan 31
Jan 31: Four things Twitter tells us about #CitizenScience (and 1,000 things it doesn't) citizensciences.net/2017/01/26/4-t… via @citizensciences

Zentrum für CS @_CitizenScience · Jan 31
Four things Twitter tells us about #CitizenScience (and 1,000 things it doesn't) citizensciences.net/2017/01/26/4-t… via @citizensciences

1 1

SciStarter liked a Tweet you were mentioned in — Jan 31
Jan 31: 4 things Twitter tells us about "Citizen Science" Cool! @tweetaurora is on here & kinda unique. citizensciences.net/2017/01/26/4-t… via @citizensciences

Ian Crabb Retweeted a Tweet you were mentioned in — Jan 31
Jan 30: Four things Twitter tells us about "Citizen Science" (and 1,000 things it doesn't) citizensciences.net/2017/01/26/4-t… via @citizensciences

Who to follow · Refresh · View all

Anna Weigel @AnnaKWeigel — Follow

David W. Hogg @davidwhogg — Follow

William Keel @NGC3314 — Follow

Find friends

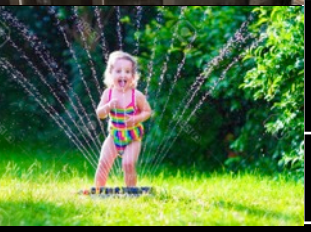
© 2017 Twitter About Help Terms Privacy Cookies Ads info Brand Blog Status Apps Jobs Advertise Businesses Media Developers

Crédit: Michel Bobillier

The nozzle – Photo – @ Bill Harby

Crédit : famveldman

Firehose	handled by 2 data providers - GNIP and DataSift sends selected data when published sends all related tweets
Twitter Stream API	sends selected data when published 1% to 40% of total related tweets
Twitter Search API	sampling of recent Tweets published in the past 7 days.



Endpoint	Resource family	Requests / window (user auth)	Requests / window (app auth)
GET account/verify_credentials	application	75	0
GET application/rate_limit_status	application	180	180
GET favorites/list	favorites	75	75
GET followers/ids	followers	15	15
GET followers/list	followers	15	15
GET friends/ids	friends	15	15
GET friends/list	friends	15	15
GET friendships/show	friendships	180	15

https://dev.twitter.com/rest/public/rate-limits

- over-representation of urban, young and educated men
- only samples of data available for free

-> check that your question fits this type of data

✓ Donald Trump's communication
✗ propagation of a worldwide phenomenon
✓ Political opinions
✗ Political opinions from elderly people
✗ Political opinions from New York's inhabitants

WORKING WITH TWITTER?

REPORT

FEBRUARY 20, 2014

MAPPING TWITTER TOPIC NETWORKS: FROM POLARIZED CROWDS TO COMMUNITY CLUSTERS

Part 2: Conversational Archetypes: Six Conversation and Group Network Structures in Twitter

BY MARC A. SMITH, LEE RAINIE, BEN SHNEIDERMAN AND ITAI HIMELBOIM

After examining thousand
distinct network structure

tic&société

tic&société

Vol. 7, N° 2 | 2ème semestre 2013
Mondes numériques : nouvelles perspectives de la recherche

Comment articuler analyse des réseaux et des discours sur *Twitter*

L'exemple du débat autour du pacte budgétaire européen

Nikos Smyrnaios and Pierre Ratinaud

revues.org

Publisher
Association ARTIC

Electronic version
URL: http://ticetsociete.revues.org/1578

Brought to you by Ur
Graduate Institute /
Genève

Article

Towards a Sociological Understanding of Social Media: Theorizing Twitter

Sociology
46(6) 1059–1073
© The Author(s) 2012
Reprints and permission: sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0038038511422553
soc.sagepub.com
SAGE

Dhiraj Murthy
Bowdoin College, USA

Abstract

This article presents the first steps towards a sociological understanding of emergent social media. This article uses Twitter, the most popular social media website, as its focus. Recently, the social media site has been prominently associated with social movements in Libya, Egypt, Tunisia, and Algeria. Rather than rush to breathlessly describe its novel role in shaping contemporary social movements, this article takes a step back and considers Twitter in historical and broad

PLOS | ONE

RESEARCH ARTICLE

"Anyone Know What Species This Is?" – Twitter Conversations as Embryonic Citizen Science Communities

Stefan Daume[1,2,3]*, **Victor Galaz**[2]

1 Faculty of Forest Sciences and Forest Ecology, Georg-August-University Göttingen, Büsgenweg 5, 37077 Göttingen, Germany, **2** Stockholm Resilience Centre, Stockholm University, SE-10691 Stockholm, Sweden, **3** Department of Biodiversity Informatics, Swedish Museum of Natural History, Box 50007, 104 05 Stockholm, Sweden

* stefan.daume@ecoveillance.org

CrossMark
click for updates

OPEN ACCESS

Citation: Daume S, Galaz V (2016) "Anyone Know What Species This Is?" – Twitter Conversations as Embryonic Citizen Science Communities. PLoS ONE 11(3): e0151387. doi:10.1371/journal.pone.0151387

Editor: Robert Guralnick, University of Colorado

Abstract

Social media like blogs, micro-blogs or social networks are increasingly being investigated and employed to detect and predict trends for not only social and physical phenomena, but also to capture environmental information. Here we argue that opportunistic biodiversity observations published through Twitter represent one promising and until now unexplored example of such data mining. As we elaborate, it can contribute to real-time information to traditional ecological monitoring programmes including those sourced via citizen science activities. Using Twitter data collected for a generic assessment of social media data in eco-logical monitoring we investigated a sample of what we denote biodiversity observations

Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors

Takeshi Sakaki
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
sakaki@biz-model.t.u-tokyo.ac.jp

Makoto Okazaki
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
m_okazaki@biz-model.t.u-tokyo.ac.jp

Yutaka Matsuo
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
matsuo@biz-model.t.u-tokyo.ac.jp

ABSTRACT

Twitter, a popular microblogging service, has received much attention recently. An important characteristic of Twitter is its real-time nature. For example, when an earthquake occurs, people make many Twitter posts (*tweets*) related to the earthquake, which enables detection of earthquake occurrence promptly, simply by observing the tweets. As described in this paper, we investigate the real-time interaction of events such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event. To detect a target event, we devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. Subsequently, we produce a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location. We consider each Twitter user as a *sensor* and apply Kalman filtering and particle filtering, which are widely used for location estimation in ubiquitous/pervasive

1. INTRODUCTION

Twitter, a popular microblogging service, has received much attention recently. It is an online social network used by millions of people around the world to remain socially connected to their friends, family members and co-workers through their computers and mobile phones [18]. Twitter asks one question, "What's happening?" Answers must be fewer than 140 characters. A status update message, called a *tweet*, is often used as a message to friends and colleagues. A user can follow other users; her followers can read her tweets. A user who is being followed by another user need not necessarily reciprocate by following them back, which renders the links of the network as directed. After its launch on July 2006, Twitter users have increased rapidly. They are currently estimated as 44.5 million worldwide[1]. Monthly growth of users has been 1382% year-on-year, which makes Twitter one of the fastest-growing sites in the world[2].

Some researchers have examined Twitter: Java et al. an-

Grandjean, *Cogent Arts & Humanities* (2016), 3: 1171458
http://dx.doi.org/10.1080/23311983.2016.1171458

cogent arts & humanities

DIGITAL HUMANITIES | RESEARCH ARTICLE

A social network analysis of Twitter: Mapping the digital humanities community

Martin Grandjean[1]*

Received: 15 November 2015
Accepted: 02 March 2016

*Corresponding author: Martin Grandjean, Department of History, University of Lausanne, Lausanne, Switzerland
E-mail: martin.grandjean@unil.ch

Reviewing editor:
Aaron Mauro, Pennsylvania State University, USA

Additional information is available at the end of the article

Abstract: Defining *digital humanities* might be an endless debate if we stick to the discussion about the boundaries of this concept as an academic "discipline". In an attempt to concretely identify this field and its actors, this paper shows that it is possible to analyse them through Twitter, a social media widely used by this "community of practice". Based on a network analysis of 2,500 users identified as members of this movement, the visualisation of the "who's following who?" graph allows us to highlight the structure of the network's relationships, and identify users whose position is particular. Specifically, we show that linguistic groups are key factors to explain clustering within a network whose characteristics look similar to a *small world*.

Subjects: Internet & Multimedia - Computing & IT; Network Theory; Sociology of Science & Technology

Keywords: digital humanities; social network analysis; Twitter; digital studies; social

ARTICLES

WHAT THE HASHTAG?

A content analysis of Canadian politics on Twitter

Tamara A. Small

Pages 872-895 | Received 23 Jul 2010, Accepted 03 Jan 2011, Published online: 19 May 2011

Download citation http://dx.doi.org/10.1080/1369118X.2011.554572

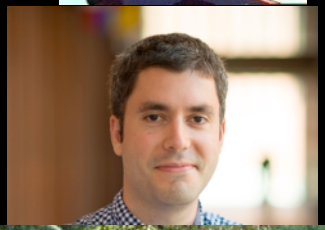
Full Article Figures & data References Citations Metrics Reprints & Permissions Get access

Abstract

Twitter is the latest social networking tool said to be reshaping politics. Twitter allows subscribers to write a 140-character status update, called a 'tweet'. This research analyzes the intersection of microblogging and Canadian politics. There is little scholarship on the political aspects of Twitter. This paper seeks to fill this gap through a content analysis of the most popular Canadian political hashtag, #cdnpoli. Designated by a 'hash' symbol (#), a hashtag is a keyword assigned to information that describes a tweet and aides in searching. With 50 million tweets per day, hashtags are central to organizing information on Twitter. Hashtags organize discussion around specific topics or events. Political hashtags came to prominence in

R & TWITTER

Name	Authors	Date	API	About
twitteR	Jeff Gentry Bioinfo, US	Aug 2016	web	not curated (see rtweet)
streamR	Pablo Barbera SHS, US	Feb 2015	stream	improved by tweet2r
tweet2r	Pau Aragó Pablo Juan SIG, SP	Jun 2016	stream	Access Twitter's filter, sample, user streams -> performs descr. + exploratory space-time analyis
rtweet	Michael W. Kearney SHS, US	Sept 2016	web & stream	Search Twitter, get trends, user info, timelines, status, friends, followers, search users, etc.
graphTweets	John Coene private, China	May 2016	web	Based on twitteR. Creates a network of citation by subsetting @tags in tweets retrieved.



Home Notifications Messages Search Twitter Tweet

get_timeline()

TWEETS 179 FOLLOWING 230 FOLLOWERS 372 LIKES 46 MOMENTS 0

Edit profile

Citizen Sciences
@CitizenSciences

Rethinking Science and Public Participation

citizensciences.net
Joined October 2015

40 Photos and videos

Tweets Tweets & replies Media

Citizen Sciences @CitizenSciences · Jan 26

Where are you on the map? 4 things #Twitter tells us about #citizenscience on our new blog post: citizensciences.net /2017/01/26/4-t … ET

4 4

You Retweeted

Barbara Prainsack @BPrainsack · Jan 20

Come see Vanessa Heggie,Javier Lezaun, @dana_mahr1 and others @KingsCollegeLon 2 March: conference on CitizenScience

Cr...
In biomedicine, initiatives subsumed under the heading of "citizen science" are highly diverse, ranging from projects run by academic institutions to private comp...
globalhealthandsocialmedicine.com

Your Tweet activity

Your Tweets earned **3,327 impressions** over the last **28 days**

View your top Tweets

Who to follow · Refresh · View all

Jeyhan @Jeyhan Follow

Kelle Cruz @kellecruz Follow

CitizenScienceGames @CS... Follow

Find friends

Trends · Change

Home Notifications Messages Search Twitter Tweet

get_friends()
get_followers()
get_favorites()

TWEETS 179 | FOLLOWING 230 | FOLLOWERS 372 | LIKES 46 | MOMENTS 0 | Edit profile

Citizen Sciences
@CitizenSciences

Rethinking Science and Public Participation

citizensciences.net
Joined October 2015

40 Photos and videos

Tweets | Tweets & replies | Media

Citizen Sciences @CitizenSciences · Jan 26
Where are you on the map? 4 things #Twitter tells us about #citizenscience on our new blog post: citizensciences.net /2017/01/26/4-t … ET

4 4

You Retweeted

Barbara Prainsack @BPrainsack · Jan 20
Come see Vanessa Heggie,Javier Lezaun, @dana_mahr1 and others @KingsCollegeLon 2 March: conference on CitizenScience

Cr...
In biomedicine, initiatives subsumed under the heading of "citizen science" are highly diverse, ranging from projects run by academic institutions to private comp...
globalhealthandsocialmedicine.com

Your Tweet activity

Your Tweets earned **3,327 impressions** over the last **28 days**

View your top Tweets

Who to follow · Refresh · View all

Jeyhan @Jeyhan Follow

Kelle Cruz @kellecruz Follow

CitizenScienceGames @CS... Follow

Find friends

Trends · Change

Home Notifications Messages Search Twitter Tweet

lookup_users()

TWEETS 179 FOLLOWING 230 FOLLOWERS 372 LIKES 46 MOMENTS 0

Edit profile

Citizen Sciences
@CitizenSciences

Rethinking Science and Public Participation

citizensciences.net

Joined October 2015

40 Photos and videos

Tweets Tweets & replies Media

Citizen Sciences @CitizenSciences · Jan 26

Where are you on the map? 4 things #Twitter tells us about #citizenscience on our new blog post: citizensciences.net /2017/01/26/4-t … ET

4 4

You Retweeted

Barbara Prainsack @BPrainsack · Jan 20

Come see Vanessa Heggie,Javier Lezaun, @dana_mahr1 and others @KingsCollegeLon 2 March: conference on CitizenScience

Cr...

In biomedicine, initiatives subsumed under the heading of "citizen science" are highly diverse, ranging from projects run by academic institutions to private comp...

globalhealthandsocialmedicine.com

Your Tweet activity

Your Tweets earned **3,327 impressions** over the last **28 days**

View your top Tweets

Who to follow · Refresh · View all

Jeyhan @Jeyhan Follow

Kelle Cruz @kellecruz Follow

CitizenScienceGames @CS... Follow

Find friends

Trends · Change

Home Notifications Messages Search Twitter Tweet

stream_tweets()
search_tweets()
search_users()

TWEETS 179 | FOLLOWING 230 | FOLLOWERS 372 | LIKES 46 | MOMENTS 0

Edit profile

Citizen Sciences
@CitizenSciences

Rethinking Science and Public Participation

citizensciences.net
Joined October 2015

40 Photos and videos

Tweets | Tweets & replies | Media

Citizen Sciences @CitizenSciences · Jan 26
Where are you on the map? 4 things #Twitter tells us about #citizenscience on our new blog post: citizensciences.net /2017/01/26/4-t … ET

4 4

You Retweeted

Barbara Prainsack @BPrainsack · Jan 20
Come see Vanessa Heggie,Javier Lezaun, @dana_mahr1 and others @KingsCollegeLon 2 March: conference on CitizenScience

Cr...
In biomedicine, initiatives subsumed under the heading of "citizen science" are highly diverse, ranging from projects run by academic institutions to private comp...
globalhealthandsocialmedicine.com

Your Tweet activity

Your Tweets earned **3,327 impressions** over the last **28 days**

View your top Tweets

Who to follow · Refresh · View all

Jeyhan @Jeyhan Follow

Kelle Cruz @kellecruz Follow

CitizenScienceGames @CS... Follow

Find friends

Trends · Change

Home Notifications Messages Search Twitter Tweet

getEdges()

TWEETS 179 | FOLLOWING 230 | FOLLOWERS 372 | LIKES 46 | MOMENTS 0

Edit profile

Citizen Sciences
@CitizenSciences

Rethinking Science and Public Participation

citizensciences.net
Joined October 2015

40 Photos and videos

Tweets | Tweets & replies | Media

Citizen Sciences @CitizenSciences · Jan 26
Where are you on the map? 4 things #Twitter tells us about #citizenscience on our new blog post: citizensciences.net /2017/01/26/4-t … ET

4 4

You Retweeted

Barbara Prainsack @BPrainsack · Jan 20
…ggie,Javier Lezau…, @dana_mahr1 a…d others @KingsCollegeLon 2…March: conference…

Cr...
In biomedicine, initiatives subsumed under the heading of "citizen science" are highly diverse, ranging from projects run by academic institutions to private comp...
globalhealthandsocialmedicine.com

Your Tweet activity

Your Tweets earned **3,327 impressions** over the last **28 days**

View your top Tweets

Who to follow · Refresh · View all

Jeyhan @Jeyhan Follow

Kelle Cruz @kellecruz Follow

CitizenScienceGames @CS... Follow

Find friends

Trends · Change

EXAMPLES

1. Cf. the demonstration with 2017Fevrier_Tancoigne_Twitter.R

2. See http://citizensciences.net/2017/01/26/4-things-twitter…-citizen-science/

CONCLUSION

- over-representation of urban, young and educated men
- only samples of data available for free

-> check that your question fits this type of data

- carefully design your list of keywords
- bring external knowledge to make sense of what you find