Collecting Twitter data with R

(and some reflexions on the limits and use of this type of data)
TWITTER?
• March 21, 2006
• 695,750,000 registered users
• 342,000,000 active users
• 9,100 tweets per second

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

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

PEW RESEARCH CENTER
Citizen Sciences
@CitizenSciences
Rethinking Science and Public Participation

citizensciences.net
Joined October 2015

40 Photos and videos

Tweets
179
FOLLOWING
230
FOLLOWERS
372
LIKES
46
MOMENTS
0

Tweets

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

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

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

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

Get started
Rosa Arias and 3 others followed you

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
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: 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 GS @_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

SciStarter liked a Tweet you were mentioned in
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 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
The nozzle – Photo – @ Bill Harby
<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Resource family</th>
<th>Requests / window (user auth)</th>
<th>Requests / window (app auth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET account/verify_credentials</td>
<td>application</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>GET application/rate_limit_status</td>
<td>application</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td>GET favorites/list</td>
<td>favorites</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>GET followers/ids</td>
<td>followers</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET followers/list</td>
<td>followers</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET friends/ids</td>
<td>friends</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET friends/list</td>
<td>friends</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET friendships/show</td>
<td>friendships</td>
<td>180</td>
<td>15</td>
</tr>
</tbody>
</table>

Firehose handled by 2 data providers - GNIP and DataSift sends selected data when published.

Twitter Stream API sends selected data when published.

Twitter Search API sampling of recent Tweets published in the past 7 days.
• 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?
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 thousands of Twitter conversations, the researchers identified six distinct network structures that emerge in structured Ways of talking and discussing in conversations. This structure is one of the most useful and efficient ways to visualize the social network, as it is clearly visible from the outside.

Comment articuler analyse des réseaux et des discours sur Twitter
L'exemple du débat autour du pacte budgétaire européen

Nikos Smyrniatis and Pierre Ratinaud

Towards a Sociological Understanding of Social Media: Theorizing Twitter

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 sociological terms. This article is not intended to provide empirical evidence or a fully formed sociological model of Twitter but to lay the ground work for a more thorough analysis of the complex and evolving social phenomena of social media.
RESEARCH ARTICLE

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

Stefan Daume¹,²,³*, Victor Galaz²

1 Faculty of Forest Sciences and Forest Ecology, Georg-August-University Göttingen, Büsengasse 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

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 ecological 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 computing, to track the path of the event.

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 worldwide1. Monthly growth of users has been 1382% year-on-year, which makes Twitter one of the fastest-growing sites in the world2.

Some researchers have examined Twitter: Java et al. analyzed Twitter in late 2007. They described the social...
A social network analysis of Twitter: Mapping the digital humanities community

Martin Grandjean

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 media; digital humanities; network analysis
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

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 events like the 2008 US presidential election. In Canada, politicians use the power of hashtags to mobilize support.
R & TWITTER
<table>
<thead>
<tr>
<th>Name</th>
<th>Authors</th>
<th>Date</th>
<th>API</th>
<th>About</th>
</tr>
</thead>
<tbody>
<tr>
<td>twitteR</td>
<td>Jeff Gentry Bioinfo, US</td>
<td>Aug 2016</td>
<td>web</td>
<td>not curated (see rtweet)</td>
</tr>
<tr>
<td>streamR</td>
<td>Pablo Barbera SHS, US</td>
<td>Feb 2015</td>
<td>stream</td>
<td>improved by tweet2r</td>
</tr>
<tr>
<td>tweet2r</td>
<td>Pau Aragó, Pablo Juan SIG, SP</td>
<td>Jun 2016</td>
<td>stream</td>
<td>Access Twitter's filter, sample, user streams -&gt; performs descr. + exploratory space-time analysis</td>
</tr>
<tr>
<td>rtweet</td>
<td>Michael W. Kearney SHS, US</td>
<td>Sept 2016</td>
<td>web &amp; stream</td>
<td>Search Twitter, get trends, user info, timelines, status, friends, followers, search users, etc.</td>
</tr>
<tr>
<td>graphTweets</td>
<td>John Coene private, China</td>
<td>May 2016</td>
<td>web</td>
<td>Based on twitteR. Creates a network of citation by subsetting @tags in tweets retrieved.</td>
</tr>
</tbody>
</table>
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

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

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
get_friends()
get_followers()
get_favorites()
stream_tweets()
search_tweets()
search_users()
getEdges()
1. Cf. the demonstration with 2017Fevrier_Tancoigne_Twitter.R

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