

Social Media and Political Participation

Lab 3

*sk5350@nyu.edu*¹

January 6, 2017

¹Adapted from Pablo Barberá and Drew Dimmery

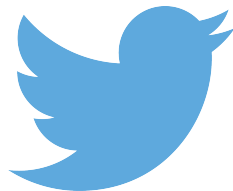
Today

- Twitter: what is it? Main features.
- Introduction to the Twitter API
- Capturing your own collection of tweets
- Analyzing Twitter data
- Tweet coding: training and details about task
- In-class exercise: collect and analyze your own Twitter data

Twitter

Twitter's numbers

- 320+ million monthly active users
- 500+ million tweets are sent everyday
- 18% of online U.S. adults use Twitter
- 25% of young adults in US (18-29) report using Twitter actively
- 52% of Twitter users get news through this platform
- 95% of Members of U.S. Congress have a Twitter account
- 77% of the governments of U.N. member countries have a presence on Twitter



Twitter's main features

The screenshot shows the Twitter interface with the following elements:

- Navigation Bar:** Home, Moments, Notifications, Messages, and a search bar.
- User Profile (Sean Kates):**
 - Profile picture: A man with a beard.
 - Name: Sean Kates
 - Username: @SKates5350
 - Tweets: 35
 - Following: 298
 - Followers: 22
- Trends:**
 - #Star: 45.3K Tweets
 - #SeizeTheDayBy: 5,861 Tweets
 - #ThursdayThought: 11.6K Tweets
 - #BLMKidnapping: 293K Tweets
 - #BadGrittyReboots
 - #NationalBirdDay: 1,011 Tweets
 - #MarijuanaFunFacts: 4,855 Tweets
 - The Anti-Inauguration
- What's happening?**
 - Retweeted by Brendan Nyhan: Adam Smith (@asmith83) · 2h. Five weeks ago today Trump said he'd hold a press conference to talk about his conflicts and still no plan.
 - Donald J. Trump (@realDonaldTrump): I will be holding a major news conference in New York City with my children on December 15 to discuss the fact that I will be leaving my ...
 - Retweeted by (@Rob Ford)): John Pollock (@John_Pollock22) · 1h. Hard to disagree w/ @stephenkb here, maybe #TheresaMay is being v clear and we don't like the answer? #Brexit newstatesman.com/politics/stagg...
- Who to follow:**
 - Donna Yates (@DrDonna...)
 - Josepharthur (@josephart...)
 - Emine Deniz (@ed947nyu)
- Footer:** © 2017 Twitter, About, Help, Terms, Privacy, Cookies, Ads info, Brand, Blog, Status, Apps, Jobs, Businesses, Media, Developers.

Twitter's main features

- Users send messages of up to 140 characters, called **tweets**
- User name (**screen names**) start with an @ sign.
- Each individual can choose to **follow** other users, which will make their tweets appear on that individual's **timeline**
- Other features:
 - hashtags** Words or phrases prefixed with the # symbol that are used to group tweets by topic
 - @-replies** Tweets that begin with the @ symbol followed by a user name (public messages)
 - retweets** Re-publication of another user's content with an indication of its original author
 - mentions** Action of including the screen name of another user in a tweet
 - trending topics** Popular hashtags or phrases

The offline effects of Twitter

- ① Twitter metrics can predict real-world outcomes:
 - Box-office revenue, spread of flu, happiness and general mood, epicenter of earthquakes... even the winner of 'American Idol'.
- ② Studies show that different Twitter metrics were correlated with election results in many countries.
 - BUT: "the predictive power of Twitter regarding elections has been greatly exaggerated" (Gayo-Avello, 2012)
- ③ Social media solve collective action problems, facilitate information diffusion, and thus foster spread of protest.
 - Arab Spring: "The revolutions were tweeted?"
 - Necessary or sufficient cause? Lack of rigorous empirical work
- ④ Twitter and word-of-mouth marketing

Twitter and social science research

Most studies on the effects of social media so far have used Twitter data. Why?

- Presence of many influential actors (journalists, politicians, celebrities...). Spillover effects.
- Effort by political campaigns to generate users' engagement
- Public nature: facilitates access to tweets with the API, and generates dynamics of competition and public expression
- Academic research has shown connection to offline behavior

Twitter API

Twitter API

API = *Application Programming Interface*

Four different methods to collect tweets:

- ① **Filter stream:** tweets filtered by keywords.
 - Example: tweets mentioning “obama” and “biden”
- ② **Geo stream:** tweets filtered by location
 - Example: tweets sent from the Arabian peninsula
- ③ **Sample stream:** 1% random sample of tweets
- ④ **Timeline:** tweets sent by a given user
 - Example: tweets sent by @nytimes

Important: except for the last option, tweets can only be downloaded in real time (as they are being published)

Anatomy of a tweet



Anatomy of a tweet

Tweets are stored in JSON format:

```
{ "created_at": "Wed Nov 07 04:16:18 +0000 2012",
  "id": 266031293945503744,
  "text": "Four more years. http://t.co/bAJE6Vom",
  "source": "web",
  "user": {
    "id": 813286,
    "name": "Barack Obama",
    "screen_name": "BarackObama",
    "location": "Washington, DC",
    "description": "This account is run by Organizing for Action staff.
      Tweets from the President are signed -bo.",
    "url": "http://t.co/8aJ56Jcemr",
    "protected": false,
    "followers_count": 40873124,
    "friends_count": 654580,
    "listed_count": 202495,
    "created_at": "Mon Mar 05 22:08:25 +0000 2007",
    "time_zone": "Eastern Time (US & Canada)",
    "statuses_count": 10687,
    "lang": "en" },
  "coordinates": null,
  "retweet_count": 783488,
  "favorite_count": 295026,
  "lang": "en"
}
```

Collecting Twitter Data

The R script `lab3_collecting_tweets.R` shows how to:

- Install R package to download tweets
- Open an OAuth token and authenticate
- Collect tweets filtering by keywords and location
- Collect a random sample of tweets
- Download all tweets sent by a given user

Analysis of Twitter data

Using the grep function

grep allows you to search for any word inside a text expression.

There are two variants of this command. The first one, `grepl` (with an `l` at the end), returns `TRUE` or `FALSE` depending on whether the text contains that word. For example:

```
> tweet = "four more years"
> grepl("year", tweet)
[1] TRUE
> tweet = "Four More Years"
> grepl("year", tweet)
[1] FALSE
> grepl("year", tweet, ignore.case=TRUE)
[1] TRUE
```

If you set `ignore.case=TRUE`, it will not distinguish between lower and upper case.

Using the grep function

The second variant, `grep` (without an `l` at the end), work for text vectors with more than one element, and returns the position of the elements that contain that word. For example:

```
> tweets = c("four more years",  
             "obama is reelected for another four years")  
> grep("year", tweets)  
[1] 1 2  
> grep("obama", tweets)  
[1] 2  
> grep("more", tweets)  
[1] 1
```


Collecting Twitter Data

The R script `lab3_analyzing_tweets.R` shows how to:

- Open a file with tweets in JSON format
- Analyze key variables about tweets: language, device, country, user characteristics, whether they mention specific words...
- Visualize tweet text with a word cloud
- Visualize geolocated tweets on a map

Tweet Coding Exercise

Tweet Coding Exercise

- Hate Speech Tweets from Trump and Hillary Collections:
- Two parts:
 - 1 In-class training:
 - Go to: bit.ly/inclasscode and create an account
 - We will code a few tweets in class
 - 2 Weekend assignment
 - Go to: bit.ly/hatespeechcode
 - Spend 90 minutes coding tweets (Number of Tweets will vary....but not by too much 😊)
 - Due by Monday at 10am

In-class exercise

In-class exercise: collecting and analyzing Twitter data

Create your own R script (with comments) that:

- ① Downloads one minute of tweets about a celebrity or politician
 - ② Runs different commands to answer the following questions:
 - ① In what language are tweets mentioning this person written?
 - ② What does the most retweeted tweet say about this person?
 - ③ Which tweet was sent by the person with the most followers?
- optional Create a word cloud with these tweets. What do you learn?

And send it to me via email (sk5350@nyu.edu)