Social Media and Political Participation

Lab 3

pablo.barbera@nyu.edu

January 8, 2014

Today

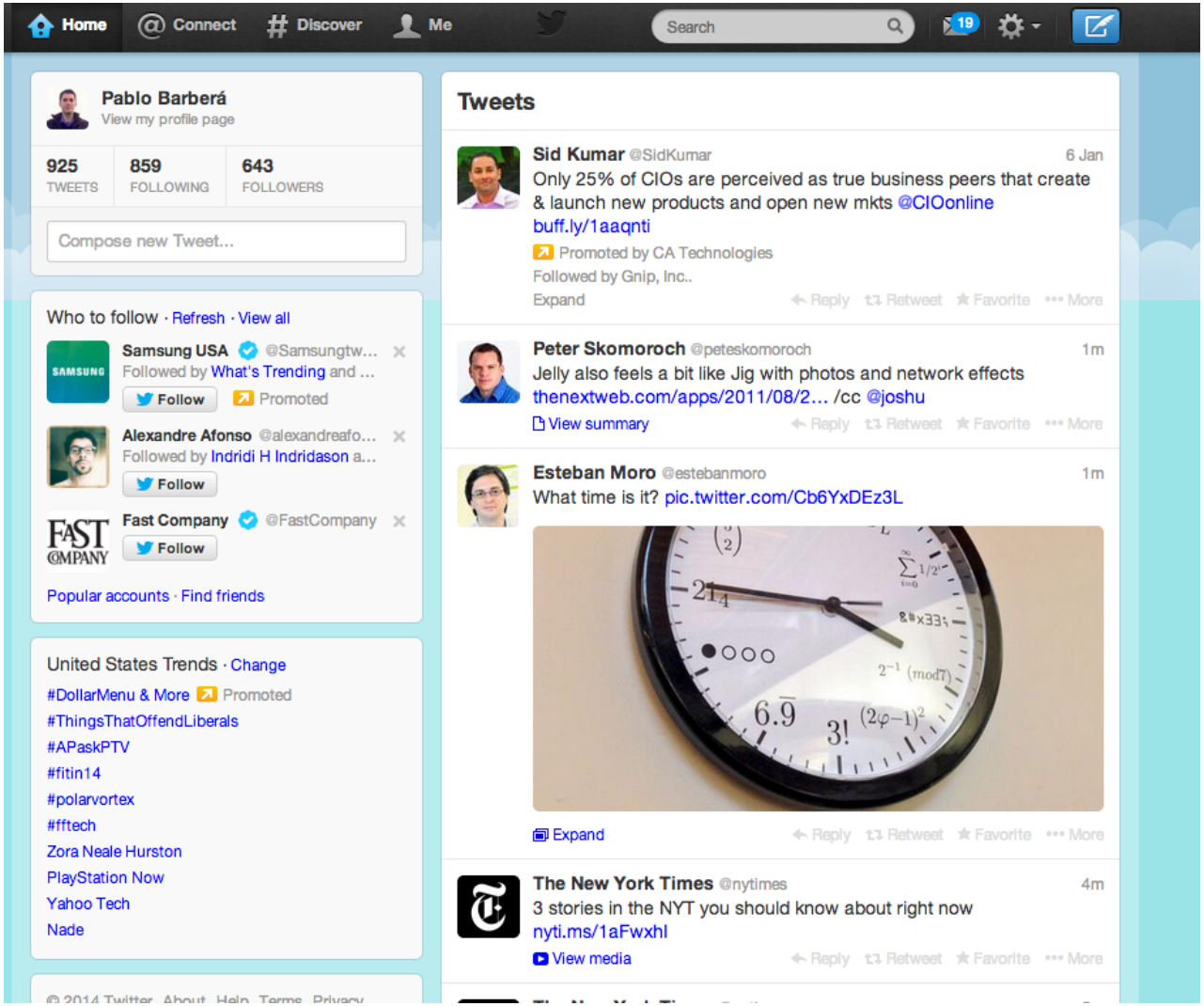
- Twitter: what is it? Main features.
- Introduction to the Twitter API
- Capturing your own collection of tweets
- Analyzing Twitter data
- In-class exercise: collect and analyze your own Twitter data

Twitter

Twitter's numbers

- 200+ million monthly active users
- 400+ 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



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

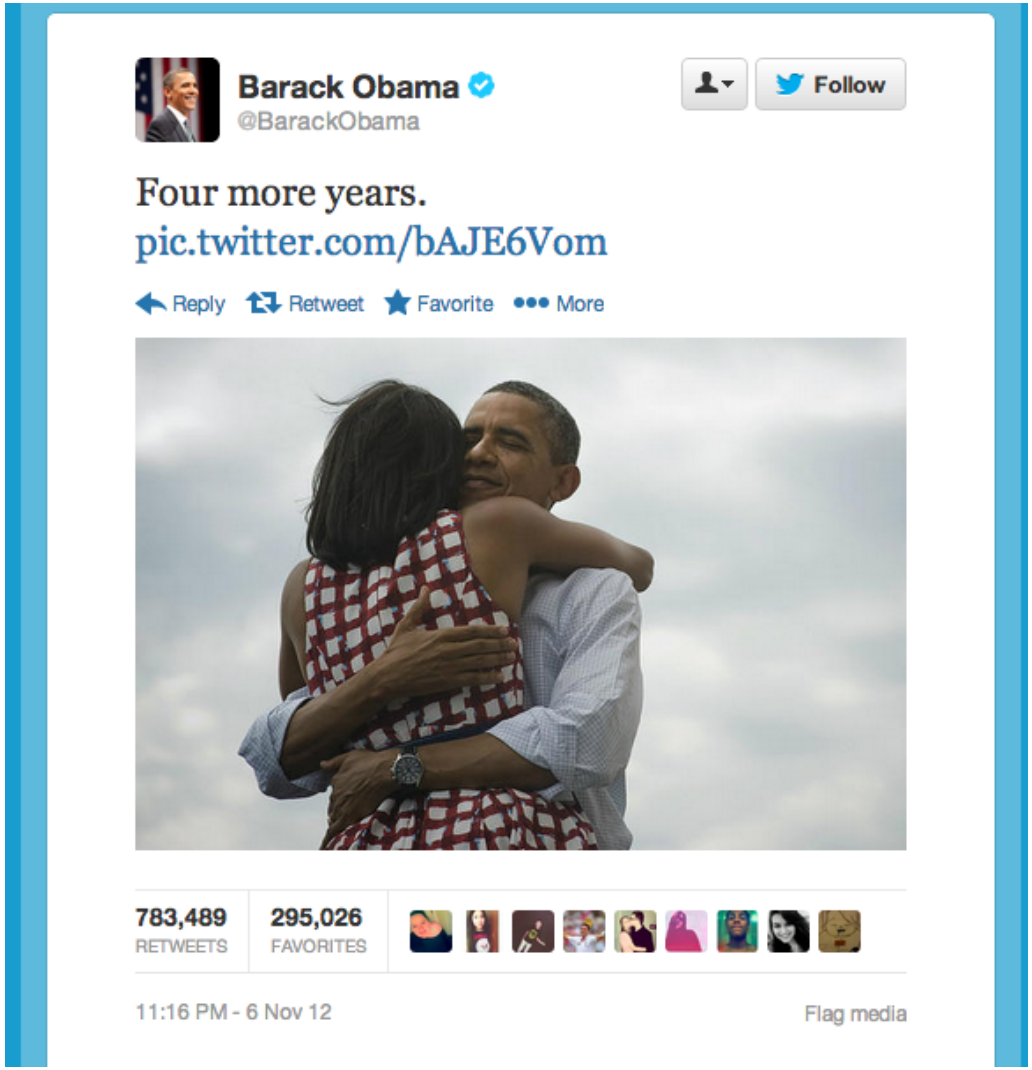
API = *Application Programming Interface*

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

In-class exercise

In-class exercise: collecting and analyzing Twitter data

Create your own Twitter account, follow Prof. Tucker (`@j_a_tucker`), the SMaPP lab (`@SMaPP_NYU`) and me (`@p_barbera`). Then, find an interesting tweet and retweet it.

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 (`pablo.barbera@nyu.edu`) before the end of the day