Structures of Facebook Open Data and Data on Server

Keng-Chi Chang

National Taiwan University

2017-10-17

Facebook Graph API

- Only data related to fan pages are publicly available
  - Posts on fan pages
  - Reactions to / comments of / public shares of these posts
  - Users do the above
  - Fan page likes fan pages
- A full list of variables can be found through Graph API Documentation

Page

Ex. Donald J. Trump

- page_id: “153080620724”
- page_name: “Donald J. Trump”
- page_fan_count: 22,813,525
- page_url: `www.facebook.com/DonaldTrump` or `/[page_id]`
- page_category: “Public Figure”

Post (Link / Photo / Video / Status)

Ex. 9GAG: Official White House Photographer Reveals His Favourite Photos Of Obama

- post_id: “15308620724” or “21785951839_15308620724”
- post_type: “link”
- post_name: “Official White House Photographer Reveals His Favourite Photos Of Obama”
- post_description: “Click to see the pic and write a comment…”
- post_caption: “9gag.com”
- post_link: `http://9gag.com/gag/ajqEV90?ref=fbp`
- post_reactions: 1297326, post_likes: 1149630
- post_comments: 20093, post_shares: 209506
- post_created_time: “2016-11-11T07:35:00+0000” (ISO 8601)

Reaction

LIKE / WOW / HAHA / SAD / ANGRY / THANKFUL

- post_id: “57972945858_10154109988750859”
- user_id: “766918176681835”
- user_name: “Trent Porter”
- reaction_type: “LIKE”

Comment

- comment_id: “10154022206161680” or “10154022159491680_10154022206161680”
- comment_message: “I’m getting really sick of seeing Ann Coulter-Lite’s crazed, glassy-eyed face plastered all over the place, and Trump hasn’t even been sworn in yet.”
- user_id: “100011100251277”
- post_id: “62317591679_10154022159491680”
- comment_created_time: “2016-11-22T05:47:18+0000”

On Server: Page

- Top *1000 pages* talking about Trump & Clinton in August 2016
  - Weight by likes:comments:shares = 1:7:14
  - Include major news outlets, puglic figures, interest groups
- US national *politician* fan page
  - Former (last one), candidate, and present listed in Wikipedia
  - Senators, Representatives, Governors

Note on Pages

- 366,840,068 unique users ever liked a post from above pages in 2015 and 2016
- 29,410,568 unique users ever liked a post from national politicians in 2015 and 2016, we call them *US political users*
- Overlapping pages: Tim Kaine, Bernie Sanders, U.S. Senator Bernie Sanders, Elizabeth Warren, U.S. Senator Elizabeth Warren, Ted Cruz, Rand Paul, Governor Jan Brewer, Al Franken

On Server: Reaction

- 1000 page:
  - Repeated capture every 20 minutes
    - 2016-09-29 to 2016-11-21
    - Record the first timestamp when observing one's reaction
  - *Likes* by *US political users* for posts from 2015-01-01 to 2016-11-30
- Politician: 2015-01-01 to 2016-11-30
  - *Likes* by *US political users* for posts from 2015-05-01 to 2016-11-30
  - Remove reactions on overlapping pages from politician folder
  - For reactions on these pages, go to 1000 page

On Server: Post & Comment

- Post:
  - 1000 page: 2015-01-01 to 2017-04-08
  - Politician: 2015-01-01 to 2016-11-30
- Comment:
  - Comment on post of 1000 page from 2015-01-01 to 2016-11-30

Directories

- Reference: Gentzkow and Shapiro, Chapter 4
- `build`: Data cleaning part
  - `input`: Data download from Google BigQuery
  - `output`: Data after basic cleaning & combining for people to use
- `analysis-[some-project]`: Data analysis part
  - `input`: Symbolic link pointing to `build/output`
  - `code`: Code to run data analysis
  - `temp`: Other data source that feeds into your code, logs, …
  - `output`: Figures, tables, …
- Just copy `analysis-[some-project]` to wherever you want
  - Rename it, links are preserved, set it as working directory
  - Code in `code`, read from `input` and `temp`, export to `output`

Why This Structure?

- No need to copy data, they are extremely huge
- Will not sync large data to your local computer if use Dropbox
- Easy to (and should always) use relative paths in your code

```r
library(tidyverse)
setwd("~/analysis-ideology")
read_csv("input/page/1000-page-info.csv")
```

```python
import os, pandas
os.chdir("~/analysis-ideology")
pandas.read_csv("input/page/1000-page-info.csv")
```

```
cd "~/analysis-ideology"
insheet using "input/page/1000-page-info.csv", clear
```