

# 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](#)

Ex. Donald J. Trump

- page\_id: "153080620724"
  - You can see from `www.facebook.com/[page_id]`
- page\_name: "Donald J. Trump"
- page\_fan\_count: 22,813,525
- page\_url: `www.facebook.com/DonaldTrump`
- page\_category: "Public Figure"

# Post (Link / Photo / Video / Status)

Ex. White House Photographer Reveals Favourite Photos Of Obama

- post\_id: "153080620724" or "21785951839\_153080620724"
  - You can see from `www.facebook.com/[post_id]`
- post\_name: "Official White House Photographer Reveals His Favourite Photos Of Obama"
- post\_message: "Obama is the coolest president in history..."
- post\_description: "Click to see the pic and write a comment..."
- post\_caption: "9gag.com", post\_type: "link"
- 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"
  - You can see from `www.facebook.com/[user_id]`
- reaction\_type: "LIKE"
- user\_name: "Trent Porter"

# Comment

Ex. Comment on "Conway on Trump's Hamilton feud: 'Who is to say that he can't do that' "

- comment\_id: "10154022206161680" or "10154022159491680\_10154022206161680"
  - You can see from `www.facebook.com/[comment_id]`
- 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, public 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

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

```
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
```