

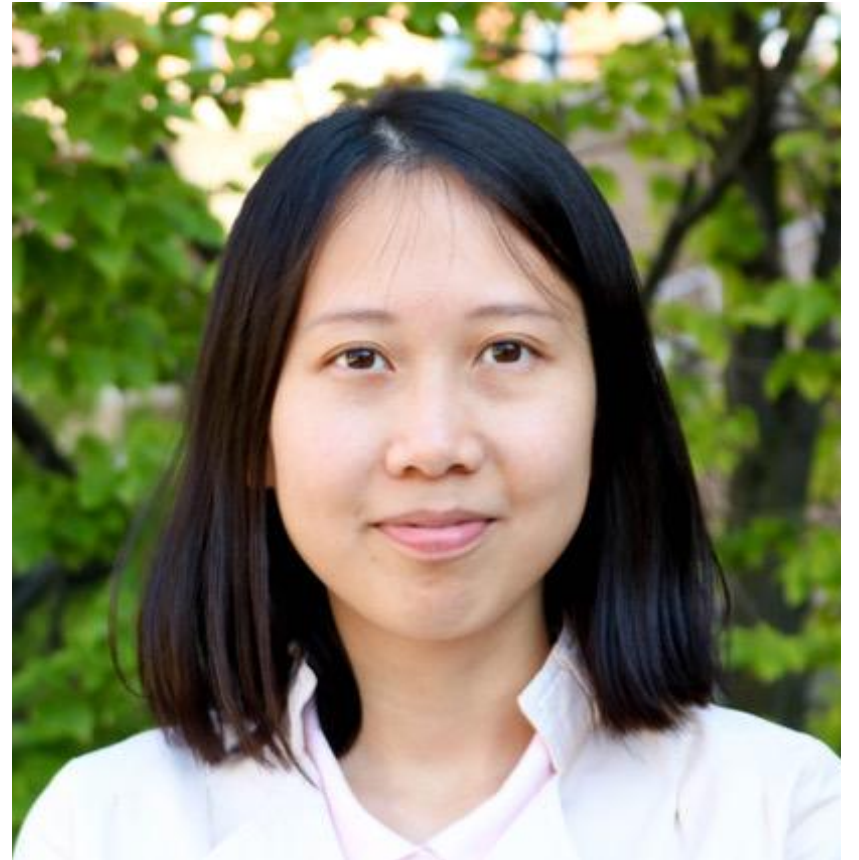
Twitter Day!

A primer to the Twitter API v2 for Academic Research

Chau Tong
Cornell University
October 27, 2021

Chau Tong (she/hers)

- Ph.D. in Mass Communications (UW-Madison)
 - Minor: Quantitative Methods
- Postdoc Associate in Department of Communication
- Collaborators at Cornell: Prof. Drew Margolin, Jeff Niederdeppe
- Specialization: Computational Social Science, Political Communication, Communication Technologies



My research focusing on Twitter

- Chiming in: A computer-assisted analysis of popular musicians' political engagement on Twitter. *Social Media and Society*.
- Disinformation performed: Self-presentation of a Russian IRA account on Twitter. *Information, Communication and Society*.
- Do I sound American? How message attributes of foreign disinformation relate to Twitter engagement (forthcoming). *Computational Communication Research*.

Goals for the workshop

- Informal learning experience!
- Pique your interest
- Help get you started

Agenda

1. Brief intro of Twitter standard API -v1
2. Twitter Academic API or API v2 – today's focus
 - How to apply : No
 - What's available in it: Yes!
 - Which API endpoints to use for your research purpose: Yes!
 - How to get data from the API v2 using R: Yes!
3. Code examples in R

Research using Twitter data

- Twitter data = public tweets provided via API:
 - Tweet-level: Tweet ID, text, creation time, public metrics :
 - Not available: protected tweets, deleted tweets, non-public metrics
 - User-level: User ID, user-name, bio description, image url
 - Not available: birthdate, political affiliation, protected accounts
- To access the APIs: need a developer account and tokens.

Standard API (v1)

- Can use this for:
 - Tweets within the last 7 days (recent search)
 - Filtered stream (real-time tweets with filtering rules)
 - Retrieve timelines, trends

- rtweet package in R is a good package for R users!

rtweet package – main functions

- Get friends/followers of an account:
 - `get_friends()`: limited to 5,000.
 - `get_followers()`: no limit. (need stable Internet connection + time)
- Search tweets containing keywords, phrases:
 - `search_tweets()`: 18,000 per 15 min (there are ways to circumvent this)
- Get timelines:
 - `get_timelines()`: 3,200 max
- Look up:
 - `lookup_users()`
 - `lookup_tweets()`

rtweet package – other functions

- `Stream_tweets()`
- `Lookup_coords()`
- `lat_lng()`
- Support for multiple languages
- Etc.

For more: <https://github.com/mkearney/rtweet>

Twitter API v2 (Academic Twitter)

- **What is it?**

- Full-archive search (historical tweets dating back to 2006)
- Higher cap per month (10M)
- Better filter options, more features (e.g., longer query length, thread conversation, tweet annotations)
- Separate application from API v1
 - Requires academic affiliation
 - Should have a project in mind
 - Might take some time for the approval process (about a week for me)

What's new in Twitter API v2

- Tweet annotations available (more on this later)
- A new **conversation_id** field to track Tweets included in a conversation
- Longer query length & more rules on search + streaming - Advanced search operators like `place_country`, `has:geo` etc in search and `bio_name`, `bio_location` etc in filtered stream
- More on Tweet objects here: <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/introduction>

Tweet annotations

- **Entities**

- **automatically detected**
- Person - Barack Obama, Daniel, or George W. Bush
- Place - Detroit, Cali, or "San Francisco, California"
- Product - Mountain Dew, Mozilla Firefox
- Organization - Chicago White Sox, IBM
- Other - Diabetes, Super Bowl 50

- **Context**

- inferred based on the Tweet text and result in domain and/or entity labels. Context annotations can yield one or many domains.
- 50 domains

3 - TV Shows

46 - Brand Category

79 - Video Game
Hardware

115 - Video Game
Conference

4 - TV Episodes

47 - Brand

84 - Book Music Genre

116 - Video Game
Tournament

6 - Sports Events

48 - Product

85 - Book Genre

117 - Movie Festival

10 - Person

49 - Product Version

86 - Movie

118 - Award Show

11 - Sport

54 - Musician

87 - Movie Genre

119 - Holiday

12 - Sports Team

55 - 56 - Actor

88 - Political Body

120 - Digital Creator

26 - Sports League

58 - Entertainment Personality

89 - Music Album

122 - Fictional Character

27- American Football Game

60 - Athlete

90 - Radio Station

130 - Multimedia Franchise

28 - NFL Football Game

65 - Interests and Hobbies
Vertical

91 - Podcast

132 - Song

35 - Politicians

66 - Interests and Hobbies
Category

92 - Sports Personality

136 - Video Game
Personality

38 - Political Race

67 - Interests and Hobbies

93 - Coach

39 - Basketball Game

40 - Sports Series

68 - Hockey Game

94 - Journalist

137 - eSports Team

DECIDING WHICH ENDPOINTS TO USE IN THE TWITTER API V2

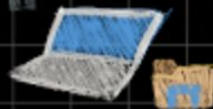
I want Tweets about something that happened in the last week

Recent Search </2/tweets/search/recent>



I want Tweets from the entire historical archive of public Tweets

Full-archive search </2/tweets/search/all>



I want to build the entire conversation thread including replies

Full-archive search </2/tweets/search/all>



I want Tweets from a user's timeline

User Tweet Timeline </2/users/:id/tweets>



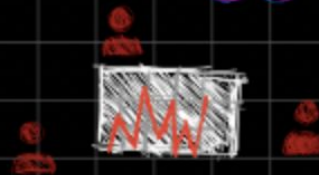
I want to collect Tweets in real-time on a breaking event

Filtered stream endpoint </2/tweets/search/stream>



I want to get follower graph for a user

Follows lookup </2/users/:id/following>
</2/tweets/:id/follows>



I want to build my Tweet dataset using Tweet IDs

Tweet lookup </2/tweets>

ID: 12345*7

<https://github.com/twitterdev/getting-started-with-the-twitter-api-v2-for-academic-research/blob/main/modules/3-deciding-which-endpoints-to-use.md>

academicwitterR package

- Pros:
 - The best R package so far yet for Academic API.
 - Built for academic social media research purposes
 - Solutions to mitigate data loss.
- Cons: for now, does not support filtered stream endpoint.
- Alternatives in Python: twarc, search-tweets-python

Now let's do some coding

Follow-up questions

Feel free to email me:

Chau Tong, PhD

Postdoctoral Associate

Department of Communication

ctt39@cornell.edu