ECPR Methods Summer School:
Automated Collection of Web and Social Data

Pablo Barberá

School of International Relations
University of Southern California
pablobarbera.com

Networked Democracy Lab
www.netdem.org

Course website:
github.com/pablobarbera/ECPR-SC103

APIs

APIs

API = Application Programming Interface; a set of structured https requests that return data in JSON or XML format.

Types of APIs:

1. RESTful APIs: queries for static information at current moment (e.g. user profiles, posts, etc.)
2. Streaming APIs: changes in users' data in real time (e.g. new tweets, new FB posts...)

APIs

API = Application Programming Interface; a set of structured https requests that return data in JSON or XML format.

Types of APIs:

1. RESTful APIs: queries for static information at current moment (e.g. user profiles, posts, etc.)
2. Streaming APIs: changes in users' data in real time (e.g. new tweets, new FB posts...)

Most APIs are rate-limited:

- Restrictions on number of API calls by user/IP address and period of time.

Connecting with an API

Constructing a REST API call:

- ▶ Baseline URL:
  `https://maps.googleapis.com/maps/api/geocode/json`
- ▶ Parameters: `?address=barcelona`
- ▶ Authentication token: `&key=XXXXX`

Response is often in JSON format.

Connecting with an API

Constructing a REST API call:
- ► Baseline URL:
  `https://maps.googleapis.com/maps/api/geocode/json`
- ► Parameters: `?address=barcelona`
- ► Authentication token: `&key=XXXXX`

Response is often in JSON format.

Authentication:
- ► Many APIs require an access key or token
- ► An alternative, open standard is called OAuth
- ► Connections without sharing username or password, only temporary tokens that can be refreshed
- ► `httr` package in R implements most cases (examples)

Other APIs

See CRAN Web Technologies Task View