

Good morning!

Please:

1. Download all files for this lesson
2. Open all **3** notebooks in Jupyter
3. Make a Twitter account (optional)
4. Log in to your Twitter account

Outline

1. APIs:
 - a. Overview
 - b. Example with Twitter
2. Scraping
 - a. Overview
 - b. Examples

What's an API?

- “Application Programming Interface”
 - “Application:” program that does things for humans
 - API: does things for other programs
- Uses
 - Get data
 - Get services

Some things with APIs

- Twitter
 - Get tweets, post them, etc.
- Google
 - Search, translate, NLP...
- Patents [<link>](#)
- New York Times
- Library of Congress
- _____?

Cautions

- Every API is different
- Read the documentation
 - Especially: rate limits, query options
- Google for example code

Neat example with Twitter

www.proporti.onl

Ethics sidebar

Randall Collins.
1998. *Sociology of Philosophies.*

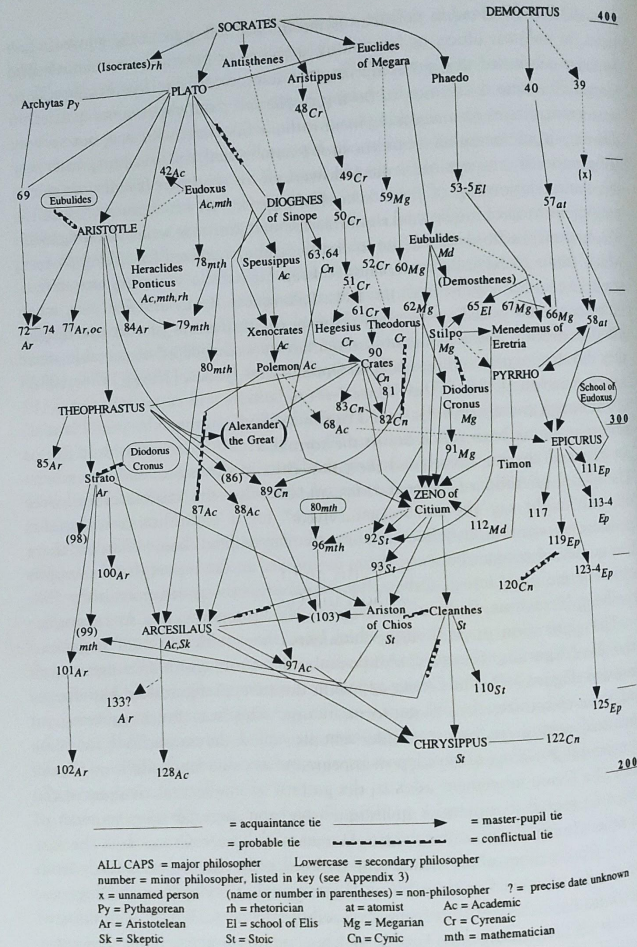


FIGURE 3.4. PROLIFERATION AND RECOMBINATION OF THE GREEK SCHOOLS, 400-200 B.C.E.

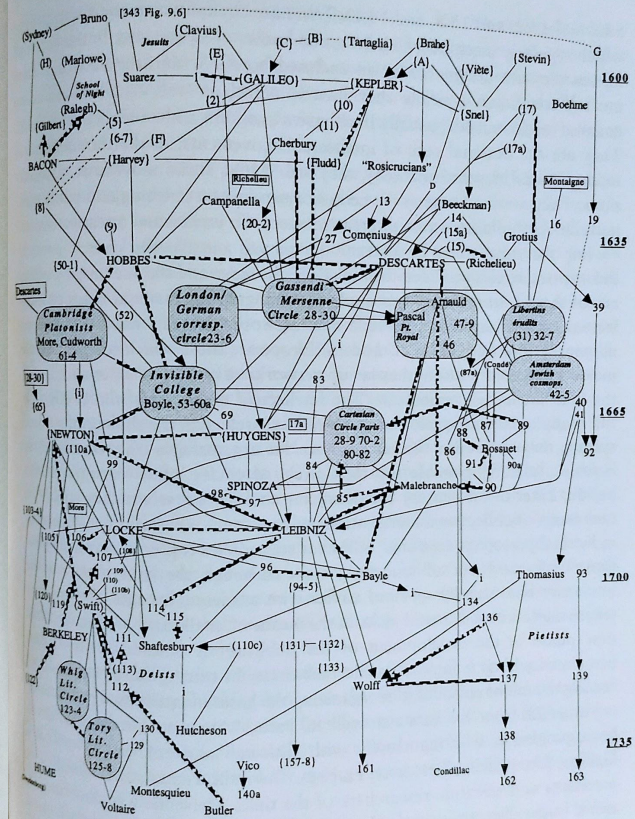


FIGURE 10.1. EUROPEAN NETWORK: THE CASCADE OF CIRCLES, 1600-1735

Ethics sidebar

Gunter Grau. 1995.
Hidden Holocaust.

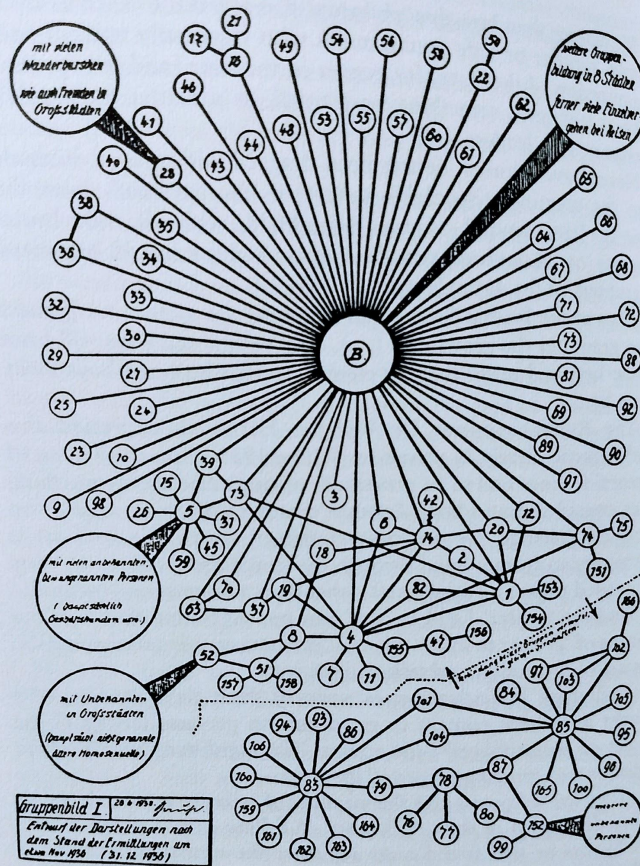


Diagram purporting to show the 'originating centre of a homosexual epidemic' and its supposed spread to other groups and towns by November 1936. From K. W. Gauhl, 'Statistische Untersuchungen über Gruppenbildung bei Jugendlichen mit gleichgeschlechtlicher Neigung unter besonderer Berücksichtigung der Struktur dieser Gruppen und der Ursache ihrer Entstehung' [Statistical Investigation of the Formation of Groups among Young People with Homosexual Tendencies, with Special Reference to the Structure of Such Groups and the Cause of their Emergence], Phil. Dissertation, University of Marburg 1940.

Twitter example

Scraping Overview

- Sometimes, there is no API.
- “Scraping:” converting web pages to usable data



Things one might scrape

- Event information
- Policy statements
- Data tables
- Faculty lists
- Public comments or posts
 - (e.g. on legislation, news)
- _____?

Cautions

1. Use the API (if it exists)
2. Every website is different
3. Read robots.txt
4. Think seriously about ethics
 - a. (OKC debacle, TOS, CAPCHA)
5. BE NICE (or get us all banned...)
6. Recursion is dangerous (exponential growth)

Scraping examples