Twitter for Social Research: Access & Disclosure

Web Archiving & Preservation Working Group: Social Media & Complex Content

Dr Luke Sloan (@DrLukeSloan)
Social Data Science Lab

tweet: @socdatalab
web: socialdatalab.net
Outline

• Background

• COSMOS (Demo)

• Understanding Twitter Data

• Exercise: Evaluating Disclosure Potential

• Discussion
My research focuses on Twitter and how social media data can be used to understand social phenomenon on it’s own, or through data linkage...

- **Who Uses Twitter?** (Sloan et al. 2015. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. Plos One 10(3), article number: e0115545. (10.1371/journal.pone.0115545)

- **Who geotags?** (Sloan and Morgan 2015. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. PLoS ONE 10(11), article number: e0142209. (10.1371/journal.pone.0142209)


- **Validating Proxies with Survey Data** (Sloan 2017. Who Tweets in the United Kingdom? Profiling the Twitter population using the British Social Attitudes Survey. Social Media + Society 3(1) (10.1177/2056305117698981)


Sloan & Quan-Haase (2017)

SAGE Handbook of Social Media Research Methods
The Social Data Science Lab

- Interdisciplinary collaboration between social and computer science
- To date we have been involved in 25 grants amounting to more than £8 million, bringing in over £4.6 million directly to the Lab
- Address the methodological, theoretical, empirical and technical challenges presented by new and emerging forms of data
- Our empirical social data science programme is complemented by a focus on ethics and the development of new methodological tools and technical/data solutions for the UK academic and public sectors
- Example projects include:
  - ‘Crime-sensing’ through social media
  - Predicting the 2015 General Election
  - Communication strategies for Twitter during food scares
  - Demographics and representation on Twitter
  - Detecting hate speech online
  - Cybersecurity, risk and safety
Core Funding (2017-2020)

• Social Data Science Lab: Methods and Infrastructure Development for Open Data Analytics in Social Research

• Funded by Economic & Social Research Council

• £705,050 (Burnap, Williams, Sloan & Rana)
• The Social Data Science Lab maintains and distributes the [ESRC](#) COSMOS Open Data Analytics software

• COSMOS is available at no cost to academic institutions and not-for-profit organisations

• A web version is currently under development...

• Hopefully you requested a download from [http://socialdatalab.net/COSMOS](http://socialdatalab.net/COSMOS)
COSMOS Demo
Understanding Twitter Data

• Do we understand what Twitter data actually is?

• Do we know how the API works?

• Do we understand what is in the JSON?

• A single tweet can come with over 150 associated ‘attributes’!

• Consider the tweet, the user and geography
Understanding Twitter Data

- Why does this matter?
- Confidentiality and/or anonymity
- Secure environment for analysis
- Archiving conditions (and terms of service)
- Sharing and re-use (and replication)
- Problems of linking Twitter and survey data
Exercise: Evaluating Disclosure Potential

• Let’s have a closer look at some of these ‘attributes’

• Taken from Sloan et al. (forthcoming)¹

• 43 attributes parsed by the StreamR package for R²

• Explanation of what is it

• Your job: to evaluate disclosure potential of individual ‘attributes’

• Work in groups, identify high risk areas


Discussion