NCRM Autumn School

Twitter Data Access

Outline

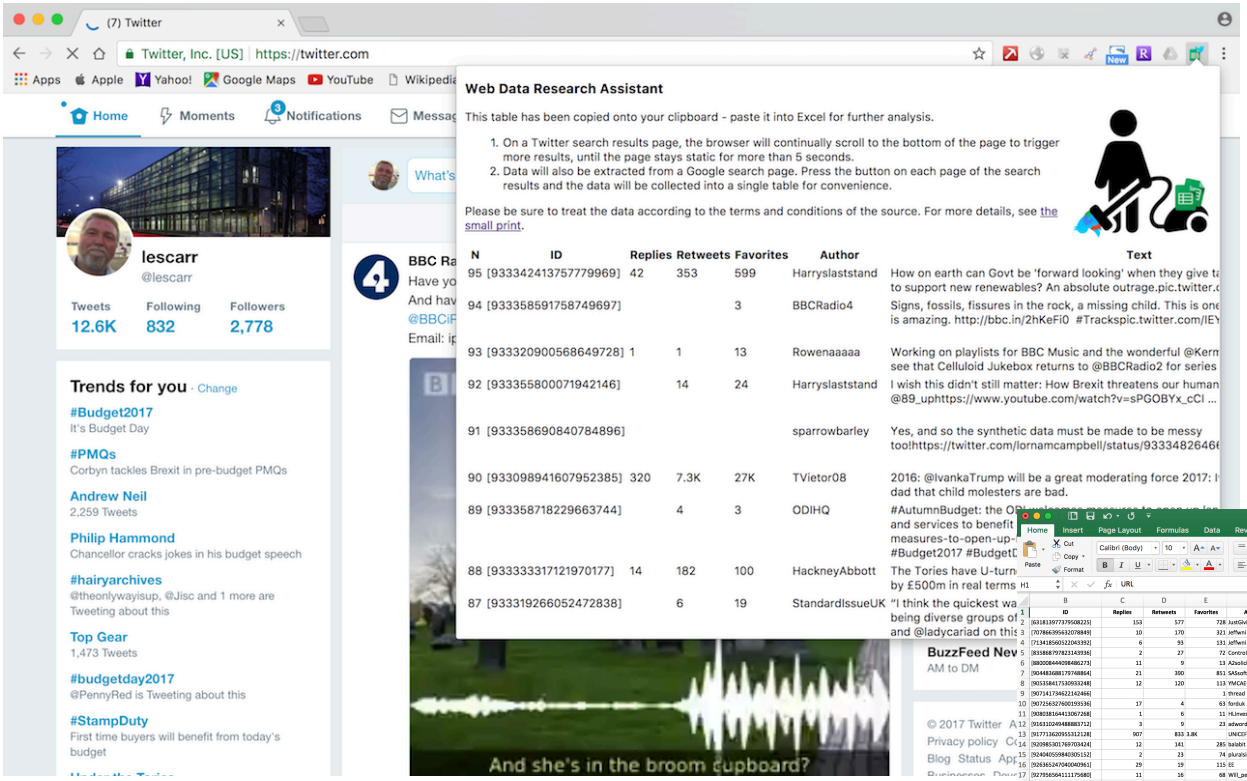
- Social media data has become an increasingly significant source of research data for University research and business stakeholders
  - Of these, Twitter has become the most used in the research and business domains
- Sources of social media data
  - DIY – low/no upfront fees but high technical & programming resources required (Twitter API)
  - Assisted – upfront and ongoing investment required to develop and maintain programs / processes, or rely on open source authors and communities.
  - Professional – (vast) annual payment for access to data gathering and value added analytical platforms.

NB There are many bespoke Twitter downloading tools, especially for highly technical researcher – programmers. The following list focuses on data gathering tools that are more accessible.

Summary of Twitter Data Options

Solution	Cost P/A	Effort Data Gathering	Effort Data Analysis	Comments
Web Data RA	Free	None	High	Extract data from a self-managed Twitter query. Access to historic data.
API	Free	High	High	Program your own access to the official API. Contemporary data only.
Google Spreadsheet	Free	None	High	No programming, high level of technical understanding. Downloads CSV. Contemporary data only.
NodeXL Pro	Modest	None	Medium	Social network quant analysis only. Contemporary data only.
EPrints Twitter Harvester	Free	None	Medium	Easy to use, provides basic analyses during collection, downloads as CSV for use in other analytical tools. Contemporary data only.
Flocker / Flow140	Free	None	Low	Dynamic retweet graph visualisation. Easy to use but limited. Contemporary data only.
Pulsar	£20k	None	Low	Data from Twitter & many social media sources, with many analytics. Download and self-analyse raw data. Access to historic data.
GNIP	£20k	None	High	Twitter data only. No analysis provided. Access to historic data.

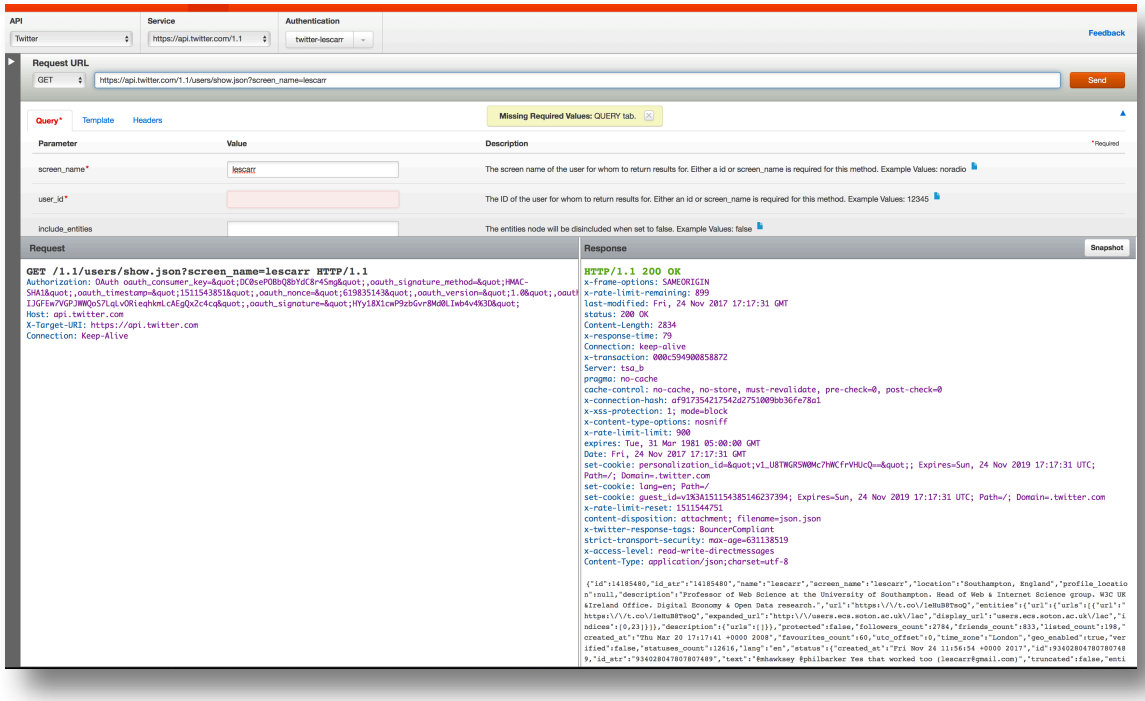
Web Data Research Assistant



- *Scrapes twitter and google data into a spreadsheet*
- *Uniquely allows free historic data capture*
- *No programming required*
- *Browser extension, one-click install*

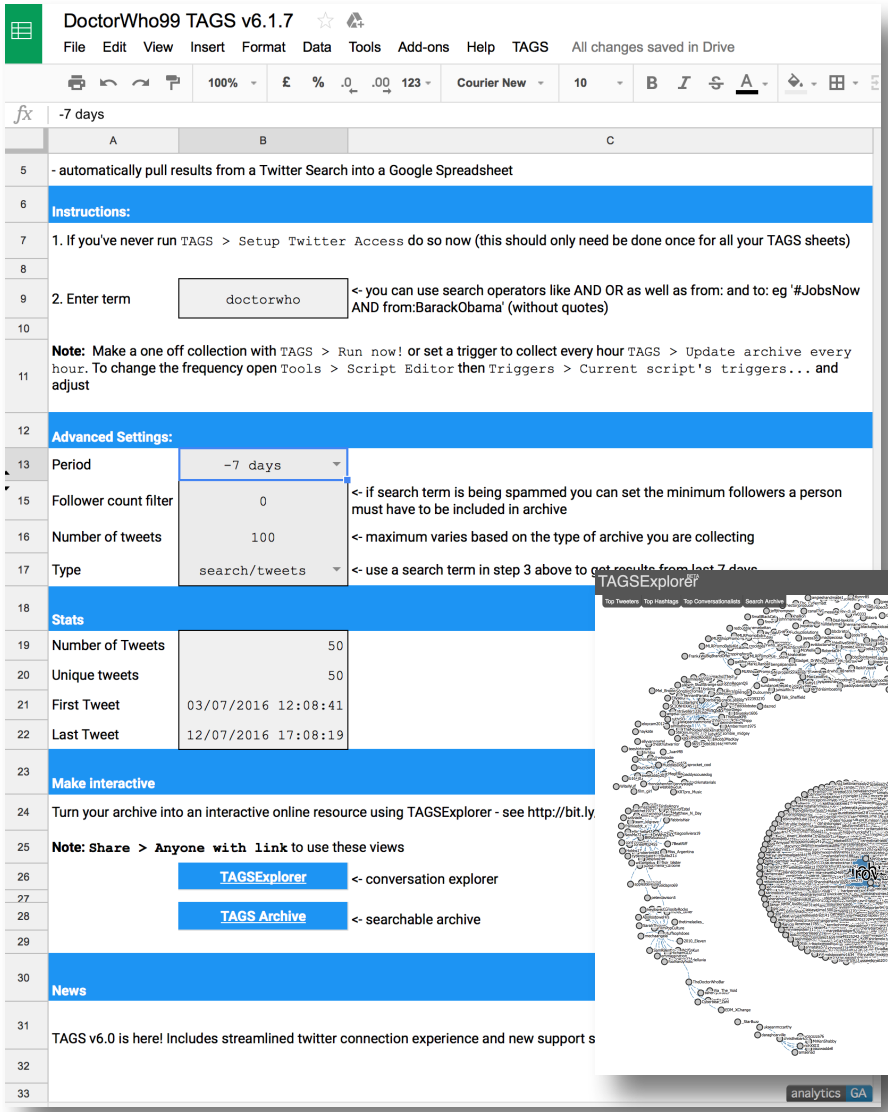
Platform API

- Explicit programmatic control of the platform through Web HTTP/REST protocols.
- Console provides helpful introduction to programming data.



https://apigee.com/console/twitter

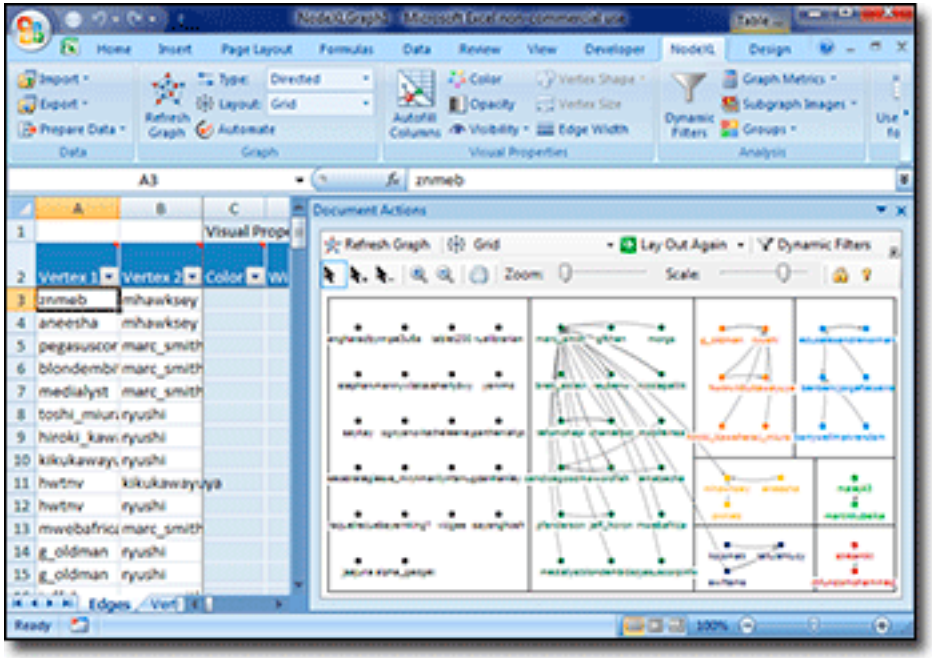
TAGS – Twitteralytics Google Spreadsheet



- tags.hawksey.info
- Easy to control
- Data in accessible spreadsheet form
- Allows current Twitter data to be collected for up to seven days
- Provides network visualisations of interactions between accounts

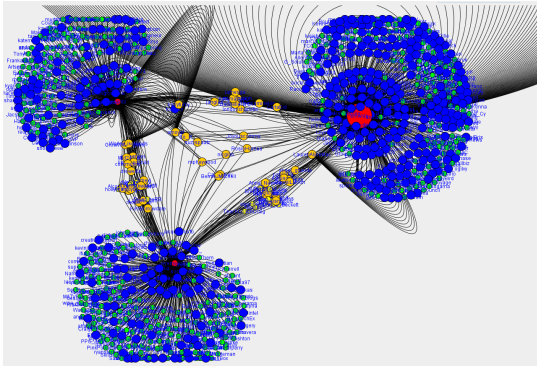
NodeXL

- NodeXL is a free, open-source template for Excel that makes it easy to explore network graphs.
- NodeXL Pro offers additional features providing easy access to social media network data streams, advanced network metrics, and text and sentiment analysis, and powerful report generation.
- NodeXL Pro can create insights into social media streams with just a few clicksx



Repositories for Collecting & Securing Research Data

Ephemeral social media about important societal issues



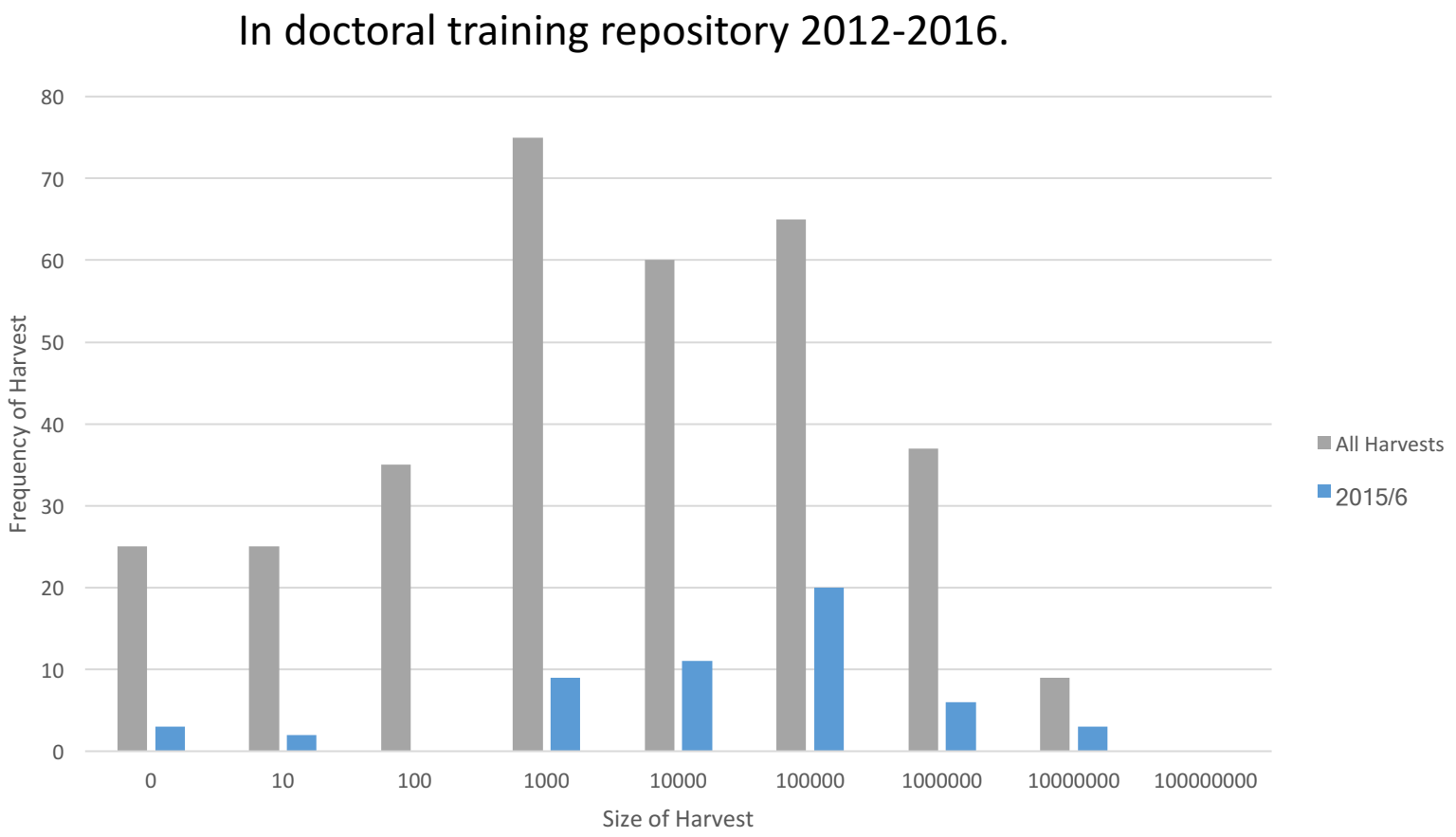
external data analysis

a Web Observatory

Large scale, high bandwidth, added value repository ingest, embedded in the research lifecycle

Available from the EPrints app store

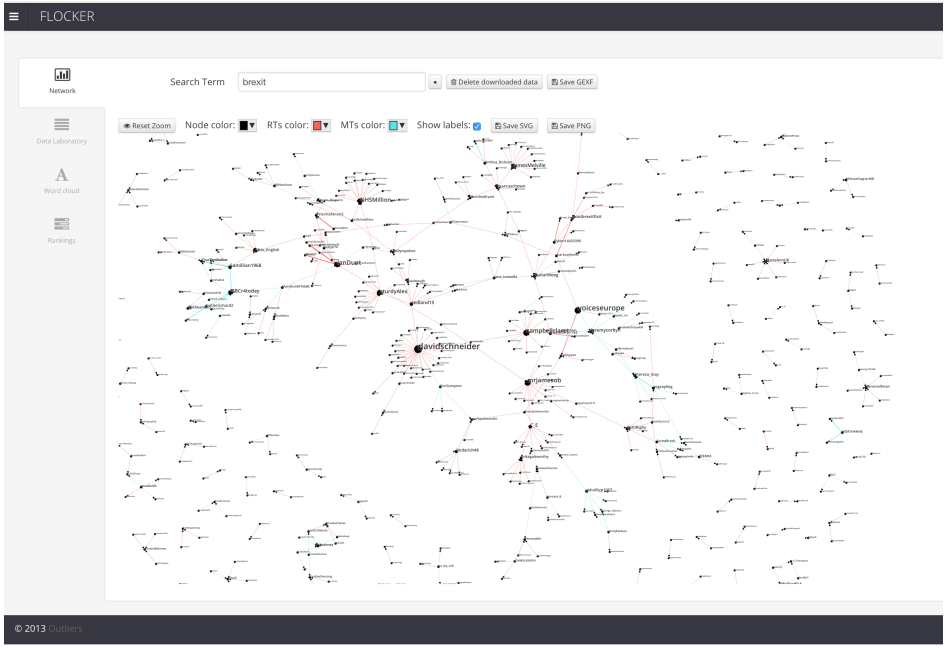
Scale of (EPrints) Twitter Harvesting



Example Captures

- Small example of a twitter stream around UCDavis http://twitter-harvester.ecs.soton.ac.uk/id/tweetstream/2
- or a larger stream about SOPA http://twitter-harvester.ecs.soton.ac.uk/id/tweetstream/9
- Used to capture >300 social media datasets for PhD analysis
  - Total 0.5 billion tweets & counting

Understanding National Conversations



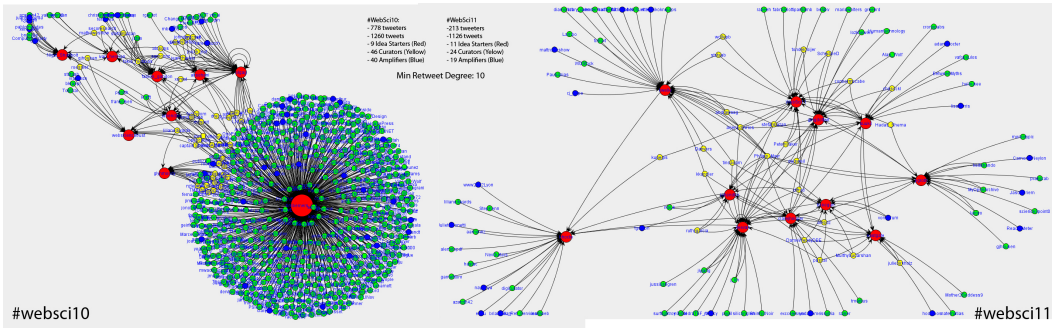
Flocker – dynamically map retweet network on a given topic.

flocker.outliers.es

Flow140 – dynamically map a retweet network on a given topic based on categories of roles.

Tinati, R., Phillipe, O., Pope, C., Carr, L., and Halford, S. (2014) Challenging social media analytics: web science perspectives. In *Proceedings of the 2014 ACM conference on Web science* (WebSci '14). ACM, New York, NY, USA, 177-181. DOI:

http://dx.doi.org/10.1145/2615569.2615690



Pulsar Social Media Analyses