

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

The image shows a browser window with a Twitter profile for @lescar. The profile includes a bio, a profile picture, and statistics: 12.6K Tweets, 832 Following, and 2,778 Followers. Below the profile are sections for 'Trends for you' and 'Top Gear'. A 'Web Data Research Assistant' window is overlaid on the browser, displaying a table of tweet data with columns for ID, Replies, Retweets, Favorites, Author, and Text. The table contains several rows of tweet information, including tweet IDs, reply counts, retweet counts, favorite counts, author names, and the full text of the tweets.

Available from the Chrome Web Store through [this link](#).

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

And she's in the broom cupboard?

Platform API

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

The screenshot displays the Apigee console interface for a REST client. At the top, the API is identified as 'Twitter', the Service as 'https://api.twitter.com/v1.1', and the Authentication as 'twitter-lescar'. The Request URL is set to 'https://api.twitter.com/v1.1/users/show.json?screen_name=lescar'. Below this, there are tabs for 'Query', 'Template', and 'Headers', with a warning for 'Missing Required Values: QUERY tab'. A table lists parameters: 'screen_name' with value 'lescar' and 'user_id' with a redacted value. The 'Request' tab shows the full HTTP request, including headers like 'Authorization: OAuth' and 'Host: api.twitter.com'. The 'Response' tab shows the HTTP/1.1 200 OK status and a large JSON body containing user profile information for 'lescar'.

<https://apigee.com/console/twitter>

TAGS – Twitteranalytics Google Spreadsheet

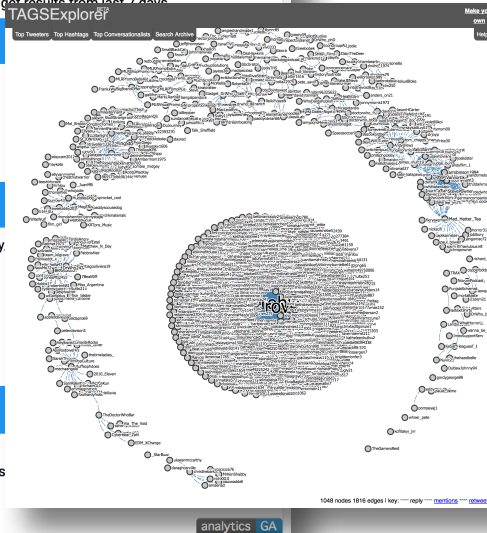
DoctorWho99 TAGS v6.1.7

File Edit View Insert Format Data Tools Add-ons Help TAGS All changes saved in Drive

100% £ % .0 .00 123 Courier New 10 B I S A

-7 days

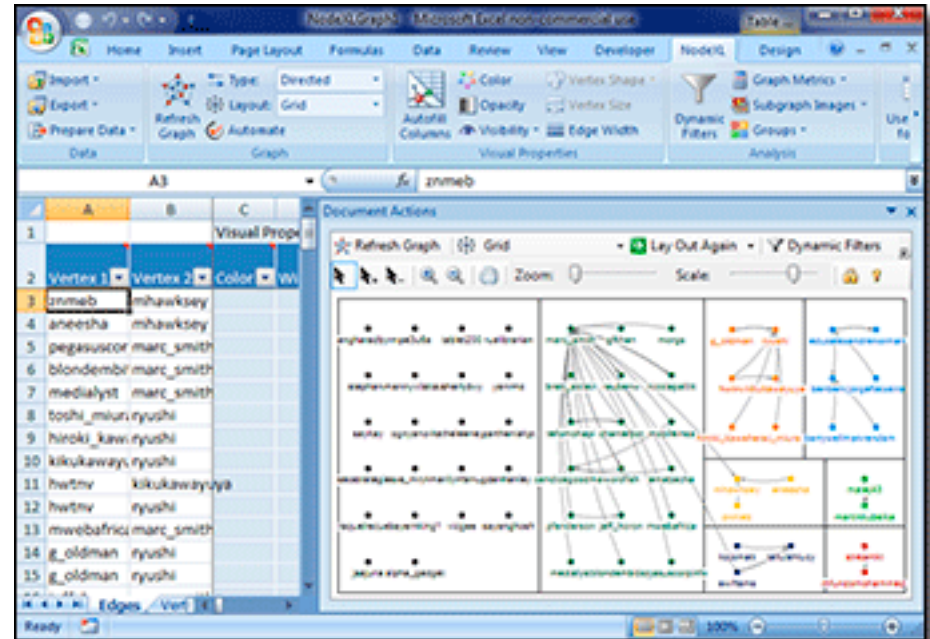
	A	B	C
5	- automatically pull results from a Twitter Search into a Google Spreadsheet		
6	Instructions:		
7	1. If you've never run TAGS > Setup Twitter Access do so now (this should only need be done once for all your TAGS sheets)		
8			
9	2. Enter term	doctorwho	<- you can use search operators like AND OR as well as from: and to: eg '#JobsNow AND from:BarackObama' (without quotes)
10			
11	Note: Make a one off collection with TAGS > Run now! or set a trigger to collect every hour TAGS > Update archive every hour. To change the frequency open Tools > Script Editor: then Triggers > Current script's triggers... and adjust		
12	Advanced Settings:		
13	Period	-7 days	
15	Follower count filter	0	<- if search term is being spammed you can set the minimum followers a person must have to be included in archive
16	Number of tweets	100	<- maximum varies based on the type of archive you are collecting
17	Type	search/tweets	<- use a search term in step 3 above to get results from last 7 days
18	Stats		
19	Number of Tweets	50	
20	Unique tweets	50	
21	First Tweet	03/07/2016 12:08:41	
22	Last Tweet	12/07/2016 17:08:19	
23	Make interactive		
24	Turn your archive into an interactive online resource using TAGSExplorer - see http://bit.ly		
25	Note: Share > Anyone with link to use these views		
26	TAGSExplorer	<- conversation explorer	
27			
28	TAGS Archive	<- searchable archive	
29			
30	News		
31	TAGS v6.0 is here! Includes streamlined twitter connection experience and new support s		
32			
33			



- 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

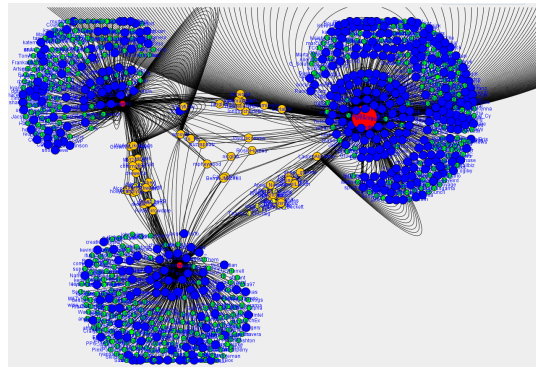
NodeXL

- 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



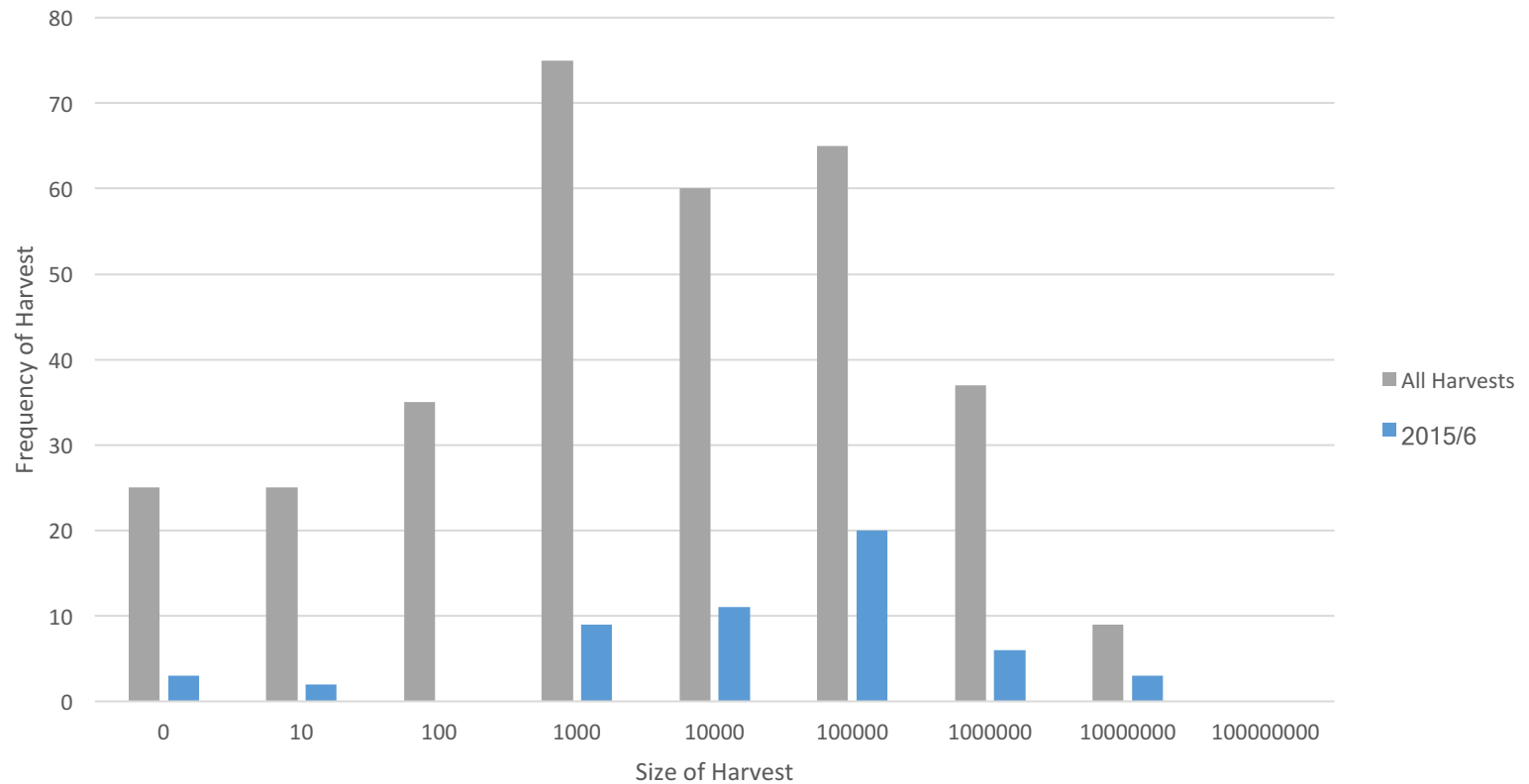
Large scale, high bandwidth, added value repository ingest, embedded in the research lifecycle



Available from the EPrints app store

Scale of (EPrints) Twitter Harvesting

In doctoral training repository 2012-2016.

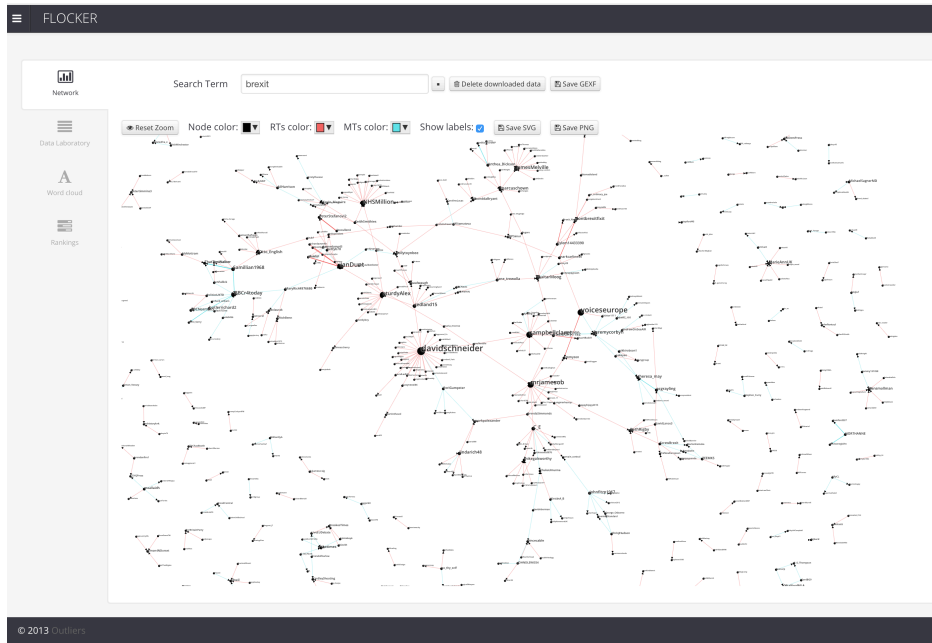


Example Captures

- 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

The screenshot displays the 'eprints repository software' interface. At the top, there are navigation links for Home, About, and Browse. Below this, a user is logged in as 'Unnamed user with email a105@ecs.soton.ac.uk'. The main content area is titled 'Twitter Feed for drwho' and shows a stream of tweets. The tweets are in chronological order and include text, user avatars, and timestamps. For example, one tweet from 'phantommasked' says 'It's not everyday you wished you lived in the Media Cascade to escape the realities of real life. #sopa #drwho'. Another tweet from 'ScottaraFR' says 'Sontaran nurses often find work at SontarBucks. Extra milk like last time? #DrWho'. The interface also features a 'Top Tweepers' section with a grid of user avatars and names, a 'Top Hashtags' section with a list of hashtags and their counts, a 'Top Mentions' section, and a 'Top Links' section with a list of URLs. At the bottom, there is a 'Tweet Frequency' section showing a bar chart of tweet counts over time, with a peak in 2011-10-06.

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., Phillippe, 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>

