

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

Web Data Research Assistant

This table has been copied onto your clipboard - paste it into Excel for further analysis.

1. On a Twitter search results page, the browser will continually scroll to the bottom of the page to trigger more results, until the page stays static for more than 5 seconds.

2. Data will also be extracted from a Google search page. Press the button on each page of the search results and the data will be collected into a single table for convenience.

Please be sure to treat the data according to the terms and conditions of the source. For more details, see the [small print](#).

N	ID	Replies	Retweets	Favorites	Author	Text
95	[933324213757779969]	42	353	599	Harryslaststand	How on earth can Govt be 'forward looking' when they give to support new renewables? An absolute outrage.pic.twitter.com/...
94	[933358591758749697]		3		BBCRadio4	Signs, fossils, fissures in the rock, a missing child. This is one is amazing. http://bbc.in/2hkeF0 #Trackspic.twitter.com/IEY...
93	[933320900568649728]	1		13	Rowenaaaaa	Working on playlists for BBC Music and the wonderful @kern see that Celluloid Jukebox returns to @BBCRadio2 for series
92	[933355800071942146]		14	24	Harryslaststand	I wish this didn't still matter: How Brexit threatens our human @89_uhhttps://www.youtube.com/watch?v=sPG0BYx_cCI ...
91	[933358690840784896]				sparrowbarley	Yes, and so the synthetic data must be to be messy toolhttps://twitter.com/ornamcambell/status/9333482646...
90	[933098941607952385]	320	7.3K	27K	TVietor08	2016: @IvankaTrump will be a great moderating force 2017: I dad that child molesters are bad.
89	[933358718229663744]	4	3		ODIHQ	#AutumnBudget: the ODIHQ and services to benefit measures-to-open-up #Budget2017 #Budget...
88	[93333337121970177]	14	182	100	HackneyAbbott	The Tories have U-turn in real terms
87	[933319266052472838]	6	19		StandardIssueUK	"I think the quickest way being diverse groups of and @ladycarriand on this"

And she's in the broom cupboard?

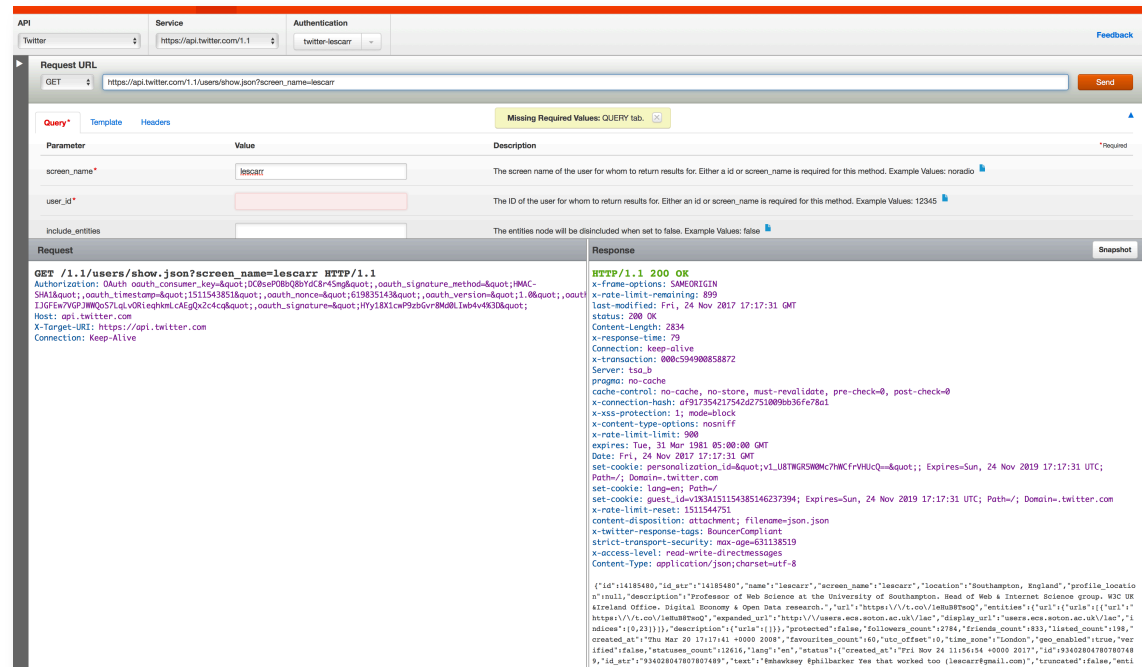
Available for the Chrome Web Store through [this link](#).

- Scrapes twitter and google data into a spreadsheet
- Uniquely provides 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 – Twitteranalytics Google Spreadsheet

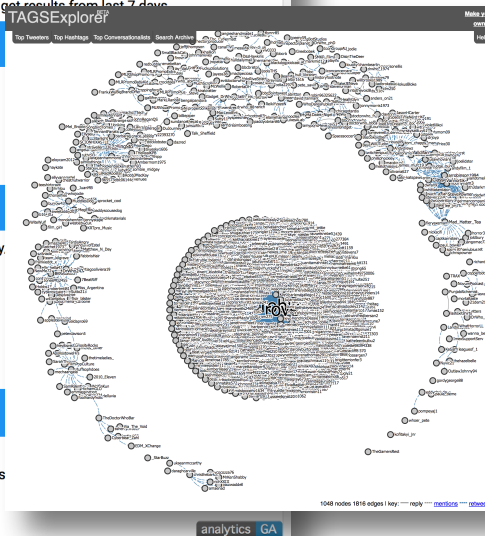
DoctorWho99 TAGS v6.1.7

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

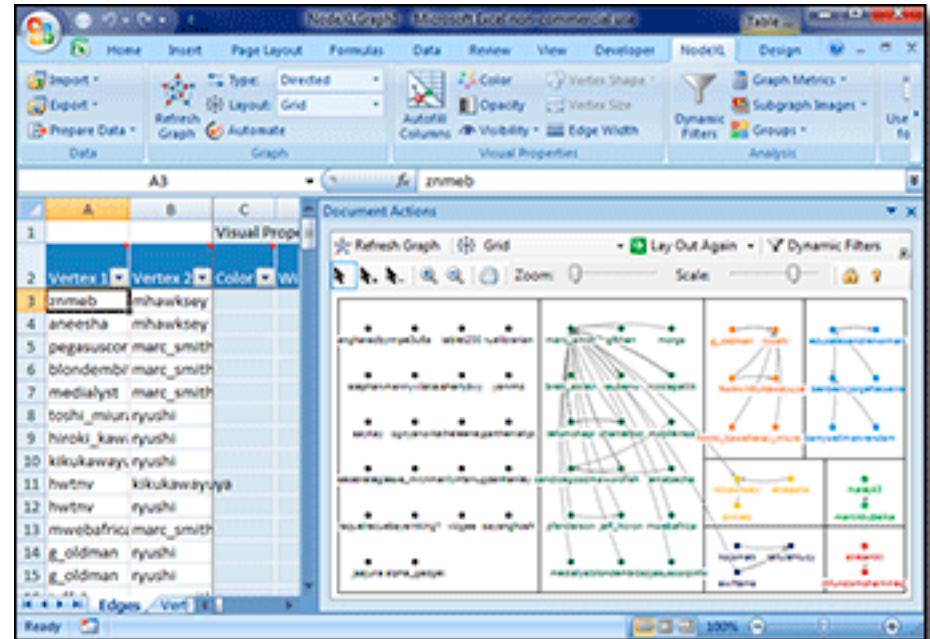
	A	B	C
5	- automatically pull results from a Twitter Search into a Google Spreadsheet		
6	<b>Instructions:</b>		
7	1. If you've never run TAGS > Setup Twitter Access do so now (this should only need to be done once for all your TAGS sheets)		
8			
9	2. Enter term	doctorwho	<- you can use search operators like AND OR as well as from: and to: eg '#JobsNow AND from:BarackObama' (without quotes)
10			
11	<b>Note:</b> 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	<b>Advanced Settings:</b>		
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	<b>Stats</b>		
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	<b>Make interactive</b>		
24	Turn your archive into an interactive online resource using TAGSExplorer - see <a href="http://bit.ly">http://bit.ly</a>		
25	<b>Note:</b> Share > Anyone with link to use these views		
26	<a href="#">TAGSExplorer</a> <- conversation explorer		
27			
28	<a href="#">TAGS Archive</a> <- searchable archive		
29			
30	<b>News</b>		
31	TAGS v6.0 is here! Includes streamlined twitter connection experience and new support s		
32			
33			



- [tags.hawksey.info](http://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 clicks

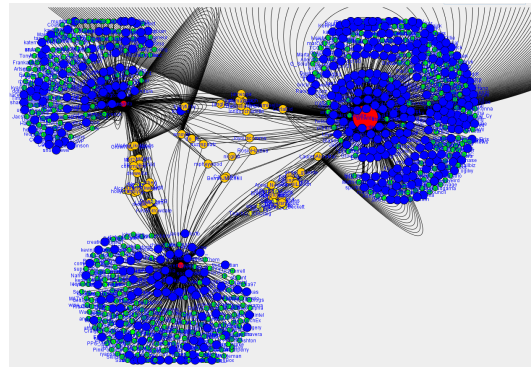




# Repositories for Collecting & Securing Research Data



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

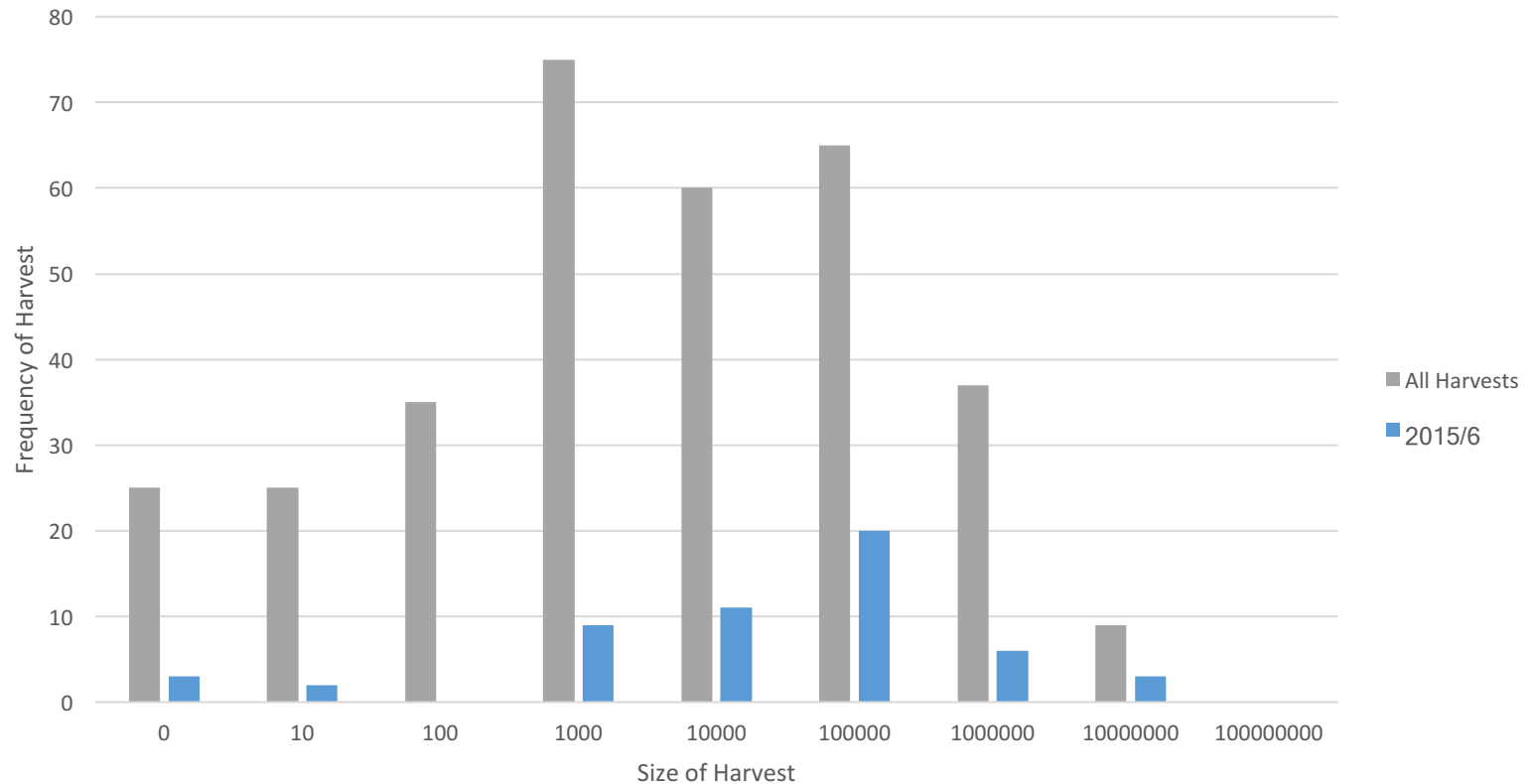


Available from the EPrints app store



# Scale of (EPrints) Twitter Harvesting

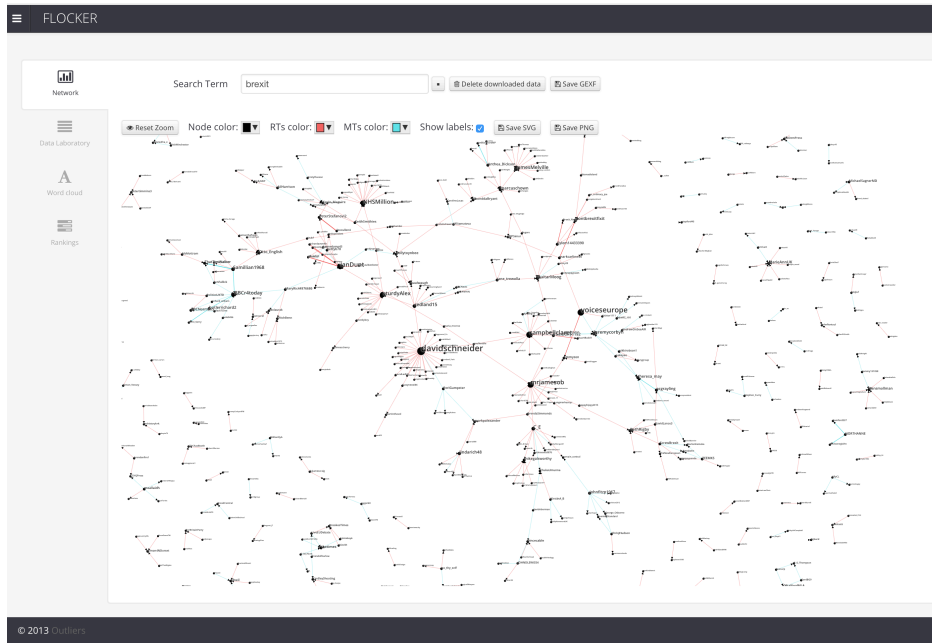
In doctoral training repository 2012-2016.



- Used to capture >300 social media datasets for PhD analysis
  - Total 0.5 billion tweets & counting

[e]prints

# Understanding National Conversations



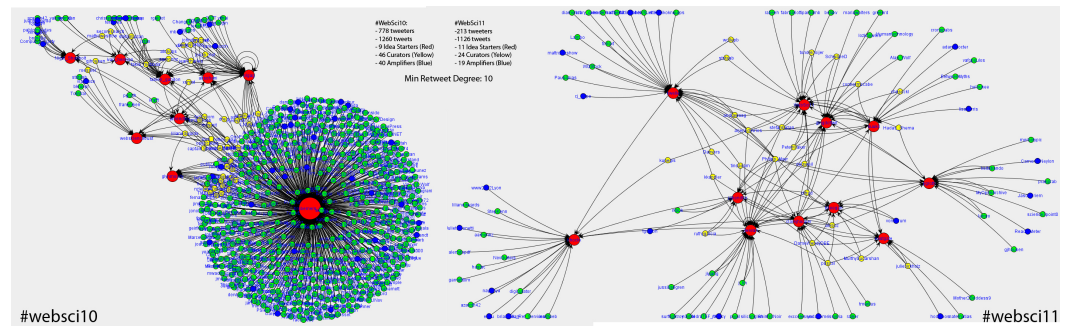
Flocker – dynamically map  
retweet network on a given topic.

*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



24,780 Results (0 Facebook)

- [illegible]

