

Using Twitter as a data source: an overview of social media research tools (updated for 2017)

LSE blogs.lse.ac.uk/impactofsocialsciences/2017/05/08/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-updated-for-2017/

5/8/2017

*Following his initial post on this topic in 2015, **Wasim Ahmed** has updated and expanded his rundown of the tools available to social scientists looking to analyse social media data. A number of new applications have been released in the intervening period, with the increasing complexity of certain research questions also having prompted some tools to increase their data retrieval functionalities. Although platforms such as Facebook and WhatsApp have more active users, Twitter's unique infrastructure and the near-total availability of its data have ensured its popularity among researchers remains high.*



This post is aimed at social sciences researchers who want to capture and analyse social media data, and it provides a useful collection of resources related to methods and practical tools which can be used for this purpose.

A lot has changed since I published my [2015 edition](#) of this post, with even more software applications with the function of retrieving and analysing social media data having been released. Additionally, a number of social listening tools have continued to gain popularity among digital marketers looking to gain insight from consumers.

There remains a number of different methods of analysing social media data. Take text analytics, for example, which can include using [sentiment analysis](#) to place bulk social media posts into categories of a particular topic, such as positive, negative, or neutral. Or [machine learning](#), which can automatically assign social media posts to a number of different topics.

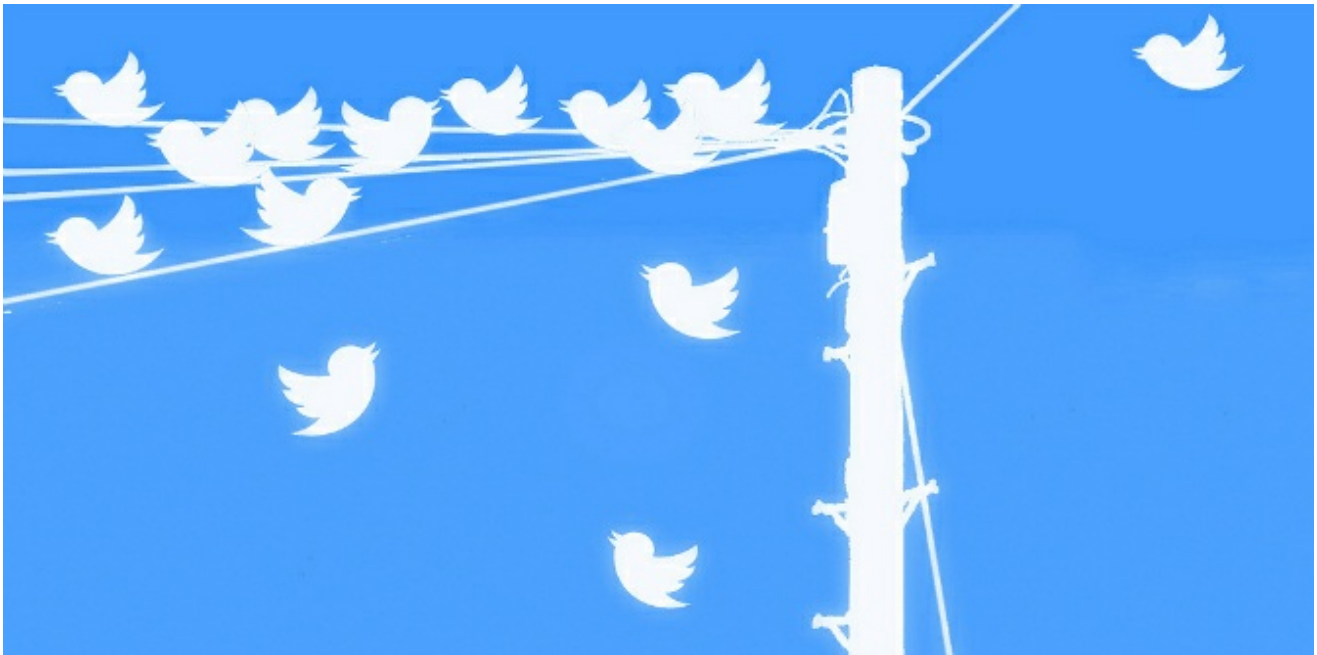


Image credit: [Multiple Tweets Plain](#) by mkhmarketing. This work is licensed under a [CC BY 2.0 license](#).

There are other methods such as [social network analysis](#), which examines online communities and the relationships between them. A number of qualitative methodologies also exist, such as [content analysis](#) and [thematic analysis](#), which can be used to manually label social media posts.

In industry, there has been much focus on gaining insight into users' personalities, through services such as [IBM Watson's Personality Insights service](#), for instance. This uses linguistic analytics to derive intrinsic personality insights, such as emotions like anxiety, self-consciousness, and depression. This information can then be used by marketers to target certain products; for example, anti-anxiety medication to users who fit the personality characteristic of being anxious. A list of personality models can be seen [here](#).

Computational methods can often save time for researchers dealing with large datasets or looking to combine efforts; i.e. humans and machines working together to tackle and analyse data. I would highly recommend reading the following paper, "[Social media analytics: a survey of techniques, tools and platforms](#)" (Batrinca and Treleven, 2015), which provides an overview of some of the methods that can be used to analyse social media data.

In both academia and industry there has been a shift towards research projects and research questions which require more than the simple retrieval of data. More complex questions are being asked which require access to more metadata. So a number of tools have started to increase their data retrieval functionalities for the number of data points that can be retrieved.

For my PhD work I reviewed many methods and opted to use a number of computational techniques to locate and eliminate duplicate and near-duplicate tweets to reduce the volume of data I was working with. I used [DiscoverText](#) to do this. I then applied the methodology of [thematic analysis](#), which involved reading through thousands of tweets in order to generate nodes and themes from them. [Read more on my approach here](#).

Popularity of Twitter

The popularity of using Twitter for social media research, both in academia and in industry, remains high; no other platform has attracted as much attention from academics. However, Twitter is not the most popular platform in terms of monthly active users, being ranked at eighth in the overall list (see Figure 1). Facebook and WhatsApp are the top two. However, many of the platforms with the highest number of monthly active users do not make their data available on a similar scale to Twitter.

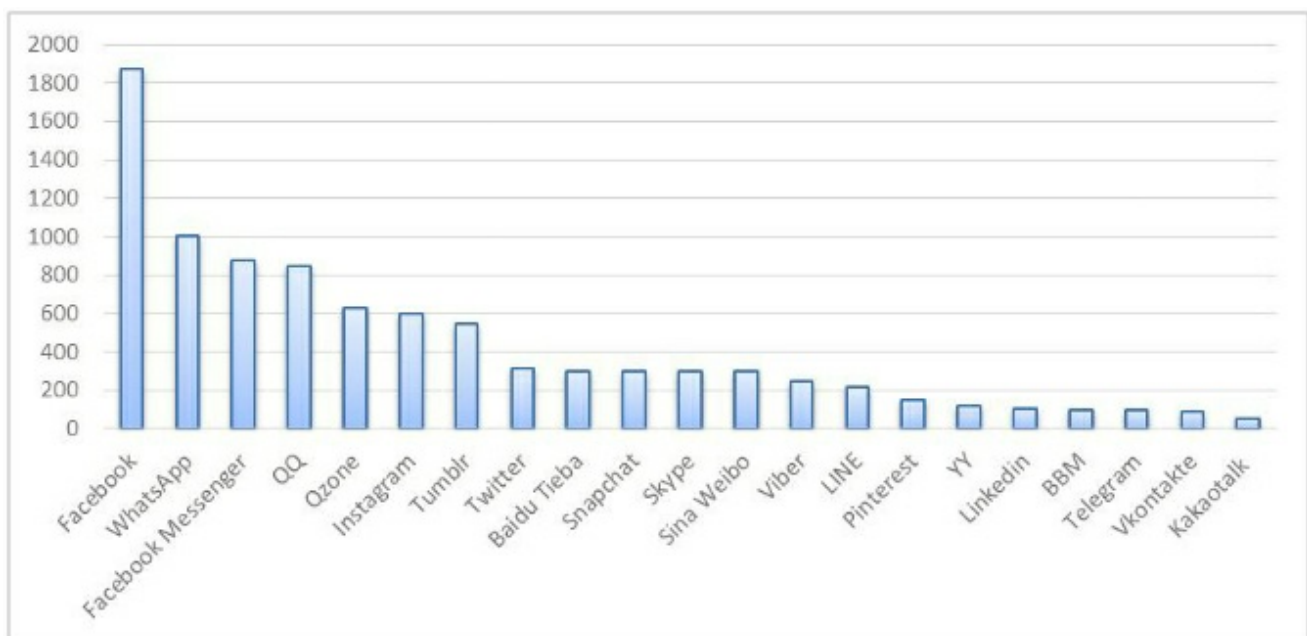


Figure 1: Number (in millions) of monthly active users across social media platforms. Created using data powered by [statista](#).

It can be argued that there is no other social media platform with an infrastructure like Twitter. Twitter is unique in the sense that it has an infrastructure which allows any user to be able to follow another user, and it provides almost 100% of

its data through its APIs. With such a large number of monthly active users, Twitter is likely to remain popular for social media and industry research.

Developments on ethics and training

When I originally published [my 2015 post](#), I received a number of queries and concerns related to the ethical implications of using social media data. In 2016, the [Academy of Social Sciences](#) and the [NSMNSS](#) network held an event solely focused on social media research ethics (you can read more about it [here](#)). You can also read a follow-up post of mine related to [social media research ethics](#), and view a [recorded conference presentation](#) in which I discuss the ethical challenges of my project.

After the previous post was published, issues and concerns were raised over social scientists potentially lacking the skills to analyse social media data. It has been nice to see a number of training events have been held in order to upskill social scientists. For example, the [Social Research Association](#) and the [NSMNSS](#) network held an event which provided an introduction to social media tools (more about that [here](#)).

So, what are some of the tools available to social scientists looking to retrieve and analyse social media data? In the table below I provide an overview of some the tools that require no prior technical and/or programming skills, updated and expanded for 2017:

An overview of tools for 2017

Tool	OS	Download and/or access from	Platforms*
Audiense (offers 14 day trial)	Web-based	https://buy.audiense.com/trial/new	Twitter
Boston University Twitter Collection and Analysis Toolkit (BU-TCAT)	Web-based	http://www.bu.edu/com/research/bu-tcat	Twitter
Chorus (free)	Windows (Desktop advisable)	http://chorusanalytics.co.uk/chorus/request_download.php	Twitter
COSMOS Project (free)	Windows MAC OS X	http://socialdatalab.net/software	Twitter
DiscoverText (offers 3 day trial)	Web-based	http://discovertext.com	Twitter Facebook Blogs Forums Online news platforms Ability to import

Tool	OS	Download and/or access from	Platforms*
Echosec	Web-based	https://www.echosec.net	Instagram Twitter Foursquare Panoramio AIS Shipping Sina Weibo Flickr YouTube VK
Followthehashtag	Web-based	http://www.followthehashtag.com	Twitter
IBM Bluemix	Web-based	https://www.ibm.com/cloud-computing/bluemix	Twitter
Mozdeh	Windows (Desktop advisable)	http://mozdeh.wlv.ac.uk/installation.html	Twitter
Netlytic	Web-based	https://netlytic.org	Twitter Facebook YouTube Instagram RSS Feed
NodeXL	Windows	http://nodexl.codeplex.com	Twitter YouTube Flicker
NVivo	Windows and MAC	http://www.qsrinternational.com/product	Twitter Ability to import
Pulsar Social	Web-based	http://www.pulsarplatform.com	Twitter Facebook topic data Online blogs
SocioViz	Web-based	http://socioviz.net	Twitter
Trendsmap	Web-based	https://www.trendsmap.com	Twitter
Twitonomy	Web-based	http://www.twitonomy.com	Twitter

Tool	OS	Download and/or access from	Platforms*
Twitter Archiving Google Spreadsheet (TAGS)	Web-based	https://tags.hawksey.info	Twitter
Visibrain	Web-based	http://www.visibrain.com	Twitter
Webometric Analyst	Windows	http://lexiurl.wlv.ac.uk	Twitter (with image extraction capabilities) YouTube Flickr Mendeley Other web resources

***It is always best to check with the developers of tools as there may be additional platforms that they can access. Moreover, some tools provide users with the ability of importing data into the applications from external sources.**

A number of the tools provided in the table have been tested and used by me over a number of years, and the vast majority of these chiefly handle data from Twitter. It would be nice to have academic and social listening tools to retrieve data from other social media platforms, such as Facebook, Instagram, and Amazon, and also [dark social media platforms](#) such as WhatsApp. However, this may not be possible because these applications are not likely to provide all of their data to developers as Twitter does. Moreover, there may be ethical implications of accessing data from dark social media platforms.

Other applications are available but these require programming knowledge and/or were not tested as part of this post. These include:

- [DMI-TCAT \(free\)](#)
- [Crimson Hexagon \(commercial\)](#)
- [Brandwatch \(commercial\)](#)
- [Leximancer](#)

Moreover, there are a number of advanced data analysis and statistical applications which can be used to analyse social media data, such as:

- [R](#)
- [SPSS](#)
- [KNIME](#)
- [Weka](#)
- [Tableau](#)
- [Gephi](#)

These packages should be researched when deciding which application is to be used for a project. I'd also like to mention

[The Digital Methods Initiatives list of tools](#) , and [Ryerson University's list of tools from its Social Media Lab](#).

In future, we should begin to ask questions regarding the types of research made possible by using tools that do not require end users to hold technical knowledge. Moreover, we should seek to better understand the types of questions more technical tools can address. Consequently, developers of tools should seek to liaise with social scientists at the development phase, to allow for the possibility of new features based on social sciences research questions.

[Phillip Brooker](#), Research Associate at the University of Bath, has noted that it is important to understand how software packages work in order for researchers to better inform their research practices. I highly recommend reading Phillip's entry on the [NSMNSS blog](#), about the [Programming as Social Science \(PaSS\)](#) network he has helped to establish.

Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.

About the author

[Wasim Ahmed](#) is a PhD candidate at the Information School at the University of Sheffield and an expert social media consultant at [Sonic Social Media](#). Wasim has a very successful [research blog](#), which includes posts about key trends and issues within social media but also covers more practical posts on using tools to capture and analyse social media data. Wasim is a keen Twitter user ([@was3210](#)), and will be happy to answer any technical (or non-technical!) questions you may have.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.