Using Twitter as a data source an overview of social media research tools (2021)

In this edition of **Wasim Ahmed's** *long-running series on using Twitter data as a research tool, Wasim considers the significance of the newly introduced Twitter 'academic research product track' and the different ways in which Twitter and social media data have been used to analyse the global COVID-19 pandemic, as well as updating his regular list of resources for researchers looking to use Twitter data in their research.*

When I wrote the original version of this post back in 2015, and the revised versions in 2017 and 2019, I wasn't sure how long Twitter would provide access to its data. This was because after a string of public scandals other platforms such as Facebook had been closing or limiting access.

Fast-forward to 2021, and something big has happened within the social media research space. Twitter has released a new product track, the *'academic research product track'*. This allows academic researchers **free access** to the complete archive of historical public tweets (by historical data we mean tweets posted in the past). This is significant news because for many researchers without a large budget or with limited time, historical data has until now been out of reach.

You can read more about the launch of this product track here. Some of the key benefits compared to what was available before is that the academic product track allows researchers to pull in 10,000,000 tweets (yes, 10 million!) per month, it also provides free access to the full-archive search.

To gain access to this Academic Research Product Track you would need to complete a developer application. Jessica Garson, Twitter Developer Advocate, has put together this tutorial of getting started with R and the Twitter API. My research into this product track has identified that the following python client provides access to the V2 academic API as well as twarc2. However, both do require programming knowledge.

Twitter has to be given great credit for launching this track and the work that has been put into it. There were fears that access could be cut back at any time a move which would have left this data to be analysed exclusively by Twitter or other private entities. However, these fears have (for now) receded, because this new academic track provides the strongest indication yet that Twitter is keen to continue allowing academics access to data.

In my previous posts, I have dealt with information on methods that researchers can use to analyse data, and these remain relevant. However, over the past year the greatest change for social media and Twitter based research, rather than being technological, has been social. The impact of the COVID-19 pandemic has created vast amounts of data on Twitter and in so doing opened up entirely new avenues for social media research.

Twitter data has been used to study COVID-19 from a range of perspectives, generating many novel insights. Notably, it has been used to identify misinformation networks, to examine public views towards various issues related to COVID-19, it has also been increasingly used by scholars to conduct epidemiological related research.
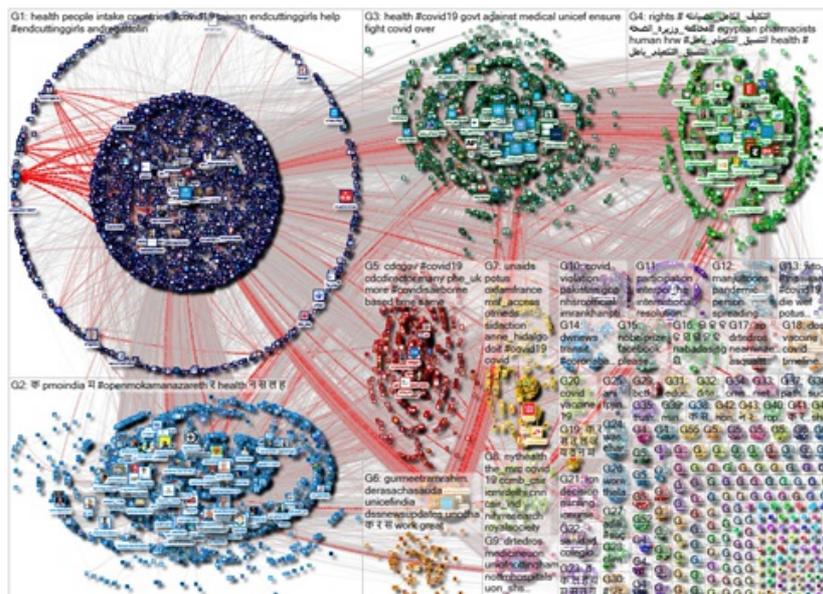
Fig.1: Social Network Graph of @WHO

In my own research and alongside others, I have explored conspiracies shared on Twitter such as the *5G and COVID conspiracy* as well as the *Film Your Hospital* conspiracy theory. These methods have become vital, in particular to the study of misinformation shedding new light on key stakeholders, online sources of misinformation and the way in which it is shared across Twitter.

In collaborative work, I have also explored some of the positives of social media during the pandemic. For instance, using Twitter data to develop understandings about how people communicated about masks during COVID-19, highlighting how Twitter was used to share positive views towards masks an encourage users to wear them.

In the table below, I provide a revised overview of some tools that can retrieve Twitter data and/or have the ability to important data for further analysis.

Overview of Tools (Sorted from Free, Limited Free and Paid)

Tools (Link)	OS	Platforms*	Free, Limited Free or Paid
Chorus	Windows	Twitter	Free
Webometric Analyst	Windows	Twitter (with image extraction capabilities), YouTube, Flickr, Mendeley, Other web resources	Free
Twitter Archiving Google Spreadsheet (TAGS)	Web-based	Twitter	Free
Mozdeh	Windows	Tiwtter	Free
COSMOS	Windows and Mac OS	Twitter	Free
Gephi	Windows, Mac OS, and Linux	Import data from other platforms	Free
KH Coder	Windows, and Mac OS	Import data for quantitative content analysis, text mining, and computational linguistics.	Free
#LancsBox: Lancaster University corpus toolbox	Windows, Mac OS, and Linux	Import and analyse language data.	Free

Tools (Link)	OS	Platforms*	Free, Limited Free or Paid
QDA Miner	Windows, and Mac OS	Facebook, Twitter, Reddit, YouTube, RSS	Limited Free
Word Stat	Windows, and Mac OS	Facebook, Twitter, Reddit, YouTube, RSS	Limited Free
Netlytic	Web-based	Twitter, YouTube, RSS Feed	Limited Free
NodeXL	Windows + other operating systems via virtual machine or cloud edition.	Twitter, YouTube, Flickr, Wikipedia	Limited Free
Twitonomy	Web-based	Twitter	Limited Free
Trackmyhashtag	Web-based	Twitter	Limited Free
Trendsmap	Web-based	Twitter	Limited Free
Socioviz	Web-based	Twitter	Limited Free
Audiense	Web-based	Twitter	Limited Free
Brand24	Web-based	Twitter, Facebook, Instagram, Blogs, Forums.	Limited Free
NodeXL Pro Insights	Web-based (PowerBI report template)	Twitter	Paid
Keyhole	Web-based	Twitter	Paid
Visibrain	Web-based	Twitter, Facebook, Instagram, blogs, and online press.	Paid
NVivo	Windows and Mac OS	Twitter, with the ability to import	Paid
Echosec	Web-based	Twitter, Instagram, Foursquare, Panoramio, AIS Shipping, Sina Weibo, Flickr, YouTube, VK	Paid
Social Elephants	Web-based	Twitter, Facebook, Instagram, YouTube	Paid
Symplur (Healthcare Focus)	Web-based	Twitter	Paid
Pulsar Social	Web-based	Twitter, Facebook topic data, Online blogs	Paid
Brandwatch	Web-based	Twitter, Facebook, YouTube, Instagram, Sina Weibo, VK, QQ, Google+, Pinterest, Online blogs	Paid
Social Bakers	Web-based	Twitter, Facebook, Instagram, YouTube, LinkedIn, Google+, VK, and Pinterest.	Paid
Talkwalker	Web-based	Twitter, and 10+ other social media platforms.	Paid

Please note some tools may allow access to other platforms and the ability to import your own data.

There are also other tools such as Botometer which allow the ability to detect bots on Twitter and also another tool Follower Audit that allows users to examine the followers of a particular account to see if they are bots. An interesting non-Twitter (paid) tool that is worth mentioning is CrowdTangle which can provide access to Facebook, Instagram, and Reddit data.

In addition to studying Twitter, scholars have designed and conducted research on other platforms and services all across the internet such as Web forums and blogs. For non-programmers, there are tools such as Scrape Storm which is an AI-powered visual web scraper and claims to be able to retrieve data from almost any platform.

For other tools that can help researchers conduct advanced data analysis and statistical analysis it could be worth exploring the likes of R, SPSS, KNIME, Weka, Tableau, PowerBI, and Leximancer. These tools often have packages and extensions that can be used to analyse Twitter data in unique ways and provide additional insights into social media data.

Finally, it is also worth mentioning that the tool NodeXL now has the ability to import tweets using tweet IDs. This means that if researchers are able to locate datasets of tweet IDS across the Web then these can be used to import data into NodeXL.

For instance, this is a collection of 354 million tweets related to Coronavirus or COVID-19 collected from March 3, 2020 and December 3, 2020. There is also a collection of 41.8 million tweets related to the BlackLivesMatter Movement and Counter Protests from 2013 to 2020. For researchers interested in a particular topic a quick Google search may reveal large tweet ID datasets that can be used to retrieve the original tweets with all their metadata.

In recent years social media research has moved from the fringes to become a more complementary source of data. It is now regularly used in addition to interview and survey-based data, as well as by study online communities to examine themselves.

Due to the popularity of social media research, I have also found myself providing talks on this subject to a far wider range of audiences across a range of different disciplines and levels of experience. I'm happy to provide training and workshops on social media research, so feel free to get in touch!

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our Comments Policy if you have any concerns on posting a comment below.

Featured image Credit: Adapted from Elisa Riva via Pixabay. In text image reproduced with permission of the author.