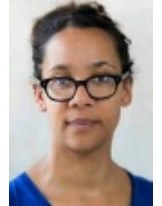


Research Resilience: Why academics and funders alike should care about #RIPTwitter

*Twitter is under close scrutiny these days with news that its timeline could be subject to further algorithmic control. **Farida Vis** looks at what such dramatic changes could mean for research. There is a great need for both funding councils and researchers to better understand the potential impact of these data and platform politics. Strategies must be developed to encourage lesser reliance on a single social media data source.*



Over the weekend a number of Twitter users were in uproar over potential changes to the way the platform shows content. The furore, discussed at length on the #RIPTwitter hashtag, was over the possible imminent introduction of an algorithmic timeline. This would essentially make Twitter more like Facebook in that it would show users the most relevant content (according to what criteria is not entirely clear) rather than simply the most recent content in chronological order. The current way in which the Twitter timeline works already contains a combination of most recent and a 'while you were away' feature, which was introduced just over a year ago, with [Twitter highlighting](#) then that: 'with a few improvements to the home timeline we think we can do a better job of delivering on that promise without compromising the real time nature of Twitter.'

The most recent changes were first reported on [Buzzfeed](#) on Friday night, and though not confirmed by Twitter, following debates on the #RIPTwitter hashtag, a six-tweet direct response from Twitter's current CEO, Jack Dorsey followed. The first tweet addressing the issue has been shared and cited the most, probably because its final sentence highlights that '[w]e never planned to reorder timeliness next week'. The fourth tweet in this series is also worth noting, especially as it connects closely to the tone of the statement above about the introduction of the 'while you were away' feature last year. In this tweet, Dorsey highlights the importance of the real-time feel of the platform: 'And we're going to continue to refine it to make Twitter feel more, not less, alive!' So not exactly denying the suggested timeline change outright. It is of course impossible to please all users all of the time and while users may object to drastic platform changes, it remains difficult to know if they will eventually accept them. My interest now is to do with what such dramatic changes could mean for research.



Image credit: A Twitter Banner Draped Over The New York Stock Exchange For Twitter's IPO (CC BY)

So why does any of this matter to academic researchers? And why should academic researchers keep an eye on platform developments? It matters for a number of reasons. Considerable research in recent years has focused on Twitter, largely because of the platform's open and partially free data ecosystem, along with a user-base that predominantly shares data publicly. As a result, research on and tool development around this platform has developed rapidly. One element that's highlighted specifically in relation to Twitter is the platform's attractiveness for research in relation to doing research in (near) 'real-time'.

Unlike Facebook or other platforms, Twitter has an open data ecosystem that essentially gives access to researchers in five different ways: three are free and two are paid for. The two paid for ones (real-time firehose and historical firehose) are mainly used by industry due to the cost implications. These latter two essentially provide a full sample of Twitter data while the other three provide a range of different kinds of samples, which – although more limited – can still be used to great effect by a lot of academic researchers. Because of the nature of the platform, very good for analyzing and understanding events in 'real-time', important research agendas have developed based on these specific conditions. It is less well understood and appreciated that these conditions may prove to be part of a specific stage in the platform's development.

Despite many evident and hard to ignore limitations (how can one 'generalise' what is happening on Twitter and relate this to wider populations, that is to say: what can Twitter data really tell us about the social world?) research on Twitter has flourished. Whilst critiques about the importance about the need to study platforms beyond Twitter (not least given its 'limited' user base compared to other platforms, including newer ones like Instagram) the take-up of research on other platforms has been limited to date. Funding councils have also put a lot of their proverbial eggs in the Twitter basket and 'real-time' social media analytics. Again the talk may be of 'streams' and 'real-time' data more generally, but when 'social media' is mentioned it is fairly safe to translate this as 'Twitter data'.

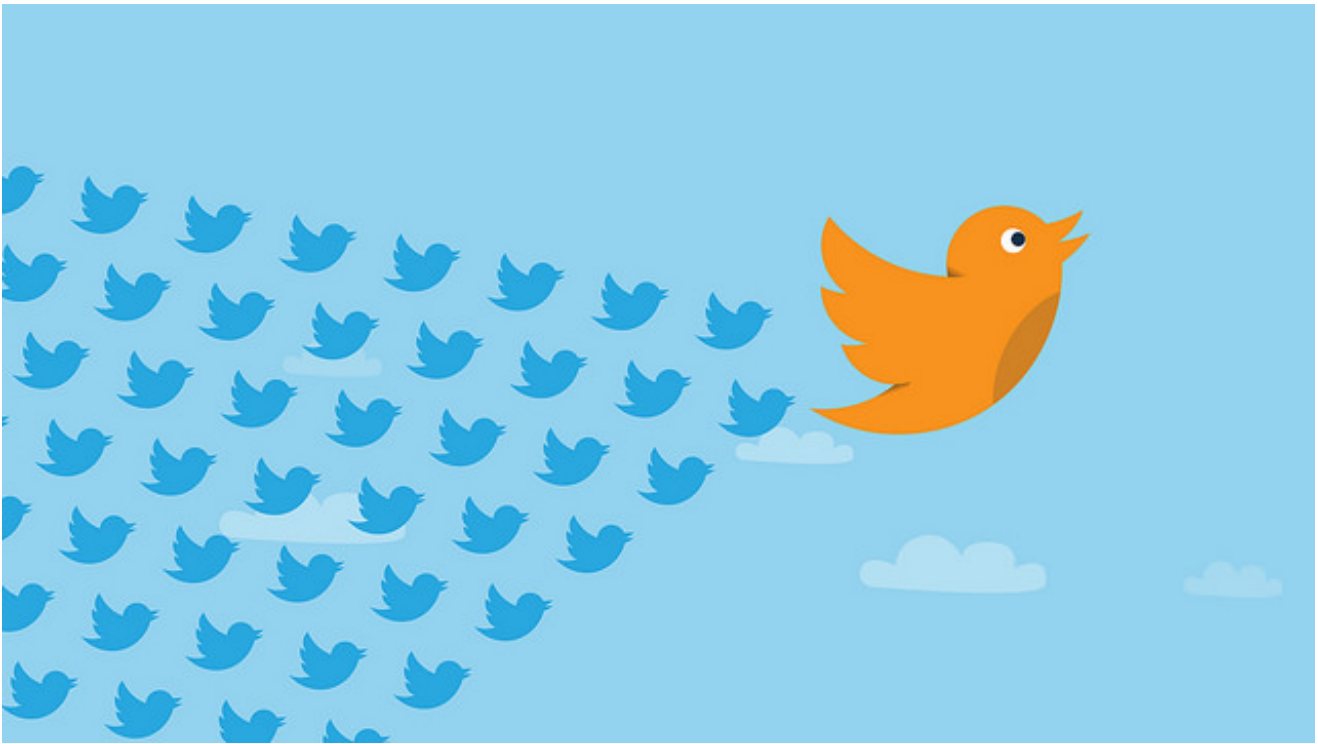


Image credit: Alan O'Rourke Twitter-birds-social-media-leader-crowd CC BY

With the Twitter timeline changes dismissed for the time being, what should be taken from this is the need for both funding councils and researchers to better understand the potential impact of these ***data and platform politics*** and their potential impact for research. More than that, these timeline changes are not the first clear signs from the last twelve months that should have already alerted researchers to the need to be more data and platform resilient in their research.

Here are a few important reasons from the last twelve months as to why and how researchers should develop 'resilience' when it comes to Twitter data:

Deal with Foursquare: change in how users share locations.

In March 2015 [Twitter announced a new partnership with Foursquare](#), which is significant to the emerging research around online location and should, from an industry perspective, be seen as part of ongoing attempts to identify and generate data revenue streams. This deal also saw Twitter location-sharing move closer to the more privacy focused, flexible, location-sharing settings offered by Instagram. Essentially Twitter announced it would encourage users to share their location data in new ways: moving away from fixed, *context poor*, precise lat/long geo-coordinates to *context rich* relational data about place.

Why does this matter for researchers?

In part what has been identified of particular importance in a research context, is the perceived ability to use data derived from Twitter and be able to connect a user to a precise offline location (geo-coordinates). This new focus on place, highlights not simply a renewed focus and interest in place-making; how this happens and is facilitated and indeed encouraged on the platform, but more than that; this new deal highlights the possibility to articulate the value of specific places; [places the user cares about](#). A key reason this change matters for research is that any research agenda that has been built on the ability to harvest exact lat/long coordinates from users is likely to be affected (see [this](#) developers thread for example, discussing the 'sudden' reduction in geo data). With users no longer being asked to opt in to share their location with *every tweet*, but instead are encouraged to share information about *place*,

and at the level of *single tweet*, this has potentially significant consequences for research. And whilst it was quite well understood that only 1% of users previously opted in to share all their location data, all of the time, [this gave access to live location data at a hitherto unprecedented scale](#), just not one that was very commercially viable in the end.

Twitter takes data sales in-house, creates data monopoly.

In April 2015 [Twitter announced that it would close off firehose access to third party developers](#). A year earlier, the largest reseller of Twitter data, Gnip was bought by Twitter –and some speculated at the time this was connected to [Apple's purchase of analytics firm Topsy](#)— in order to take analytics and data sales in house. Apple has [recently closed down](#) Topsy, two years after purchase. Many at the time and since have argued that it would at that point be 'inevitable' [that Twitter would want to control the sale and access to their own data more](#). There is much more to say about this of course, but let's focus on one key aspect.

Why does this matter for researchers?

It matters because it could be an important indicator that what is (free) and reasonably accessible today, might not be tomorrow. So whilst in the current data landscape, *on which so much research hinges*, a significant amount of Twitter data is freely available, this might not be the case forever. More than that, the way in which this decision was made and executed around closing off access to third party developers is also significant: many were taken by surprise by its swiftness. Again this speaks to the need for academic researchers to consider how resilient they would be to such structural data access changes were they to directly impact on them. How quickly would they be able to bounce back?

Thunderbolts, hearts, and stars

Other recent feature changes and introductions are also worth mentioning. These include the introduction of [Twitter Moments](#) (signified by a thunderbolt icon), which curates tweets (done by a team of editors) around key events in order for users to easily find content (beyond scrolling through hashtags), as well as the introduction of hearts, abolishing the much-loved 'favourite' feature (signified by a star). In a research context such feature changes can have all sorts of implications for data collection and analysis. Take for example a news context: would users 'heart' a tweet about a recent terrorist attack in the same way they might have 'faved' it for later reading or for use in a news story? The way in which users engage with these features should of course be taken on board in analysis, but is harder to do as changes are rolled out and happen during data collection and user communities don't adopt new features as expected. Feature changes have long represented a significant methodological challenge to researchers.

So what can be done about this for researchers?

If the Twitter timeline indeed changed and becomes an algorithmic timeline, this could have disastrous consequences for research, both for individual research projects (PhD projects especially), wider research agendas, including those of funding councils. Researchers therefore need to better understand how such shifts by key social media platforms can potentially impact current research trajectories. We should openly explore the research possibilities these platform-enabled data shifts close off, and at the same time we should focus on new possibilities and opportunities for different kinds of research they in turn could open up.

Fundamentally, we should critically focus on what kinds of research questions these new social data landscapes could address. But finding out about these issues can be tricky. It might become necessary for funding councils to collectively set up a 'Social Data Council' or similar so that the academic research community can be kept informed, has a place to go for up to date information, share best practice and collectively develop research approaches that more firmly foreground the increased necessity to develop strategies that mean lesser reliance on a single social media data source.

Update added 11 Feb 2016: As it turned out Jack Dorsey's carefully worded non-denial over the weekend was exactly that, as a new timeline feature was [announced](#) yesterday (on the same day as the Q4 2015 earnings were [announced](#) and [scrutinised](#)). In the [short post](#) it was explained that for now users could switch on the new feature themselves in settings, but that in coming weeks the new timeline feature would be switched on automatically. So essentially opt-in for now, but then 'default' and opt-out. I will use a future post to go in more detail about what this could mean for research, (and how to deal with this) but for now I think this is how it complicates Twitter research: aside from having to deal with sampling issues (an already well-known and well-documented problem), academic researchers will now also have to find ways for dealing with algorithmic bias and try to overcome how – in a research context – you might analytically distinguish (and it is unclear at this stage if you would be able to) those users that have opted out of the new timeline feature from those who are using it. Time will tell if the new algorithmic timeline will become the platform's norm, with users simply accepting it, but if for a long time some do and some don't, this will be tough to deal with. Researchers would have to find additional ways to best describe what they think is going on in their data and add additional caveats to their results.

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

Farida Vis is a Faculty Research Fellow, based in the Information School at The University of Sheffield. During 2012-2015 her Fellowship focused on 'Big Data and Social Change'. The final two years of her Fellowship will examine 'The Futures of the Visual Web'. She is the Director of the [Visual Social Media Lab](#), sits on the [World Economic Forum's Global Agenda Council on Social Media](#) and was recently elected to the Board of Directors of the [Big Boulder Initiative](#). She is a frequent public speaker and tweets as [@flygirltwo](#)

- Copyright 2015 LSE Impact of Social Sciences - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.