

2018-06-12

Twitter: an emerging source for geographical study

Fearnley, F

<http://hdl.handle.net/10026.1/11683>

10.1080/00167487.2018.12094044

Geography

Geographical Association

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Author's copy of the pre-referred manuscript

Cite as:

Fearnley F and Fyfe RM (2018) Twitter: an emerging source for geographical study *Geography* 103, 97-101

Twitter: an emerging source for geographical study

Fearnley, F* and Fyfe RM

School of Geography, Earth and Environmental Sciences, University of Plymouth, Plymouth, PL4 8AA

*corresponding author: francesca.fearnley@students.plymouth.ac.uk

Introduction

In 2004, the term 'Web 2.0' was coined to describe the new way in which users were consuming the Internet (Kaplan and Haenlein, 2010). It was on this new platform that the evolution of social media took place and consequently, the production of user generated content on a scale not previously seen (Kaplan and Haenlein, 2010). Twitter is among the array of social media sites and is currently the fastest growing and most popular microblogging site in the world (Atefeh and Khreich, 2015).

Twitter was founded in 2006 and offers users the ability to post and read short messages known as 'tweets' conveying their thoughts and opinions. In 2018 Twitter reported 330 million active monthly users (Twitter, 2018). In September 2017, Twitter successfully trialled and implemented longer character limits for some languages, rising from 140 characters to 280 characters. Krishnamurthy et al. (2008) determined that Twitter is most popular in European countries such as Germany, the UK, France, Spain and Italy, but note it is gaining popularity outside of Europe too, evidenced by the fact that Twitter is now available in 33 different languages, including those that use non-Latin character sets such as Arabic (Weller et al., 2014).

Social media platforms have emerged as a potential source for academic study (Zimmer and Proferes, 2014). Tweets have come to be regarded as "sensory values". Twitter users are viewed as sensors that can detect, report and respond to their situation, whether this be social, political, economic, or reflecting global events. This potential corpus of information that might be usefully studied was predicted by Goodchild (2007), who coined the concept of volunteered geographical information (VGI), with citizens acting as sensors and co-producing and disseminating geographical data. Twitter users provide "real time [commentary], responding to, and amplifying the impact of current events" (Williams et al., 2013, p.385). Thus, Twitter fits into the wider concept of 'big data' (Kitchen, 2014). It is huge in volume, has high velocity (data is generated in real-time), it is diverse and it has the potential to be exhaustive (complete) in scope.

Consequently, interest in Twitter-based research now spans numerous disciplines, as tweets provide a rich dataset for researchers to better understand the world (Williams et al., 2013). Since 2006, over 300 billion tweets have been created and with the relative ease in which tweets can be collated and analysed, the opportunity to study is vast (O’Leary, 2015). Between 2009 and 2016, the number of studies on Twitter returned within the search engine Google Scholar rose from just over 200,000 to 1.5 million (academic papers before 2007 did not relate to the microblogging site, as Twitter was founded in 2006). Miller (2011) extends the argument further, declaring that Twitter offers an unprecedented opportunity to study humans and their interactions. Alongside disciplines such as Business, Economics and Computer Science, Geography is an example of one such discipline which is utilising this unusual resource as opportunity for study.

All information is collected through Twitter’s API (Application Programming Interface), where researchers and developers can use simple code to collect tweets from ‘databases’. There are many different ways in which information associated with the collected tweets can be used for academic research, and consequently, the type of data that can be extracted. Cairns (2013) has specifically argued that the most compelling studies using social media are ones which use geolocation in their analysis, making Geography a highly relevant discipline for using Twitter in its research. There are two ways researchers can collect location-based tweets, either through a self-defined user location, stated in the user’s profile or alternatively, through a geotagging facility which produces a set of associated coordinates (a feature which users have to ‘opt-in’ to utilise). Twitter, then, provides researchers with a high “volume of...data that are time-stamped and...precisely located” providing “researchers with useful information about activities and opinions across time and space” (Lansley and Longley, 2016, pp.85). Research can range from a content analysis, to sentiment analysis and geolocation-based investigations.

Research utilising geolocated tweets spans a range of geographical interest areas, and it is not possible to review them all here. Instead focus will be placed on two contrasting applications: tracking of real-time events that pose a threat to human life, and research examining how social media is used in the dissemination and spread of public/private emotion and sentiment.

Mapping natural hazards through Twitter

The inherent spatial nature of Tweets, and the tendency of users to post ‘information’ about themselves and their lives in real-time means that there is tremendous potential for understanding real-time risks in both space and time. It has been possible to map earthquake incidence by mining

archived posts. Earle et al. (2011), for example, has demonstrated that Twitter can be used in combination with United States Geological Survey (USGS) earthquake response methods to quickly deliver warnings and information. Twitter was used to predict the occurrence of earthquakes over a five-month period through tracking the use of the word 'earthquake'. A total of 48 earthquakes were identified through Twitter, and whilst this was dramatically fewer than the ~5200 detected by the USGS in the same time period, Twitter was shown to have had many advantages, including first-hand commentary and a faster detection rate in poorly instrumented regions of the world. This highlighted the potential for the use of social media as a tool in disaster response and information delivery. Crooks et al. (2013) demonstrated the ability to track the spatial and temporal impact of an individual earthquake, examining a 5.8 magnitude event on the east coast of the United States. Over 1,100 geocoded tweets that included "#earthquake" or "#quake" recorded the event within ten minutes of its occurrence, with the first post within a minute. The geographical spread of these tweets reflected the spread of the seismic waves. Whilst the focus here has been on earthquakes, considerable research has also been undertaken on wildfire risk and hazard, for example, Kent and Capello (2012) examined social media posts relating to situational awareness during a wildfire in Wyoming, USA, demonstrating that there was a strong spatial relationship between the distance to wildfires and the frequency of social media messages. Whilst there is a risk of Twitter commentators from outside wildfire risk areas posting, and thus obfuscating patterns, the majority of wildfire-related posts were from within the immediate local area.

Dissemination and spread of knowledge on Twitter

In the case of earthquake or wildfires, well-established methods (e.g. networks of seismographs) already exist to track and map these events. The advance that social media such as Twitter offers is to also assess the impact on these events on the individual (mediated through a 280-character message). In situations of political unrest, or in response to regional or global events, it is possible to explore the spread and diffusion of ideas and knowledge. Crampton et al. (2013) tracked the spatial and temporal diffusion of a specific hashtag created in response to rioting in Kentucky, USA in 2012, through the analysis of associated coordinates. The researchers found that news of the riots spread quickly to areas outside of the city. More importantly, the authors identified that classic distance decay functions could be applied to the distribution of information: the level of tweeting about the riots decreased as distance from the city increased. This exemplified Tobler's first law of Geography, in that everything is related to everything else, but that near things are more related than those that are further away (Tobler, 1970). Technological advances have allowed us to see this patterns holds true in both social

and geographical distance (Sui and Goodchild, 2011). Other locally-focused events can be seen to spread through the local community and then far beyond with Crawford (2010) demonstrating this diffusion in relation to protests in Iran in 2010. Bruns et al. (2012) build on this, demonstrating that it is reactions to emotionally charged crises which quickly spread far beyond the local boundaries, suggesting that fundamentally local events do impact those further away.

Other studies have taken a different approach, instead focusing on the content of Tweets rather the locations they are sent from. Essentially a form of self-categorising, hashtags, noted by the '#' symbol, are used to create an index of messages which relate to a particular theme, making topics easy to follow. If a hashtag exceeds an expected level of activity within a certain time frame, then it has the potential to become a 'trending topic' on Twitter (Naaman et al., 2011). Often, large events develop their own event-specific hashtag, most recently demonstrated by the '#MeToo' movement, a hashtag used on social media to demonstrate the extent of sexual assault and harassment in the workplace (Smartt, 2017). The content of Tweets which contain a certain hashtag can also be used as the basis of academic investigation.

Bruns et al. (2012) examined the composition of tweets, specifically containing "#Qldfloods", sent in response to flooding in South East Queensland in 2011. 50—60% of Tweets were retweets, showing the importance of sharing situational information using the platform. Links to further information were present in between 30 and 40% of tweets. They also note that heavily retweeted accounts included those of the emergency services and media organisations, which, they suggest play "a leading role in disseminating timely and relevant information", with users acting as a community to amplify this information (Bruns et al., 2012, p.8). Developing this further, Heverin and Zach (2012) have declared this the 'critical-stage' of a crisis. Drawing from theories from psychology, they suggest that information-sharing is critical in allowing individuals to comprehend and make sense of a situation.

Appropriate usage of social media as a geographical data source

There are, of course, limitations and ethical considerations that need to be taken into consideration if using Twitter as the basis of research. The first relates to participant consent and privacy concerns. To some, messages collected from social media are perceived as alternatives to public surveys (Lansley and Longley, 2016), which usually require consent from participants before being used in academic work. In the case of Twitter, it is not possible to collect the consent of users whose Tweets are used in research, as the number of tweets used in investigations can range anywhere up to five billion (Zimmer and Proferes, 2014). Additionally, some academics argue that, because data collected through

Twitter's APIs are public information, no special considerations need to be made (Zimmer and Proferes, 2014). Others disagree, arguing individuals have the "reasonable expectation that one's tweet stream will be practically obscure within thousands (if not millions) of tweets" (Zimmer, 2010, p.1). To address this issue, many academics carry out analysis on an aggregate level, making sure that no individual author can be identified, ensuring anonymity. So, whilst tweets may be considered as 'volunteered geographical information' *sensu* Goodchild (2007), it may be better to consider these as (in)voluntary geographical information, as users are not active participants within the academic study.

A further limitation, especially important to consider from a Geographical perspective, is that concerning geolocation-based research. As mentioned, location can either be determined through a location stated in the user's bio, or through associated metadata containing precise coordinates, if the user has opted-in to utilise this facility. The number of tweets available via the latter method are substantially fewer. Although an estimated 80% of users access Twitter through a mobile device (Lansley and Longley, 2016) (many of which are smart phones and GPS enabled and therefore theoretically have the ability to enable locations to be added to tweets) users have to 'opt-in', which means roughly only 1% of tweets have accurate associated coordinates (Gaffney and Puschmann, 2014). Additionally, the self-defined location is often incomplete or inaccurate, and does not necessarily have to be an actual location (Bakshi, 2011). Research projects must bear this in mind, adapting their methodology if the number of tweets returned is too low. Further, the Twitter API will only return a randomised 1% sample of all tweets posted; thus whilst datasets may be very large, they are not 'complete'.

On the other hand, if these limitations can be taken into consideration and addressed, Twitter is a highly accessible, unusual, but rich resource of geographical information, which has allowed researchers the opportunity to develop and gain new geographical insights that just would not have existed without the development of social media. From detecting earthquakes, to providing evidence that older Geographic theories are still relevant in the 21st century, to building communities and disseminating disaster-related information more effectively, Twitter, as a social media site, is capable of adding context and commentary to situations at a speed never previously seen and revealing insights which would not otherwise be available.

Reference List

- Atefeh, F. and Khreich, W. (2015) 'A Survey of Techniques for Event Detection in Twitter', *Computational Intelligence*, 31(1), pp.132-164
- Bakshi, H. (2011) Framework for crawling and local event detection using twitter data. Available at: <https://rucore.libraries.rutgers.edu/rutgers-lib/33565/> (Accessed: 17.02.2018)
- Bruns, A., Burgess, J., Crawford, K. and Shaw, F. (2012) #qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods. Available at: <http://eprints.qut.edu.au/48241/1/floodsreport.pdf> (Accessed: 08.02.2018).
- Cairns, I. (2013) Mapping Travel, Languages and Mobile OS Usage with Twitter Data. Available at: <https://blog.gnip.com/twitter-visualizations-languages-travel/> (Accessed: 13th October 2016).
- Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M. and Wilson, M.W. (2013) 'Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb', *Cartography and Geographic Information Science*, 40(2), pp.130-139
- Crawford, K. (2010) Everything's gone green. Available at: <http://www.abc.net.au/news/2009-06-24/28486> (Accessed: 08.02.2018).
- Crooks, A., Croitoru, A., Stefanidis, A. and Radzikowski, J. (2013) '#Earthquake: Twitter as a Distributed Sensor System', *Transactions in GIS*, 17 (1), pp.124-147
- Earle, P.S., Bowden, D.C. and Guy, M. (2011) 'Twitter earthquake detection: earthquake monitoring in a social world', *Annals of Geophysics*, 54(6), pp.708-715
- Gaffney, D. and Puschmann, C. (2014) 'Data Collection on Twitter', in Weller, K., Bruns, A., Burgess, J., Mahrt, M. and Puschmann, C. (eds.) *Twitter and Society*. New York: Peter Lang, pp.55-69
- Goodchild, M. (2007) 'Citizens as sensors: the world of volunteered geography', *GeoJournal*, 69(4), pp.211-221
- Heverin, T. and Zach, L. (2012) 'Use of microblogging for collective sense-making during violent crises: A study of three campus shootings', *Journal of the Association for Information Science and Technology*, 63(1), pp.34-47
- Kaplan, A.M. and Haenlein, M. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media', *Business Horizons*, 1 (53), pp.59-68

Kent, J.D. and Capello, H.T. (2012) 'Spatial patterns and demographic indicators of effective social media content during the Horseshoe Canyon fire of 2012', *Cartography and Geographic Information Science*, 40(2), pp.78-89

Kitchen, R. (2014) 'Big data, new epistemologies and paradigm shifts', *Big Data and Society* 1, pp.1-12

Krishnamurthy, B., Gill, P. and Arlitt, M. (2008) 'A few chirps about Twitter', *Proceedings of the First Workshop on Online Social Networks*. Seattle, USA. 17 – 22 August, doi: 10.1145/1397735.1397741

Lansley, G. and Longley, P.A. (2016) 'The geography of Twitter topics in London', *Computers, Environment and Urban Systems*, 58, pp.85-96

Miller, G. (2011) 'Social Scientists Wade Into the Tweet Stream', *Science*, 333 (6051), pp.1814-1815

Naaman, M., Becker, H. and Gravano, L. (2011) 'Hip and trendy: Characterizing emerging trends on Twitter', *Journal of the Association for Information Science and Technology*, 62(5), pp.902-918

O'Leary, D. (2015) 'Twitter Mining for Discovery, Prediction and Causality: Applications and Methodologies', *Intelligent Systems in Accounting, Finance and Management*, 22 (3), pp.227-247

Smartt, N. (2017) Sexual Harassment In the Workplace in A #MeToo World. Available at: <https://www.forbes.com/sites/forbeshumanresourcescouncil/2017/12/20/sexual-harassment-in-the-workplace-in-a-metoo-world/#55dca7e95a42> (Accessed: 08.02.2018).

Sui, D. and Goodchild, M. (2011) 'The convergence of GIS and social media: challenges for GIScience', *International Journal of Geographic Information Science*, 25 (11), pp.1737-1748

Tobler, W. R. (1970) 'A computer movie simulating urban growth in the Detroit region', *Economic Geography*, 46, pp.234–40

Twitter (2018) Twitter Q4 and FY 2017 Earnings Report. Available at: <https://investor.twitterinc.com/results.cfm> (Accessed: 20.02.2018).

Weller, K., Bruns, A., Burgess, J., Mahrt, M. and Puschmann, C. (2014) 'Twitter and Society: An Introduction', in Weller, K., Bruns, A., Burgess, J., Mahrt, M. and Puschmann, C (eds.) *Twitter and Society*. New York: Peter Lang, pp.xxix

Williams, S., Terras, M. and Warwick, C. (2013) 'What do people study when they study Twitter? Classifying Twitter related academic papers', *Journal of Documentation*, 69 (3), pp.384-410

Zimmer, M. (2010) Is it ethical to harvest public Twitter accounts without consent? Available at: <http://michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public--twitter-accounts-without-consent/> (Accessed: 20.02.2018)

Zimmer, M . and Proferes, N. (2014) 'A topology of Twitter research: disciplines, methods, and ethics', *Aslib Journal of Information Management*, 66(3), pp.250-261