



Selected Papers of AoIR 2016:
The 17th Annual Conference of the
Association of Internet Researchers
Berlin, Germany / 5-8 October 2016

BEYOND HASHTAGS: COLLECTING AND ANALYSING CONVERSATIONS ON TWITTER

Brenda Moon
Queensland University of Technology

Nicolas Suzor
Queensland University of Technology

Ariadna Matamoros-Fernandez
Queensland University of Technology

Abstract

In this paper we examine a series of techniques to enhance the collection and analysis of conversations on Twitter. We start from the position of seeking to understand how ordinary discussions about particular issues or controversies are unfolding on the social media platform. A key limitation of studies of topics on Twitter that rely on searching for a keyword or hashtag is that they may miss important sections of conversations around issues that do not match the keywords selected (Rambukkana, 2015, Bruns & Burgess, 2015). The Tracking Infrastructure for Social Media Analysis (TrISMA) (Bruns, Burgess & Banks et al., 2016) captures tweets of 2.8m Australian users on a continuing basis providing a comprehensive dataset that we can use to find reply chains which do not include the hashtag or keyword we are studying. Many existing methods of exploring Twitter data do not present conversation chains in a linked format (Bruns, 2012, Grant, Moon, & Busby Grant, 2010); we investigate how using network visualization might help researchers better understand the qualitative content and context of conversations.

Keywords: Twitter, social media, big data, issue mapping, Uber

Introduction

In this paper, we examine a series of techniques to enhance the collection and analysis of conversations on Twitter. We start from the position of seeking to understand how ordinary discussions about particular issues or controversies are unfolding on the social media platform. A key limitation of studies of topics on Twitter that rely on searching for a keyword or hashtag is that they may miss important sections of conversations around

Suggested Citation (APA): Moon, B., Suzor, N., Matamoros-Fernandez, A. (2016, October 5-8). *Beyond Hashtags: Collecting and analysing conversations on Twitter*. Paper presented at AoIR 2016: The 17th Annual Meeting of the Association of Internet Researchers. Berlin, Germany: AoIR. Retrieved from <http://spir.aoir.org>.

issues that do not match the keywords selected (Rambukkana, 2015, Bruns & Burgess, 2015). The Tracking Infrastructure for Social Media Analysis (TrISMA) (Bruns, Burgess & Banks et al., 2016) captures tweets of 2.8m Australian users on a continuing basis providing a dataset that we can use to find reply chains which do not include the hashtag or keyword we are studying.

Many existing methods of exploring Twitter data do not present conversation chains in a linked format (Bruns, 2012, Grant, Moon, & Busby Grant, 2010); we investigate how using network visualization might help researchers better understand the qualitative content and context of conversations.

Identifying tweets

We develop a similar method to Lorentzen and Nolin's (2015) approach to collect a broader set of conversations on Twitter. Using the TrISMA dataset, we first identify all tweets that match a given keyword, and then also identify all tweets that respond to or were responded to by such keyword tweets. We trace these linked conversations recursively, to extract the full chains of conversations in which a search keyword has been mentioned at least once. Where Lorentzen and Nolin (2015) were limited by considering conversations based on the Twitter in-reply-to chains in a single direction from their hashtag tweets or their 5,000 most active participants, we are able to recursively move both ways along the reply chain, including tweets that are replies to tweets we have identified. This is not possible using the Twitter API. Our study is limited by the boundaries of the Australian Twittersphere as defined by TrISMA; conversation chains will be broken if they extend to accounts outside the TrISMA dataset, even if subsequent replies occur inside it.

Analysing reply chains

This paper will present the results of our initial case study of Australian controversies over the legitimization of the Uber ride-sharing service in Australia. We seek to collect all tweets that form part of the conversation around the keyword 'uber' in the Australian twittersphere since 2013, and zoom in from there to analyse specific controversies. We conduct a comparative study of two datasets – an original set that contains only tweets that match a keyword, and an expanded set that includes all conversations that reference those initial tweets. The first dataset – tweets matching the keyword 'uber' from January 2013 through to November 2015, comprises 176,941 tweets. The second dataset, which also includes tweets we can identify as part of the conversation around the first (using `in_reply_to_tweet_id`), comprises an additional 30,210 tweets.

The conversation chains are visualised as a directed network and a range of network metrics are calculated, including those used by Lorentzen and Nolin (2015), to allow comparison with their results.

We then apply issue mapping to social media analysis (Marres, 2015; Marres & Moats, 2015) to evaluate the advantages and limitations of these methods for extracting conversations from Twitter data. While work on issue mapping has largely used web data to trace the relationships among actors, themes and objects in the discussion of

matters of concern (Rogers & Marres, 2000; Severo & Venturini, 2015), we will use social media data to study how issue publics emerge through their engagement with specific controversies or themes surrounding Uber. Issue mapping using digital methods and informed by controversy analysis provides a useful methodology to account for popular and everyday modes of online participation in matters of concern (Burgess & Matamoros-Fernandez, 2016).

Our initial results show that the conversations we collected differ significantly from the tweets that match our keywords. From a preliminary textual analysis of small samples of tweets, we note that the tweets that contain the keyword 'Uber' seem to be reflect themes and frames dominated by the tech media, lobbyists and campaigners while the tweets that are found through our reply chain approach appear to be much more diverse. They include debates explicitly about the controversy by established players in the issue. Importantly, they also include the much less readily visible ordinary voices of individuals in conversations that are not highly politicised.

We seek to investigate the hypothesis that network metrics and visualisations of conversations may assist in mapping the actors, themes and objects involved in Twitter discussions of matters of concern. We will seek to understand how including conversations that do not explicitly mention 'uber' or related keywords may be able to add a richness to the data that was previously invisible. We suggest that this method appears to enable richer analysis of ordinary voices around a controversy. Not only it highlights some additional issues as the controversy plays out, but it also appears to surface new actors who do not necessarily seek to participate in visible public debates by using a common hashtag. We think that these discussions are of great value in identifying new issues and actors that help understand public communication on social media. Not only does this method highlight some additional issues as the controversy plays out, but it also appears to surface a greater proportion of everyday discussions by ordinary users who do not necessarily seek to participate in visible public debates by using a common hashtag. We think that these discussions will often be of very great value in understanding ordinary discussions around controversial issues.

Acknowledgments

This research was supported by infrastructure provided through the Australian Research Council LIEF project Tracking Infrastructure for Social Media Analysis, and the ARC Future Fellowship grant Understanding Intermedia Information Flows in the Australian Online Public Sphere. Researchers have also received funding from an Australian taxi industry body to study the peer economy.

References

- Bruns, A. (2012). How Long Is A Tweet? Mapping Dynamic Conversation Networks On Twitter Using Gawk And Gephi. *Information, Communication & Society*, 15(9), 1323–1351. <http://doi.org/10.1080/1369118X.2011.635214>
- Bruns, A., and Burgess, J. 2015. Twitter hashtags from ad hoc to calculated publics. In *Hashtag Publics: The Power and Politics of Discursive Networks*, N. Rambukkana, Ed. New York: Peter Lang, 13-28
- Bruns, A., Burgess, J., Banks, J., Tjondronegoro, D., Dreiling, A., Hartley, J., ... Sadkowsky, T. (2016). TrISMA: Tracking Infrastructure for Social Media Analysis. Retrieved from <http://trisma.org/>
- Burgess, J., & Matamoros-Fernandez. (2016). Mapping sociocultural controversies across digital media platforms: One week of #gamergate on Twitter, YouTube and Tumblr. *Communication, Research & Practice* [forthcoming].
- Grant, W. J., Moon, B., & Busby Grant, J. (2010). Digital Dialogue? Australian Politicians' use of the Social Network Tool Twitter. *Australian Journal of Political Science*, 45(4), 579–604. <http://doi.org/10.1080/10361146.2010.517176>
- Lorentzen, D. G., & Nolin, J. (2015). Approaching Completeness: Capturing a Hashtagged Twitter Conversation and Its Follow-On Conversation. *Social Science Computer Review*. <http://doi.org/10.1177/0894439315607018>
- Marres, N. (2015). Why Map Issues? On Controversy Analysis as a Digital Method. *Science, Technology & Human Values*, 40(5), 655–686. <http://doi.org/10.1177/0162243915574602>
- Marres, N., & Moats, D. (2015). Mapping Controversies with Social Media: The Case for Symmetry. *Social Media + Society*, 1(2), 1–17.
- Rambukkana, N. (2015). *Hashtag Publics: The Power and Politics of Discursive Networks*. Peter Lang Publishing, Incorporated.
- Rogers, R., & Marres, N. (2000). Landscaping climate change: A mapping technique for understanding science and technology debates on the World Wide Web. *Public Understanding of Science*, 9(2), 141–163.
- Severo, M., & Venturini, T. (2015). Intangible Cultural Heritage Webs: comparing national networks with digital methods. *New Media & Society*, 1461444814567981.
- Venturini, T. (2012). Building on faults: How to represent controversies with digital methods. *Public Understanding of Science*, 21(7), 796–812. <http://doi.org/10.1177/0963662510387558>