

# Researching News Discussion on Twitter: New Methodologies

---

*Axel Bruns & Jean Burgess*

*ARC Centre of Excellence for Creative Industries and Innovation*

*Queensland University of Technology*

*Brisbane, Australia*

*a.bruns / je.burgess @ qut.edu.au – @snurb\_dot\_info / @jeanburgess – <http://mappingonlinepublics.net/>*

## Abstract

*Twitter* has become a major instrument for the rapid dissemination and subsequent debate of news stories, and comprehensive methodologies for systematic research into news discussion on *Twitter* are beginning to emerge. This paper outlines innovative approaches for large-scale quantitative research into how *Twitter* is used to discuss and cover the news, focussing especially on #hashtags: brief identifiers which mark a tweet as taking part in an established discussion.

## Introduction

*Twitter* has become an important addition to the toolboxes of journalists and journalism researchers (Ahmad, 2010). Its trajectory has mirrored that of previous social media platforms: just as blogs became established as tools both for occasional first-hand reporting and for extensive follow-on curation, commentary, and discussion of news stories (cf. Bruns, 2006), so have *Twitter's* uses been extended beyond everyday “lifesharing” and interpersonal communication to similar journalistic, para-journalistic, and quasi-journalistic activities (Kwak *et al.*, 2010; Subasic & Berendt, 2011). Even accompanying claims of “over-hype” or concerns about *Twitter's* possible adverse affects on journalistic practice (Farhi, 2009) signal widespread awareness and take-up, and indicate that for journalists and journalism researchers, when it comes to *Twitter*, there is something at stake. This paper addresses the challenges and opportunities for researchers as they study how *Twitter* is used in the context of specific journalistic activities. First, however, we sketch the possible uses of *Twitter* in relation to news events.

First and most obviously, *Twitter* is used for first-hand reporting of events as they occur; the simple format of *Twitter* messages and the near-ubiquitous accessibility of the network (with tweet-via-SMS functionality as ultimate fallback) combine to make live tweeting a more important practice on *Twitter* than comparable live activities have been for previous platforms. Such activities now include not only the reporting of events by actual eyewitnesses on the ground, from live coverage of the emergency water landing of a plane on New York's Hudson River (Subasic & Berendt, 2011) to the inadvertent (and only retrospectively discovered) live account of the raid on Osama bin Laden's compound by an Abbottabad local (Hill, 2011), but also second-hand live discussion of unfolding events as they are covered by other media – such as the worldwide public sharing of news on the Japanese earthquake and tsunami in March 2011, which generated some 140,000 tweets per hour containing the word ‘tsunami’ in the hours immediately following the earthquake (see Bruns, 2011a) or the appearance of Rupert and James Murdoch before the British Parliament's Culture, Media, and Sports Committee in July 2011 (Richards *et al.*, 2011).

An key driver here is the ease with which additional materials (links, photos, video, audio) can also be shared. This aids both the dissemination of first-hand, user-generated material documenting unfolding events as directly experienced by the user, and the sharing of secondary material in the form of links, screen captures, or even photos of TV screens. These additions extend *Twitter's* affordances far beyond the 140 character limit,

adding a rich media layer to the tweets themselves – and these multimedia materials often also make their way into mainstream media coverage.

Second, and consequently, *Twitter* is also used widely for ongoing discussion – and instant evaluation – of newsworthy events. This is triggered by mainstream media reporting, first-hand coverage by *Twitter* users themselves, or established interests of specific communities of *Twitter* users; it employs processes similar to what Bruns (2005) has described as “gatewatching”: highlighting, sharing, and evaluating relevant material released by other sources, in order to develop a more comprehensive understanding. What is shared, and how it is framed in discussion, provides an interesting alternative to conventional ‘vox pops’; ‘what *Twitter* thinks’ about an issue has become a staple feature of mainstream media coverage. One important aspect of news discussion practices on *Twitter* is the *curating* of information related to specific stories: the more active *Twitter* users frequently engage in gathering and sharing what they perceive to be relevant materials, for example tweeting links to further information (or retweeting relevant posts of other users) to their own followers or to *Twitter* communities formed around topical hashtags.

Finally, *Twitter*’s news coverage also consists of significant amounts of broader commentary on current events, reflecting mainly the senders’ own perspectives and intended more as markers of those perspectives than as formal contributions to debate. Here, perhaps, the “ambient” function of *Twitter*, highlighted by both Hermida (2010) and Burns (2010), emerges most clearly: for the average user, the majority of their *Twitter* use on any day may be taken up with non-news-related communication – much of it mundane and phatic – with *Twitter* as an ambient, always-on, always-in-the-background medium akin to ambient background music (Crawford, 2009). However, when important news breaks and spreads across the Twittersphere, shifts in tone and topical focus of incoming tweets may cause that user to pay attention to the story (again, much as a sudden shift in musical style may cause background music no longer to appear merely ambient).

Shifts from ambient to central are measured, simplistically, by *Twitter*’s own ‘trending topics’, and used by some newsrooms as additional measures of the newsworthiness of specific stories; more detailed analysis not only of such trends, but also of the relative attention paid to particular sources, political and other actors, or to the information shared by different *Twitter* users, is also possible. A recurring feature of the *Guardian*’s coverage of the *News of the World* voicemail hacking scandal, for example, have been visualisations of how *Twitter* users’ mentions of key figures and organisations in the scandal unfolded over time (see e.g. Richards *et al.*, 2011). Here, especially, journalism practitioners and researchers share a common interest: the development of more sophisticated methods and metrics for describing, analysing and representing *Twitter*’s response to news events.

## Twitter as Social Networking Site and Information Stream: Followers and Hashtags

*Twitter* is the most prominent example of a recent shift in social media, which has seen the convergence of explicit networking practices (‘friending’, ‘following’, interpersonal communication) with original content (‘broadcasting’ of updates), and large-scale information sharing and propagation. It is *through* the social network that news and information spreads: *Twitter* is both a social networking site and an ambient information stream. This convergence (cf. Kwak *et al.*, 2010) underpins *Twitter*’s significance for journalism; any evaluation of user activities on *Twitter* must reflect on the structural aspects of this convergence. Explicit networking structures among users determine – to an important extent, though not exclusively – which tweets are visible to what subset of the total userbase (now over 200 million worldwide; cf. White, 2011).

Overall, *Twitter*’s communicative structure is determined by two overlapping and interdependent networks – one long-term and relatively stable, based on follower-followee relationships; one relatively short-term and emergent, based on shared interest in a topic and coordinated by a common hashtag. First, *Twitter* users are able to ‘follow’ one another: all tweets originating from the followee will automatically be visible to the follower, in an update feed combining tweets from all followed users.<sup>1</sup> Such relationships are usually based on a longer-term interest in updates from the followee; the overall network structure shows

unsurprising tendencies to cluster around key interests or attributes shared by communities of followers (Java *et al.*, 2007), much as in other social networks (e.g. Adamic & Glance, 2005; Kelly & Etling, 2008). Our own investigation shows that even an as yet incompletely mapped network of some 440,000 Australian *Twitter* users already tends towards clustering around shared professional and cultural interests (Bruns, 2011c).

To the extent that *Twitter* users consciously understand this network structure, their responses to newsworthy events address and interact with their own immediate community of followers, and may also attempt to overcome the barriers dividing specific clusters in the network. The imagined immediate audience for any one tweet is likely to be the user's network of followers; receiving responses from, or being retweeted by, these followers is seen as possible and perhaps even desired. Indeed, some tweets carry explicit encouragements to 'pls RT', asking for messages to be spread well beyond the user's own follower network.

Similarly, while retweeting practices vary widely in meaning and intention (boyd *et al.*, 2010), they reflect implicit understandings of *Twitter*'s network structure, recognising that unaided, original messages will reach only a limited number of users, and that further passing-along amplifies their visibility. Retweeting users may even see themselves as information brokers, bridging distinct communities of interest by passing on tweets from one network cluster to another. Retweeting can also be interpreted as an implicit endorsement for message and sender, unless additional commentary is added by the retweeter during retweeting; especially where original messages stem from *Twitter* users who are already widely visible, the primary intention may well be commentary rather than dissemination, however.

*Twitter* follower/followee structures are far from static: following another user only takes a few clicks. Nonetheless, it such structures are unlikely to fluctuate wildly for any one user: substantial changes to one's list of followees also significantly change the focus of one's incoming update stream. Analogous to blogrolls, *Twitter* follower/followee networks represent the long-term interests, rather than the short-term foci, of individual *Twitter* users.

Follower/followee networks, however, are also overlaid by another mechanism for coordinating *Twitter* communication: hashtags. These brief keywords or abbreviations, prefixed by the hash symbol, are included in tweets – for example, #auspol for discussing of Australian politics, or #NotW to track the *News of the World* scandal. Hashtags are simply entered manually by users as they tweet; they can be created *ad hoc*, and emerge almost instantaneously as news breaks. Hashtags make topical tweets more visible: drawing on *Twitter*'s search functionality, users can find (and subscribe to) all tweets marked with the same hashtag, regardless of whether these tweets originate from established followees or previously unknown users.<sup>ii</sup>

Hashtags enable users to communicate with an *ad hoc* community around the hashtag topic without needing to establish mutual follower/followee relationships with any of the other participants. It is this flexibility and ability to rapidly form discursive communities around breaking news which underlies *Twitter*'s recognition as a platform for news dissemination and discussion. A drawback of the *ad hoc* and non-supervised emergence of hashtags, however, is that competing hashtags may emerge (for example, #eqnz and #nzeq for coverage of the Christchurch earthquakes in 2010/2011), or that the same hashtag may be used for vastly different events simultaneously (for instance, #spill for the 2010 BP oil spill in the Gulf of Mexico, and for the leadership change in the Australian Labor Party on 23 June 2010). *Twitter* users themselves will often work to resolve such conflicts as soon as they are identified.

Using a hashtag can be seen as an explicit attempt to address an imagined community of users – and the network of *Twitter* users formed this way is separate from follower/followee networks. However, the two network layers overlap: hashtagged tweets are visible *both* to the sender's established followers, *and* to anyone else following the hashtag conversation. Each user participating in a hashtag conversation potentially acts as a bridge between the hashtag community and members of their own follower network, therefore. But not all users posting *to* a hashtag conversation also *follow* that conversation: they may use a hashtag to make their tweets visible to others, but may themselves focus only on tweets coming from their established followers. Conversely, not all relevant conversations following on from hashtagged tweets will themselves carry the hashtag: explicitly hashtagging a response to a previous hashtagged tweet may be understood as

performing the conversation in front of a wider audience, by comparison with the more limited visibility of non-hashtagged responses.<sup>iii</sup>

Finally, we note a unique communicative feature of the *Twitter* platform: excepting 'private' accounts, tweets are generally visible also to non-members. Such users may visit the *Twitter* Website to view tweets from specific accounts, or use *Twitter's* search functionality to find all tweets containing particular keywords or hashtags. This further adds to the value of hashtags as a mechanism for coordinating news discussion and information curation.

## Methods for Researching *Twitter*

For journalism scholars and practitioners, there is a clear need to develop further methods for researching *Twitter's* response to news and current events. *Twitter* provides significant levels of access to data on user activities through its Application Programming Interface (API): an interface which is designed predominantly for use by *Twitter* clients, but can also be used for tracking current activity by users, or on specific keywords and hashtags. However, there are substantial limitations to what is available directly through the API, and these limitations have been gradually tightened as *Twitter* seeks further revenue. Changes made late in 2010 (Melanson, 2011) mean that even for the purposes of publicly funded, non-commercial research, it is no longer possible to gain access to the full 'firehose' of *all* tweets, or to substantial subsets of this full feed; tracking the ongoing public activities of more than 5,000 *Twitter* users at a time is now only possible by working with *Twitter's* licenced third-party API provider Gnip, at a cost well beyond the funding available to most research projects. This effectively rules out academic studies that track the current activities and thematic interests of large, representative samples of *Twitter* users.<sup>iv</sup>

A more immediately achievable direction for research into news and current events is to track keywords and/or hashtags. The *Twitter* API enables users to automatically capture any tweets containing given keywords (including hashtags), with comparatively few limitations, and tools for doing so are readily available to researchers; of these, the leading open-source solution to date is *yourTwrapperkeeper* (*yTK*).<sup>v</sup> With minor modifications to improve data export (see Bruns, 2011b), *yTK* can be used to track a substantial number of keywords simultaneously; further processing and evaluation of these datasets reveals many important patterns.

Such research draws on various additional tools: the data processing tool Gawk, statistics packages such as Excel, textual processing software like Leximancer or WordStat, and the open-source package Gephi for network analysis and visualisation. This paper provides an overview of approaches to analysing these datasets; it does not intend to act as a hands-on methodological guide. However, our project Website at <http://mappingonlinepublics.net/> offers substantial practical advice on processing *yTK* data, and access to the custom-made open source software tools we have developed.

Here, we outline the insights that this data-driven approach provides. We acknowledge that a keyword- or hashtag-based study of *Twitter* activity is necessarily limited, and will not capture tweets of the discussion which do not explicitly include the chosen terms. Therefore, the data must be understood as a reasonably representative sample rather than a comprehensive dataset of activities around an event – hashtag datasets in particular are weighted considerably towards the most engaged subset of *Twitter* users (those engaged enough to include the hashtag in their tweets), and thus towards a comparatively élite group. This does not invalidate such studies any more than studies of the tenor of letters to the editor in a leading newspaper would be invalidated by that newspaper's specific audience composition. Additionally, the overall make-up of the *Twitter* userbase, and its lack of correlation with overall demographic patterns, also means that views of *Twitter* users can never simply be regarded as representative for the general population.

## Development over Time

The simplest form of analysis is the study of activity patterns over time. This charts the number of tweets made during any one period, possibly breaking them down into a number of categories. Fig. 1, for example,

depicts the daily number of tweets made to the #ausvotes hashtag in the lead-up to and immediate aftermath of the 2010 Australian federal election, showing a gradual ramping-up of activity and a major spike on election day; another minor early spike occurs on 25 July, during the televised debate between the party leaders.<sup>vi</sup>

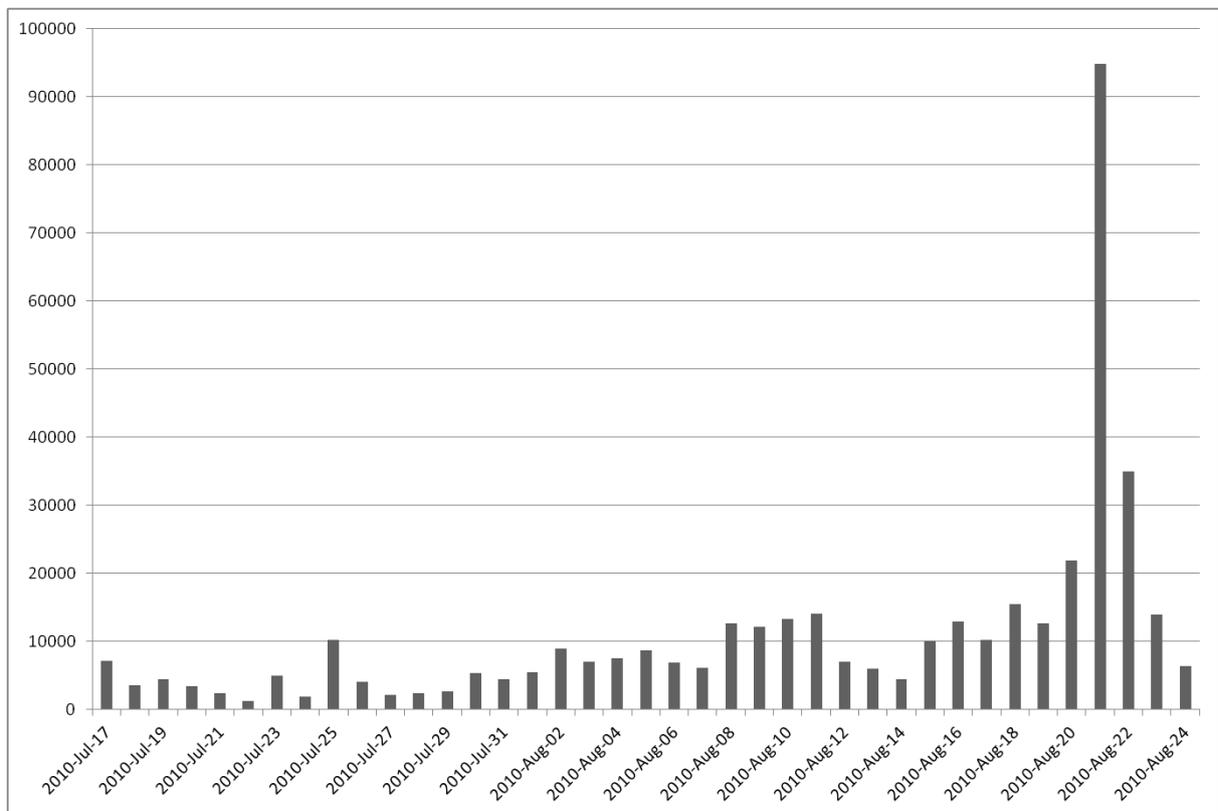


Figure 1: #ausvotes tweets, 17 July - 24 Aug. 2010

By contrast, Fig. 2 shows minute-by-minute activity under the #royalwedding hashtag, following the wedding of Prince William and Kate Middleton on 29 April 2011; here, activity is broken down into the overall number of tweets (in black), retweets, @replies, and tweets containing URLs (various shades of grey). This graph points to the importance of television, with significant increases in volume at 7:00 and 8:15 GMT, as major networks switched to their live coverage; it also highlights key moments of the day: activity spikes around 12:25, for example, during the newlyweds' first kiss.

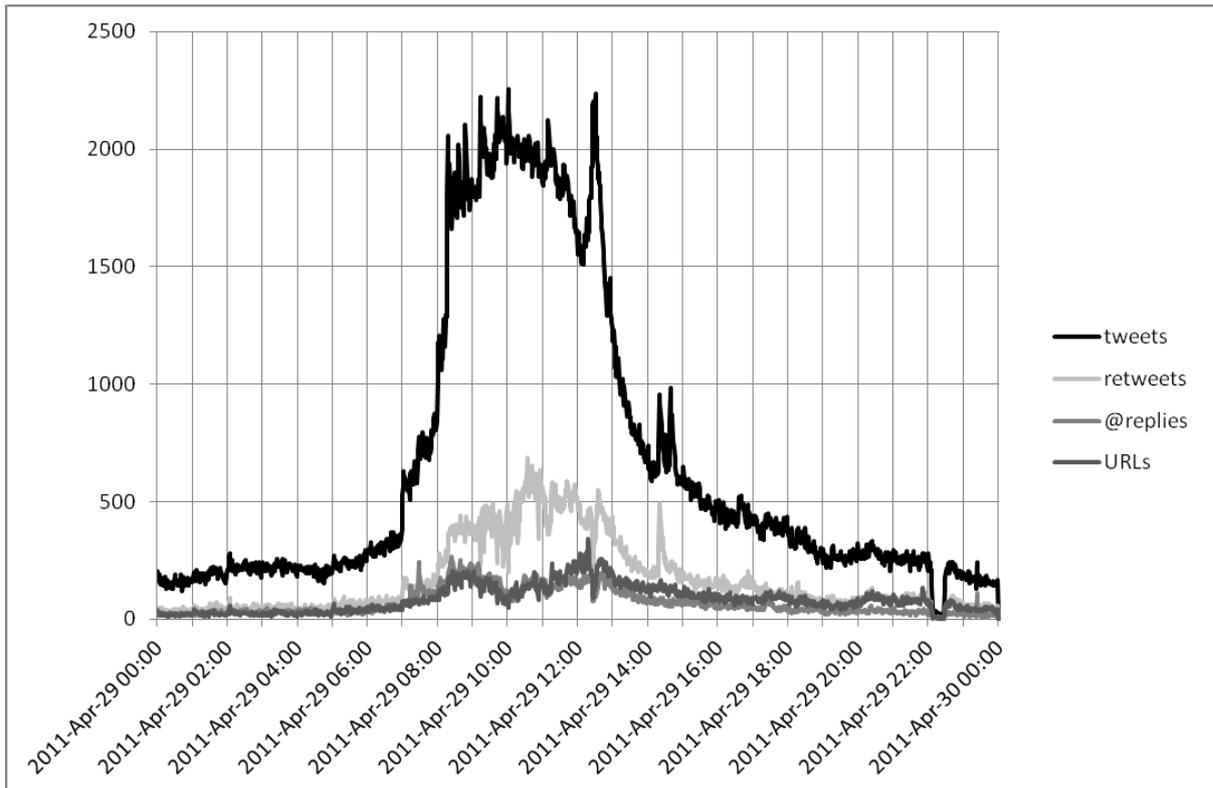


Figure 2: #royalwedding tweets, 29 Apr. 2011

These simple metrics provide clear insights into what moments had particular resonance; they may be correlated with key developments in a story. Inferences about users' media practices may also be drawn: both figs. 1 and 2 point to the use of *Twitter* as a backchannel for television, for example, and to the continuing resonance of mainstream media tropes in the coverage of these events (debates and election night broadcasts, vows and kisses) even in social media environments.

## Key Users

Further, it is useful to identify key participants in the discussion. Given obvious ethical concerns with highlighting activities of individual users, the goal here is not to engage in detailed profiling of individuals, but to establish the overall community structure. In doing so, it is usually less important to examine the total number of tweets *sent* by each user (the volume of contributions made does not provide a reliable approximation of the impact of those tweets), but rather to focus on the number of responses (that is, public @replies) and retweets *received*.<sup>vii</sup>

It should also be expected that most messages received by an account in a hashtag conversation are retweets rather than @replies: as noted, @reply responses to hashtagged tweets will not usually include the hashtag, unless respondents deliberately choose to make their response visible to the wider hashtag community. Much of the follow-on conversation around the hashtagged topic will therefore be absent from the dataset.

An analysis of responses and retweets provides a useful indication of the overall visibility of each account: as discussed, retweets are a means of amplifying the reach of a tweet, and thus of increasing the visibility of a tweet and its sender, while the @replies received by a user can be seen as a direct result of the visibility of their tweets, and thus of the user themselves. Fig. 3, for example, shows the most visible users participating in #eqnz in the immediate aftermath of the second Christchurch earthquake on 22 Feb. 2011. The field is led by the *Twitter* account of newspaper *NZ Herald*, which received over 9,000 retweets and @replies, while key emergency authorities (including the Canterbury Earthquake Authority and Christchurch City Council) also

feature prominently. This indicates that 'official' accounts are able to establish themselves as authoritative sources of information, even in the open environments of social media.

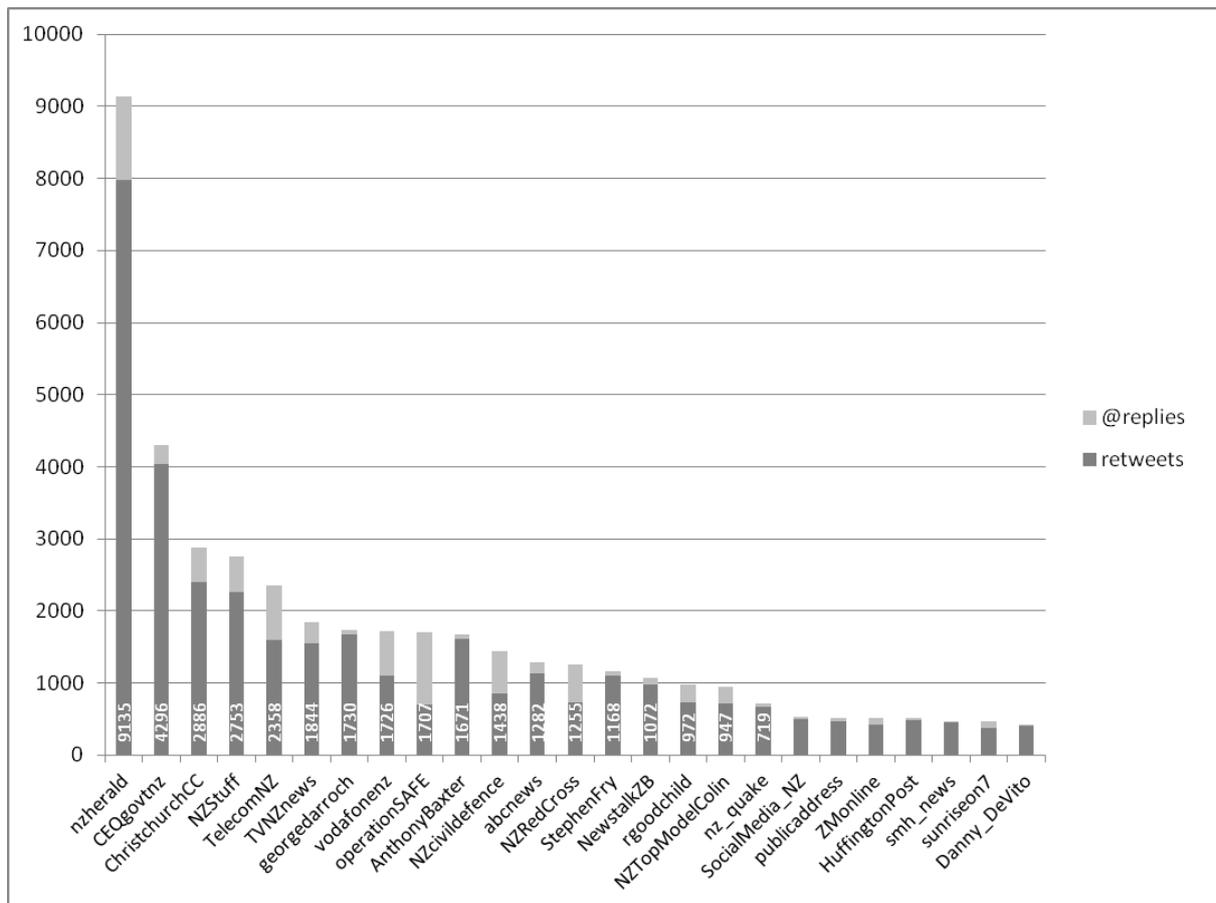


Figure 3: Key users in #eqnz, 22 Feb. - 7 Mar. 2011

For such official accounts, more in-depth further analysis is also ethically acceptable. Fig. 4, for example, examines the performance of the Queensland Police Service Media Unit *Twitter* account (@QPSmedia) during the southeast Queensland flood crisis on 10-16 January 2011. Similar to the Christchurch experience, a handful of official accounts emerged as the most visible, led here by @QPSmedia; as fig. 4 indicates, the relatively limited number of @QPSmedia tweets to #qldfloods (dashed line, peaking at 17 on 11 January) were amplified very significantly through retweets (resulting in 601 manual retweets that same day). This clearly indicates the impact of retweeting on the visibility of messages from official accounts.

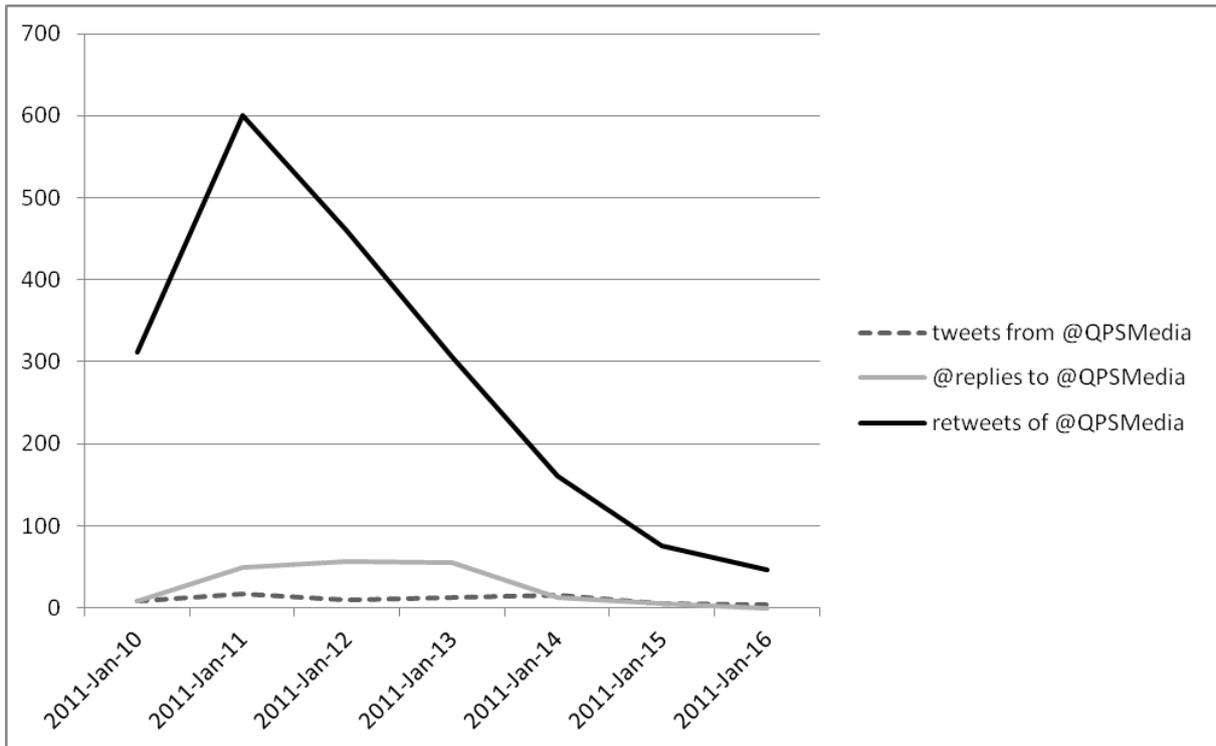


Figure 4: @QPSmedia during #qldfloods, 10-16 Jan. 2011

### Mentions of Key Concepts and Key Actors

Much like tweets and retweets can be tracked over time, it is also possible to examine the presence of names and themes. This requires preparatory work to identify the key terms or named actors occurring across the entire corpus; researchers can then define bundles of terms to be tracked. Fig. 5, for example, shows a cumulative count of #ausvotes mentions of the two main candidates for Australian Prime Ministership during the 2010 campaign, and indicates that their numbers are virtually equal until 10 August; from that day, following a controversial television appearance, Opposition Leader Tony Abbott (in black) takes a clear lead.

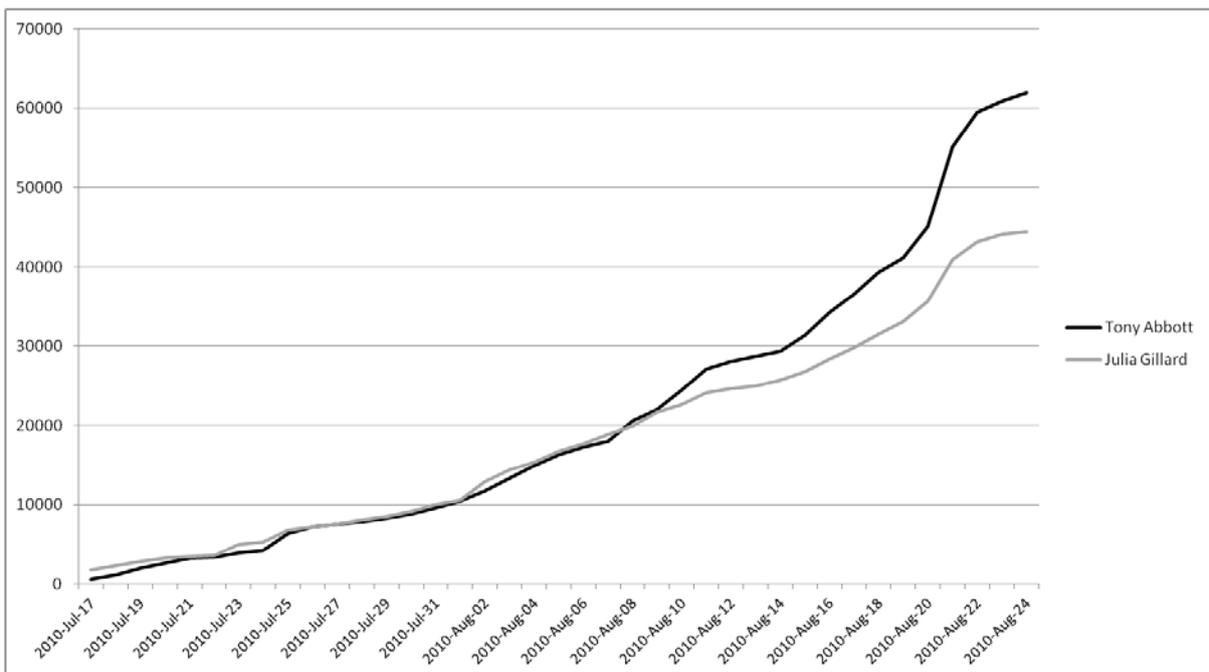


Figure 5: Leader names on #ausvotes, 17 July - 24 Aug. 2010

Similarly, Fig. 6 tracks five bundles of keywords related to the key themes of the campaign, normalising mentions to 100% for each day. What emerges here is both an overall bias in the *Twitter* conversation on #ausvotes, in comparison with mainstream media coverage, towards ICT themes, and a shift in discursive focus after 10 August: from the mandatory Internet filter proposed by the Labor government to the National Broadband Network project opposed by its conservative challengers, and criticised by Abbott in his television interview.

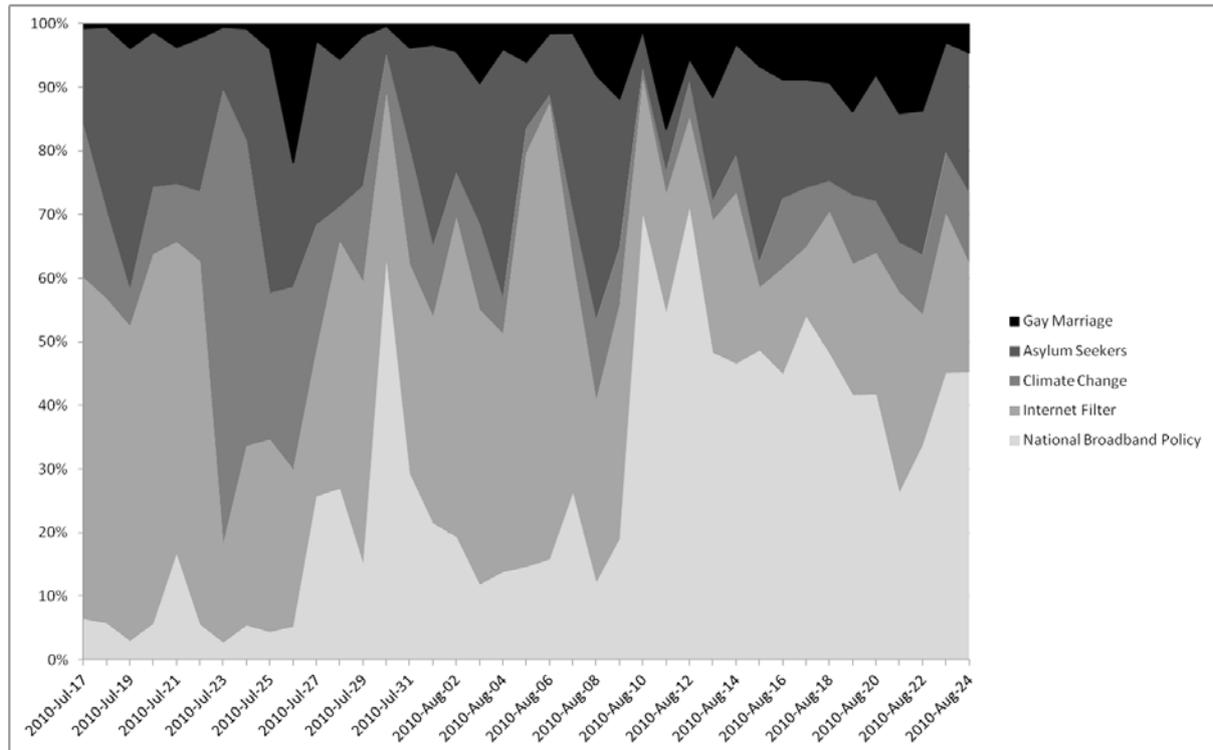


Figure 6: Key #ausvotes themes, 17 July - 24 Aug. 2010

### Advanced Network Analysis

Finally, network analysis provides further possibilities. One approach here is the analysis of the (temporary) social networks evident in response and retweeting patterns; such analysis supports the statistical identification of leading accounts, discussed above. Social network analysis adds various other metrics: it points to participants, for example, with high 'betweenness' – that is, importance as connectors of otherwise distant parts of the network; such users may not be major net sources or recipients of information, but are highly instrumental in ensuring the movement of information *across* the network.

Fig. 7, for example, shows the network of retweets and @replies in the #spill hashtag, discussing rumours of a leadership challenge in the Australian Labor Party during 23 June 2010. While not actively participating, the account of then Prime Minister @KevinRuddPM is featured prominently as an addressee of @replies, while some journalists are prominent as senders and receivers of @replies and retweets. It is even possible to visualise developments over time in a series of snapshots or as a dynamic animation of the available social network data (see Bruns, 2011d).



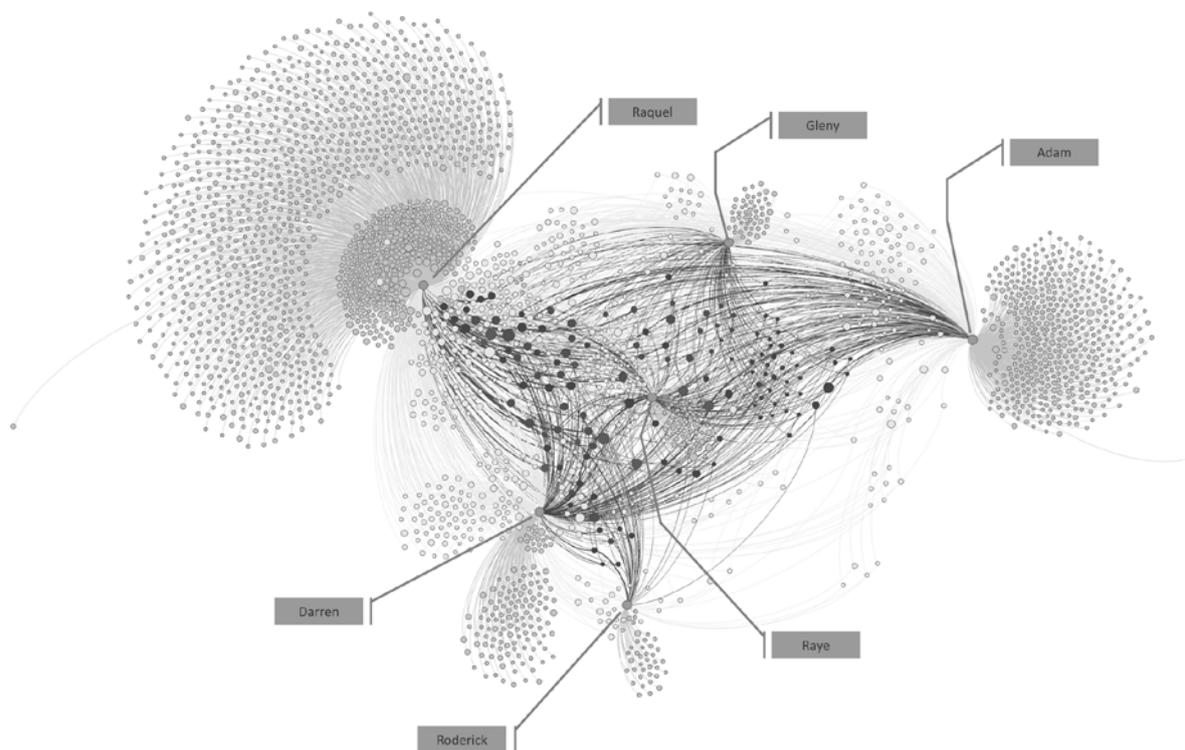


Figure 8: Mentions of #GoBackSBS participants

## Conclusion

The approaches outlined here are readily available to researchers, and provide detailed insights into how *Twitter* users engage in newsworthy events. Such observations should not be seen as representative for society at large, any more than letters to the editor are representative, but how the *Twitter* userbase reacts to events is relevant in its own right. Especially where our methods can be used to examine the interrelation between *Twitter* use and other forms of (mainstream) media participation, there are significant further benefits for scholars: compared to studies which build on such relatively artificial devices as media usage surveys or diaries, tracking *Twitter* activities provides a more *in situ* picture of media engagement, comparatively unaffected by the act of observation. Additionally, many of these analyses can be performed virtually in real time; this enables researchers to respond very quickly to current events, and to examine popular responses on *Twitter* even while events are still unfolding.

This research does rely on the continued availability of data through the *Twitter* API – and researchers are treated as no more than uninvited guests by *Twitter* at this point, with no guarantees of continued access and a push by *Twitter* towards the commercialisation of large-scale data access (Steele, 2011). It would be most unfortunate for this trend to continue: commercialisation would effectively rule out much publicly funded research into *Twitter* use, but it is precisely these scholarly studies which clearly document the important role which *Twitter* now plays in public communication.

Even if *Twitter* were to become an increasingly hostile environment for researchers, however, the time spent developing *Twitter* research methods remains time well spent. It is inconceivable for any future social media platform not to offer an API comparable to that of *Twitter*, and while technologies may change, the methods which we have outlined here will be transferable – *mutatis mutandis* – to the study of other online social networks as well.

## Index of Figures

URLs for higher-resolution, colour versions of the figures in this paper:

- 1: <http://www.mappingonlinepublics.net/2010/11/24/election-2010-the-view-from-twitter/>
- 2: <http://www.mappingonlinepublics.net/2011/08/12/twitter-and-the-royal-wedding-pt-1-something-processed/>
- 3: <http://www.mappingonlinepublics.net/2011/07/12/emergency-management-conference/>
- 4: <http://www.mappingonlinepublics.net/2011/07/12/emergency-management-conference/>
- 5: <http://www.mappingonlinepublics.net/2010/11/24/election-2010-the-view-from-twitter/>
- 6: <http://www.mappingonlinepublics.net/2010/11/24/election-2010-the-view-from-twitter/>
- 7: <http://www.mappingonlinepublics.net/2010/11/24/election-2010-the-view-from-twitter/>
- 8: <http://www.mappingonlinepublics.net/2011/07/07/tweeting-at-the-tv-some-observations-on-gobacksbs/>

## References

- Adamic, Lada, and Natalie Glance (2005) "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog." 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Chiba, Japan, 10 May 2005. <http://www.blogpulse.com/papers/2005/AdamicGlanceBlogWWW.pdf>, accessed 1 Aug. 2011.
- Ahmad, Ali Nobil (2010) "Is Twitter a Useful Tool for Journalists?" *Journal of Media Practice*, 11(2), pp. 145-55.
- boyd, danah, Scott Golder and Gilad Lotan (2010) "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter." Proceedings of 43rd Hawaii International Conference on System Sciences, IEEE. <http://doi.ieeecomputersociety.org/10.1109/HICSS.2010.412>, accessed 1 Aug. 2011.
- Bruns, Axel (2005) *Gatewatching: Collaborative Online News Production*. New York: Peter Lang.
- (2006) "The Practice of News Blogging." In Axel Bruns and Joanne Jacobs (eds.), *Uses of Blogs*. New York: Peter Lang, pp. 11-22.
- (2011a, 10 Feb.) "Twitter Events in Perspective (updated)." *Mapping Online Publics*, <http://www.mappingonlinepublics.net/2011/02/10/twitter-events-in-perspective/>, accessed 1 Aug. 2011.
- (2011b, 21 June) "Switching from Twapperkeeper to yourTwapperkeeper." *Mapping Online Publics*, <http://www.mappingonlinepublics.net/2011/06/21/switching-from-twapperkeeper-to-yourtapperkeeper/>, accessed 1 Aug. 2011.
- (2011c, 4 Aug.) "A First Map of Australia." *Mapping Online Publics*, <http://www.mappingonlinepublics.net/2011/08/04/a-first-map-of-australia/>, accessed 1 Aug. 2011.
- (2011d). "How Long Is a Tweet? Mapping Dynamic Conversation Networks on Twitter Using Gawk and Gephi." *Information, Communication & Society*, 17 Nov. 2011.
- Bruns, Axel, and Jean Burgess (2011) *Mapping Online Publics*, <http://mappingonlinepublics.net/>, accessed 1 Aug. 2011.
- Burns, Alex (2010, May) "Oblique Strategies for Ambient Journalism." *M/C Journal* 13(2). <http://journal.media-culture.org.au/index.php/mcjournal/article/view/230>, accessed 1 Aug. 2011.
- Crawford, Kate (2009) "Following You: Disciplines of Listening in Social Media." *Continuum: Journal of Media & Cultural Studies* 23(40): pp. 525-35.
- Hermida, Alfred (2010, May). "From TV to Twitter: How Ambient News Became Ambient Journalism." *M/C Journal* 13(2). <http://journal.media-culture.org.au/index.php/mcjournal/article/view/220>, accessed 1 Aug. 2011.
- Hill, Amelia (2011) "Osama bin Laden Killed - How a Live Blogger Captured the Raid." *The Guardian*. <http://www.guardian.co.uk/world/2011/may/02/osama-bin-laden-live-blogger>, accessed 1 Aug. 2011.

- Farhi, Paul (2009) "The Twitter Explosion." *American Journalism Review* April/May. [http://ajr.org/article\\_printable.asp?id=4756](http://ajr.org/article_printable.asp?id=4756), accessed 1 Aug. 2011.
- Java, Akshay, et al. (2007) "Why We Twitter: Understanding Microblogging Usage and Communities." Paper presented at *WebKDD/SNA-KDD '07*, New York. <http://portal.acm.org/citation.cfm?id=1348556>, accessed 1 Aug. 2011.
- Kelly, John, and Bruce Etling (2008) "Mapping Iran's Online Public: Politics and Culture in the Persian Blogosphere." Boston: Berkman Center for Internet & Society. [http://cyber.law.harvard.edu/publications/2008/Mapping\\_Irans\\_Online\\_Public](http://cyber.law.harvard.edu/publications/2008/Mapping_Irans_Online_Public), accessed 1 Aug. 2011.
- Kwak, Haewoon, et al. (2010) "What Is Twitter, a Social Network or a News Media." *WWW'10: 19th International Conference on World Wide Web*. <http://portal.acm.org/citation.cfm?id=1772751>, accessed 1 Aug. 2011.
- Melanson, Mike (2011, 11 Feb.) "Twitter Kills the API Whitelist: What It Means for Developers & Innovation." *ReadWriteWeb*. [http://www.readwriteweb.com/archives/twitter\\_kills\\_the\\_api\\_whitelist\\_what\\_it\\_means\\_for.php](http://www.readwriteweb.com/archives/twitter_kills_the_api_whitelist_what_it_means_for.php), accessed 1 Aug. 2011.
- Raymond, Matt (2010, 25 Apr.) "How Tweet It Is!: Library Acquires Entire Twitter Archive." *Library of Congress Blog*. <http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/>, accessed 1 Aug. 2011.
- Richards, Jonathan, et al. (2011, 13 July) "How Twitter Tracked the News of the World Scandal." *The Guardian*. <http://www.guardian.co.uk/media/interactive/2011/jul/13/news-of-the-world-phone-hacking-twitter>, accessed 1 Aug. 2011.
- Steele, Julie (2011) "Data Markets Aren't Coming, They're Already Here." *O'Reilly Radar*. <http://radar.oreilly.com/2011/01/data-markets-resellers-gnip.html>, accessed 1 Aug. 2011.
- Subasic, Ilija, and Bettina Berendt (2011) "Peddling or Creating? Investigating the Role of Twitter in News Reporting." In P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, and V. Mudoch (eds.), *Advances in Information Retrieval*. Berlin: Springer, pp. 207-13.
- White, Charlie (2011) "Reaching 200 Million Accounts: Twitter's Explosive Growth." *Mashable*. <http://mashable.com/2011/07/16/twitter-accounts-200-million/>, accessed 1 Aug. 2011.

<sup>i</sup> Generally, it is possible for a user to follow any other *Twitter* user; however, users may set their accounts to 'private', in which case they will have to approve any follow requests.

<sup>ii</sup> One remaining limitation is that hashtagged tweets from accounts marked as 'private' will not be included in the search results.

<sup>iii</sup> Hashtags are also used in other contexts – for example as ironic markers of one's mood (#tired) or to highlight key terms (#Australia) –, but our focus in the following discussion is purely on *topical* hashtags.

<sup>iv</sup> *Twitter* has entered into an agreement with the U.S. Library of Congress to make its full 'firehose' feed available to the Library, for access by researchers after a six-month embargo (Raymond, 2010); however, the embargo means that research into *current* coverage and news discussion remains impossible, and modalities of access to the dataset at the Library are as yet unknown.

<sup>v</sup> See <http://your.twapperkeeper.com/>

<sup>vi</sup> Higher-resolution, colour versions of all figures presented in this paper are available online – see the Index of Figures below for URLs.

<sup>vii</sup> We note that *yourTwapperkeeper* only captures *manual* retweets ("RT @user ..."), not retweets made using *Twitter's* more recent 'retweet button' functionality. This is not necessarily a drawback: because they can be edited before sending, manual retweets serve significantly more conversational functions than 'button' retweets; for example, users will often retweet part of an earlier message in order to add their own, original commentary. 'Button' retweets, on the other hand, constitute merely a verbatim passing-along of the original message, with retweeting users unable to include any additional comments with the retweeted message. While tracking the amount of button retweets for each message might provide an interesting additional dimension to the analysis, it does not have significant relevance to the analysis of actively *discursive* interaction.