Towards a Typology of Hashtag Publics: A Large-Scale Comparative Study of User Engagement across Trending Topics

Axel Bruns, Brenda Moon, Avijit Paul, and Felix Münch
Digital Media Research Centre
Queensland University of Technology
Brisbane, Australia
a.bruns / brenda.moon / a1.paul / felix.muench @ qut.edu.au

with contributing authors Jan Schmidt, Fabio Giglietto, Steven McDermott, Till Keyling, Xi Cui, Steffen Lemke, Isabella Peters, Athanasios Mazarakis, Yu-Chung Cheng, and Pailin Chen

Introduction

From #sandiegofire to #egypt, from #kony2012 to #illridewithyou, hashtags have become one of *Twitter*'s most prominent and enduring features. Although themselves based on the channel tags of Internet Relay Chat (IRC), and popularised to the point that even social media market leader *Facebook* has experimented with implementing hashtags on its platform – with limited success –, hashtags are particularly well suited to the communicative affordances of the *Twitter* platform: any *Twitter* user can create new hashtags and post hashtagged messages to their followers and the overall *Twitter* user community; hashtags are now automatically highlighted in tweets, and made clickable, by the *Twitter* Website and most tweeting apps and tools, and thus provide instant access to the full stream of other tweets using the same hashtag; and hashtags thereby enable *Twitter* users – and even non-registered visitors to the *Twitter* site – to discover the public posts of a very wide range of other contributors to the platform, without first needing to follow these users.

Existing research has therefore especially highlighted the utility of hashtags in rapidly bringing together *ad hoc* publics (Bruns & Burgess, 2015) with a shared interest in specific events, issues, and topics; indeed, the study of this function is at the centre of the recent collection *Hashtag Publics* (Rambukkana, 2015), which focusses especially on documenting the various political uses of hashtags. The hashtag studies which collections such as this present can be understood as representing the dominant stream of *Twitter* (and arguably, of social media) research to date, largely because the current technical constraints on data gathering that are imposed by the Application Programming Interfaces (APIs) provided by *Twitter* make it considerably easier to track and capture hashtag (and, by extension, keyword) streams than any other form of *Twitter* data. In essence, hashtag datasets thus constitute the low-hanging fruit in social media data, which has led to an abundance of research building on such datasets, compared to a relatively dearth of studies drawing on less instantly accessible sources (Burgess & Bruns, 2015).

Even within the field of hashtag studies, in fact, a particular category of hashtags appears to dominate: much of the published work has focussed especially on hashtags relating to key events in news and politics, with comparatively less work on the many other approaches to using hashtags. (No detailed bibliometric studies documenting these patterns exist to date, but *Google Scholar*, for instance, in early 2016 finds some 16,300 articles referencing "hashtag" and "news", but only just over 4,000 each referencing "hashtag" and "entertainment" or "hashtag" and "sports", and 2,000 examining "hashtag" and "meme".) This current state of the literature is problematic for a number of reasons: first, it is known at least anecdotally that hashtags now serve a wide range of purposes: amongst them, the well-understood function of assembling an *ad hoc* public around a key issue; gathering of a community of practice engaging in shared, possibly concurrent activities (such as attending a live entertainment or sporting event, or using *Twitter* as a backchannel to radio and TV broadcasts); attempting to create and promote a (playful or serious) meme that is virally distributed across local, national, and global *Twitter* networks; or introducing a point of emphasis that – similar to an emoticon or emoji – carries a stronger semiotic charge than a word alone would be able to do. A more thorough investigation of any and all of these uses of hashtags – and of the many others not included in the list above – has yet to be conducted, and would shed considerably more light on the full range of contemporary hashtag uses now evident on *Twitter*.

Second, even in relation to those hashtag uses which have already been studied in considerable detail, there is still a notable absence of comparative studies that examine the similarities and differences between specific cases – for example, across the hashtags used to track various election campaigns or natural disasters. An article by Bruns & Stieglitz (2012) begins this work, but draws on the limited range of datasets then available to the researchers; it serves as a basis for the present work, but needs to be updated with a collection of newer and more diverse hashtag datasets in order to capture the full breadth of hashtag uses that have been established by now.

Third, hashtags represent only a specific subset of all of the communicative layers provided by *Twitter* (Bruns & Moe, 2014), and arguably do not even constitute the most prominent of the key forms of communication on the platform: that honour must surely go to the layer of communicative exchanges enabled by the networks of follower/followee relationships between accounts, which determine the majority of information flows on *Twitter*. There is therefore a strong need to put hashtag use into better perspective also by comparing the patterns of user engagement around topical hashtags with the broader patterns of activity relating to these topics outside of the hashtags themselves – however methodologically difficult such work may turn out to be. We stress here that none of these critiques are meant to belittle extant scholarly research in this field; such work has been important and valuable in its own right, and has produced rich insights into the uses of *Twitter* across a range of cases and contexts. However, it remains necessary to now take the next steps and consolidate such work by developing more robust comparative and longitudinal approaches to *Twitter* research.

This is important especially also because *Twitter* itself continues to change and evolve. Users who have joined the platform more recently may not have been socialised into using it – and in particular, hashtag functionality – in the same way as the early adopters, for instance; their expectations of hashtag use will be coloured more strongly by media reporting about hashtag events (from crisis event hashtags such as #qldfloods through viral responses to political gaffes such as #bindersfullofwomen to major international TV events such as #eurovision) than by long-term experience, and they may also seek to replicate the affordances they have come to know from other leading social media platforms, such as *Facebook*. Twitter, Inc.'s own interventions as it seeks to make the platform more palatable to such new users – and in doing so risks alienating seasoned users by breaking some of the unwritten rules they had established – also affect the future trajectory of hashtags and other *Twitter* functionality: across successive site redesigns, the relative visibility of *Twitter*'s trending topics list or the development of the "While You Were Away" function, for instance, also affect the prominence of leading hashtags, and may lead to more or less user engagement with these hashtags (at least by those users who are accessing *Twitter* through its Website or official apps).

This article, then, continues and extends the work begun by Bruns & Stieglitz (2012) by adding a substantial number of new cases to the comparative analysis begun in that article. In particular, we focus on the most notable pattern identified by the 2012 study: it observed that across the hashtag datasets it evaluated, measures of the percentage of retweets and of the percentage of tweets containing URLs in each dataset clearly clustered around two focal points. On the one hand, "acute events" (from natural disasters to political unrest) reliably contained some 40-75% URL tweets and some 35-65% retweets; on the other hand, "media events” (from major sports and entertainment broadcasts to election-night political coverage) usually generated only some 0-20% URL tweets and 15-35% retweets (illustrated in fig. 1). This suggested a very different approach to engaging in hashtag activities across these two categories of cases: a communal “audiencing” (cf. Fiske, 1992) of media events where users participate in hashtagged posting, but comparatively rarely share additional external information in the form of new URLs or amplify other contributors’ tweets by retweeting, and a dedicated “gatewatching” (cf. Bruns, 2005) of acute events where users actively seek out additional material and share it by posting new URLs into the hashtag, and where they help increase the visibility of already available material by frequently retweeting those tweets they deem to be important to others.

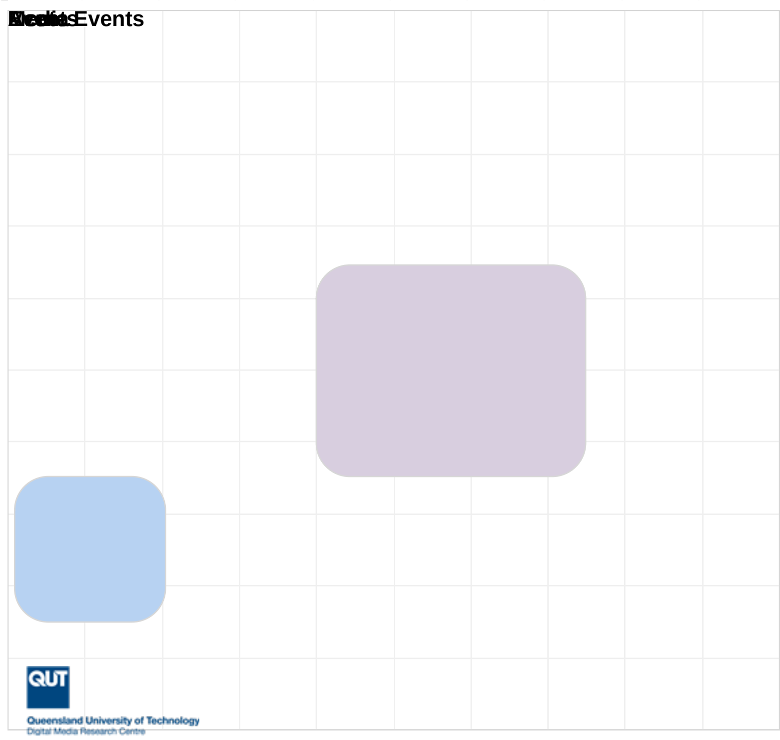

Fig. 1: Illustration of key hashtag types as identified in Bruns & Stieglitz (2012)

The identification of these two distinct user behaviours, related to different event types, raises the possibility that future hashtag events may be categorisable as specific event types on the basis of such patterns alone, if such user activity patterns remain reasonably stable and even become habitualised over the longer term; however, given the fluctuations in the userbase and the evolution of use practices which we would expect to see for any online platform, including *Twitter*, this is by no means guaranteed. Also, the 2012 paper necessarily drew only on the hashtag datasets readily available to the authors, and was thus influenced strongly by their longer-term topical research interests; datasets on hashtag topics not covered in the original article may lead to the discovery of new, similarly distinct activity patterns around other event types, or may undermine any neat distinctions between media, acute, and other events.

Our article therefore builds on the original collection of datapoints used by Bruns & Stieglitz (2012), and adds to this the corresponding metrics for a significant number of new datasets, drawn from a range of sources. In particular, we are indebted to the contributions of a number of colleagues in the Association of Internet Researchers who responded to an open call to contribute datapoints from their own archival datasets, and we wish to acknowledge them as contributing authors to this paper; any errors in the analysis and interpretation of the overall collection of datapoints are ours and ours alone, however.

Data Sources and Methodology

For this study we considered the same set of metrics used in Bruns & Stieglitz (2012), with particular focus on figure 4 of that article:

- the number of tweets in the hashtag dataset;
- the percentage of retweets in the hashtag dataset (including button retweets as well as manual retweets using the RT @user, MT @user, HT @user, via @user, or "@user formats);
- the percentage of tweets in the hashtag dataset that contain URLs (which we will describe as 'URL tweets' in the subsequent discussion).

We have supplemented the datasets used by Bruns & Stieglitz (2012) with new datasets captured using a range of different approaches. In addition to *yourTwapperkeeper* (2012), the open-source platform for tracking and capturing *Twitter* hashtag and keyword datasets that was used in that article, we also used another, similar open-source platform, *TCAT* (Borra & Rieder, 2014), to collect several of the new datasets. Where *yourTwapperkeeper* uses both the *Twitter* search and streaming API functionality, *TCAT* only uses the *Twitter* streaming API to gather data; however, both platforms capture, in real time, any tweets containing the keywords (including hashtags) selected by the operator.

Datasets from *yourTwapperkeeper* and *TCAT* were further supplemented with data drawn from the *TrISMA – Tracking Infrastructure for Social Media Analysis* (Bruns et al., 2015): a new facility that gathers, on an ongoing basis, all public tweets posted by the population of some 2.8 million Australian *Twitter* accounts identified by Bruns, Burgess, & Highfield, (2014). At the time of writing, *TrISMA* includes accounts identified by September 2013, and so will miss out on any activities by users who joined *Twitter* more recently, and it will only track activities by Australian accounts – but in spite of these limitations, it provides important insights especially on hashtags and keywords used predominantly by Australian participants and offers a rare opportunity to explore *Twitter* activities around themes and topics which had not been tracked in real time as they happened. These additional data sets were chosen by exploring the top hashtags in the *TrISMA* dataset in 2015, and selecting especially those hashtags that increased the diversity of the areas of communication covered in our analysis, or could provide a 2015 update on similar hashtags from earlier years.

Finally, to further increase the diversity of topics covered, contributions of summary statistics for *Twitter* keyword and hashtag datasets were solicited from researchers on the AoIR mailing list, and we acknowledge the assistance of Fabio Giglietto in setting up a Google form for the submission of these datapoints. We are very grateful to all of the researchers who have provided such datapoints, even if we have not been able to feature all of their contributions here, and we include them as contributing authors to this paper.

For each dataset selected here, summary statistics on the number of tweets, percentage of retweets and percentage of URLs are listed in Table 1 in the Appendix to this paper, as are brief details on each dataset, its timeframe of coverage, and its provenance. Ideally, of course, it would be preferable to have gathered each dataset using identical tools and methods; however, given the large number of datasets covered here, the six-year period over which they were gathered, and the considerable changes both to the *Twitter* API and to the tools available for gathering *Twitter* data during that timeframe, this is unrealistic. We have taken every care possible to ensure that such differences in the mode of gathering data do not unduly affect our analysis here, and for that reason also focus only on three relatively simple metrics for each dataset: number of tweets, percentage of retweets, and percentage of URL tweets; these should not be overly affected by the specific features of each data gathering method. The analysis which follows still remains somewhat skewed towards the specific research interests both of our team and of our external data contributors, but we hope to have significantly extended the explorative results reported by Bruns & Stieglitz (2012).

Analysis

A scatterplot showing the complete collection of 192 datasets discussed in this article is presented in fig. 2; this sizes each mark according to the total number of tweets contained in the dataset (up to a maximum of one million tweets), and colours them according to the starting year of data collection for each dataset (ranging here from 2010 to 2015). In order to test both whether the patterns for media and acute events identified by Bruns & Stieglitz (2012) still hold, and to explore whether additional hashtag types may be able to be added to the rudimentary typology emerging from that article, we employed an abductive analytical approach (Dixon, 2012): this involved, first, an open-minded exploration of the patterns emerging from our data; second, the formulation of working hypotheses to explain these patterns; and third, the iterative testing and revision of these hypotheses, over several rounds of analysis, by engaging yet more deeply with the datasets. This lengthy iterative process involved a close reading of the most prominent tweeting patterns in selected datasets, especially for outlier datasets which could not easily be explained through the emerging hypotheses. Available space in this article does not permit us to document the entire abductive process here, and we will focus instead on the final results of this exercise – but we pay particular attention to these outliers in the discussion that follows.

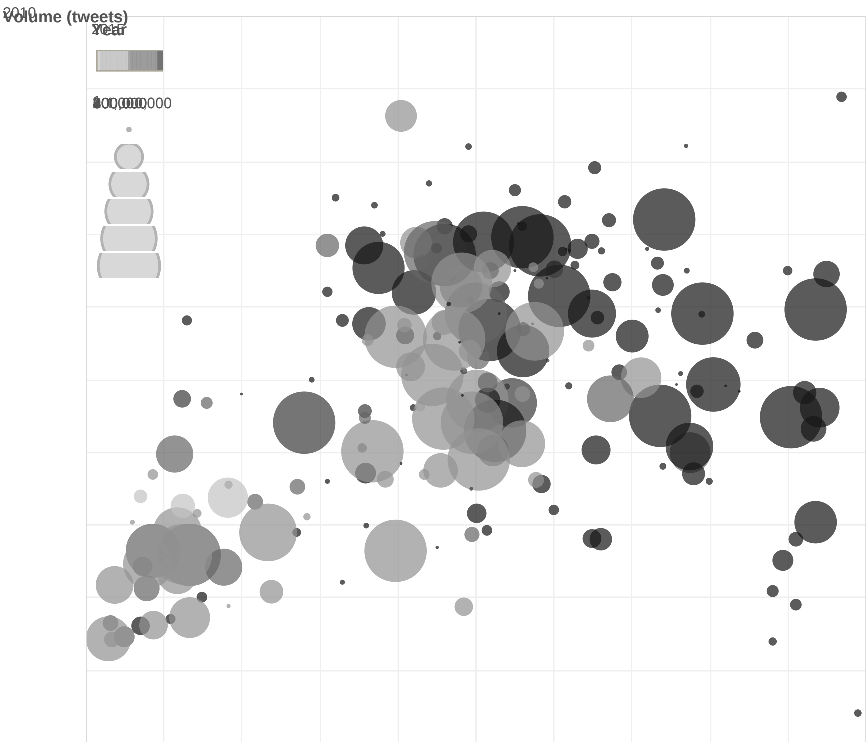

Fig. 2: Overview of all datasets contained in the analysis (datapoints sized by volume of tweets, colour by start date of data capture)

Already it appears evident from this graph that certain clustering patterns continue to persist; we will explore the principles that appear to govern such clusters in the following discussion. Further, there also appears to be a general trend towards greater percentages of both retweets and URL tweets over the years: many of the 2015 hashtag and keyword datasets (shown here in darker grey) are situated towards the top right-hand side of the graph, indicating higher percentages. In the discussion below, we explore further whether this should be seen as a genuine trend, or is a side effect of a greater number of datasets becoming available in recent years as *Twitter* research, and the data capture technologies used for such research, have matured.

We begin our analysis by testing first for the two categories – Acute Events and Media Events – already identified in Bruns & Stieglitz (2012). We proceed further by exploring a number of other categories that emerge from our extended collection of datasets.

Acute Events

First, the pattern of acute events datasets containing substantial percentages of retweets and URL tweets continues to hold firm even with the addition of further datapoints, as fig. 3 shows. A substantial majority of such datasets contained between 35% and 75% retweets, and between 40% and 80% URL tweets. This points to a stable tendency for *Twitter* participants to engage in gatewatching activities both within and beyond *Twitter*: they are posting new information, linked through embedded URLs, into the hashtag (or keyword) conversation, and retweet the material already available within *Twitter* itself.

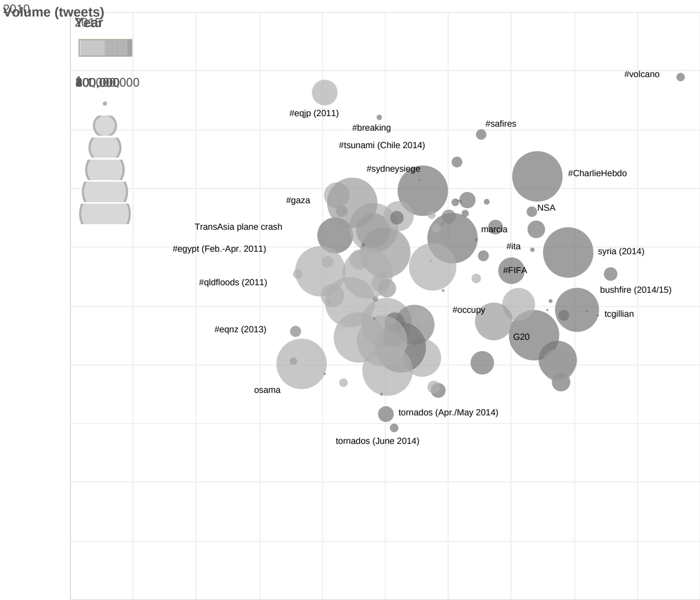

Fig. 3: Patterns for acute events datasets

Within this, recent (2014 and 2015) datasets appear especially likely to contain high percentages of URL tweets; the majority of datasets from these years feature 60% or more URL tweets in their collections. This may indicate an overall shift in the positioning of *Twitter* in media processes especially during crisis events: news organisations and journalists, as well as other sources of key information, may now be considerably more likely to post links to breaking news stories immediately to *Twitter* as a key medium for live coverage, giving the *Twitter* community a greater range of URLs to share and retweet. Emergency management organisations are also considerably better prepared for providing crucial information through *Twitter*: our 2011 #qldfloods dataset contained only 36% URL tweets, for instance, while recent Australian tropical cyclones #tcita and #tcdylan, as well as the keyword dataset 'Marcia' (for tropical cyclone Marcia) featured between 61% and 68% URL tweets, respectively.

Additionally, we also note that several of the datasets from 2014 and 2015 that contained high percentages of URL tweets were keyword- rather than hashtag-based collections (for terms such as NSA, Syria, bushfire, cyclone, and terremoto); this may also impact on the patterns we have observed. For instance, tweets sharing news articles will often include the article title – which in turn would be very likely to feature a keyword such as 'Syria', but not as likely to feature #syria as a hashtag.

Although the acute events category generally represents datasets with substantial retweet percentages, it is also notable that the *keyword* datasets within the category tend to feature a significantly lower percentage of retweets than acute events *hashtags*; the majority of acute events keywords attract fewer than 50%

retweets. This is evidence of the considerable importance of acute events hashtags in increasing the visibility of tweets covering such events, and thus of the continuing utility of hashtags as a mechanism for discovering and tracking breaking news stories: it appears from our analysis that *Twitter* users are more likely to discover and retweet hashtagged than non-hashtagged tweets. For the 2014 earthquake and tsunami in Chile, for instance, keyword datasets such as ‘Chile’ and ‘terremoto’ contained only 43% and 49% retweets, respectively, while the #tsunami dataset for the same event period contained some 74% retweets. One hashtag which receives extraordinarily many retweets even compared to the rest of the acute events category, unsurprisingly, is the generic hashtag #breaking, which contained some 82% retweets; by contrast, the hashtag dataset #bigwet contained only 35% – demonstrating that it may be used by Australian *Twitter* users as an off-hand way of referring to flood and storm events, but that operational information is found in, and retweeted from, other hashtags.

Finally, the hashtag #volcano, with some 97% URL tweets and 89% retweets, proved to be a substantial outlier in our collection of datasets. This appeared driven by the considerable aesthetic appeal of volcano photography: rather than as a means of alerting at-risk populations to new volcanic activity (which would most likely occur in location-specific hashtags), the generic, world-wide #volcano hashtag is used predominantly to share the latest and greatest images of volcanic eruptions, posted by freelance photographers and nature magazines and Websites.

Overall, then, in spite of such exceptions, the general definition of acute events datasets tends to hold, and remains stable even when the scope is widened from hashtag to keyword datasets; indeed, even two datasets tracking @mention activity around the @abcemergency and @abcfarnorth accounts that were prominent information sources during recent cyclone events in north Queensland match the overall acute events pattern. Compared to the patterns observed in Bruns & Stieglitz (2012), however, there appears to be a further increase in the percentage of URL tweets contained in recent datasets, which may point to changing *Twitter* usage patterns.

Media Events

For the most part, our analysis also points to the continued existence of a distinct category of media events hashtags, which are distinguished by very low percentages of URL tweets and relatively limited retweeting (fig. 4). It should be noted that what Bruns & Stieglitz (2012) described as “media events” really constitute broadcast events, from major TV shows and series to one-off broadcasts including events as diverse as Eurovision and election-night coverage. Here, audiences tend to use *Twitter* as a second-screen channel, enabling them to comment on and respond to what they see on television – and TV producers are increasingly closing this feedback loop by including selected tweets as on-screen inserts during live broadcasts.

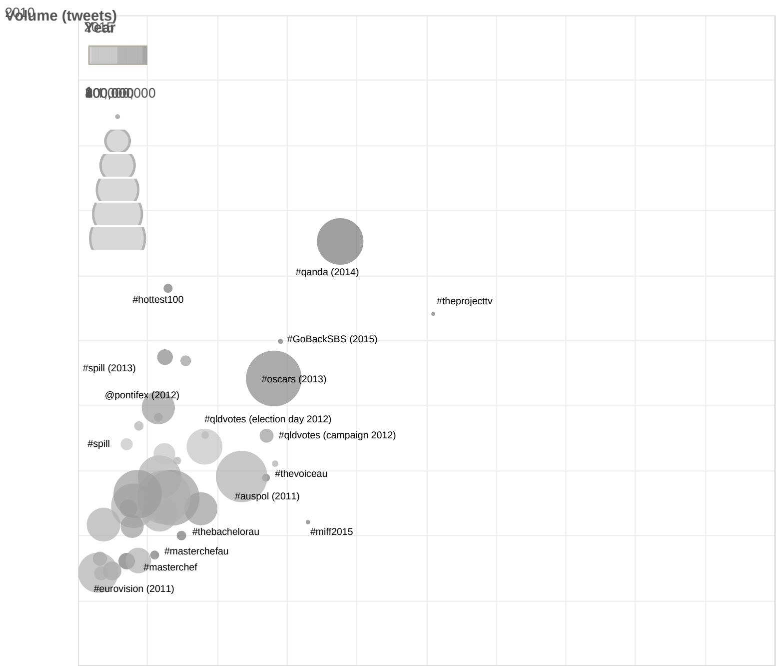

Fig. 4: Patterns for media events datasets

Many of the datasets included in this paper follow the established patterns of containing few URL tweets and relatively few retweets. Even 2015 datasets such as #thebachelorau, #thevoiceau, or #masterchefau remain below 30% on both measures, which indicates that at least at the level of broadcast television in Australia, there does not seem to be any substantial change in the way that most audiences engage with the shows they watch. Some domestic television shows do depart from the general pattern, however: first, at 47% and 50%, respectively, the hashtag for SBS's reality TV-style series retracing the experiences of asylum seekers, *Go Back to Where You Came From* (cf. Sauter & Bruns, 2014) features a considerably greater percentage of retweets during the 2012 and 2015 seasons than in 2011 (38%), and more than other broadcasts included here; the 2015 season also attracted a greater percentage of URL tweets. This may indicate a greater level of audience engagement with this heatedly debated public issue than with generic entertainment programming. Similarly, comedic Australian current affairs show *The Project*'s #theprojecttv hashtag departs entirely from the media events pattern by attracting some 51% of URL tweets and 54% retweets. This places it squarely in the acute events category, and may be explained by the fact that the show addresses current events and is therefore more closely aligned with breaking news stories than everyday entertainment television content. This repositioning is most pronounced for the weekly Australian political talkshow *Q&A*: in 2011, #qanda features only 4% URL tweets and 22% retweets, while by 2014, it attracts some 38% URL tweets and 65% retweets.

Further, there may be a trend for the broadcasts of recent entertainment industry events, such as the Academy Awards and the Video Music Awards, to similarly shift towards the acute event category. While the #oscars hashtag in 2011 sat squarely within the media events category (with 8% URL tweets and 25% retweets), the 2013 #oscars hashtag attracted 28% URL tweets and 44% retweets. Unfortunately we do not have access to comparable datasets for 2014 and 2015, but the 2014 event in particular would be likely to have attracted even more URL tweets and retweets, as it featured the famous 'group selfie' of movie stars that quickly became one of the most retweeted images ever posted to *Twitter* (Smith, 2014). As *Twitter* has by now become a core tool for promoting such events, we would expect to see many more major awards shows and similar mass broadcasts to appear within the acute events space, rather than remaining amongst more generic broadcast television hashtags. By contrast, notably, SBS's domestic hashtag for its delayed Australian telecast of the Eurovision Song Contest, #sbseurovision (cf. Highfield et al., 2013) is not affected by such developments, as it remains deliberately detached from the mainstream #eurovision and #esc hashtags used for the live broadcast across Europe and thus does not benefit from the promotion efforts of the European host broadcasters – the metrics for #sbseurovision hardly change between 2012 and 2015.

Several other datasets also share similarities with the media event category. @mention activities around then-Pope Benedict's @pontifex account, at the time of its creation in late 2012, resemble those for other media events, with a comparatively high percentage of 40% retweets. At 12%, the low percentage of URLs being shared may be explained by the fact that, as a breaking news event native to *Twitter*, there was no need to share anything other than the @pontifex handle itself to alert users to this development, however; were such @mentions of the Pope's account considered as equivalent to URL tweets, the dataset may instead fall into the acute events category. The 2015 #hottest100 hashtag (for the annual countdown of the previous year's best songs, as voted by listeners of Australian youth radio station Triple J) similarly falls into the media events category at least by virtue of the low percentage of only 13% URL tweets included in the dataset, while the 58% retweets position it as an outlier; this substantial volume of retweets may be driven in part by Triple J's own livetweeting of the Hottest 100 countdown. The 2015 Melbourne International Film Festival's #miff2015 hashtag, finally, contains a limited volume of retweets but – at 33% – a higher percentage of URL tweets, which may point to the promotional uses of the hashtag by festival organisers. Overall, therefore, the media events category remains most consistent for quotidian second-screen engagement with televised content, while events involving other media forms (radio, cinema) appear to follow divergent patterns, and major international broadcast events are increasingly coming to resemble acute events.

Political Events

Following the precedent set in Bruns & Stieglitz (2012), we initially categorised election night broadcasts as media events; metrics for the 2010 #ausvotes hashtag, and for Australian state elections such as #nswvotes in 2011, and #qldvotes in 2012, on their respective election days, certainly fit the media events pattern. More recent election datasets no longer exhibit the same activity patterns, however: #qldvotes on election day 2015 contained 33% URL tweets and 58% retweets, while #nswvotes on election day 2015 contained 45% URL tweets and 56% retweets. This closely resembles the metrics for acute events, and we interpret it as pointing to a significant shift in the media mix for political engagement: while even in 2012, the election night TV broadcast may still have been the core shared media text for tracking the election outcome as it emerged, the substantial number of news organisations, journalists, politicians, and politically active users now congregating on *Twitter* and in similar social media spaces has positioned these social media platforms as much more important channels in their own right – with key hashtags especially central. Even election broadcasts in recent years now frequently report on rumours and political statements circulating first on *Twitter*, in fact.

At the same time, the continuing influence of television is felt in the fact that URL tweet percentages during the widely televised election days remain lower than they are across the remainder of hashtag datasets such as #qldvotes and #nswvotes – with many viewers still gathered around the shared media text of the TV broadcast, there is comparatively less need to share additional URLs on *Twitter*. It should be noted in this context that the vast majority of our political datasets depict activity patterns from Australia only, however: developments in other political and media systems may have proceeded considerably differently.

In Australia, at any rate, we have been able to identify a range of recent datasets – beyond election nights themselves – which exhibit similar activity patterns, and we classify them here as covering political events (fig. 5). This constitutes almost certainly a sub-category of acute events; for the purposes of our analysis, however, it is more sensible to treat them separately in our discussion. Other events in this category, then, include political crises and scandals such as the controversies over the reintroduction of Australian knighthoods (#knightsanddames, 'Prince Philip') or over plans for random visa checks on the streets of Melbourne (#borderforce, #borderfarce), as well as several recent attempts to replace then-Australian Prime Minister Tony Abbott. Such leadership challenges to a sitting Prime Minister by their own party room – usually driven by backbench anxieties over flagging opinion poll ratings that may foreshadow a loss of power at the next election – are known in Australian political parlance as 'spills', and thus hashtagged #spill or (referencing the political party in question) #libspill. Notably, these recent events differ considerably from *Twitter* activity around the Labor leadership spill events in 2010 and 2013, which featured a very low percentage of URL tweets (7% and 12%, respectively) compared to the February 2015 attempt (49%) and the successful September 2015 spill (36%) in the Liberal Party; we regard this as a clear indication of a shift in the Australian media landscape, where URLs with information about breaking news events in politics, as well as photos and image memes relating to these issues, are now much more readily and immediately shared than even just a few years ago.

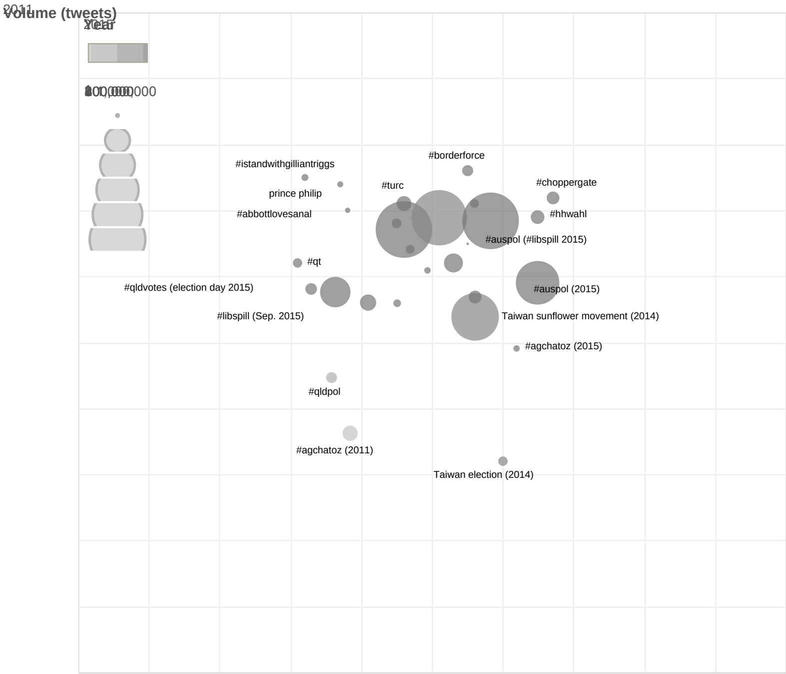

Fig. 5: patterns for political event datasets

Compared to other acute events, these political events datasets tend to feature an even more substantial percentage of retweets – commonly in the 60-70% range. This may also point to the existence of a number of well-established, long-term hashtags that act as gathering grounds for interested *Twitter* users even when no more event-specific hashtag has yet emerged, and which provide a ready stream of topical tweets to be retweeted. Activities within well-known general Australian political hashtag #auspol fall into the political events category especially around the delivery of the controversial first Abbott/Hockey budget in 2014 and the Liberal spill attempt in February 2015, for example. (Outside of such moments of heightened tension, #auspol contains more URL tweets and fewer retweets: during the period of April to June 2015, it attracted 65% URL tweets and 59% retweets.)

Another long-standing and at least semi-political hashtag, #agchatoz (for discussions of farming and of agriculture policy in Australia), operates somewhat differently and attracts considerably lower percentages of retweets, perhaps because of its more specialised focus; however, consistent with patterns observed for other hashtags in this analysis, between 2011 and 2015 it, too, has substantially increased its percentage of URL tweets (from 38% to 62%). In the absence of a sufficient number of datasets covering such cases, international comparisons remain difficult; however, the #hhwahl dataset (for the regional election in Hamburg, in February 2015) shows patterns that are broadly compatible with the Australian observations. By contrast, the 2014 Taiwan presidential election dataset exhibits different patterns, with some 60% URL tweets but only 32% retweets; this, however, may be due to the specific mode of gathering data for this study, which tracked tweets containing the (Chinese) names of candidates and parties rather than following an election-specific hashtag; indeed, it is the very absence of a hashtag (which could make tweets more visible to other users) that may explain the low number of retweets observed in this case.

Overall, then, as we have noted, the political events covered by the datasets we have included here may well be seen as constituting simply a subset of the overall acute events category. The crossover between these categories is demonstrated here especially by two international cases from 2014: the Taiwan sunflower movement and the Hong Kong umbrella movement, both of which involved major demonstrations over several weeks or months, calling for political change. Such protests are both acute and political. However, by studying this group of political datasets in its own right we have been able to uncover a number of notable developments in user practices around these events: the very fact that these events are now following acute event patterns, rather than constituting simply another form of media(ted) event, is significant in its own right.

Sports Events

As was the case with earlier political events, Bruns & Stieglitz (2012) largely included the sports events covered in that article within the media events category; with few URL tweets and few retweets (but often nonetheless a substantial volume of tweets), it was evident that the audiencing patterns around shared televisual texts that were observed at the time for TV shows and election night broadcasts also applied to these sporting events. Our analysis here paints a somewhat different picture, however: on the one hand, we are able to include a number of longer-term datasets covering entire sporting seasons, which behave rather differently; on the other, more recent televised sports events also show *Twitter* activity patterns that depart notably from earlier events (fig. 6).

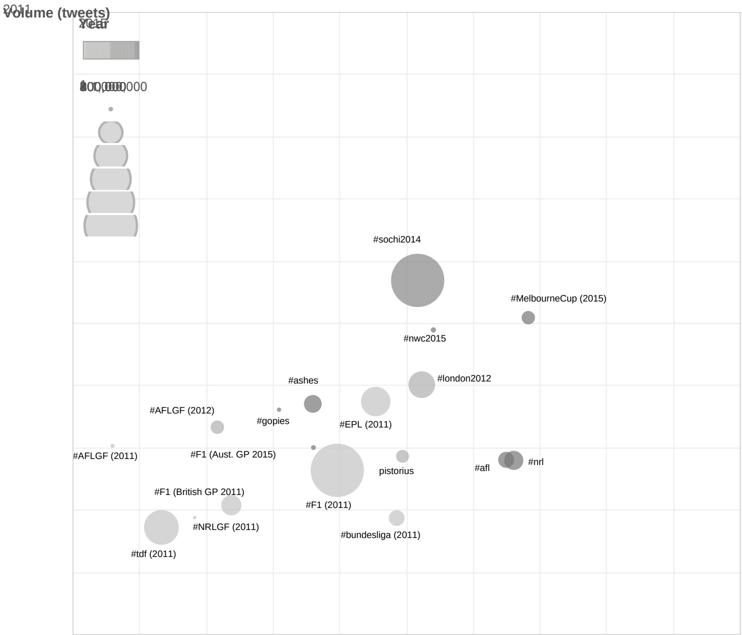

Fig. 6: patterns for sports event datasets

The 2011 events included in the earlier paper are clearly located in the bottom left-hand side of the graph; notably, there are significant shifts even between the 2011 and 2012 AFL Grand Finals already, with the latter already generating considerably more URL tweets. We see corresponding shifts between activities around the #F1 hashtag for the 2011 British Grand Prix and the 2015 Australian Grand Prix, as the percentages for both URL tweets and retweets increase. Other recent major sporting events similarly feature substantial growth in both types of tweets: hashtags such as those relating to the 2014 Sochi Winter Olympics and the 2015 Ashes, Netball World Cup, and Melbourne Cup are all positioned much closer to the acute events zone in our graph, and even the lead-up activity to a comparatively early event, the 2012 London Olympics, also falls into this group. As with the gradual shift of political events towards the acute events zone, we consider this to be evidence of shifting engagement patterns that are likely also to be driven by the changes in how sporting events themselves are now embracing social media for promotion and outreach, as well as by changing fan engagement practices.

Further, another group of datasets are distinguished by a similarly high percentage of 40-70% URL tweets, but a lower percentage of only around 30% retweets. These datasets cover the season-long discussion of specific sports from Formula 1 to the Bundesliga and English Premier League in 2011, as well as the NRL and AFL seasons in 2015 (which feature the highest percentage of URL tweets within this group). Such hashtags for ongoing coverage and fan discussion of these sports are less inherently connected to high-profile media coverage, and instead draw on a broader range of media and fan texts; that the 2015 hashtags feature greater percentages of URL tweets may indicate the greater range of such texts now available to be shared. The lower percentage of retweets across these hashtags, on the other hand, may point to the fact that a smaller number of participants is engaged and prepared to retweet materials across the entire season, compared to the larger audiences which form around specific matches, races, and other key events. We must note here that sports fans' *Twitter* practices may well vary widely across sporting codes and national boundaries; more work will need to be done, therefore, to examine whether these patterns for a small selection of sports hold across a wider and more diverse range of sports fans' social media engagement activities.

(We have also included here a dataset of *Twitter* discussion around Oscar Pistorius's killing of Reeva Steenkamp, which exhibited similar patterns; this may be due to Pistorius's status as a Paralympic athlete, and the resulting attention from a global sports audience.)

Keyword Hashtags

Our analysis in this article also covers a number of datasets for hashtags which we describe as 'keyword hashtags': these are hashtags for a range of locations (such as #sydney, #melbourne, or #brisbane), or for comparatively generic terms such as #job, #data, or #climatechange that are unlikely to serve as the focal point for topical discussion communities (fig. 7).

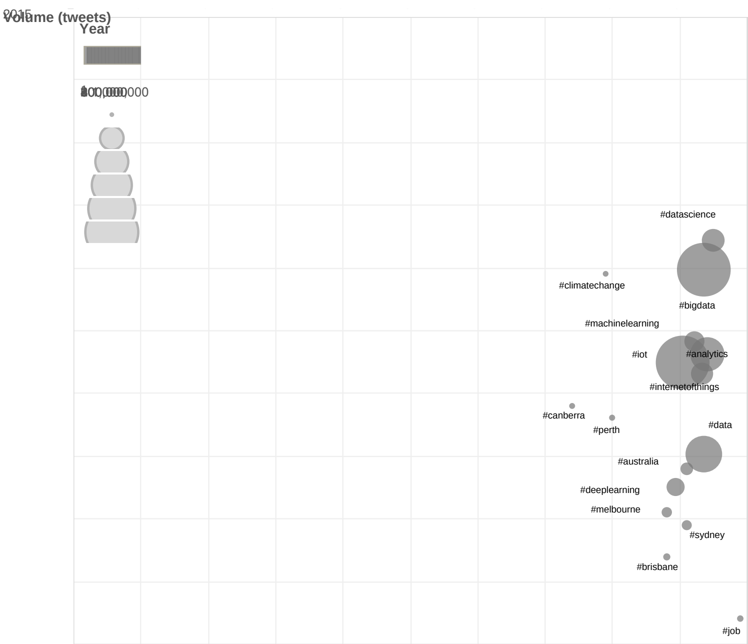

Fig. 7: patterns for keyword hashtag datasets

Such datasets are universally distinguished by very high percentages of around 90% URL tweets, indicating that they are largely used for sharing links relevant to these topics – and in fact that hashtags in these tweets may be used more as a form of emphasis ("New #job in #datascience available in #brisbane…") than with the intent to attract an *ad hoc* issue public around these topics. Indeed, the different levels of retweeting which we observe for these hashtags may serve as an indicator of the extent to which *Twitter* users are monitoring these hashtags for information: it is notable that the more topically specific hashtags (such as #climatechange or #bigdata) attract considerably greater percentages of retweets (well above 40%), while the more generic terms, such as #job or #brisbane, remain at a much lower level of retweeting.

Meme Hashtags

Finally, our analysis also observes the patterns around a group of hashtags which we classify as meme hashtags: these are terms which often emerge rapidly in response to a specific issue or topic, often expressing a particular sentiment in response to current domestic or international events. As Fig. 8 shows, such hashtags do not exhibit any unified patterns of user engagement; rather, we see them as often inheriting some of the attributes of the types of issues they respond to.

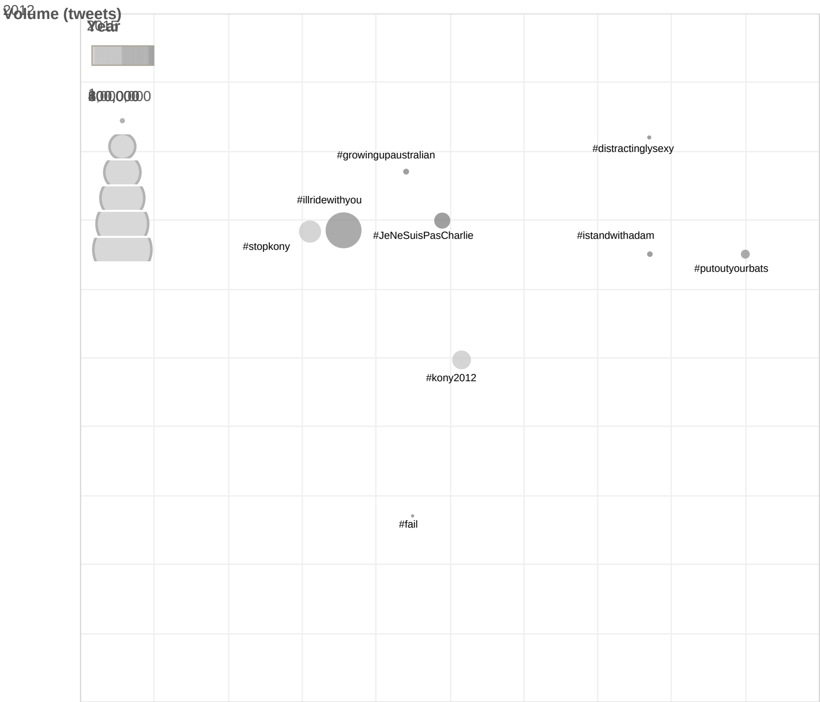

Fig. 8: patterns for meme hashtag datasets

So, for instance, hashtags such as #illridewithyou (expressing solidarity with Muslim Australians in the wake of the 2014 Sydney Siege) or #JeNeSuisPasCharlie (a counter-response to the #JeSuisCharlieHebdo hashtag after the 2015 Charlie Hebdo attacks) fall broadly into the acute events category, much like the events they respond to. The earlier #kony2012 and #stopkony hashtags similarly essentially set out to create an acute event in their own right, and are place there.

By contrast, #putoutyourbats (a response to the sudden death of cricketer Phil Hughes) and #istandwithadam (responding to the racist booing of AFL star Adam Goodes) depart from this pattern by featuring a much higher percentage or URL tweets: this is related largely to the practice of sharing photos of cricket bats in tribute to Hughes, or selfies featuring messages of support for Goodes, within these hashtags. A third such hashtag, #distractinglysexy, responds to Nobel laureate Tim Hunt's comments about working with female scientists, and similarly features a substantial percentage of photo tweets of female researchers, which are in turn also retweeted substantially. The more generic #fail, on the other hand, does not receive the same volume of engagement, on either metric – but it also operates differently from the keyword hashtags as it contains a far lower percentage of URL tweets.

Conclusion

Our study of a considerably expanded collection of hashtag and keyword datasets from 2010 to 2015 has confirmed some of the patterns observed in the earlier article by Bruns & Stieglitz (2012), but also identified a number of new categories of hashtag use, as well as pointed out some significant changes in the way existing hashtag categories are being used. Central to many of the uses we have observed here is the practice of sharing URLs and resharing existing tweets through retweeting, which is especially prominent both in the context of acute crisis events, and for political and sporting events that unfold live in a manner that resembles more critical acute events. This documents *Twitter*'s by now well-established role as the leading social media platform for the live coverage of developing stories, across a wide range of news categories.

The fact that many of the types of events that the previous paper classified as media events, with comparatively low percentages of URL tweets and retweets, must now be repositioned within the acute events category and similar categories, as they now feature substantially more URL tweets and retweets, points to a significant shift in the use practices for *Twitter* over recent years; this shift is almost certainly related to the growing adoption and use of *Twitter* by relevant stakeholders in such events (including media organisations, journalists, politicians, sporting bodies, sportspeople, celebrities, and others) and to changing approaches by everyday users as they engage with such public actors, as well as to the evolving technological frameworks for using *Twitter* (which have enabled the growing use of selfies and other photos, of graphical meme, of Vines and other video content, and of other embedded content). Media coverage of the role of *Twitter* in specific events also contributes to a feedback loop that makes it more likely that the platform will be used again in similar ways in future events of a similar type.

Further, our analysis also points to the existence of a number of other patterns for hashtag use, which will require further study. Long-term fan engagement – for example season-long participation in sports-related hashtags – appears to proceed notably differently from the brief flurry of heightened activity that surrounds pivotal moments such as individual matches or events; this requires further study. Memes also appear to unfold in a variety of ways (which may be linked to the types of topics they address) – and given the limited number of major memes we were able to include in our analysis, additional research on these patterns is needed. Finally, what we have defined as keyword hashtags constitute a very different way of using hashtags – largely for emphasis rather than to institute an issue public –, and the uses and utility of such hashtags remain to be explored in greater detail still.

We are acutely aware that the analysis presented here is limited by the range of hashtag and keyword datasets we had access to, and that in spite of the considerable extension of existing comparative approaches which this article presents our coverage is more detailed in some areas (acute events, media, sports) than in others. This article is therefore an interim report on research in progress: we will continue to add further datasets to this comparison on an ongoing basis; eventually, we hope that this will result in a yet more comprehensive overview of hashtag types. Similarly, while we have sought to include global hashtags and hashtag phenomena from other regional Twitterspheres where possible (especially also with the help of our contributing authors elsewhere), there remains a distinctly Australian flavour to the collection of hashtags whose activity patterns we have reviewed here. This means that a further expansion of the present study with a view to taking in a more internationally diverse range of datapoints would be desirable, in order to detect any possible local and regional variations in *Twitter* use practices. On the basis of the datapoints we have presented here, no particularly strong regional variations have emerged to date; rather, what does potentially affect tweeting patterns is the size of a hashtag dataset (and thus of its contributor base) – and this size is in turn often related to the extent to which a hashtag has reached a global rather than merely local or national audience. Some of this may also be related to disruptive *Twitter* activities, of course: global, trending hashtags are more likely to attract spambots, for instance, whose high volume of tweets could affect the hashtag's overall activity metrics if they contain URLs or retweets.

Finally, we must also continue to explore opportunities for expanding our analysis beyond hashtags themselves, for two major reasons: to document how the patterns of *Twitter* activity around hashtags and keywords covering the same topic differ (and thus, to better understand the impact that the use of a topical hashtag can have); and to move *Twitter* and social media research beyond an overemphasis on hashtags as the most prominent – and most easily captured – form of public communication through social media. Hashtags may be prominent in the social media research literature, but our own research – using the *TrISMA* infrastructure – has shown that even the most consistently prominent Australian hashtag, #auspol, usually accounts for only just under one per cent of the average of more than 900,000 tweets posted each day by the 2.8 million Australian *Twitter* accounts whose public posts we currently track. Hashtags remain an important feature, but everyday *Twitter* activities are distributed across a much wider range of practices – hashtagged and non-hashtagged – than the available literature covers.

Acknowledgments

Research for this article was supported by the Australian Research Council through ARC Future Fellowship grant *Understanding Intermedia Information Flows in the Australian Online Public Sphere*, and by infrastructure provided through the Australian Research Council LIEF project *Tracking Infrastructure for Social Media Analysis*. A number of additional hashtag datapoints for our analysis were provided by members of the Association of Internet Researchers community in response to an open call for crowdsourced contributions. We thank our colleagues for their contributions, and acknowledge them here as contributing authors.

References

Borra, E., & B. Rieder. (2014). Programmed method: Developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*, *66*(3), 262-278. http://doi.org/10.1108/AJIM-09-2013-0094

Bruns, A. (2005). *Gatewatching: Collaborative Online News Production*. New York: Peter Lang.

Bruns, A., & J. Burgess. (2015). Twitter hashtags from ad hoc to calculated publics. In N. Rambukkana (ed.), *Hashtag Publics: The Power and Politics of Discursive Networks* (pp. 13-28). New York: Peter Lang.

Bruns, A., J. Burgess, & T. Highfield. (2014). A "big data" approach to mapping the Australian Twittersphere. In P.L. Arthur & K. Bode (eds.), *Advancing Digital Humanities: Research, Methods, Theories* (pp. 113-129). Houndmills: Palgrave Macmillan.

Bruns, A., & H. Moe. (2014). Structural layers of communication on Twitter. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (eds.), Twitter and Society (pp. 15-28). New York: Peter Lang.

Bruns, A., & S. Stieglitz. (2012). Quantitative approaches to comparing communication patterns on Twitter. *Journal of Technology in Human Services*, *30*(3-4), 160-185. http://doi.org/10.1080/15228835.2012.744249

Bruns, A., J. Burgess, J. Banks, D. Tjondronegoro, A. Dreiling, J. Hartley, T. Leaver, A. Aly, T. Highfield, R. Wilken, E. Rennie, D. Lusher, M. Allen, D. Marshall, K. Demetrious, & T. Sadkowsky. (2015). *TrISMA: Tracking Infrastructure for Social Media Analysis*. http://www.trisma.org/.

Burgess, J., & A. Bruns. (2015). Easy data, hard data: The politics and pragmatics of Twitter research after the computational turn. In G. Langlois, J. Redden, and G. Elmer (eds.), *Compromised Data: From Social Media to Big Data* (pp. 93-111). New York: Bloomsbury Academic.

Dixon, D. (2012). Analysis tool or research methodology: Is there an epistemology for patterns? In D.M. Berry (ed.), *Understanding Digital Humanities* (pp. 191-209). Basingstoke: Palgrave Macmillan.

Fiske, J. (1992). Audiencing: A cultural studies approach to watching television. *Poetics*, *21*(4), 345-359.

Highfield, T., S. Harrington, & A. Bruns. (2013). Twitter as a technology for audiencing and fandom: The #Eurovision phenomenon. *Information, Communication & Society*, *16*(3), 315-39. doi:10.1080/1369118X.2012.756053

Rambukkana, N. (ed.). (2015). *Hashtag Publics: The Power and Politics of Discursive Networks*. New York: Peter Lang.

Sauter, T., & A. Bruns. (2014). Tweeting the TV event, creating "public sphericules": Ad hoc engagement with SBS's *Go Back to Where You Came From* – season two. *Media International Australia* 152, 5-15.

Smith, Caspar L. (2014). Ellen DeGeneres' Oscars selfie beats Obama retweet record on Twitter. *The Guardian*, 3 Mar. 2014. http://www.theguardian.com/film/2014/mar/03/ellen-degeneres-selfie-retweet-obama

YourTwapperkeeper. (2012). https://github.com/jobrieniii/ yourTwapperKeeper

Appendix

Table 1 provides a full overview of all datasets used in our analysis:

Dataset	Tweets	URLs	Retweets	Topic	Start Date	End Date	Source
Acute Events							
#0zapftis	26,158	58%	63%	Spy scandal, Germany	11.10.2011	31.10.2011	yTK
#bigwet	3,717	49%	35%	Natural disaster	16.02.2013	10.03.2013	yTK
#BlackLivesMatter	104,648	63%	68%	Political protests	3.05.2015	10.05.2015	crowdsourced
#blatter	72,528	52%	65%	FIFA scandal	1.06.2015	8.06.2015	yTK
#breaking	11,158	49%	82%	General breaking news	1.08.2015	31.08.2015	TrISMA
#CharlieHebdo	1,318,157	74%	72%	Terrorist attack	1.01.2015	15.02.2015	yTK
#chch	24,400	58%	65%	Christchurch (NZ) earthquake	22.02.2011	28.02.2011	yTK
#cooktown	1,553	55%	71%	Tropical cyclone	11.04.2014	21.04.2014	yTK
#cyclonedylan	1,025	48%	48%	Tropical cyclone	27.01.2014	4.02.2014	yTK
#earthquake (2011)	358,737	52%	65%	Sendai (Japan) earthquake	11.03.2011	24.03.2011	yTK
#egypt (Feb.-Apr. 2011)	1,242,731	40%	56%	Political unrest	28.02.2011	1.02.2011	yTK
#egypt (Feb.-Nov. 2011)	6,277,782	50%	39%	Political unrest	26.02.2011	26.11.2011	yTK
#eqjp (2011)	259,952	40%	86%	Sendai (Japan) earthquake	11.03.2011	24.03.2011	yTK
#eqnz (2010)	27,211	43%	46%	Christchurch (NZ) earthquake	17.09.2010	4.09.2010	yTK
#eqnz (2011)	156,940	46%	58%	Christchurch (NZ) earthquake	22.02.2011	7.03.2011	yTK
#eqnz (2012)	22,621	35%	41%	Christchurch (NZ) earthquake	1.12.2011	30.01.2012	yTK
#eqnz (2013)	48,506	36%	46%	Christchurch (NZ) earthquake	1.07.2013	30.09.2013	yTK
#eqnz (Jan. 2014)	5,324	47%	60%	New Zealand earthquake	19.01.2014	26.01.2014	yTK
#eqnz (June 2011)	29,485	43%	37%	Christchurch (NZ) earthquake	26.06.2011	13.06.2011	yTK
#FIFA	277,011	70%	56%	FIFA scandal	1.06.2015	8.06.2015	yTK
#gaza	1,931,111	45%	67%	Gaza conflict	16.11.2012	29.11.2012	yTK
#GazaUnderAttack	462,188	48%	63%	Gaza conflict	15.11.2012	22.11.2012	yTK
#idf	53,813	43%	66%	Gaza conflict	16.11.2012	28.11.2012	yTK
#irene	64,315	58%	36%	Hurricane Irene	27.08.2011	17.09.2011	yTK
#ita	7,967	73%	60%	Tropical Cyclone	9.04.2014	20.04.2014	yTK
#libya (2011)	3,825,272	47%	55%	Political unrest	26.02.2011	26.11.2011	yTK
#libya (2012 - Benghazi)	131,188	50%	53%	Political unrest	1.09.2012	20.09.2012	yTK
#londonriots	212,213	42%	52%	London / UK riots	8.08.2011	21.08.2011	yTK
#norway	63,244	56%	48%	Terrorist attack	24.07.2011	9.08.2011	yTK
#nswfires	78,581	60%	65%	Natural disaster	15.10.2013	1.11.2013	yTK
#nswflood	1,006	41%	51%	Natural disaster	25.01.2013	5.02.2013	yTK
#occupy	560,560	67%	47%	Political protests	19.12.2011	19.04.2012	yTK
#occupywallstreet	885,174	58%	57%	Political protests	27.09.2011	27.11.2011	yTK
#qldfloods (2011)	35,658	36%	55%	Natural disaster	10.01.2011	16.01.2011	yTK
#riotcleanup	53,381	41%	58%	London / UK riots	8.08.2011	21.08.2011	yTK
#safires	43,683	65%	79%	Natural disaster	1.01.2015	15.01.2015	yTK

#sandy	3,458,507	50%	59%	Hurricane Sandy	27.10.2012	5.11.2012	yTK
#straddie	795	41%	38%	Natural disaster	3.01.2014	17.01.2014	yTK
#sydneyfires	3,355	59%	53%	Natural disaster	17.10.2013	31.10.2013	yTK
#sydneysiege	1,025,015	56%	69%	Terrorist attack	14.12.2014	16.12.2014	TCAT
#sydneystorm	43,160	73%	66%	Natural disaster	20.04.2015	30.04.2015	yTK
#syria (2011)	5,230,025	49%	44%	Political unrest	26.03.2011	26.11.2011	yTK
#syria (2012)	425,697	78%	40%	Political unrest	10.09.2012	25.09.2012	yTK
#tcdylan	3,694	64%	61%	Tropical cyclone	26.01.2014	7.02.2014	yTK
#tcfletcher	305	76%	49%	Tropical cyclone	2.02.2014	6.02.2014	yTK
#tchadi	167	82%	49%	Tropical cyclone	8.03.2014	11.03.2014	yTK
#tcita (2014)	23,077	61%	68%	Tropical cyclone	9.04.2014	20.04.2014	yTK
#tornadoes (1)	12,313	49%	51%	Natural disaster	10.05.2013	26.05.2013	yTK
#tornadoes (3)	13,250	66%	68%	Natural disaster	20.04.2014	5.05.2014	yTK
#tsunami (Chile 2014)	45,600	61%	74%	Chile earthquake and tsunami	1.04.2014	5.04.2014	yTK
#tsunami (Japan 2011)	948,640	48%	63%	Sendai (Japan) earthquake	11.03.2011	11.04.2011	yTK
#ukriots	126,664	49%	54%	London / UK riots	8.08.2011	21.08.2011	yTK
#vicfire (2013)	2,086	57%	58%	Natural disaster	13.01.2013	5.02.2013	yTK
#vicfires (2014)	20,223	63%	66%	Natural disaster	9.01.2014	24.02.2014	yTK
#volcano	27,991	97%	89%	Natural disaster	25.05.2014	10.06.2014	yTK
#wafire (2014)	2,328	62%	68%	Natural disaster	25.06.2014	24.07.2014	yTK
#wikileaks (2011)	422,635	71%	50%	Politics	26.02.2011	26.11.2011	yTK
#wikileaks (Assange arrest)	35,451	64%	55%	Political crisis	1.09.2011	7.09.2011	yTK
@abcemergency (TC Ita)	1,458	62%	68%	Natural disaster	10.04.2014	19.04.2014	yTK
@abcfarnorth (TC Ita)	1,510	59%	64%	Natural disaster	12.04.2014	18.04.2014	yTK
Bin Laden	3,987,919	46%	45%	Osama bin Laden killing	2.05.2011	2.06.2011	yTK
bushfire (2014)	46,071	78%	48%	Natural disaster	6.01.2014	15.02.2014	yTK
bushfire (2014/15)	72,396	86%	55%	Natural disaster	25.12.2014	30.01.2015	yTK
chile	2,437,684	53%	43%	Chile earthquake and tsunami	1.05.2014	10.06.2014	yTK
cyclone (Apr. 2014)	86,250	59%	36%	Natural disaster	5.04.2014	20.04.2014	yTK
cyclone (Feb./Mar. 2015)	581,536	77%	41%	Natural disaster	15.02.2015	1.04.2015	yTK
cyclone (Oct. 2014)	215,771	65%	40%	Natural disaster	1.10.2014	20.10.2014	yTk
drought	161,711	52%	47%	Natural disaster	10.05.2014	25.05.2014	yTK
G20	1,066,295	74%	45%	G20 summit, Brisbane	1.11.2014	25.11.2014	TCAT
grexit	1,039,679	61%	61%	Political crisis in Greece	3.07.2015	31.07.2015	yTK
idf	258,003	42%	69%	Gaza conflict	16.11.2012	28.11.2012	yTK
Malcolm Fraser	46,683	66%	58%	Malcolm Fraser death	20.03.2015	20.03.2015	TCAT
marcia	85,306	68%	63%	Tropical Cyclone	16.02.2015	24.02.2015	TCAT
meteor (2013)	629,022	55%	47%	Meteor strike in Russia	15.02.2013	17.02.2013	yTK
NSA	121,592	74%	63%	NSA scandal	5.06.2015	4.07.2015	TCAT
osama	4,167,804	37%	40%	Osama bin Laden killing	2.05.2011	2.06.2011	yTK
sandy	2,020,814	50%	47%	Hurricane Sandy	5.11.2012	29.11.2012	yTK
steve jobs	562,411	56%	41%	Steve Jobs death	7.10.2011	7.11.2011	yTK
stradbroke	6,089	76%	51%	Natural disaster	20.05.2014	19.06.2014	yTK
syria (2014)	4,116,020	79%	59%	Civil war	15.01.2014	30.07.2014	yTK
tcgillian	565	84%	48%	Tropical Cyclone	7.03.2014	26.03.2014	yTK

terremoto	762,228	80%	49%	Chile earthquake and tsunami	23.04.2015	20.05.2015	yTK
tornados (Apr./May 2014)	97,952	50%	32%	Natural disaster	25.04.2014	5.05.2014	yTK
tornados (June 2014)	29,027	51%	29%	Natural disaster	16.06.2014	29.06.2014	yTK
TransAsia plane crash (several keywords combined)	508,666	42%	62%	Plane crash	4.02.2015	16.02.2015	crowdsourced
tsunami (2011)	4,246,019	45%	51%	Sendai (Japan) earthquake	11.03.2011	11.04.2011	yTK
wildfire (2014)	134,625	78%	37%	Natural disaster	11.05.2014	2.06.2014	yTK
Media Events							
#angryboys	63,333	3%	14%	TV sitcom	12.05.2011	31.07.2011	yTK
#auspol (2011)	854,019	23%	29%	General Australian politics	8.02.2011	8.12.2011	yTK
#ausvotes (2010 campaign)	415,511	18%	34%	Australian federal election	17.07.2010	25.08.2010	yTK
#ausvotes (2010 election day)	151,855	12%	33%	Australian federal election	20.08.2010	22.08.2010	yTK
#budget (2011)	13,616	28%	31%	Australian federal budget	10.05.2011	16.05.2011	yTK
#esc (2012)	171,159	8%	21%	Eurovision Song Contest	21.05.2012	27.05.2012	yTK
#esc2012 (2012)	97,607	7%	24%	Eurovision Song Contest	10.05.2012	1.06.2012	yTK
#eurovision (2011)	520,543	3%	14%	Eurovision Song Contest	9.05.2011	15.05.2011	yTK
#eurovision (2012)	753,995	9%	26%	Eurovision Song Contest	21.05.2012	27.05.2012	yTK
#ge11 (2011)	28,468	9%	37%	UK general election	26.02.2011	26.02.2011	yTK
#GoBackSBS (2011)	25,080	12%	38%	Political TV series	21.06.2011	24.06.2011	yTK
#GoBackSBS (2012)	36,384	15%	47%	Political TV series	26.08.2012	2.09.2012	yTK
#GoBackSBS (2015)	8,436	29%	50%	Political TV series	27.07.2015	30.07.2015	TrISMA
#hottest100	26,190	13%	58%	Triple J Hottest 100	26.01.2015	26.01.2015	TrISMA
#masterchef	210,773	9%	16%	Reality TV	1.05.2011	8.08.2011	yTK
#masterchefau	25,456	11%	17%	Reality TV	1.06.2015	30.06.2015	TrISMA
#miff2015	6,528	33%	22%	Melbourne International Film Festival	1.08.2015	31.08.2015	TrISMA
#mkr	63,866	3%	16%	Reality TV	13.02.2012	31.03.2012	yTK
#mw3	413,922	12%	23%	Computer game launch	1.11.2011	30.11.2011	yTK
#nswvotes (2011)	19,781	14%	32%	New South Wales state election	25.03.2011	25.04.2011	yTK
#oscars (2011)	639,251	8%	25%	Entertainment event	27.02.2011	27.02.2011	yTK
#oscars (2013)	1,277,505	28%	44%	Academy Awards event	22.02.2013	28.02.2013	yTK
#qanda (2011)	366,209	4%	22%	Political TV talkshow	21.02.2011	21.11.2011	yTK
#qanda (2014)	699,450	38%	65%	Political TV talkshow	1.07.2014	30.07.2015	yTK
#qldvotes (campaign 2012)	62,774	27%	35%	Queensland state election	19.02.2012	26.04.2012	yTK
#qldvotes (election day 2012)	17,456	18%	36%	Queensland state election	23.03.2012	25.03.2012	yTK
#royalwedding	926,527	12%	26%	British royal wedding	29.04.2011	29.04.2011	yTK
#sbseurovision (2012)	112,745	5%	14%	Eurovision Song Contest	25.05.2012	27.05.2012	yTK
#sbseurovision (2015)	87,757	7%	16%	Eurovision Song Contest	27.07.2015	30.07.2015	TrISMA
#spill	46,937	7%	34%	Australian political leadership crisis	23.06.2010	24.06.2010	yTK
#spill (2013)	79,665	12%	47%	Australian political leadership crisis	27.06.2013	26.06.2013	yTK
#thebachelorau	29,263	15%	20%	Reality TV	1.08.2015	31.08.2015	TrISMA
#theprojecttv	4,866	51%	54%	Comedic political talkshow	1.08.2015	31.08.2015	TrISMA
#thevoiceau	19,686	27%	29%	Reality TV	1.08.2015	31.08.2015	TrISMA
@pontifex (2012)	354,952	12%	40%	Papal Twitter account launched	9.12.2012	17.12.2012	yTK
eurovision (2012)	1,248,729	13%	26%	Eurovision Song Contest	21.05.2012	27.05.2012	yTK
masterchef	609,714	12%	29%	Reality TV	1.05.2011	8.08.2011	yTK

masterchef (2012)	353,142	18%	24%	Reality TV	1.06.2012	30.07.2012	yTK
Political Events							
#abbottlovesanal	8,892	38%	70%	Political controversy	1.08.2015	31.08.2015	TrISMA
#agchatoz (2011)	72,124	38%	36%	Recurring farming discussion	10.05.2011	30.04.2012	yTK
#agchatoz (2015)	12,911	62%	49%	Recurring farming discussion	1.04.2015	30.06.2015	TrISMA
#alpconf2015	23,718	47%	64%	Australian Labor Party conference	1.07.2015	31.07.2015	TrISMA
#auspol (#libspill 2015)	1,310,608	58%	68%	Australian political leadership crisis	1.02.2015	28.02.2015	yTK
#auspol (2015)	592,174	65%	59%	General Australian politics	1.04.2015	30.06.2015	TrISMA
#auspol (budget 2014)	954,081	51%	69%	Australian federal budget	1.05.2014	30.05.2014	yTK
#borderfarce	23,508	56%	71%	Political controversy	1.08.2015	31.08.2015	TrISMA
#borderforce	37,802	55%	76%	Political controversy	1.08.2015	31.08.2015	TrISMA
#choppergate	50,603	67%	72%	Political controversy	1.07.2015	30.09.2015	TrISMA
#hhwahl	58,283	65%	69%	Hamburg regional election	19.01.2015	1.03.2015	crowdsourced
#istandwithgilliantriggs	15,077	32%	75%	Political meme	1.02.2015	28.02.2015	TrISMA
#knightsanddames	2,128	55%	65%	Political controversy	26.01.2015	8.02.2015	TrISMA
#libspill (Feb. 2015)	82,223	41%	56%	Australian political leadership crisis	1.02.2015	20.02.2015	TrISMA
#libspill (Sep. 2015)	288,468	36%	58%	Australian political leadership crisis	14.09.2015	15.09.2015	TCAT
#marriageequality	29,236	45%	68%	Political controversy	1.08.2015	31.08.2015	TrISMA
#nswvotes (campaign 2015)	51,482	56%	57%	New South Wales state election	1.03.2015	31.03.2015	TrISMA
#nswvotes (election day 2015)	18,026	45%	56%	New South Wales state election	28.03.2015	28.03.2015	TrISMA
#qldpol	36,534	36%	45%	General Queensland politics	15.01.2012	30.04.2012	yTK
#qldvotes (campaign 2015)	112,137	53%	62%	Queensland state election	1.01.2015	31.01.2015	TrISMA
#qldvotes (election day 2015)	41,936	33%	58%	Queensland state election	31.01.2015	31.01.2015	TrISMA
#qt	27,090	31%	62%	Parliamentary question time	1.08.2015	31.08.2015	TrISMA
#spill (Feb. 2015)	13,099	49%	61%	Australian political leadership crisis	1.02.2015	20.02.2015	yTK
#turc	68,573	46%	71%	Trade Union Royal Commission	1.08.2015	31.08.2015	TrISMA
prince philip	11,337	37%	74%	Political controversy	26.01.2015	8.02.2015	TrISMA
Taiwan election (2014) (names of candidates/parties**)**	27,968	60%	32%	Taiwan presidential election	7.01.2014	21.01.2014	crowdsourced
Taiwan sunflower movement (2014) (several keywords)	705,628	56%	54%	Political protest	18.03.2014	29.04.2014	crowdsourced
Hong Kong umbrella movement (2014) (several keywords)	1,603,849	46%	67%	Political protest	24.08.2014	17.12.2014	crowdsourced
Sports Events							
#afl	90,397	65%	28%	Australian Football League season	1.04.2015	30.09.2015	TrISMA
#AFLGF (2011)	6,135	6%	30%	Football final	1.10.2011	2.10.2011	yTK
#AFLGF (2012)	63,686	22%	33%	Football final	25.09.2012	15.10.2012	yTK
#ashes	110,793	36%	37%	Ashes cricket series	1.07.2015	31.08.2015	TrISMA
#bundesliga (2011)	87,474	49%	19%	German Bundesliga season	5.08.2011	8.12.2011	yTK
#EPL (2011)	306,472	45%	37%	English Premier League season	4.08.2011	19.12.2011	yTK
#F1 (2011)	1,095,271	40%	26%	Formula One season	24.06.2011	24.11.2011	yTK
#F1 (Aust. GP 2015)	8,402	36%	30%	Formula One Grand Prix	13.03.2015	15.05.2015	TrISMA
#F1 (British GP 2011)	143,697	24%	21%	Formula One Grand Prix	5.07.2011	15.07.2011	yTK

#gopies	6,512	31%	36%	Australian Football fan hashtag	1.08.2015	31.08.2015	TrISMA
#london2012	250,836	52%	40%	Lead-up activity to the London Olympics	30.06.2011	30.04.2012	yTK
#MelbourneCup (2015)	62,949	68%	51%	Major horse race	2.11.2015	4.11.2015	yTK
#nrl	129,511	66%	28%	National Rugby League season	1.03.2015	30.09.2015	TrISMA
#NRLGF (2011)	4,182	18%	19%	Football final	1.10.2011	2.10.2011	yTK
#nwc2015	9,669	54%	49%	Netball World Cup	1.08.2015	31.08.2015	TrISMA
#sochi2014	1,533,301	52%	57%	Sochi Winter Olympics	1.02.2014	1.04.2014	yTK
#tdf (2011)	427,467	13%	17%	Tour de France	4.07.2011	26.07.2011	yTK
pistorius	58,560	50%	29%	Oscar Pistorius arrest	3.09.2012	30.09.2012	yTK
Keyword Hashtags							
#analytics	403,748	94%	46%	General keyword	27.09.2015	30.06.2015	crowdsourced
#australia	54,929	91%	28%	General keyword	1.08.2015	31.08.2015	TrISMA
#bigdata	1,040,644	94%	60%	General keyword	27.09.2015	30.06.2015	crowdsourced
#brisbane	17,589	88%	14%	General keyword	1.08.2015	31.08.2015	TrISMA
#canberra	12,157	74%	38%	General keyword	1.07.2015	31.07.2015	TrISMA
#climatechange	11,304	79%	59%	General keyword	1.08.2015	31.08.2015	TrISMA
#data	463,725	94%	30%	General keyword	27.09.2015	30.06.2015	crowdsourced
#datascience	179,621	95%	64%	General keyword	27.09.2015	30.06.2015	crowdsourced
#deeplearning	113,215	90%	25%	General keyword	27.09.2015	30.06.2015	crowdsourced
#internetofthings	169,557	93%	43%	General keyword	27.09.2015	30.06.2015	crowdsourced
#iot	1,051,596	91%	45%	General keyword	27.09.2015	30.06.2015	crowdsourced
#job	14,373	99%	4%	General keyword	1.06.2015	30.06.2015	TrISMA
#machinelearning	139,813	92%	48%	General keyword	27.09.2015	30.06.2015	crowdsourced
#melbourne	37,011	88%	21%	General keyword	1.08.2015	31.08.2015	TrISMA
#perth	12,678	80%	36%	General keyword	1.07.2015	31.07.2015	TrISMA
#sydney	34,731	91%	19%	General keyword	1.08.2015	31.08.2015	TrISMA
Meme Hashtags							
#distractinglysexy	4,728	77%	82%	Viral meme	1.06.2015	30.06.2015	TrISMA
#fail	2,894	45%	27%	General keyword	1.08.2015	31.08.2015	TrISMA
#growingupaustralian	10,004	44%	77%	Viral meme	1.07.2015	31.07.2015	TrISMA
#illridewithyou	370,984	36%	68%	Sydney siege political meme	15.12.2014	15.12.2014	TCAT
#istandwithadam	8,880	77%	65%	Anti-racism Australian Football meme	1.08.2015	31.08.2015	TrISMA
#JeNeSuisPasCharlie	74,047	49%	70%	Charlie Hebdo attacks politicial meme	7.01.2015	11.01.2015	crowdsourced
#kony2012	101,425	52%	50%	Viral political campaign	8.03.2012	21.03.2012	yTK
#putoutyourbats	23,137	90%	65%	Tribute meme for cricketer Phil Hughes	27.11.2014	30.11.2014	TrISMA
#stopkony	140,958	31%	68%	Viral political campaign	8.03.2012	21.03.2012	yTK