Towards More Systematic *Twitter* Analysis: Metrics for Tweeting Activities

Axel Bruns
ARC Centre of Excellence for Creative Industries and Innovation
Queensland University of Technology
Brisbane, Australia
a.bruns@qut.edu.au / @snurb_dot_info / http://mappingonlinepublics.net/

Stefan Stieglitz
Department of Information Systems
University of Münster
Münster, Germany
stefan.stieglitz@uni-muenster.de / http://www.wi-kuk.de

Abstract
Twitter is an important and influential social media platform, but much research into its uses remains centred around isolated cases – e.g. of events in political communication, crisis communication, or popular culture, often coordinated by shared hashtags (brief keywords, prefixed with the symbol '#'). In particular, a lack of standard metrics for comparing communicative patterns across cases prevents researchers from developing a more comprehensive perspective on the diverse, sometimes crucial roles which hashtags play in *Twitter*-based communication. We address this problem by outlining a catalogue of widely applicable, standardised metrics for analysing *Twitter*-based communication, with particular focus on hashtagged exchanges. We also point to potential uses for such metrics, presenting an indication of what broader comparisons of diverse cases can achieve.

Keywords
Twitter, social media, hashtags, metrics, communicative exchanges

Towards More Systematic *Twitter* Analysis: Metrics for Tweeting Activities

Axel Bruns
ARC Centre of Excellence for Creative Industries and Innovation
Queensland University of Technology
Brisbane, Australia
a.bruns@qut.edu.au / @snurb_dot_info / http://mappingonlinepublics.net/

Stefan Stieglitz
Department of Information Systems
University of Münster
Münster, Germany
stefan.stieglitz@uni-muenster.de / *http://www.wi-kuk.de*

Introduction

Twitter is an important new channel for public communication; it has been studied especially in contexts of political communication (Bruns & Burgess, 2011; Christensen, 2011; Harlow & Harp, 2012; Larsson & Moe, 2011; Lotan, 2011; Small, 2011; Stieglitz & Dang-Xuan, 2012b), crisis communication (Bruns *et al.*, 2012; Hughes & Palen, 2009; Mendoza *et al.*, 2010; Palen *et al.*, 2010), brand communication (Krüger *et al.,* 2012; Stieglitz & Krüger, 2011), engagement around shared experiences which use *Twitter* as a backchannel – e.g. television shows (Deller, 2011) or conferences (Dröge *et al.*, 2011; Weller *et al.*, 2011) – and everyday interpersonal exchanges (boyd *et al.*, 2009; Papacharissi, 2011; Marwick & boyd, 2011).

Much scholarly work on *Twitter* to date focusses on conversations coordinated by hashtags: brief keywords or abbreviations, prefixed by the symbol '#', included in order to make tweets more easily searchable amongst all *Twitter* message traffic. Hashtags enable *Twitter* users (and non-registered visitors to the *Twitter* Website) to follow real-time feeds of all messages containing the hashtag; this is notable in cases from #eqnz (for the 2010/11 earthquakes in Christchurch, New Zealand) to #royalwedding (for the April 2011 British royal wedding), but constitutes an everyday practice which also marks much more minor events and discussions. Beyond such coordinating uses, hashtags can also be more idiosyncratic discursive markers – for example to indicate approval (#win, #ftw) or disapproval (#facepalm, #headdesk).

Hashtags also aid *Twitter* research by making communicative exchanges comparatively easy to track. Tracking general activity across a large userbase over a long time is difficult, resource-intensive, and expensive: it would require a somehow representative choice of users from the total *Twitter* userbase, whose updates should be tracked; access to their tweetstreams; and archiving of their tweets for a set timeframe. This is possible for a handful of users – it's the *Twitter* Application Programming Interface (API) provides free access to up to 5,000 user streams – but will produce results which are too sample-specific to support generalisations about overall *Twitter* activity patterns. Tracking substantially larger numbers of users requires substantially more complex infrastructure (to capture, store, and process a very high volume of tweets; cf. Bruns & Liang, 2012), and generates significant cost as such high-volume access is only available through third-party data resellers such as Gnip (cf. Melanson, 2011).[1]

Tracking hashtagged tweets, even where they constitute a substantial amount of traffic, remains a manageable and low-cost alternative. Using the open source *yourTwapperkeeper* (2011; also see Bruns & Liang, 2012) or similar tools, researchers subscribe to one or multiple hashtag feeds (or keywords, without the hash symbol) and retrieve a stream of all matching tweets through the *Twitter* API. These datasets can be analysed subsequent to the event, issue, or topic they discuss, or even while it continues, to extract information about the shape of the conversation, identify the main participating users, examine major themes, highlight key links to external resources, and establish other, more context-specific facts about the exchange.

As the body of *Twitter* scholarship grows, more comprehensive comparative approaches to hashtagged communicative exchanges become possible. Absent from current literature, however, are clear and transferable definitions of key metrics which may be applied to the analysis of communication within hashtag communities. Only once such metrics are established and widely used will it become possible to compare the observations about one hashtagged event with another, for example to highlight how uses of *Twitter* have changed from the September 2010 Christchurch earthquake to the highly destructive aftershock in February 2011; how competing television shows manage to mobilise their *Twitter* audiences; or how different brands are engaging with fans and critics through hashtagged promotions.

This article presents a catalogue of standard, replicable metrics for studying hashtagged *Twitter* conversations. We outline metrics which examine the total activity and visibility of individual participants; metrics which establish the temporal flow of conversation, and of specific forms of conversation; and metrics which combine these aspects to examine the relative contributions of specific, more or less active, user groups during each unit of time. We also outline additional metrics which could address further, more specific research questions. We point to open source tools we have developed for these purposes, to make it easier for researchers to apply such metrics to their own datasets and thereby build a larger evidence base which can be used to compare hashtagged communicative events and chart the overall, longer-term development of *Twitter* as a platform for communication.

Data Access and Data Formats

A full discussion of capturing *Twitter* data around specific hashtags is beyond our scope (but see Bruns & Liang, 2012, for greater detail); a brief introduction to available tools, and to the datasets they generate, is nonetheless useful as background. The open source tool *yourTwapperkeeper* constitutes a straightforward solution to capturing *Twitter* data for hashtags and keywords; *yTK* utilises the *Twitter* API to ingest all tweets which match the tracking criteria (containing selected hashtags or keywords). *Twitter* offers two relevant API components: the search API, used to retrieve *past* tweets matching the criteria, within the search window available for *Twitter* searches (which, depending on the frequency with which the search term occurred in recent tweets, covers a period ranging from a few days to several weeks); and the streaming API, used to subscribe to a continuing stream of *new* tweets matching the criteria, delivered via the API as soon as they become available. *yourTwapperkeeper* mainly relies on the streaming API, using the search API to fill any temporary gaps caused by any interruptions to its connection to the streaming API.

In the past, one limitation of this approach has been that the streaming API did not capture all retweets matching the tracking criteria. Specifically, it did not deliver retweets made with *Twitter*'s 'retweet button' (cf. Bruns, 2012), since these do not constitute distinct messages in *Twitter*'s internal data structures; instead, the API delivered only 'manual' retweets (following the common form "RT @*user* [original message]" or its variations). In early 2012, *Twitter* changed its API approach, and 'button' retweets are now converted on the fly to a pseudo-manual "RT @*user* [original message]" syntax before delivery through the API. An occasional

[1] Some very early *Twitter* research projects continue to enjoy free large-scale data access under a grandfather clause (see e.g. Jürgens & Jungherr, 2011) – but *Twitter* has stopped granting new fee waivers.

side effect is the appearance of tweets in the data which – by prefixing "RT @*user*" to a button-retweeted message – extend beyond *Twitter*'s 140-character message limit.

It should be noted that there is no guarantee at the API nor *yTK* end of the process that *all* tweets matching the tracking criteria will be captured by this process: temporary interruptions may cause gaps in transmission which even a secondary check through the search API cannot fill. Additionally, because the *Twitter* API constitutes the only avenue of large-scale access to *Twitter* data which is available to researchers, there are few opportunities for independent verification of data fidelity: spot checks can be performed by searching for the hashtag through the *Twitter* Website, but such checks are ineffective for large datasets. A more complex solution is running multiple instances of *yourTwapperkeeper* or similar tools on separate servers, comparing and correlating the datasets they have gathered; this is well beyond the scope of most projects, however. Rather, researchers need to accept a (small) margin of error in their data captures, and treat the resulting datasets as close approximations of the total amount of hashtag activity, but not as entirely exhaustive representations.

yourTwapperkeeper provides the tweet text, as well as additional metadata, including the sending user's name and numerical ID, the time of posting, geolocation information (where available), and various data points which relate to the sender's *Twitter* profile settings:

- **archivesource:** API source of tweet (twitter-search or twitter-stream)
- **text:** contents of tweet
- **to_user_id:** numerical ID of tweet recipient (for @replies)
- **from_user:** screen name of tweet sender
- **id:** numerical ID of tweet
- **from_user_id:** numerical ID of tweet sender
- **iso_language_code:** code (e.g. en, de, fr, ...) of sender's default language
- **source:** name or URL of tool used for tweeting (e.g. Web, Tweetdeck, ...)
- **profile_image_url:** URL of tweet sender's profile picture
- **geo_type:** form in which sender's geographical coordinates are provided
- **geo_coordinates_0:** first element of geographical coordinates
- **geo_coordinates_1:** second element of geographical coordinates
- **created_at:** tweet timestamp in human-readable format
- **time:** tweet timestamp as numerical Unix timestamp

Of these data points, *text*, *from_user*, and *time* are the key sources for the metrics we introduce here. Other data points are less useful, for various reasons: *to_user_id* is not necessarily provided, even for tweets which mention another user in an @reply; this is due to differing implementations of @replying in different *Twitter* clients. *from_user_id* is unreliable, at least for older datasets, due to the different API functions which *yourTwapperkeeper* and similar tools draw on: for technical reasons, search and streaming API functions sometimes use different numerical user IDs.[2] (The alphanumeric username *from_user* of any given user may also change over the course of a dataset, if the *Twitter* user changed their username during this time, but this is comparatively less likely.) *iso_language_code* provides an indication of the interface language chosen by the *Twitter* user in their profile settings, but does not relate to the language of individual tweets; tweets by a German user posting in English on occasion would uniformly be marked as German, for example. *geo_type* as well as *geo_coordinates_0* and *geo_coordinates_1* could provide useful information about the geographical location of users, but geolocation functionality is used at present only by a very small minority of users, even in situations – such as natural disasters – where such information may be useful; it is impossible to generalise from this small evidence base to the overall userbase for a hashtag, therefore. *source*, finally, may be useful at

[2] Our thanks to *yourTwapperkeeper* developer John O'Brien III for pointing out this problem; see http://code.google.com/p/twitter-api/issues/detail?id=214 for details.

least in specific cases, where it is sensible to distinguish groups of users posting from different devices – in crisis situations, for example, it may be relevant to distinguish between desktop- and mobile-based clients to identify users who may be tweeting from the scene; occasionally, even more specific platform choices may also be of interest.[3]

yourTwapperkeeper stores these data in a MySQL database, to be exported in various formats. A simple but flexible approach to subsequent processing, which we employ in our work, is to export in comma- or tab-separated formats, process using the programmable command-line open source software Gawk, which can be used to filter datasets and extract key metrics, and visualise metrics as data graphs in standard spreadsheet softwares such as Excel.[4] Other approaches use equivalent processing tools, for example in the programming language R (e.g. Gentry, 2012), or programmable spreadsheet tools such as Google Docs (e.g. Hawksey, 2012). We do not intend to evaluate the relative advantages of these approaches; rather, the following discussion sets out a range of standard metrics which can be implemented across these research technologies while retaining compatibility.

Research technology choices also depend on the size of the datasets gathered. Spreadsheets and MySQL databases of tweets do not scale well, causing increasing problems in handling data as they grow beyond a few tens or hundreds of thousands of tweets (and related metadata). Larger datasets (containing millions of tweets) require more advanced solutions which utilise state-of-the-art NoSQL database solutions and/or horizontally scalable cloud-based architectures (see Bruns & Liang, 2012, for possible technology models). The datasets used as examples in this paper range from over 10,000 tweets (#qldfloods, #eqnz) to several millions (#tsunami), stored in MySQL databases.

Once a dataset has been gathered, a further step is data preparation. It may be necessary to remove spam messages if those messages, unrelated to the content of interest, constitute a significant share of collected tweets. Further, researchers must decide whether retweets might also be removed. Often, this type of messages will dominate a dataset; removing retweets could be relevant where research focusses on the discursive elements within a specific communicative context. But retweeting can be considered an important instrument for information sharing, and therefore of high importance for understanding the characteristics of the overall communicative context; additionally, retweets themselves may have inherently discursive functions if they contain additional comments added by the retweeting user. Such decisions must be made (and documented) on a case-by-case basis.

Metrics for Hashtag Datasets

There are three key areas of metrics which we suggest are of general use in the study of hashtag datasets: metrics which describe the contributions made by specific users and groups of users; metrics which describe overall patterns of activity over time; and metrics that combine these aspects to examine the contributions by specific users and groups over time. Further, more specific metrics may also be established, but these soon become substantially more case-specific, and are no longer useful for a comparison of patterns across different cases. We discuss these areas in turn, and provide examples of how these metrics may be utilised for the study of individual hashtags as well as for comparative work across hashtags.[5]

---

[3] News coverage of the 2011 UK riots suggested that Blackberry messaging was used to incite the riots, for example (see Halliday, 2011); here, it may be useful to examine whether Blackberry users showed divergent trends in their participation in the #londonriots or #ukriots hashtags.

[4] This is documented in detail at http://mappingonlinepublics.net/; also see Bruns & Liang, 2012; Bruns, 2011.

[5] A Gawk script for generating these metrics from *yourTwapperkeeper* datasets is available at http://mappingonlinepublics.net/2012/01/31/more-twitter-metrics-metrify-revisited/

User Metrics

Metrics about user activity within the hashtag dataset provide an obvious starting-point for any analysis. We begin by distinguishing two broad areas: metrics about a user's *activities*, and metrics about their *visibility* within the overall community[6] of hashtag participants.

Activity metrics begin with a simple count of the **tweets sent** by each user. This provides useful information about their relative commitment to the hashtagged exchange: usually, a few users will be highly active contributors, while others are present only because they retweeted a hashtagged message on occasion, perhaps without noticing the hashtag in the tweet they passed along. Additionally, by analysing the content of each tweet through basic pattern matching, more detailed patterns of tweeting activity for each user emerge: first, we break down the total number of tweets sent into **original tweets sent** (tweets which are simply original statements, without mentioning other users) and **@mentions sent** (tweets which refer to other users by their username, prefixed by the '@' symbol). The @mentions sent may be separated into **genuine @replies sent** (tweets which contain "*@user*", but no indication that the message is a retweet of an earlier post by *user*) and **retweets sent** (tweets which are in the format "RT *@user* [original message]" or equivalent[7]). Retweets sent may also be divided – if with some margin of error, given the range of retweet formats – into **unedited retweets sent** (tweets which begin with "RT *@user*" or equivalent) and **edited retweets sent** (tweets which contain "RT *@user*" or equivalent but do not start with it). Finally, it is also useful to count for each user the number of tweets sent which contain URLs (by pattern matching for "http://", as well as for other transport protocols if required). This information about **URLs sent** provides a useful indication of the amount of external resources a user is introducing into, or retweeting from, the hashtag conversation.

Visibility metrics, by contrast, draw on pattern matching to extract any @mentions of *Twitter* users from the tweets in the dataset. This results in a count of the total **@mentions received** by the user; we describe this as an indicator of *visibility* because it acts as a measure of the extent to which other users have taken note of and gone to the trouble of replying to or mentioning the user. Simple *activity* – tweeting frequently, or @mentioning other users frequently – does not necessarily mean that those other users will take notice of or engage with an active user; being @mentioned, by contrast, implies a process of evaluation on part of those users who do the mentioning. @mentions received can again be separated into **genuine @replies received** (messages which – following the same format parsing as above – are not retweets) and **retweets received** (messages which do include "RT *@user*" or equivalent); retweets received may in turn be distinguished into **unedited retweets received** and **edited retweets received**. Further, any one tweet may include multiple @mentions of other users; thus, tweets may need to be parsed more than once in order to correctly evaluate all @mentions. Hashtagged tweets may also @mention users who did not themselves post to the hashtag; the lists of *active* and *visible* users are usually not entirely homologous, therefore.

Fig. 1 illustrates this approach with an example from the #qldfloods hashtag which covered the January 2011 floods in south-east Queensland, Australia (see Bruns *et al.*, 2012, for detail): it shows clear differences between the most active and the most visible users within the dataset, and highlights diverging tweeting patterns between these users. The majority of most active accounts, for example, are of individual users who merely retweeted important crisis information; none of these accounts themselves receive substantial @mentions or retweets from the hashtag community. By contrast, many of the most visible users are accounts of major news or emergency organisations, led by the Queensland Police Service (@QPSMedia); these were

[6] For want of a better term, we use "community" loosely here: overall, hashtag participants may act more or less strongly as a *genuine* community. Indeed, our metrics provide useful measures for the extent to which they do so.

[7] Various equivalent conventions for marking messages as retweets now exist, and should be tested for in this context: in addition to RT *@user* [message], formats such as MT *@user* [message] (for 'manual tweet'), [message] (via *@user*), and "*@user* [message]" (username and original message enclosed in quotation marks) are also common, with or without added comments from the retweeting user. At present, a pattern matching expression which searches for *RT @user*, *MT @user*, *via @user*, and *"@user* will identify the vast majority of retweets; as *Twitter* formatting conventions evolve, additional variants may need to be added, however.

comparatively less active in their own tweeting, but received a very substantial number of @mentions, mostly as retweets.

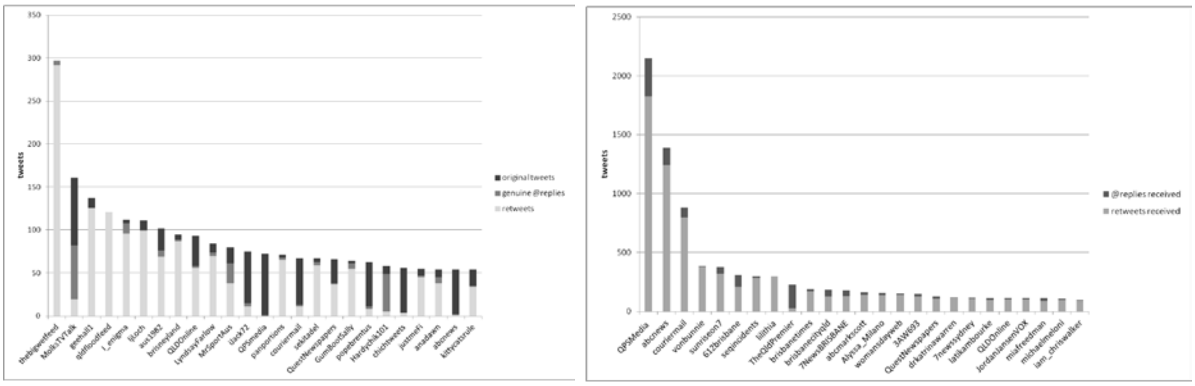

Fig. 1: Most active and most visible contributors to #qldfloods, 10-16 Jan. 2011

Useful comparisons between activity and visibility metrics are also possible: most simply, we may establish a **@mentions received:tweets sent** ratio which points to the relative impact of messages from a specific user have on the overall hashtag conversation. A user tweeting frequently but receiving few replies would have a ratio well below 1, indicating limited impact; a user whose few tweets were widely @replied to or retweeted would have a ratio well above 1, indicating significant impact.[8] In exploring such metrics, however, it is also important to note that not all @replies to hashtagged messages are themselves hashtagged; the hashtag-focussed approach we outline here will systematically underestimate follow-on communication, therefore.

This example, and similar analyses of other hashtags, usually identify a long-tail distribution pattern for both user activity and user visibility: a handful of leading users are disproportionately active or visible by comparison with the vast majority of their peers. Beyond generating *individual* user metrics, therefore, it is also useful to distinguish users into two or more *groups*, based on their relative activity or visibility within the hashtag conversation. While other models are also worth exploring, common approaches utilise a 90/10 or 90/9/1 division (Tedjamulia, 2005): the top 10% most active or most visible users are placed in a different group from the remaining 90%, potentially with a further distinction between the top 1% and the next most active 9%. Based on our metrics, this is easily possible: the total lists of active or visible users can simply be ranked according to their activity or visibility, with percentile group divisions introduced at appropriate points.

On this basis, a number of additional metrics can then be established, simply by aggregating the per-user metrics for each group. For each group, we may add up the total number of **tweets sent by the percentile**, as well as the total number of the specific categories of tweets sent (original tweets, @mentions, genuine @replies, retweets, unedited retweets, edited retweets, and tweets containing URLs); we may do likewise with per-user visibility metrics to generate the total number of **@mentions received by the percentile** (and sub-categories: genuine @replies, retweets, unedited and edited retweets received). Such aggregate figures are also available for the entire userbase, and provide a useful baseline against which the relative contribution of each percentile may be compared.

[8] Depending on context, it may be necessary to remove outliers from this calculation. This may include users who tweeted once but received substantial retweets or @replies, or users who did not participate in the hashtag, but were @mentioned frequently. Conversely, such unusual cases may themselves be worthy of closer investigation.

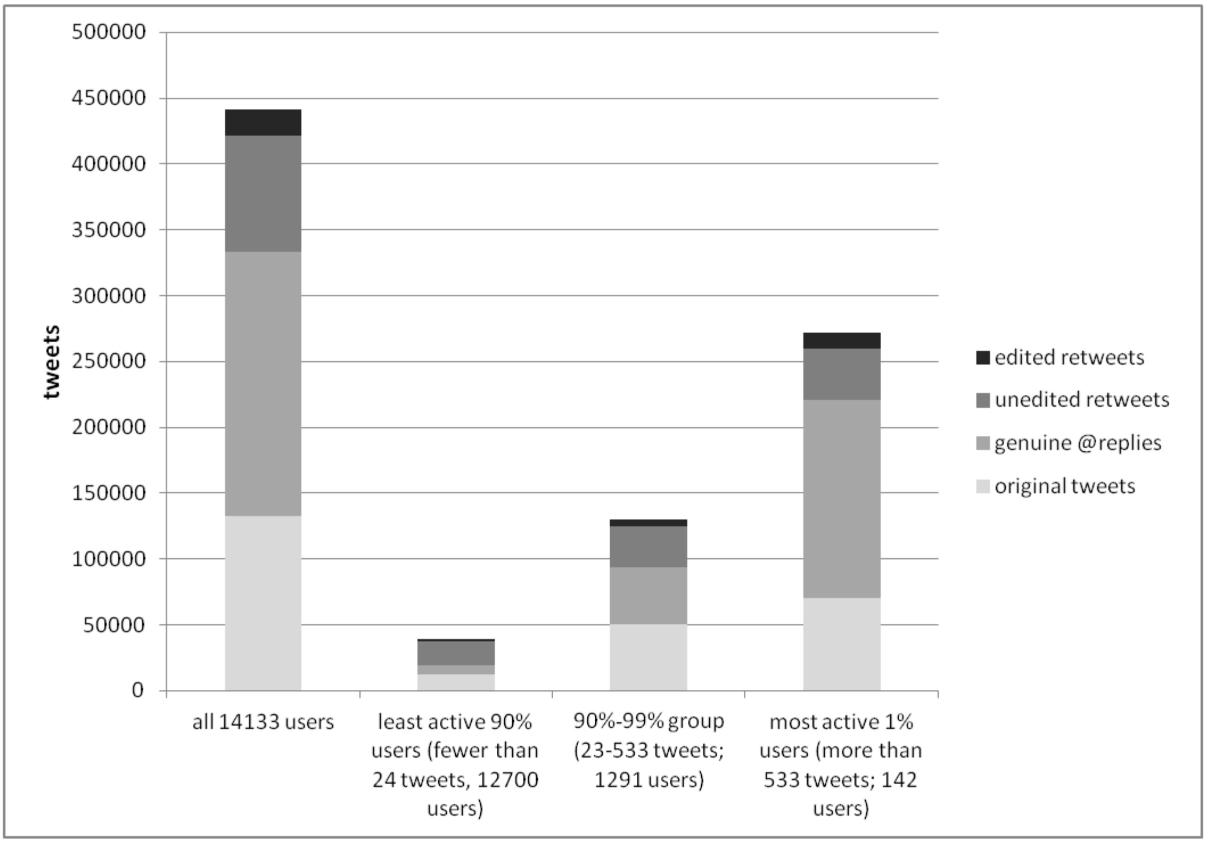

Fig. 2: Activity of different user percentiles in #auspol, Feb.-Aug. 2011

Fig. 2 demonstrates this distribution of tweeting activity for #auspol, the standard hashtag for political discussion in Australia, over the course of six months in 2011. Of the more than 440,000 hashtagged messages in this dataset, well more than half were posted by the 1% most active users, while the long tail of participants remains mostly inactive; this characterises this hashtag community as constituted of a handful of highly engaged “political junkies” (Coleman, 2003), whose messages are at times retweeted by outsiders, perhaps unaware that in doing so they contribute to the #auspol hashtag. Obvious differences in specific activity patterns are also visible: the lead users form a highly discursive community (more than 55% of their tweets are genuine @replies), while the less active percentiles mainly retweeting (more than 51% of the tweets by the least active 90% of users are retweets). The leaders send an unusually large amount of genuine @replies, given that @replies to hashtagged tweets often do not themselves contain the hashtag – the #auspol leadership group, then, appears to deliberately ‘perform’ its conversations in front of the wider userbase, by hashtagging @replies.

Temporal Metrics

While user-based metrics are valuable for analysing the overall shape of the userbase of a specific hashtag, for highlighting especially active or visible contributors, and for examining whether hashtags are used mainly for posting original thoughts, for engagement within the community, or for sharing information, a second major group of metrics emerges from a breakdown of the total dataset not by user, but by time. Which unit of time is most useful here depends on the underlying timeframe: for hashtags relating to short-term events (from live sports to television shows), minute-by-minute analysis might be appropriate; for longer-term activities (such as election campaigns or unfolding crises), day-by-day timeframes may make more sense; for long-term phenomena from brand communication to military conflicts even a month-by-month analysis may generate useful results.

Regardless of the specific unit of time, our approaches to generating temporal metrics remain the same. They begin, by counting the number of **tweets per period of time**, to establish the total volume of hashtagged communication during each unit of time. As before, these tweets can be distinguished separate categories once again: this results in metrics for the number of original tweets, @mentions, genuine @replies, edited and unedited retweets, and tweets containing URLs during each period. We can also identify the total number of **unique users active per period of time**, to establish not only the total volume of tweets, but also whether the hashtag saw a substantial influx or exodus of participants at specific points. Finally, a combination of these metrics also provides an indication of the **tweets sent by each user per period of time**: this serves as a metric of user productivity and engagement, indicating – regardless of the size of the active userbase at the time – when those users present generated the largest amount of tweets per capita.

Fig. 3 shows *Twitter* activity per minute over the course of the wedding day of Prince William and Kate Middleton, on 29 April 2011 (GMT), in the #royalwedding hashtag. It documents the correlation of global hashtag activity with the start and end of key live television broadcasts: tweet volumes increase markedly at 8:00, 9:00, and 9:15, as various networks switch to live coverage, and decline again when these shows end after the ceremony. The graph also highlights key events during the day, and changing *Twitter* activity patterns around these events: most notably, there is a sharp spike in activity towards the end of the ceremony, as the newlyweds step onto the Buckingham Palace balcony around 13:30 and kiss; that moment also sees a corresponding drop in the number of @replies, retweets, and URLs being exchanged. A compelling interpretation of this pattern is that the immediate response of *Twitter* users participating in #royalwedding is simply to send original tweets expressing their emotional reactions: a collective “ah” which does not rely on @replies or retweets. Conversely, the number of URLs being shared trends upwards over the course of the day, as increasing amounts of photos, videos, and news updates are being shared.

This strong correlation between *Twitter* activity patterns and the minute-by-minute choreography of the live event indicates that the vast majority of users tweeting to the #royalwedding hashtag were watching the event live, through television broadcasts or online videostreams. Had a substantial number of #royalwedding participants viewed the event on a time delay (by some hours, or event just by a few minutes), user responses to the key moments would have been spread across a wider span of minutes and hours, and could not have resulted in the very sharp spikes in activity which fig. 3 clearly shows. This demonstrates *Twitter*’s utility as a backchannel for live events (Harrington *et al*., 2012); however, this primacy of live viewing may not always apply: during less time-critical events, or in contexts where *Twitter* is not used mainly as a backchannel for mainstream media, the dynamics of *Twitter* activity which our methods observe will be driven by factors other than the exogenous influence of television.

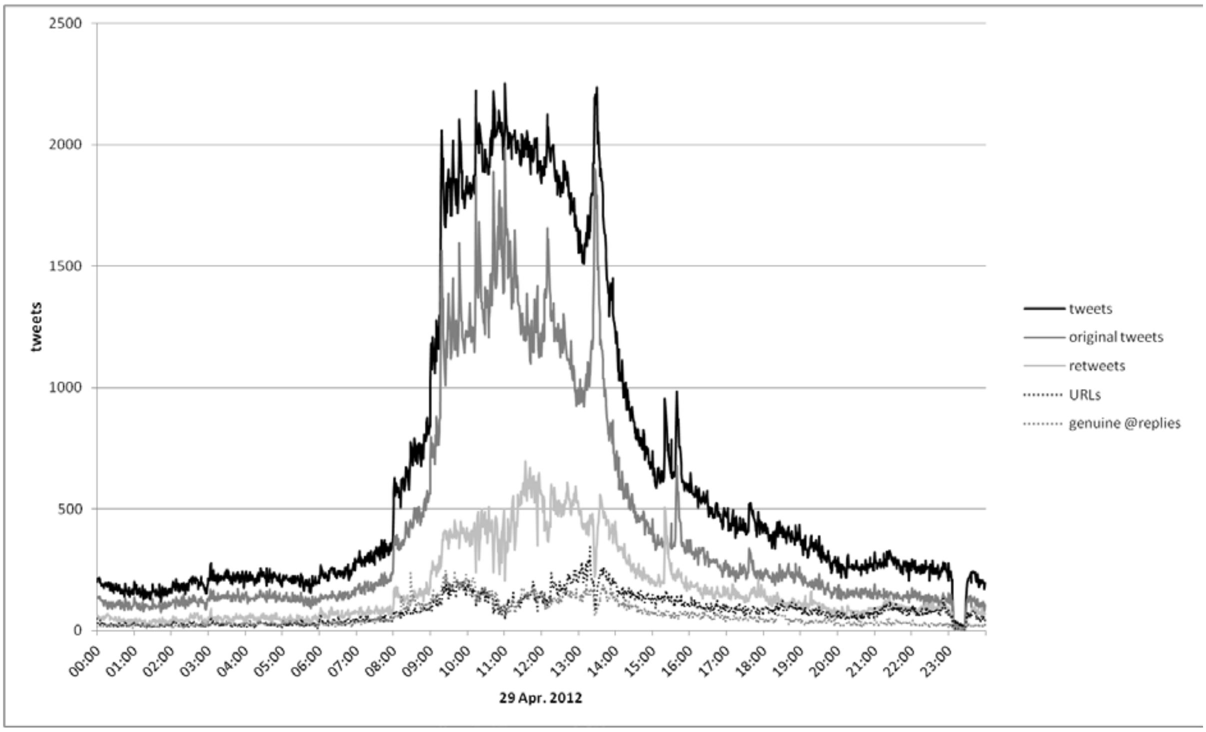

Fig. 3: minute-by-minute activity in the #royalwedding hashtag, 29 Apr. 2011 (times in GMT)

Combined Metrics

Such temporal metrics are simple but powerful, especially where they pinpoint moments of especially heightened activity to be selected for further, in-depth qualitative analysis. They may be extended by combining them with the user and user percentile metrics introduced above: this is particularly useful where obvious differences between the activities of leading user groups and the more random contributions by less active users have already been identified. During moments of heightened activity, for example, the voices of an otherwise leading group may be overwhelmed by the contributions of normally less active users from the long tail of the userbase; our distinction of these groups, above, enables an analysis of their respective activities separately from one another.

For each of the user percentiles, it is possible to determine two key metrics: the number of **currently active users from the percentile for each time period**, and the number of **tweets posted by users from the percentile for each time period**. It is important to note here that the percentiles are established on the basis of user activity patterns across the *entire* timeframe covered by the hashtag dataset, rather than being calculated afresh for each individual unit of time; the latter approach would provide a changing leaderboard of user activity for each moment (which, for short units of time, would resemble an almost random interchange of leading users), but cannot provide insight into what contribution a more long-term, comparatively stable leadership group is making at any one point. By contrast, by establishing overall percentile groups for the entire dataset first, and then testing the relative contribution of these percentile groups for any one period of time, we track at what points their voices come to dominate the discussion, or are pushed back by the greater activity of the other groups. (Additionally, of course, it is also possible again to break down the total number of tweets posted by each percentile group per time period into its constituent categories: original tweets, @mentions, genuine @replies, edited and unedited retweets, and tweets containing URLs.)

Fig. 4 demonstrates these metrics for the #eqnz dataset of *Twitter* activity following the 22 Feb. 2011 earthquake in Christchurch, New Zealand (see Bruns & Burgess, 2012, for in-depth analysis). Against the total numbers of tweets and of unique users for each day of the fortnight following the disaster, it plots the relative contribution to the total volume of tweets which was made by each of three user percentiles (the top 1% of lead users, the next 9% of active users, and the remaining 90% of least active users over the entire fortnight). It clearly shows that the leading users are responsible for a comparatively small percentage of all tweets in the days immediately following the quake: they contribute fewer than 20% of all tweets on day 1, for example. From day 5, however, they begin to dominate #eqnz, contributing around half of all tweets; combined with the second most active group, these top 10% of #eqnz contributors come to generate around 80% of all tweets. These shifts move in close conjunction with the overall number of unique users participating in the hashtag each day, which indicates that a substantial number of users who were active during the immediate aftermath of the earthquake exited the hashtag after a few days, leaving behind a smaller and more active leadership group only. (We would further hypothesise, but cannot prove from these observations alone, that the users who leave are more likely to be international users who took a news interest in the Christchurch event, while those who remain will be local users sharing information about how to cope with the disaster and its consequences.)

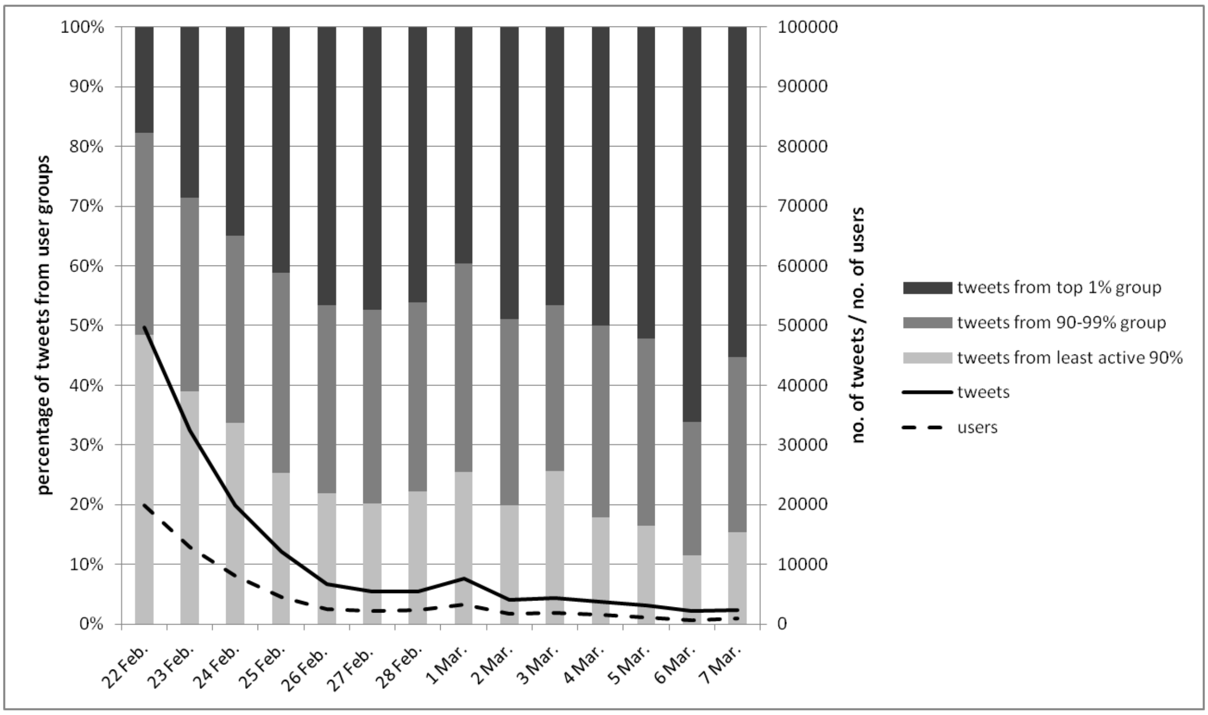

Fig. 4: Numbers of tweets and unique users in #eqnz during 22 Feb.-7 Mar. 2011, and breakdown of tweeting activity into percentiles of more and less active users

Additional Metrics

The user, temporal, and combined metrics above constitute a flexible, universally applicable toolbox for the overall analysis of hashtag activities, whose use makes it possible to more clearly and effectively compare such patterns across diverse hashtags. Additional, more context-specific metrics may also be established, and we outline a handful of such cases here as an indication of the further possibilities.

First, as noted earlier, additional data points provided by the *Twitter* API may be connected with the general metrics we have introduced. The total userbase may be divided according to the devices and *Twitter* clients from which users post, or according to the language settings of their profiles, in order to explore whether there are any obvious differences in activity across these markers of distinction. Similarly, where a sufficient amount of geolocated tweets are available, these may be treated separately from the non-geolocated component of the dataset.

Second, in addition to applying pattern-matching to every tweet in the dataset in order to detect standard syntactical conventions which identify @replies or retweets, the same techniques can also recognise relevant keywords or names; while the entire dataset is defined by the presence of one common hashtag, these may also include any secondary hashtags that are present in tweets. Similarly, in addition to merely identifying the presence of URLs, any short URLs (*t.co*, *bit.ly*, etc.) can be resolved to their eventual destinations, and classified as required. Using automated tools, such tests can be applied to each tweet in the dataset; where the original dataset is small or where a workable representative sample can be selected, manual coding approaches can also be employed. What specific keywords or other elements are of interest will depend strongly on the thematic context of the research project, and we do not explore such matters in detail here; however, these thematic patterns in the data can be usefully combined with the catalogue of general metrics which we have introduced.

Finally, where additional external data can be obtained (for example from an examination of the *Twitter* profiles of active users), these may be combined with our metrics. Which users are most active or most visible in the hashtag conversation could be compared, for instance, with information on which of the users participating in the hashtag have the largest number of followers or followees, or have been active *Twitter* users for the longest amount of time; similarly, research could investigate the comparative performance of individual or institutional accounts in a range of communicative contexts on *Twitter*.

Based on the suggested user, temporal, and combined metrics it is possible to extract specific subsets from the overall dataset for further in-depth examination; for example, researchers may wish to examine the communication practices of lead users, or to study the communicative exchanges which take place over a specific timeframe of interest (such as the peak of a crisis event). To investigate these subsets of the overall dataset, various additional research methods could be employed. Given the wide variety of possible strategies and approaches, we sketch out the most important methodological approaches only briefly:

- *Social Network Analysis* is a widely used methodological approach which helps to describe the structure of the entire communicative network, but can also be adapted to identify specific nodes (Kleinberg, 1999; Wasserman & Faust, 1994). However, social network analysis is often used only to generate static snapshots while neglecting the network's dynamics (Lin *et al*., 2008; Bruns, 2011). A small but growing number of studies explicitly analyse social media networks (Cha *et al*., 2010; Wu *et al*., 2011).
- *Sentiment Analysis* enables us to manually or automatically classify tweets with regard to their emotionality (e.g. positive or negative). For online communication, Huffaker (2010) provides evidence of sentiment diffusion, showing how messages containing positive (or negative) emotions and words are likely to receive verbal responses which also express positive (or negative) emotions. Studies have also found that emotionally-charged content is more likely to be shared by online users (e.g. Berger & Milkman 2012; Stieglitz & Dang-Xuan, 2012a). Instruments which might be used to automatically categorise tweets with respect to their sentiments include LIWC or SentiStrength.
- *Manual Content Analysis* and *Genre Analysis* are used to develop an insight into the actual content of tweets. Following such approaches, a sample of messages is manually categorised and clustered, based on a coding scheme which enables researchers to extrapolate findings to the overall dataset (Riemer, 2009).

Conclusion: Standard Metrics for Increased Comparability across Hashtags

Even without such further extensions, however, one key benefit of establishing a toolkit of standard metrics for the analysis of *Twitter* hashtag datasets are the enhanced opportunities for comparison and correlation across a range of diverse case studies. Such comparison can point to similarities and differences in patterns of *Twitter* use, and may lead to the identification of a number of common genres of hashtag-based communication on *Twitter*; as the *Twitter* platform and its uses evolve, and as the overall userbase develops, we may also be able to track these shifts over time.

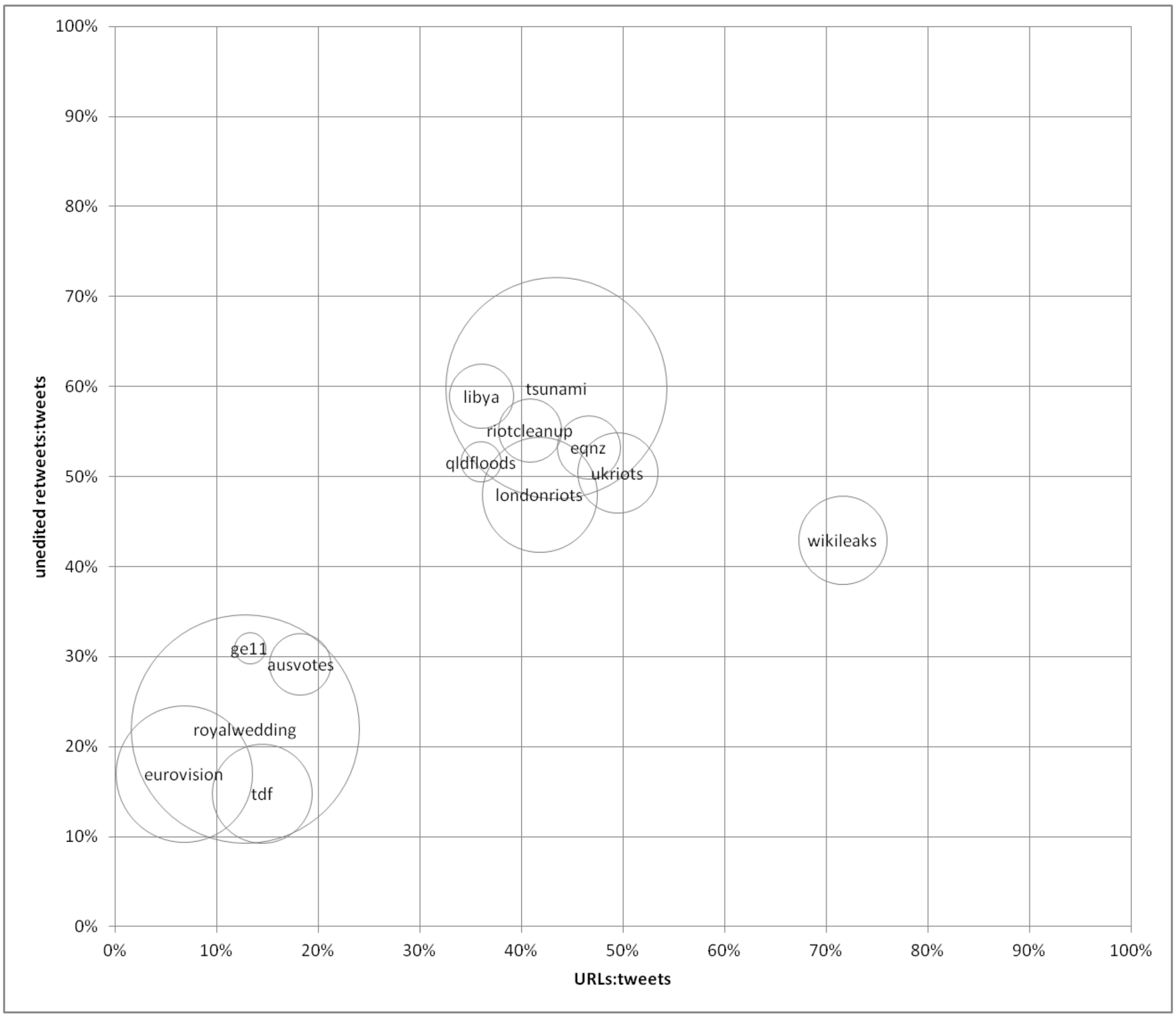

Fig. 5: User activity patterns for selected *Twitter* hashtags (size indicates total number of contributors)

A full outline of these possibilities is well beyond the scope of this article, but our early work in this area is encouraging, and the establishment of mutually compatible metrics should encourage greater scholarly exchange. As an early indicator of possible developments in this area we close by comparing, in Fig. 5, two key metrics from our toolbox across a diverse range of hashtag datasets (also cf. Bruns & Stieglitz, forthcoming): for each hashtag, it charts the percentage of URLs in all tweets against the percentage of unedited retweets in all tweets. What emerges from this graph are two obvious clusters of similarly-behaved hashtags: the first of these, in the centre, contains crisis- and emergency-related hashtags from #qldfloods through #eqnz to #tsunami (for the March 2011 tsunami in Japan), the 2011 UK riots hashtags #londonriots, #ukriots, and #riotcleanup, and #libya (for the Libyan popular revolt in 2011); this group contains substantial numbers of URLs as well as retweets, pointing to a use of *Twitter* for gatewatching (Bruns, 2005) – that is, for finding, sharing, and re-sharing relevant information on these topics (also cf. Starbird & Palen, 2010). Such similarities are notable even though the hashtag datasets included here differed markedly in timeframe (from a few days for the UK riots to several months for the Libyan conflict) and userbase (from some 15,000 users for #qldfloods to nearly half a million for #tsunami).

A second cluster of similar activity patterns contains the hashtags #ge11 (for the 2011 general election in Ireland), #ausvotes (for the 2010 Australian federal election), #tdf (for the 2011 edition of the Tour de France), #eurovision (for the 2011 Eurovision Song Contest), and #royalwedding. These hashtags contain substantially fewer URLs and retweets, and proportionally more original tweets, and are otherwise linked by the fact that they are each closely related to widely televised events (in the case of the two election-related hashtags, the bulk of user activity took place on the election nights, alongside mainstream television coverage of the tallying process). Again, similarities between these hashtags emerge even in spite of their differences in timeframe and userbase. Here, then, *Twitter* is used largely as a backchannel for national or international television coverage, resulting in a substantially reduced need to share links to additional information or to retweet the messages of other users.

We chose our metrics on the respective percentages of URLs and retweets in the datasets as the basis for fig. 5 because they relate specifically to the two key practices of user engagement with information which are postulated by the gatewatching model: finding information online (i.e. identifying and posting relevant URLs), and sharing information with other users (i.e. retweeting relevant messages). This demonstrates how the quantitative approach to analysing *Twitter* data at large scale, which the methods and metrics which we have introduced here make possible, can generate clear evidence of communicative patterns on *Twitter*. Approaches of mass data analysis can be used to test what must otherwise remain theoretical hypotheses or conceptual models, based at best on small-scale observation. Fig. 5 proves the presence of gatewatching practices for a specific subset of (crisis-related) hashtags – and in doing so demonstrates the value of combining computational, 'big data' research methods with traditional media and communication research.

That said, it is too early to state with certainty that this analysis points to two emerging genres of *Twitter* use – crisis communication and audience backchannel – but if further case studies confirm these clustering tendencies, that conclusion will become increasingly inescapable. We include the case of #wikileaks as a further outlier in this graph, to indicate that yet other patterns of activity are possible; the still more heightened use of links to further information seems appropriate for this countercultural phenomenon (cf. Lindgren & Lundström, 2011), and we intend to compare this case with #occupy and similar international political protest movements to examine whether it points to a third major genre cluster. We also note that Fig. 5 draws on only two data points from the broader catalogue of metrics we have outlined in this article; further work must also examine whether the correlations between these diverse hashtag cases change if different overall metrics are compared, or whether similarities in user activities are as strong when only specific percentiles of the total userbase are included. Another point of interest is whether comparable metrics can be found in communicative exchanges on other social media platforms such as *Facebook*, *Google+*, or *Yammer*.

Finally, research which utilises our catalogue of metrics can contribute both to scholarly discussion (by providing information about the structures and mechanisms of public communication) and to professional applications (by supporting enterprises or political actors in adopting and utilising social media more successfully). Much more work remains to be done, but our standard toolkit of metrics provides an important starting point.

References

boyd, d., Golder, S., & Lotan, G. (2010, 6 Jan.). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *Proceedings of the 43rd Annual Hawaii International Conference on System Science*. Retrieved from http://www.danah.org/papers/TweetTweetRetweet.pdf

Berger, J., & Milkman, K. (2012). What Makes Online Content Viral?. *Journal of Marketing Research 49*(2), 192-205. doi: 10.1509/jmr.10.0353

Bruns, A. (2012). Ad Hoc innovation by users of social networks: The case of Twitter. *ZSI Discussion Paper* 16 (2012). Retrieved from https://www.zsi.at/object/publication/2186

———. (2011, 17 Nov.). How long is a tweet? Mapping dynamic conversation networks on Twitter using Gawk and Gephi. *Information, Communication & Society*.

———. (2005). *Gatewatching: Collaborative Online News Production*. New York: Peter Lang.

———, & Liang, E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday*, *17*(4). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3937/3193

———, & Burgess, J. (2011). #ausvotes: How Twitter covered the 2010 Australian federal election. *Communication, Politics & Culture*, *44*(2), 37-56.

———, & Burgess, J. (2012, 18 Apr.). Local and global responses to disaster: #eqnz and the Christchurch earthquake. Paper presented at the Australia New Zealand Disaster & Emergency Management Conference, Brisbane. Retrieved from http://snurb.info/files/2012/Local%20and%20Global%20Responses%20to%20Disaster.pdf

———, Burgess, J., Crawford, K., & Shaw, F. (2012). *#qldfloods and @QPSMedia: Crisis communication on Twitter in the 2011 South East Queensland floods.* Brisbane: ARC Centre of Excellence for Creative Industries and Innovation. Retrieved from http://cci.edu.au/floodsreport.pdf

———, & Stieglitz, S. (forthcoming). Quantitative approaches to comparing communication patterns on Twitter. *Journal of Technology in Human Services*.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring user influence in *Twitter*: The million follower fallacy. *Proceedings of the International AAAI Conference on Weblogs and Social Media* (ICWSM).

Christensen, C. (2011). Twitter revolutions? Addressing social media and dissent. *Communication Review 14*(3), 155-157.

Coleman, S. (2003). A tale of two houses: The House of Commons, the *Big Brother* House and the people at home. *Parliamentary Affairs 56*(4), 733-58.

Deller, Ruth. (2011). Twittering on: Audience research and participation using Twitter. *Participations 8*(1). Retrieved from http://www.participations.org/Volume%208/Issue%201/deller.htm

Dröge, E., Maghferat, P., Puschmann, C., Verbina, J., & Weller, K. (2011). Konferenz-Tweets. Ein Ansatz zur Analyse der Twitter-Kommunikation bei wissenschaftlichen Konferenzen. Proceedings of the 12th International Symposium for Information Science. Verlag Werner Hülsbusch. Retrieved from http://ynada.com/pubs/isi2010.pdf

Gentry, Jeff. (2012). twitteR: R based Twitter client. Retrieved from http://cran.r-project.org/web/packages/twitteR/

Halliday, Josh. (2011, 8 Aug.). London riots: How BlackBerry Messenger played a key role. *The Guardian*. Retrieved from http://www.guardian.co.uk/media/2011/aug/08/london-riots-facebook-twitter-blackberry

Harlow, S., & D. Harp. (2012). Collective action on the Web. *Information, Communication & Society 15*(2), 196-216.

Harrington, S., Highfield, T., & Bruns, A. (2012). More than a backchannel: Twitter and television. In J.M. Noguera (Ed.), *Audience Interactivity and Participation*. Brussels, Belgium: COST Action Transforming Audiences, Transforming Societies.

Hawksey, M. (2012, 11 Jan.). Twitter Archiving Google Spreadsheet TAGS v3. *JISC CETIS MASHe: The Musing of* Martin Hawksey (EdTech Explorer). Retrieved from http://mashe.hawksey.info/2012/01/twitter-archive-tagsv3/

Huffaker, D. (2010). Dimensions of leadership and social influence in online communities. *Human Communication Research*, *36*(4), 593-617.

Hughes, A., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management 6*(3-4), 248-60.

Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter. *Proceedings of the ACM WebSci'11*, June 14-17 2011, Koblenz, Germany, 1-5.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal ACM*, *46*(5), 604–632.

Krüger, N., Stieglitz, S. & Potthoff, T. (2012). Brand Communication on Twitter - A Case Study on Adidas. *Proceedings of the 16th Pacific Asia Conference on Information Systems*, Paper 343.

Larsson, A.O., & Moe, H. (2011). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*. doi:10.1177/1461444811422894

Lin, Y.-R., Chi, Y., Zhu, S., Sundaram, H., & Tseng B.L. (2008). Facetnet: A framework for analyzing communities and their evolutions in dynamic networks. WWW '08: *Proceeding of the 17th international conference on World Wide Web*, 685-694.

Lindgren, S., & Lundström, R. (2011) Pirate culture and hacktivist mobilization: The cultural and social protocols of #wikiLeaks on Twitter. *New Media & Society 13*(6), 999-1018.

Lotan, G., Ananny, M., Gaffney, D., & boyd, d. (2011). The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication 5*, 1375-1405.

Marwick, A.E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114-133.

Melanson, M. (2011, 11 Feb.). Twitter kills the API whitelist: What it means for developers & innovation. *ReadWriteWeb*. Retrieved from http://www.readwriteweb.com/archives/twitter_kills_the_api_whitelist_what_it_means_for.php

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis: Can we trust what we RT? 1st Workshop on Social Media Analytics (SOMA '10), Washington, DC.

Palen, L., *et al*. (2010). Twitter-based information distribution during the 2009 Red River Valley flood threat. *Bulletin of the American Society for Information Science and Technology*, *36*(5), 13-17.

Papacharissi, Z. (2011). Conclusion: A networked self. In Z. Papacharissi (ed.), *A networked self: Identity, community and culture on social network sites*. New York: Routledge, 304-318.

Riemer, K., & Filius, S. (2009). Contextualizing media choice using genre analysis. *Business & Information Systems Engineering*, *1*(2), 164-176.

Small, T. (2011). What the hashtag? A content analysis of Canadian politics on Twitter. *Information, Communication & Society 14*(6), 872-95.

Starbird, K., & Palen, L. (2010). Pass it on?: Retweeting in mass emergency. *Proceedings of the 7th International ISCRAM Conference*. Seattle.

Stieglitz, S., & Krüger, N. (2011). Analysis of sentiments in corporate twitter communication – A case study on an issue of Toyota. *Proceedings of the 22nd Australasian Conference on Information Systems* (ACIS), Paper 103.

Stieglitz, S., & Dang-Xuan, L. (2012a). Impact and diffusion of sentiment in political communication: An empirical analysis of public political Facebook pages. *Proceedings of the 20th European Conference on Information Systems (ECIS)*, Barcelona, Spain.

Stieglitz, S., & Dang-Xuan, L. (2012b). Social Media and Political Communication - A Social Media Analytics Framework. *Social Network Analysis and Mining*, doi:10.1007/s13278-012-0079-3, Springer.

Tedjamulia, S.J.J., Dean, D.L., Olsen, D.R., & Albrecht, C.C. (2005). Motivating content contributions to online communities: Toward a more comprehensive theory. *Proceedings of the 38th Annual Hawaii International Conference on System Science*.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Weller, K., Dröge, E., & Puschmann, C. (2011). Citation analysis in Twitter: Approaches for defining and measuring information flows within tweets during scientific conferences. #MSM2011 – 1st Workshop on Making Sense of Microposts. Retrieved from http://files.ynada.com/papers/msm2011.pdf

Wu, S., Hofman, J., Mason, W., & Watts, D. (2011). Who says what to whom on Twitter. *Proceedings of the WWW'11 Conference*.

yourTwapperkeeper (2011). Retrieved from http://your.twapperkeeper.com/