International Journal of Social Research Methodology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsrm20

Reading the riots on Twitter: methodological innovation for the analysis of big data

Rob Procter [a] , Farida Vis [b] & Alex Voss [c]
[a] Manchester eResearch Centre , Manchester University , Manchester , UK
[b] Information School , Sheffield University , Sheffield , UK
[c] School of Computer Science , St Andrews University , St Andrews , UK
Published online: 03 Apr 2013.

International Journal of Social Research Methodology, 2013
Vol. 16, No. 3, 197–214, http://dx.doi.org/10.1080/13645579.2013.774172

Reading the riots on Twitter: methodological innovation for the analysis of big data

Rob Procter[a]*, Farida Vis[b] and Alex Voss[c]

[a]*Manchester eResearch Centre, Manchester University, Manchester, UK;* [b]*Information School, Sheffield University, Sheffield, UK;* [c]*School of Computer Science, St Andrews University, St Andrews, UK*

(*Received 14 June 2012; final version received 21 January 2013*)

For social scientists, the widespread adoption of social media presents both an opportunity and a challenge. Data that can shed light on people's habits, opinions and behaviour is available now on a scale never seen before, but this also means that it is impossible to analyse using conventional methodologies and tools. This article represents an experiment in applying a computationally assisted methodology to the analysis of a large corpus of tweets sent during the August 2011 riots in England.

Keywords: big data; social media; computational social science; Twitter; crisis communication; rumour; riot

Introduction

The rapid growth over the past 10 years of the Web as a publishing tool, and the recent explosion of social media such as blogs (and micro-blogs such as Twitter) and social networking sites (such as Facebook) presents both an opportunity and a challenge to social researchers.

The past 10 years has seen a large-scale investment in the development of more powerful research methods, digital infrastructure and tools with the aim of making it possible to tackle new and more complex and interdisciplinary research challenges (Atkinson et al., 2009; Halfpenny & Procter, 2010; Halfpenny, Procter, Lin, & Voss, 2009). This article represents an experiment in applying these methods and tools to the analysis of a corpus of tweets sent during the August 2011 riots in England.

We begin with a review of recent studies looking at the role of social media in crisis situations. We follow this with a description of the methodology and tools we developed to analyse the corpus of tweets and present some of our findings to illustrate their potential. We conclude with a discussion of limitations of the study and a summary of how we are planning to address them, including the infrastructure and tools we are now developing to analyse even larger social media corpora.

*Corresponding author. Email: rob.procter@warwick.ac.uk. Current address: Department of Computer Science, University of Warwick

Crisis communications and social media

The role different communication platforms play during crisis situations is now a significant area of research whose growth mirrors the expansion over the past 10 years in forms of communication technologies (Allan, 2006; Barsky, Trainor, & Torres, 2006; Bruns, 2006; Bruns, Burgess, Crawford, & Shaw, 2012; Mendoza, Poblete, & Castillo, 2010; Vis, 2009). Comparing three crisis situations in 2005, including Hurricane Katrina, Thelwall and Stuart (2007) examined how different communication technologies were used. New communication technologies were seen as especially useful for sharing information and fact-finding in the initial stages of the event, after which mainstream media outlets were able to deal more successfully with covering the aftermath.

The August 2011 riots in England began as an isolated incident in Tottenham, London, on 6th August. They quickly spread across London and to other cities in England and gave rise to levels of looting, destruction of property and violence not seen in England for more than 30 years. Eventually, after five days, the riots ceased. The causes have been attributed to many factors (Lewis et al., 2011; Morrell, Scott, McNeish, & Webster, 2011). Perhaps most surprising was the claim made by some politicians in the immediate aftermath that social media such as Twitter had played a key role.[1]

Twitter is a micro-blogging site set up in 2006 that allows users to post messages (‘tweets’) of up to 140 characters. A recent estimate puts the number of UK Twitter users at 10 million.[2] Unlike social media platforms such as Facebook, Twitter’s friendship model is directed and non-reciprocal. Users can follow whomever they wish, but those they follow do not have to follow them back. When one user follows another, the latter’s tweets will be visible in the former’s ‘tweet timeline’. It is not necessary, however, to follow another user to access tweets: by default, Twitter is an open platform, tweets are public and can be discovered through Twitter search tools. The exception is the *direct message* (DM), which is private, and can be seen only by the follower to whom it is sent. Users can reference other users through the *mention* convention, where a user name, prefixed with ‘@’, is included anywhere in a tweet. A user, thus referenced, will see the tweet in their tweet timeline. Another important Twitter convention is the retweet. By either pressing the retweet button or by copying the original tweet and putting ‘RT’ in front of it, users can forward tweets from others to their own followers. In this way, tweets can propagate through users’ follower networks. One final important Twitter convention is the *hashtag*, which is distinguished by prefixing a string of text with‘#’. Hashtags provide a way for users to label a tweet with a topic, enabling them to co-create a fluid and dynamic structure within the tweet timeline that facilitates information discovery: anyone searching for the hashtag can see what everyone else is saying about this topic.

In the following sections, we first detail the methods and tools we developed to analyse the riots corpus. We then present case studies to illustrate how using these methods and tools in combination enables us to describe in some detail how Twitter was used during the riots. We conclude by outlining some areas for further work.

Methodology

The Twitter corpus was provided to the Guardian Newspaper and its collaborators under an agreement with Twitter. The sampling frame was public tweets sent during the period 1 pm on 6 August and 8 pm on 17 August 2011. The corpus was defined by those tweets matching one or more of 54 hashtags drawn up by the team of Guardian journalists who covered the riots. The resultant corpus contains 2.6 Million tweets and 700,000 distinct user accounts. User profiles for all the accounts in the corpus were also provided by Twitter.

Analysing the corpus raised some challenging methodological issues. In particular, the volume makes it impossible to analyse using conventional media research methods and tools. To address this problem, we began by exploring the use of natural language processing (NLP) techniques. However, experiments led us to conclude that significant further development would be needed for these techniques to reach a level of performance compatible with human interpretation of tweet content (see, e.g. Black, Procter, Gray, & Ananiadou, 2012).

The methodology we subsequently developed makes use of less sophisticated computational tools to expose underlying structures in the corpus, which enabled us to identify potentially significant fragments. The content of these fragments were then analysed using established qualitative methods. The methodology is based on the classic two-step flow model of communication, highlighting how information flows from 'opinion leaders' to others (Katz & Lazarsfeld, 1955; Wu, Hofman, Mason, & Watts, 2011). To map this model onto the corpus, we built a computational tool to group source tweets and their retweets into 'information flows' (Lotan et al., 2011). Ranking information flows by size (i.e. number of retweets) provides a simple way of determining their relative significance, which would be important for deciding where to focus subsequent content analysis.

We built a database from the corpus and used its search tools to identify and extract information flows whose content matched criteria that we had established could be relevant for the topics that we wished to analyse. Examining results of different combinations of search terms (hashtags are useful but not sufficient in themselves) enabled us to identify false positives (irrelevant flows) and reduce false negatives (missing, but nevertheless relevant flows). We also ranked authors of tweets by the numbers they sent, the numbers of mentions they received and their follower count.

To understand how Twitter was being used, we developed a code frame (Krippendorff, 2004) for tweet content (see Appendix) to categorise information flows (e.g. a report of an event, a comment about a report, a request for information, etc.) and used the resultant groupings to explore how people were using Twitter in the context of a given topic. The source of an information flow was of particular interest for understanding how Twitter users reacted to rumours. To help with this analysis, we developed an actor types code frame (see Appendix) and used this to categorise accounts with more than 500 mentions.[3]

Computational tools and infrastructure

Managing 2.6 million tweets and their associated metadata is not easily done using traditional desktop computer tools. We therefore imported the data set into a relational database management system, which allowed us to query it efficiently. To provide easy access to the dataset for the researchers, we developed a virtual research environment (VRE) (Voss & Procter, 2009) using the continuous integration tool Jenkins[4] to provide a web-based user interface and the capability to run analysis jobs, as well as manage their configurations and output files. The VRE thus

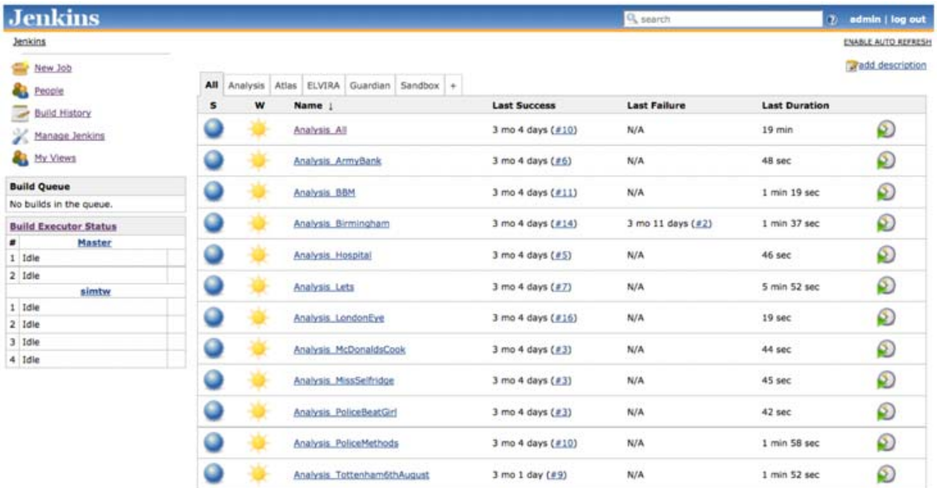

Figure 1. VRE user interface.

provided a complete provenance record of all analyses conducted. The VRE user interface is shown in Figure 1.

Each type of analysis was translated into a script that would issue the necessary SQL statements to extract data from the database and convert it into a suitable human-readable format or a format that could be further processed using simple desktop tools.

The main function of the VRE was the information flow analysis, which matches retweets to their source tweets. This involves a comparison of each retweet with all tweets previously sent by the user identified as the sender of the source tweet. The outcome of the comparison is a similarity measure. We used the Levenshtein distance (Navarro, 2001), which indicates how similar the retweet is to a candidate source tweet. We determined empirically that a distance of 30 was a good cut-off, meaning that we allowed up to 30 individual character differences between the original and its retweet. This is to allow for the changes that users retweeting a tweet sometimes make to add their own comments or to ensure that the resulting retweet, together with the attribution, still fits into the 140-character format. For those information flows, we analysed in depth, the quality of the matching was manually checked.

In contrast to the analysis scripts, which are executed by a single server within a reasonable timeframe, the retweet analysis was computationally expensive as it involves calculating the Levenshtein distance between a large number of pairs of tweets. To make this feasible, we used 16 instances (virtual servers) on the St Andrews University StACC cloud to provide the computational resources. With this level of resourcing, the task was completed within a day.

Code frames

We began tweet code type frame development by making lists, including examples of tweets for topics that were brought to the fore by the information flow analysis. The code frame developers initially worked separately and then compared their results. This led to the narrowing of the categories that all could agree on. These initial code frames then went through a further process of merging and refinement to produce a final code frame with the following top-level categories:

Media reports highlighted tweets that were either sent from a mainstream media account or a journalist working for a mainstream media organisation reporting a news story. We included tweets from other accounts pointing to mainstream media coverage and providing a link to the story. The link to the story is important as we see this as a measure of 'reliability'. We therefore did not include tweets where people might simply say 'I saw this on ITV news', but did not provide the link to the news item.

Pictures reflected that many Twitter users use services like twitpic[5] to upload and link to images in their tweets. We felt that this represented a type of information that was distinct from, for example, media reports and so deserved its own category.

Rumours coded for tweets that make claims or counterclaims about events, but did not provide any way of checking this information as well as for tweets that make reference to corroborating evidence (supporting or challenging). It allowed a clear distinction between tweets that are backed up by evidence and those that are not. In this category, we would typically include tweets where users highlight that they have 'heard' something, but they neglect to provide a link. Equally, counterclaims, where claims made by others might be disputed but, again, without providing a link were coded here as 'rumour'. Finally, we added a code for tweets that appear to be appealing for more information and a code for tweets that appear to be expressing a reaction to the rumour.

Having both a Media reports and a Rumour category enabled us to track stories that were initially circulating on Twitter as 'rumour', but then got picked up by the mainstream media. This then allowed us to say something about the cycle of such information and, because the tweets have both account information and a time stamp, we could get a much richer understanding of the rumour lifecycle.

Reactions coded users' responses to the riots in general and to specific riot-related events. It included different subcategories depending on the sub-corpus we looked at, whereas others were shared across most of the sub-corpora, such as anger at looters or requests for verification of information.

Developing a general tweet type code frame had a number of advantages. First, it could easily be adapted to other sub-corpora, but it was also more easily applicable to the whole corpus. Second, it meant that most of the material from smaller sub-corpora was also comparable, as we coded for the same three top-level codes, which are principally concerned with how the media reported the riots and the ways in which both the mainstream and non-mainstream media engaged with this on Twitter, as well as how ordinary Twitter users discussed and disseminated this news.

It was important that the code frames were rigorously tested and could be operationalised such that they were not only applicable to a specific sub-corpus of tweets but that they would be useful for analysing the full corpus. In order to do this, we coded several sections of different sub-corpora. In testing the individual code frames at various stages, it was important that high inter-coder reliability was established.

To test the code frames, we applied them to two sub-corpora. These were the Birmingham sub-corpus [all tweets matching the term 'birmingham', size: 50325] and the BBM sub-corpus [all tweets matching the term ‘bbm’, size: 13139]. Different coders read through the sub-corpora to identify information flows larger than 25 tweets and inductively code them for topics. We then used the results to develop and refine the code frames.

All coding of the corpus was done by at least two coders and where the two coders disagreed, a third coder arbitrated. The level of inter-coder agreement for the rumour corpora ranged from 89 to 96%. The full tweet type code frame is shown in Table A1 in the Appendix.

The actor type code frame was specifically concerned with identifying different types of actors in the corpus. It further built on the actor type coding developed by Lotan et al. (2011), who looked at different actor types in relation to Twitter use in the Tunisian and Egyptian revolutions. We adapted that code frame to our specific study, clarified some of the codes and significantly expanded on it, adding another eight actor types we felt it important to distinguish. The final actor type code frame is shown in Table A2 in the Appendix.

Being able to analyse not only which information flows were significant and how they circulated, but to also get a clear sense of who sent the source tweets allowed for a further level of analysis that could highlight the presence or absence of certain actors in particular sub-corpora or discussions.

How Twitter was used

People’s use of social media as a way of mitigating the impact of crises has been a particular feature of recent studies. There are many hundreds of examples of this in the riots corpus. We have chosen for more detailed examination one of the most

Table 1. Selected information flows for riot cleanup sub-corpus.

	Date, Time	Size	Actor type	Followers	Tweet
1	08/08/2011 22:41	110	Member of the public	266	Can we organize a #riotcleanup on social media? Clean up! Tool up with binbags, tea flasks and smiles. After a nap. #londonnriots
2	08/08/2011 23:54	1712	Non-media organisation employees	8145	#riotcleanup at Camden 11 am, Chalk farm 10 am, Roman Rd Hackney 9 am, Clapham 9 am, Peckham 10 am, Westbourne Grove 9 am
3	09/08/2011 00:36	3044	Riot account	68075	#riotcleanup – all info of cleanups @riotcleanup please RT and spread the word
4	09/08/2011 01:33	3521	Riot account	68075	#riotcleanup info stream and all info @Riotcleanup please spread the word and RT
5	09/08/2011 04:44	7320	Celebrity	1643996	Visit www.riotcleanup.co.uk for info on how and where to help if you can. #riotcleanup
6	09/08/2011 12:19	5294	Celebrity	1461290	Love this picture, these people are the REAL Great Britain: http://t.co/6E3VGje #riotcleanup @Lawcol888

compelling examples, the use of Twitter for mobilizing support for and organising the riot 'cleanup'. Table 1 shows selected information flows on this topic.[6] Many of these actors have thousands of followers (those highlighted in Table 1 below total over seven million) and their tweets get retweeted more than 31,000 times in total.

Although Table 1 shows evidence that some contributors were not taking the cleanup entirely seriously, this set of information flows reflect a mixture in more or less equal measure of appeals for help, information about cleanup activities (where to meet, at what time, etc.) and praise for the efforts of 'ordinary' people to respond positively to the challenges of dealing with the aftermath of the riots.

We coded the actor types for the top 200 accounts in the riot cleanup corpus. The results are shown in Figure 2. It is noticeable how the distribution of actor types in this sub-corpus is quite different from that of the riots corpus overall (Vis, 2012). In the latter, it is (in order) media organizations, journalists and riot accounts that dominate whereas, in the former, it is celebrities, UK Twitterati, non-(news) media employees and riot accounts.

Figure 3 shows the timeline of tweets with the #riotcleanup hashtag. We can see the impact of the interventions of actors with huge numbers of followers has on the volume of tweets on this topic:

(1) The first tweet in this sub-corpus makes public the idea of using social media to organise the riot cleanup.
(2) The idea is picked up by an activist arts group. They subsequently claimed the credit in their Twitter account profile for organising the riot cleanup.
(3) An account set-up specifically to coordinate the riot cleanup makes its first contributions and the volume of tweets begins to grow.
(4) The first actors with large numbers of followers get involved and the volume of tweets reaches significant levels. Even though, as Table 1 shows, celebrities

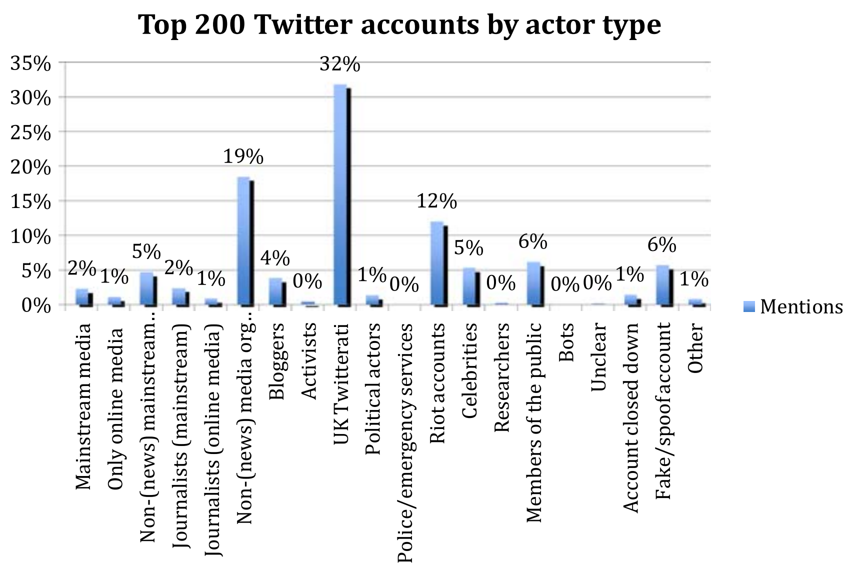

Figure 2. Top 200 Twitter accounts by actor type for riot cleanup sub-corpus.

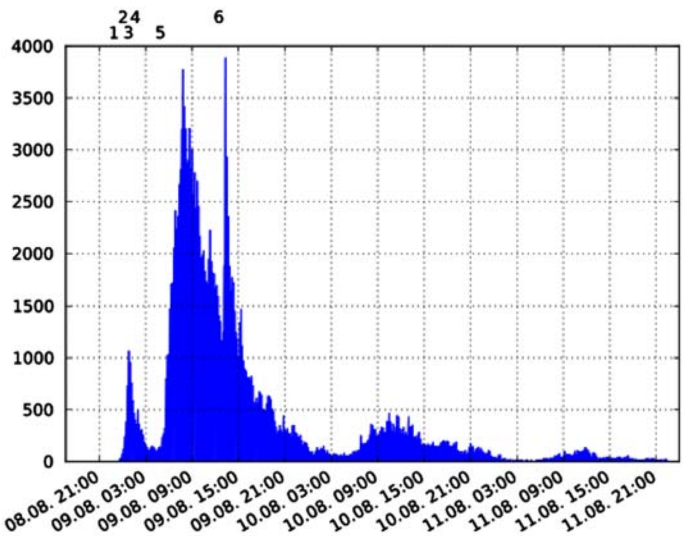

Figure 3. Timeline for riot cleanup sub-corpus. *Y* axis is the volume of tweets per 10 min interval.

do not feature very prominently in percentage terms (5%), their impact nevertheless is clearly visible.

Rumour on Twitter

Rumours are a predictable feature of any crisis and so we were interested in analysing how they emerge and circulate in a social medium such as Twitter. Other studies have used contagion or meme-based models to analyse the propagation of rumours in social media (e.g. Kaigo, 2012; Leskovec, Backstrom, & Kleinberg, 2009; Paranyushkin, 2012). We were interested, in particular, in examining the kinds of roles played by different participants in rumour discourses, including the kinds of 'conversational moves' they made as the rumour unfolded. To do this, we used the tweet type code frame to categorise the different types of conversational moves (i.e. claims, counterclaims, etc.; see Appendix, Table A2) made by actors who get involved, as well as content analysis to help determine the topic. We selected for detailed study seven rumours relating to the riots.

To illustrate our findings, we will examine in some detail the rumour that a mob of rioters was attacking Birmingham Children's Hospital (see Figure 4 and Table 2).

Figure 4 shows the timeline of the rumour, which began on 8 August and highlights some salient tweets (1–7), which we show in Table 2 and examine in more detail below.[7]

Tweets 1–3 repeat the initial rumour in various forms and generate a significant volume of retweets, especially (3), which is the source of the largest information flow in this sub-corpus. Tweets 4–7 illustrate variants of denials of the rumour, variously referring to eye witness reports (4), offering an alternative explanation for reports of police being seen near the hospital (5), relaying information from other media sources (6) and drawing parallels with other (false) rumours (7).

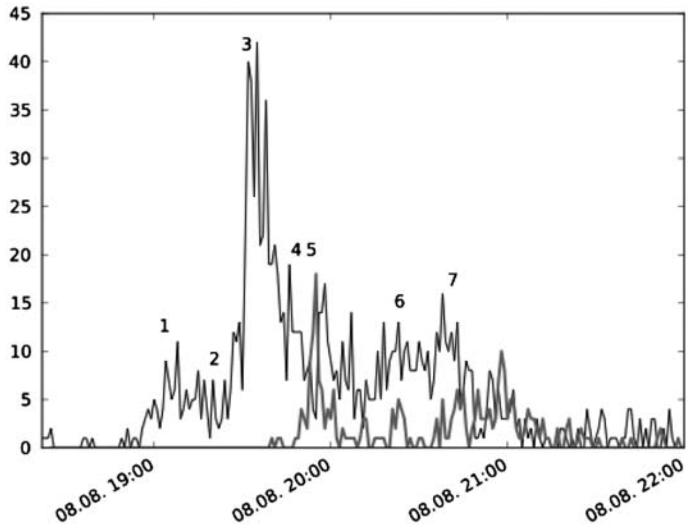

Figure 4. Timeline for Birmingham Children's Hospital rumour. *Y* axis is the volume of tweets per 10 min interval. The – represents tweets supporting the rumour and – represents tweets challenging it.

As in the riots corpus as a whole, a number of tweets in this sub-corpus contain links to other media, for example, mobile phone images (see Figure 5(a), blogs and, more rarely, other media such as newspaper websites.

Using the datasets we produced by coding the rumour sub-corpora information flows as claim, counterclaim, etc. the Guardian Interactive team created animated visualisations of each rumour's trajectory over time (see Figure 6(a) and (b), illustrating the changes over the rumour lifecycle of the weight of claim and counterclaims.[8]

Discussion

In the immediate aftermath of the riots, some politicians and media commentators were quick to blame social media, including Twitter, for their scale and spread. On the specific question of whether social media had been used to incite unlawful acts,

Figure 5. Examples of images provided in tweets as (a) evidence of a threat to Birmingham Children's Hospital and (b) that the London Eye had been set on fire.

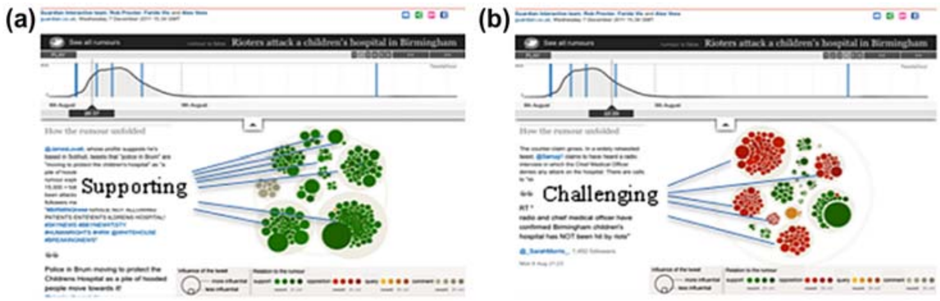

Figure 6. (a) and (b) Extracts from the visualisation of the time line of the Birmingham Children's Hospital rumour showing information flows supporting the rumour (green) and flows challenging it (red). Each individual circle represents a tweet and its size reflects the influence (i.e. the number of followers) of the actor who tweeted it. Individual tweets are grouped according to the information flow to which they belong. We can see in this example how, initially, tweets supporting the rumour (green, arrowed) dominate (Figure 6a) but, within two hours, tweets challenging it (red, arrowed) become dominant. For the interactive visualisations, see http://www.guardian.co.uk/uk/interactive/2011/dec/07/london-riots-twitter

the answer must be yes: people have been convicted of this offence.[9] However, based on the evidence available to us in the corpus, Twitter was used overwhelmingly for more positive ends and, in particular, the organisation of the riot cleanup. Furthermore, we note that the police themselves rejected the idea of closing down social media sites during such crises,[10] arguing that they are a valuable tool for information gathering, for keeping the public informed and for providing advice. However, our study does confirm the conclusions of other studies (e.g. Crump, 2011) that the police (and government agencies in general) have yet to get a grip on using social media platforms like Twitter effectively.

Table 2. Selected information flows from Birmingham Children's Hospital sub-corpus.

	Date, time	Size	Tweet
1	08/08/2011 19:03	53	Police in Brum moving to protect the Childrens' Hospital as a pile of hooded people move towards it! #birminghamriots
2	08/08/2011 19:20	39	Gangs are trying to get into Birmingham's Children Hospital. That is fucking disgusting. Have they no heart? #BirminghamRiots
3	08/08/2011 19:31	146	Rioters in Birmingham make moves for a CHILDREN's hospital, are people that low? #Birminghamriots
4	08/08/2011 19:48	42	Girlfriend has just called her ward in Birmingham Children's Hospital & there's no sign of any trouble #Birminghamriots
5	08/08/2011 19:53	53	May I remind clueless/hysterical #birminghamriots commentators that Children's Hospital sits face-face with city's central police station
6	08/08/2011 20:13	41	#birminghamriots brmb radio and chief medical officer have confirmed Birmingham children's hospital has NOT been hit by riots
7	08/08/2011 20:41	67	#birminghamriots children's hospital NOT attacked. Bull's head NOT cut off. Primark NOT on fire. Can we stop these ridic rumours now.

We can illustrate this point with our findings of how rumours propagate in Twitter. The Birmingham Children's Hospital case study reflects a pattern or trajectory common to the seven rumours that we studied (Table 2):

(1) A rumour starts with someone tweeting about the occurrence of an alleged incident.
(2) The rumour gets retweeted (see Figure 6(a). Some form of evidence – eyewitness reports, references to mainstream news sources, links to pictures (Figure 5) or to mainstream news sources on the Web, etc. – may be added as the original tweet gets retweeted and various reformulations of the rumour also begin to appear.
(3) Others begin to challenge its credibility (i.e. make a counter-claim), perhaps on the basis of logical arguments (e.g. 'it's not possible because …') or new information that throws into doubt the reliability of evidence previously offered.
(4) A consensus begins to emerge (see Figure 6(b)). Where this is that the rumour is false, it may nevertheless re-surface in the corpus as latecomers pick up the original tweet and join in.

A common feature in these rumours is that the mainstream media is seen to lag behind crowd-sourced ('citizen journalism') reports appearing in social media. For example, in the Birmingham Children's Hospital case study (see Table 2), counter-claims seem initially to be driven by a) reports coming in from plausible eyewitness sources (4: 'Girlfriend has just called …'), (b) an appeal to alternative, logical, explanations for the initial claim (5: a more mundane reason for police massing near the hospital) and, finally, reports from mainstream media (6: #birminghamriots brmb radio and chief medical officer have confirmed …). This emphasises how collaborative efforts by large numbers of 'producers' (Bruns, 2008, p. 2) can provide competing and, at times, better coverage of events than mainstream media. Similarly, we observe the absence of early involvement of the police and other emergency services in the rumour lifecycle. For example, in the Birmingham Children's Hospital case study, it was over 24 h after the initial rumour before a mainstream media organisation tweets a report sourced from the local police.

The use of links to other media, for example, mobile phone images (see Figure 5), blogs and online newspaper sites as corroborating evidence is another common feature in all seven of the rumour case studies. However, they show that this evidence cannot always be taken at face value. For example, the authenticity of the image shown in Figure 5(b) (and other, similar images) of the London Eye burning was subsequently challenged (see Table 3) by claims that it had been faked ('photoshopped') to give the impression of a blaze.

It would appear, then, that while Twitter is a fertile medium for launching rumours, it also provides robust mechanisms for self-correction (Mendoza et al., 2010; Sutton, Palen, & Shklovski, 2008). When we examine the proportion of information flows supporting or denying a rumour in these case studies, our findings are broadly consistent with those of Mendoza et al. (2010) who noted that users deal with 'true' and 'false' rumours differently: the former are affirmed more than 90% of the time, whereas the latter are challenged (i.e. questioned or denied) 50% of the time. However, though our findings do not support concerns that Twitter is intrinsically vulnerable to rumours (see Burns & Eltham, 2009), there is a case

Table 3. Tweets claiming the image of the London Eye burning was a fake.

Tweet
03:80 – Pictures of the London Eye are so far known as being fake. To be confirmed. #LondonRiots
CONFIRMED. LONDON EYE PHOTOS FAKE http://t.co/rWYaNM6 #londonriots
That picture of the London eye was most likely photoshopped by some gyp on their freshly robbed MacBook. #londonriots
London Eye Is NOT Burning (not yet anyway) this pic is #photoshopped http://t.co/2DbHdlv #LondonRiots #PrayForLondon
London eye's picture is a FAKE picture, I apologise. http://t.co/3U4n5Gm .. #FAKE #London #LondonRiots
The London Eye is NOT on fire. That pic was photoshopped. #prayforlondon #londonriots
I think your a dick if your photoshopping pictures stirring up rumours e.g.: London eye #londonriots
People are making fake photo's of the London Eye and Big Ben burning. That's sick, the #londonriots are not entertainment.
Big Ben – The London Eye and Waltham Abbey are NOT ON FIRE – the photos are fake – RT this #londonriots
The London Eye isn't on fire, neither is Big Ben. You can't burn metal and heard of Photoshop, hellooooo? #londonriots
London eye in flames is a hoax – several photos making the rounds. Great photoshop skills though 8/// #Londonriots #ukriots

for exploring how self-correction mechanisms may be amplified so that false rumours are identified more quickly.

Being able to analyse not only which information flows were significant and how they circulated – but to also get a clear sense of who propagated the original tweet – allows for a further level of analysis that may highlight the presence of certain actors, but not others. However, further work is needed on the content of the tweets in order to better understand how some of these actors were engaged with and, indeed, how they themselves engaged with other actors. Although the number of actor mentions alone tells us something about a broader picture, to understand this data better we must also look at the context and the ways in which these actors have been mentioned. As 'mentions' essentially include a range of possible Twitter uses (from original tweets, to @replies to mentions), it is important to explore this in more detail. For example, some mentions might highlight media actors that were doing a good job using Twitter in their reporting, as well as those that did not. In our study, we find examples of both.

We note that the use of Twitter as a source of social research data poses a number of methodological challenges. First, the method by which the riots corpus was collected means that we must allow for the possibility of a sampling bias, which could distort our findings. It is highly likely that there were tweets sent during the period in question that would be relevant to our investigation, but were excluded from the corpus because they did not contain any of the hashtags used to select tweets. Second, Twitter users are not representative of the population as a whole (Mislove, Lehman, Yong-Yeol, Jukka-Pekka, & Rosenquist, 2011). In general, how to avoid sampling bias in social media sources is an unresolved question (Omand, Bartlett, & Miller, 2012).

Finally, a point that is particularly relevant to any conclusions we might draw from the riots corpus as to the use of Twitter to incite or organise the riots is the fact that our sampling frame excluded DMs, which are not public.

Conclusions and further work

In this article, we have demonstrated how our methodology, with its combination of computational tools and more established content analysis methods, enabled us to conduct a detailed analysis of a large in corpus of social media data. Computational tools provided the means to expose useful structure in the corpus, which, in turn, then helped us to decide where to focus the human expertise essential for its robust interpretation and analysis.

However, we are aware of the importance of evolving our methodology to meet future challenges in the analysis of 'big data'. In particular, it is clear that we need to take advantage of continuing advances in computational techniques for analysing social media, while ensuring that we make the most effective use of methods that rely on human expertise.

First, the corpus in this study is quite modest by comparison with corpora we might expect to collect in the future. Hence, it is important that the infrastructure we use be scalable to meet expanding computational requirements. We are now in the process of developing the infrastructure so that it can be deployed on a range of different cloud-based solutions, such as Amazon's EC2[11] or Eduserve.[12] This will provide elastic scalability and allow users to retain full control of their data.

Second, we are encouraged by more recent experiments with NLP-based techniques to generate clusters of tweets based on more semantically rich notions of similarity than are captured by information flows. This will facilitate the exposure of additional, useful structure within large corpora and help to offset some of the challenges posed by growing volumes of data. However, it is unlikely (in the short-term at least) that NLP-based techniques will achieve levels of reliability that will enable them to be used 'unsupervised'. Hence, we stress the importance of benchmarking performance against a representative sample of the corpus under investigation annotated by human coders, while noting that the irregular syntax and non-standard language of micro-blogs brings new challenges for defining representativeness.

Third, and related to the previous point, if computational tools are to be applied appropriately in social research then it is vital that users fully understand their strengths and weaknesses, and how they work. Hence, it is essential that social researchers be trained in the underlying concepts of computational methods and tools so they can decide when and how to apply them (Wing, 2008).

Fourth, in seeking a better understanding of the role of platforms such as Twitter during crises, we should remember that social media are part of a much larger and complex media and information ecology, and their interrelationships need to be acknowledged.

Finally, our findings confirm that police, emergency services and government agencies face difficult problems in making effective use of social media platforms such as Twitter during crises. In particular, though our analysis of rumours on Twitter suggests that false rumours are 'self-correcting', we argue that there is a public safety case for providing information and advice via sources that the public can trust in a more timely way. We are now working with a number of UK government agencies to develop operational guidelines and policy recommendations that will help to address this challenge (Procter et al., 2013). For example, there are lessons that may be learnt from the success of the Queensland Police Service Media Unit in tackling false rumours (see Bruns et al., 2012).

Acknowledgements

We would like to thank the UK Joint Information Systems Committee (JISC) for funding this study. We would also like to thank Yana Manykhina, Marta Cantijoch of the Analysing Social Media Collaboration[13] for their contributions to the study, and Steve Cooke, Phil Brooker, Mary Bourne, Yimei Zhu, Ghalia Sarmani, Rosalynd Southern, Patricia Scalco, Daniel Martos Nebrera, Ghalia Sarmani and Stephanie Rinaldi for their invaluable assistance in coding extracts of the corpus. Finally, we would like to acknowledge the invaluable contribution of our collaborators in the Guardian Interactive team, Jonathan Richards, Alastair Dant and Mariana Santos, who created the visualizations based on the data we produced, and Guardian journalists Paul Lewis and James Ball who made this project possible.

Notes

1. www.guardian.co.uk/uk/2011/aug/12/louise-mensch-social-network-blackouts
2. http://www.guardian.co.uk/technology/2012/may/15/twitter-uk-users-10m
3. We drew on the analysis of the top 1000 accounts by mentions by Lisa Evans of the Guardian. See http://www.guardian.co.uk/news/datablog/2011/dec/08/riot-twitter-top-200
4. http://jenkins-ci.org/
5. http://twitpic.com/
6. All times are shown in Coordinated Universal Time (UTC). Local time in August 2011 was UTC + 1.
7. Account names have been anonymised.
8. www.guardian.co.uk/uk/interactive/2011/dec/07/london-riots-twitter
9. www.bbc.co.uk/news/uk-england-14557772
10. www.guardian.co.uk/uk/2011/aug/12/louise-mensch-social-network-blackouts
11. http://aws.amazon.com/ec2/
12. http://www.eduserv.org.uk/
13. analysingsocialmedia.org

Notes on contributors

Rob Procter is a professor of Social Informatics in the Department of Computer Science, University of Warwick, UK. In the past 10 years, he has contributed to innovations in digital research infrastructure, tools and methods, with a particular focus on the challenges of harnessing 'big data' for social research. Recently, he has been applying these innovations in the large-scale study of uses of social media. He has led the multidisciplinary team working with the Guardian/LSE on the Reading the Riots project. Details of subsequent projects in this area can be found at the Analysing Social Media Collaboration website (http://www.analysingsocialmedia.org). He is also a member of the recently launched Centre for Urban Science and Progress (http://www.cusp.nyu.edu). He is s co-editor with Peter Halfpenny of the forthcoming *Innovations in digital research methods*, to be published by Sage in 2014.

Farida Vis is a research fellow in the Social Sciences in the Information School at the University of Sheffield, UK. Her three-year research fellowship is on the theme of Big Data and Social Change, and her work is centrally concerned with researching social media, crisis communication, visual communication, data journalism, open data and citizen engagement. She is a co-author of the *Data journalism handbook*, sits on the editorial board of *Digital Journalism*, is a founding member of Open Data Manchester and has published her own data journalism work in The Guardian Newspaper's well-known Data Blog. Her forthcoming textbook for Sage, *Researching social media*, is written in collaboration with computer scientist Mike Thelwall, supporting the need for interdisciplinary work in this area.

Alex Voss is a lecturer in Software Engineering in the School of Computer Science at the University of St Andrews, UK. His research focuses on how modern information and communication technologies can be used to research social phenomena and how we might turn the insights generated into practical and policy interventions. As a member of the

Analysing Social Media Collaboration (http://www.analysingsocialmedia.org), he is currently developing generic tools to help social scientists collect and analyse data from Twitter and similar sources of data.

References

Allan, S. (2006). *Online news: Journalism and the Internet*. Maidenhead: Open University Press.

Atkinson, M., Britton, D., De Roure, D., Garnett, N., Geddes, N., Gurney, R., … Trefethen, A. E. (2009). Century-of-information research (CIR) – a strategy for research and innovation in the century of information. *Prometheus, 27*, 27–45.

Barsky, L., Trainor, J., & Torres, M. (2006). Disaster realities in the aftermath of Hurricane Katrina: Revisiting the looting myth. *Natural Hazards Center Quick Response Report, 84*, 1–4.

Black, W. J., Procter, R., Gray, S., & Ananiadou, S. (2012). A data and analysis resource for an experiment in text mining a collection of micro-blogs on a political topic. Proceedings of the Eighth International Conference on Language Resources and Evaluation. Istanbul.

Bruns, A. (2006). Wikinews: The next generation of online news? *Scan Journal, 3*(1). Retrieved from http://www.scan.net.au/scan/journal/display.php?journal_id=69.

Bruns, A. (2008). *Blogs, wikipedia, second life, and beyond: From production to produsage*. New York, NY: Peter Lang.

Bruns, A., 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. Available at mappingonlinepublics.net/2012/01/11/cci-report-on-qldfloods-and-qpsmedia-in-the-2011-floods/

Burns, A. & Eltham, B. (2009). Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 Election Crisis. In Papandrea, F. and Armstrong, M. (Eds.) Record of the communications policy & research forum 2009. Network Insight Pty Ltd. Available at www.networkinsight.org/events/cprf09.html/group/6

Crump, J. (2011). What are the police doing on Twitter? Social media, the police and the public. Policy and Internet 3, Article 7.

Halfpenny, P., & Procter, R. (2010). The e-social science research agenda. *Philosophical Transactions of the Royal Society A, special issue on e-Science, 368*, 3761–3778.

Halfpenny, P., Procter, R., Lin, Y., & Voss, A. (2009). Developing the UK e-social science research programme. In N. Jankowski (Ed.), *e-Research, Transformation in Scholarly Practice* (pp. 73–90). Abingdon: Routledge.

Kaigo, M. (2012). Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake. *Keio Communication Review, 34*, 19–35.

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence*. Glencoe, IL: Free Press.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining , Paris (pp. 497–506). New York, NY: ACM Press.

Lewis, P., Newburn, T., Taylor, M., Mcgillivray, C., Greenhill, A., Frayman, H., & Procter, R. (2011). *Reading the riots: Investigating England's summer of disorder*. Retrieved from http://www.guardian.co.uk/uk/series/reading-the-riots

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I, & Boyd, D. (2011). The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication* (5) Feature: 1375–1405.

Mendoza, M., Poblete, B. and Castillo, C. (2010). Twitter under crisis: can we trust what we RT? In 1st Workshop on Social Media Analytics (SOMA '10). Washington, D.C.: ACM Press.

Mislove, A., Lehman, S., Yong-Yeol, A., Jukka-Pekka, O., & Rosenquist, N. J. (2011). *Understanding the demographics of Twitter users*. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona (pp. 554–557).

Morrell, G., Scott, S., McNeish, D., & Webster, S. (2011). *The August riots in England: Understanding the involvement of young people*. National Survey Research Centre. Retrieved from www.natcen.ac.uk/study/the-august-riots-in-england

Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys, 33*, 31–88. doi: 10.1145/375360.375365

Omand, D., Bartlett, J., & Miller, C. (2012). #INTELLIGENCE. Demos, London. Retrieved from www.demos.co.uk/publications/intelligence

Paranyushkin, D. (2012). *Informational epidemics and synchronized viral contagion in social networks*. Nodus Labs. Retrieved from http://noduslabs.com/publications/text-polysingularity-network-analysis.pdf

Procter, R., Crump, J., Karstedt, S., Voss, A., & Cantijoch, M. (2013). Reading the riots: What were the Police doing on Twitter? *Policing and Society*, Special issue on policing and cybercrime.

Sutton, J., Palen, L., & Shklovski, I. (2008). *Backchannels on the front lines: Emergent uses of social media in the 2007 Southern California Wildfires*. In Proceedings of the 5th International ISCRAM Conference. Washington, DC: ISCRAM.

Thelwall, M., & Stuart, D. (2007). RUOK? Blogging communication technologies during crises *Journal of Computer Mediated Communication, 12*, 523–548.

Twitter. (2011). *One hundred million voices*. Twitter blog. Retrieved from blog.twitter.com/2011/09/one-hundred-million-voices.html

Vis, F. (2009). Wikinews reporting of Hurricane Katrina. In S. Allan & E. Thorsen (Eds.), *Citizen journalism: Global perspectives* (pp. 65–74). New York, NY: Peter Lang.

Vis, F. (2012). Reading the Riots on Twitter: who tweeted the riots? Retrieved from researchingsocialmedia.org/2012/01/24/reading-the-riots-on-twitter-who-tweeted-the-riots/

Voss, A., & Procter, R. (2009). Virtual research environments in scholarly work and communications. *Library Hi Tech Journal, special issue on Virtual Research Environments, 27*, 174–190.

Weaver, M. (2010). Iran's Twitter revolution was exaggerated, say editor, *Guardian*. Retrieved from www.guardian.co.uk/world/2010/jun/09/iran-twitter-revolution-protests

Wing, J. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A, 366*, 3717–3725. doi: 10.1098/rsta.2008.0118

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). *Who says what to whom on Twitter*. In Proceedings of the 20th international conference on World Wide Web Conference, Hyderabad (pp. 705–714). New York, NY: ACM Press.

Appendix: code frames

Table A1. Final tweet type code frame.

1. **Media reports** (media accounts or journalists working for MSM or non MSM)

1.1 Mainstream media (MSM)
1.2 Non-mainstream (non MSM)

2. **Pictures** (relying on links to images, not media reports)

2.1 Statement (simply highlighting the image)
2.2 Statement with additional comments

3. **Rumours**

3.1 Claim without evidence
3.2 Claim with evidence
3.3 Counterclaim without evidence
3.4 Counterclaim with evidence
3.5 Appeal for more information
3.6 Comment

4. **Reactions**

4.1 Anti blaming social media for the riots
4.2 Pro blaming social media for the riots
4.3 Anti-shutting down social media
4.4 Pro shutting down social media
4.5 General comments about social media and their role in the riots
4.6 Requests for verification
4.7 Critical about mainstream media
4.8 Critical about politicians/political initiatives
4.9 Highlighting (lists of) credible new sources/journalists
4.10 Appeals to identify rioters through Twitter (links to images).
4.11 Naming and shaming rioters on Twitter
4.12 Details of riot cleanups
4.13 Appeals to join riot clean ups
4.14 Riot clean up is not political
4.15 Highlighting power of Twitter/Twitter's finest hour
4.16 Appeal to not re-tweet rumours
4.17 Appeal to not make fun of the riots
4.18 Anger at rioters
4.19 Highlighting stupidity of rioters
4.20 Journalists requesting people to help (who were involved in the riots)
4.21 Police requesting people to help
4.22 Appeals to help the police/journalists
4.23 Humour
4.24 Warming people (about giving away their location in case looters see this)
4.25 Suggestions of Twitter accounts/hashtags to follow
4.26 Critical about the police
4.27 Other reaction/comments

5. **Link broken**– no longer available
6. **Other**

Table A2. Final actor type code frame.

Actor type		Description
1	Mainstream media	News and media organizations, global, national and local that have both digital and non-digital outlets.
2	Only online media	Only online media (news): Blogs, news portals or journalistic entities that exist solely online.
3	Non-(news) mainstream media orgs	Groups, companies, or organizations that are not primarily news-orientated. This may include social media sites such as Facebook, Twitter and Youtube.
4	Journalists (mainstream)	Individuals employed by MSM organizations, or who regularly work as freelancers for MSM organizations
5	Journalists (online media)	Individuals employed by web news organizations, or who regularly work as freelancers for MS new Media.
6	Non-(news) media org employees	Individuals employed by non-media organisations.
7	Bloggers	Individuals who post regularly to an established blog, and who appear to identify as a blogger on Twitter.
8	Activists	Individuals or organisations, that self-identify as an activist, or who appear to be tweeting purely about activist topics to capture the attention of others.
9	UK Twitterati	Individuals from the UK who have worldwide influence in social media circles and are widely followed on Twitter.
10	Political actors	Elected officials, known primarily for their membership of political parties or relationship to government. This includes local government, councillors.
11	Police/emergency services	
12	Riot accounts	An account that appears to have been set up especially to tweet the riots.
13	Celebrities	Individuals who are famous for reasons unrelated to technology, politics or activism.
14	Researchers	An individual who is affiliated with a university or think-tank.
15	Members of the public	Individuals who provide no link to organization or institution. The account appears to be maintained by a private citizen in their personal capacity, highlighting personal information in their bio.
16	Bots	Accounts that appear to be an automated service tweeting consistent content, usually in extraordinary volumes.
17	Unclear	An account that may fit more than one of the categories (blogger/activist/journalist) where it is difficult to distinguish which is most important.
18	Account closed down	Unable to classify actor.
19	Fake/spoof account	An account that has clearly been set up to pretend to be someone else.
20	Other	Accounts that do not clearly fit in any category.