Sentiment in Twitter events[1]

Mike Thelwall, Kevan Buckley, Georgios Paltoglou
Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK.
E-mail: m.thelwall@wlv.ac.uk, K.A.Buckley@wlv.ac.uk, G.Paltoglou@wlv.ac.uk
Tel: +44 1902 321470 Fax: +44 1902 321478

The microblogging site Twitter generates a constant stream of communication, some of which concerns events of general interest. An analysis of Twitter may, therefore, give insights into why particular events resonate with the population. This article reports a study of a month of English Twitter posts, assessing whether popular events are typically associated with increases in sentiment strength, as seems intuitively likely. Using the top 30 events, determined by a measure of relative increase in (general) term usage, the results give strong evidence that popular events are normally associated with increases in negative sentiment strength and some evidence that peaks of interest in events have stronger positive sentiment than the time before the peak. It seems that many positive events, such as the Oscars, are capable of generating increased negative sentiment in reaction to them. Nevertheless, the surprisingly small average change in sentiment associated with popular events (typically 1% and only 6% for Tiger Woods' confessions) is consistent with events affording posters opportunities to satisfy pre-existing personal goals more often than eliciting instinctive reactions.

Introduction

Social networking, blogging and online forums have turned the web into a vast repository of comments on many topics, generating a potential source of information for social science research (Thelwall, Wouters, & Fry, 2008). The availability of large scale electronic social data from the web and elsewhere is already transforming social research (Savage & Burrows, 2007). The social web is also being commercially exploited for goals such as automatically extracting customer opinions about products or brands. An application could build a large database of web sources (Bansal & Koudas, 2007; Gruhl, Chavet, Gibson, Meyer, & Pattanayak, 2004), use information retrieval techniques to identify potentially relevant texts, then extract information about target products or brands, such as which aspects are disliked (Gamon, Aue, Corston-Oliver, & Ringger, 2005; Jansen, Zhang, Sobel, & Chowdury, 2009). From a social sciences perspective, similar methods could potentially give insights into public opinion about a wide range of topics and are unobtrusive, avoiding human subjects research issues (Bassett & O'Riordan, 2002; Enyon, Schroeder, & Fry, 2009; Hookway, 2008; White, 2002).

The sheer size of the social web has also made possible a new type of informal literature-based discovery (for literature based discovery, see: Bruza & Weeber, 2008; Swanson, Smalheiser, & Bookstein, 2001): the ability to *automatically* detect events of interest, perhaps within pre-defined broad topics, by scanning large quantities of web data. For instance, one project used time series analyses of (mainly) blogs to identify emerging public fears about science (Thelwall & Prabowo, 2007) and businesses can use similar techniques to quickly discover customer concerns. Emerging important events are typically signalled by sharp increases in the frequency of relevant terms. These bursts of interest are important to study because of their role in detecting new events as well as for the importance of the events discovered. One key unknown is the role of sentiment in the emergence of important events because of the increasing recognition of the importance of emotion in awareness, recall and judgement of information (Fox, 2008, p. 242-244, 165-167, 183;

[1] Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406-418.

Kinsinger & Schacter, 2008) as well as motivation associated with information behaviour (Case, 2002, p. 71-72; Nahl, 2006, 2007a).

The research field of sentiment analysis, also known as opinion mining, has developed many algorithms to identify whether an online text is subjective or objective, and whether any opinion expressed is positive or negative (Pang & Lee, 2008). Such methods have been applied on a large scale to study sentiment-related issues. One widely-publicised study focused on the average level of sentiment expressed in blogs (as well as lyrics and US presidential speeches) in order to identify overall trends in levels of happiness as well as age and geographic differences in the expression of happiness (Dodds & Danforth, 2010). A similar approach used Facebook status updates to judge changes in mood over the year and to assess "the overall emotional health of the nation" (Kramer, 2010) and another project assessed six dimensions of emotion in Twitter, showing that these typically reflect significant offline events (Bollen, Pepe, & Mao, 2009). Nevertheless, despite some research into the role of sentiment in online communication, there are no investigations into the role that sentiment plays in important online events. To partially fill this gap, this study assesses whether Twitter-based surges of interest in an event are associated with increases in expressed strength of feeling. Within the media, a well-established notion is that emotion is important in engaging attention, as expressed for violence by the common saying, "if it bleeds, it leads" (Kerbel, 2000; Seaton, 2005), and through evidence that audiences emotionally engage with the news (Perse, 1990). It seems logical, therefore, to hypothesise that events triggering large reactions in Twitter would be associated with increases in the strength of expressed sentiment, but there is no evidence yet for this hypothesis.

Literature review

According to Alexa, and based upon its panel of toolbar users, Twitter had become the world's ninth most popular web site by October 2010 (alexa.com/topsites, accessed October 8, 2010), despite only beginning in July 2006. The rapid growth of the site may be partly due to celebrities tweeting regular updates about their daily lives (Johnson, 2009). Also according to Alexa, amongst Internet users people aged 25-44 were slightly over-represented in Twitter and those aged 55+ were much less likely to use it than average; women were also slightly over-represented (http://www.alexa.com/siteinfo/twitter.com, accessed October 8, 2010). Thus, despite the mobile phone connection Twitter is not a teen site, at least in the US: "Teens ages 12-17 do not use Twitter in large numbers, though high school-aged girls show the greatest enthusiasm" (Lenhart, Purcell, Smith, & Zickuhr, 2010).

Information dissemination and socialising with Twitter

Twitter can be described as a microblog or social network site. It is for microblogging because the central activity is posting short status update messages (tweets), via the web or a mobile phone. Twitter is also a social network site because members have a profile page with some personal information and can connect to other members by "following" them, thus gaining easy access to their content. It seems to be used to share information and to describe minor daily activities (Java, Song, Finin, & Tseng, 2007) although it can also be used for information dissemination, for example by government organisations (Wigand, 2010). About 80% of Twitter users update followers on what they are currently doing, whilst the remainder have an informational focus (Naaman, Boase, & Lai, 2010). There are clear differences between users in terms of connection patterns: although most seem to be symmetrical in terms of having similar numbers of followers to number of users followed, some are heavily skewed, suggesting a broadcasting or primarily information gathering/evangelical function (Krishnamurthy, Gill, & Arlitt, 2008). Twitter displays a low reciprocity in messages between users, unlike other social networks, suggesting that its primary function is not as a social network (Kwak, Lee, Park, & Moon, 2010), but perhaps to spread news (including personal news) or other information instead.

An unusual feature of Twitter is retweeting: forwarding a tweet by posting it again. The purpose of this is often to disseminate information to the poster's followers, perhaps in

modified form (boyd, Golder, & Lotan, 2009), and this reposting seems to be extremely rapid (Kwak et al., 2010). The reposting of the same (or similar) information works because members tend to follow different sets of people, although retweeting also serves other purposes such as helping followers to find older posts. The potential for information to flow rapidly through Twitter can also be seen from the fact that the average path length between a pair of users seems to be just over 4 (Kwak et al., 2010). Moreover, if retweeted, a tweet can expect to reach an average of 1000 users (Kwak et al., 2010). Nevertheless, some aspects of information dissemination are not apparent from basic statistics about members. For instance the most followed members are not always the most influential, but topic focus within a Twitter account helps to generate genuine influence (Cha, Haddadi, Benevenuto, & Gummadi, 2010). Moreover, an important event can be expected to trigger more informational tweeting (Hughes & Palen, 2009), which suggests that it would be possible to detect important events through the automatic analysis of Twitter. In support of this, Twitter commentaries have been shown to sometimes quite closely reflect offline events, such as political deliberations (Tumasjan, Sprenger, Sandner, & Welpe, 2010).

Another communicational feature of Twitter is the hashtag: a metatag beginning with # that is designed to help others find a post, often by marking the Tweet topic or its intended audience (Efron, 2010). This feature seems to have been invented by Twitter users, in early 2008 (Huang, Thornton, & Efthimiadis, 2010). The use of hashtags emphasises the importance of widely communicating information in Twitter. In contrast, the @ symbol is used to address a post to another registered Twitter user, allowing Twitter to be used quite effectively for conversations and collaboration (Honeycutt & Herring, 2009). Moreover, about 31% of Tweets seem to be directed at a specific user using this feature (boyd et al., 2009), emphasising the social element of Twitter rather than the information broadcasting function associated with hashtags.

In summary, there is considerable evidence that even though Twitter is used for social purposes it has significant use for information dissemination of various kinds, including personal information, and this may be its major use. It is therefore reasonable to conduct time series analyses of Tweets posted by users.

Twitter use as an information behaviour: The affective dimension

Whilst the above subsection describes Twitter and usage patterns, responding to an external event by posting a tweet is information behaviour, and therefore has an affective component, in the sense of judgements or intentions, irrespective of whether the information used is subjective (e.g., Nahl, 2007b; Tenopir, Nahl-Jakobovits, & Howard, 1991). Of particular interest here is whether individuals encode sentiment into their messages: a topic that appears to have attracted little research.

A useful theoretical construct for understanding how people may react to events is the concept of *affordances* (Gaver, 1991; Gibson, 1977, 1986), as also used by Nahl (2007b). Instead of focusing on the ostensible purpose or function of something, it also makes sense to consider what uses can be made of it to suit the goals of the person concerned. For a use to occur, however, its potential must be first perceived. In the context of Twitter, this suggests that an event reported in the media may be perceived by some Twitter users as affording an opportunity to satisfy unrelated goals, such as to create humour, show analytical skill or declare a moral perspective. Hence, whilst an emotional event might seem likely to elicit intuitive reactions, such as declarations of pleasure or disgust, this is not inevitable. This analysis aligns with the uses and gratifications approach from media studies (Blumler & Katz, 1974; Katz, 1959; Stafford, Stafford, & Schkade, 2004), which posits that people do not passively consume the media but actively select and exploit it for their own goals. Borrowing an idea from computer systems interface design, it seems that non-obvious affordances need a culture of use to support them (Gaver, 1991; MacLean, Carter, Lövstrand, & Moran, 1990) and so if there is a culture of using information in non-obvious ways for Twitter posts then this culture can be passed on by its originators to other users and could become the main explanation for the continuation of the practice.

The rest of this section describes methods for sentiment-based internet time series analysis.

Sentiment analysis of online text

Sentiment analysis is useful for research into online communication because it gives researchers the ability to automatically measure emotion in online texts. The research field of sentiment analysis has developed algorithms to automatically detect sentiment in text (Pang & Lee, 2008). Whilst some identify the objects discussed and the polarity (positive, negative or neutral) of sentiment expressed about them (Gamon et al., 2005), other algorithms assign an overall polarity to a text, such as a movie review (Pang & Lee, 2004). Three common sentiment analysis approaches are full-text machine learning, lexicon-based methods and linguistic analysis. For standard machine learning (e.g., Witten & Frank, 2005), a set of texts annotated for polarity by human coders are used to train an algorithm to detect features that associate with positive, negative and neutral categories. The text features used are typically sets of all words, word pairs and word triples found in the texts. The trained algorithm can then look for the same features in new texts in order to predict their polarity (Pak & Paroubek, 2010; Pang, Lee, & Vaithyanathan, 2002). The lexicon approach starts with lists of words that are pre-coded for polarity and sometimes also for strength and uses their occurrence within texts to predict their polarity (Taboada, Brooke, Tofiloski, Voll, & Stede, in press). A linguistic analysis, in contrast, exploits the grammatical structure of text to predict its polarity, often in conjunction with a lexicon. For instance, linguistic algorithms may attempt to identify context, negations, superlatives and idioms as part of the polarity prediction process (e.g., Wilson, Wiebe, & Hoffman, 2009). In practice, algorithms often employ multiple methods together with various refinements, such as pre-filtering the features searched for (Riloff, Patwardhan, & Wiebe, 2006), and methods to cope with changes in data over time (Bifet & Frank, 2010).

A few algorithms detect sentiment strength in addition to sentiment polarity (Pang & Lee, 2005; Strapparava & Mihalcea, 2008; Wilson, Wiebe, & Hwa, 2006), including some for informal online text (Neviarouskaya, Prendinger, & Ishizuka, 2007; Thelwall, Buckley, Paltoglou, Cai, & Kappas, in press). These work on the basis that humans can differentiate between mild and strong emotions in text. For instance, hate may be regarded as a stronger negative emotion than dislike. Sentiment strength algorithms attempt to assign a numerical value to texts to indicate the strength of any sentiment detected.

In addition to academic research, sentiment analysis is now a standard part of online business intelligence software, such as Market Sentinel's Skyttle and sysomos's Map. The direct line provided by Twitter between customer opinions and businesses has potentially valuable implications for marketing as a competitive intelligence source (Jansen et al., 2009). There are also now web sites offering free sentiment analysis for various online data sources, including tweetfeel and Twitter Sentiment.

Time series analysis of online topics

In many areas of research, including statistics and economics, a time series is a set of data points occurring at regular intervals. Time series data is useful to analyse phenomena that change over time. Whilst there are many complex mathematical time series analysis techniques (Hamilton, 1994), this review focuses on simple analyses of web data. These typically aggregate data into days to produce daily time series. International time differences can be eliminated by either adjusting all times to Greenwich Mean Time or by ignoring the time of day and recording the day in the country of origin of the data.

Several previous studies have analysed online communication from a time series perspective, revealing the evolution of topics over time. One investigation of blogs manually selected 340 topics, each defined by a proper noun or proper noun phrase. Time series data of the number of blogs per day mentioning each topic was then constructed. Three common patterns for topics were found (Gruhl, Guha, Liben-Nowell, & Tomkins, 2004): a single spike of interest - a short period in which the topic is discussed (i.e., an increase in the number of

blogs referring to it), with the topic rarely mentioned before or afterwards; fairly continuous discussion without spikes; or fairly continuous discussion with occasional spikes triggered by relevant events. It seems likely that spikes are typically caused by external events, such as those reported in the news, but some may result from the viral spreading of jokes or information generated online.

The volume of discussion of an issue online has been used to make predictions about future behaviour, confirming the connection between online and offline activities. One study used book discussions to predict future Amazon sales with the assumption that a frequently blogged book would see a resultant increase in sales. Such a connection was found, but was weak (Gruhl, Guha, Kumar, Novak, & Tomkins, 2005). A deeper text mining approach decomposed Amazon product reviews of the 242 items in the "Camera & Photo" and "Audio & Video" categories into segments concerning key different aspects of the product (e.g., viewfinder, software) and used them to estimate the value of the different features found. This information was then used to predict future Amazon sales (Archak, Ghose, & Ipeirotis, 2007). In combination with machine learning, various temporal aspects of blogs, such as posting frequency, response times, post ratings and blogger roles, have also been used with some success within a particular community of experts to predict stock market changes (Choudhury, Sundaram, John, & Seligmann, 2008). Similarly, the number of tweets matching appropriate keywords has been shown to correlate with influenza outbreaks, a non-commercial application of similar methods (Culotta, 2010). Perhaps most impressively, one system monitors Twitter in real time in Japan and uses keyword-based models to automatically identify where and when earthquakes occur, with a high probability, from Tweets about them (Sakaki, Okazaki, & Matsuo, 2010).

The studies reviewed here demonstrate that time series analysis of online data is a promising research direction and that online events can often be expected to correlate with offline events. Finally, note also that link analysis has also been used to model the online spread of information (Kumar, Novak, Raghavan, & Tomkins, 2003) but this approach is not relevant here.

Sentiment-based time series analysis of online topics

Time series analysis of online data has been combined with sentiment analysis to make predictions, extending the work reviewed above. For instance, blog post sentiment and frequency can predict movie revenues, with more positive reviews suggesting improved revenue (Mishne & Glance, 2006). Moreover, estimates of the daily amount of anxiety, worry and fear in LiveJournal blogs may predict overall stock market movement directions (Gilbert & Karahalios, 2010).

The relationship between sentiment and spikes of online interest, the topic of the current paper, has previously been investigated through an analysis of LiveJournal blog postings for a range of mood-related terms (e.g., tired, excited, drunk – terms self-selected by bloggers to describe their current mood). In order to detect changes in mood, the average number of mood-annotated postings was compared with the average for the same hour of the same day of the week over all previous weeks for which data was available (Balog, Mishne, & Rijke, 2006). Note that the same method would not work in Twitter since it does not have the same mood annotation facility (at least as of October 2010). To find the cause of identified mood changes, word frequencies for posts associated with the change in mood were compared to a reference set to identify unusually common words. Although a full evaluation was not conducted, the method was able to identify major news stories (e.g., Hurricane Katrina) as the *causes* of the mood changes. The goal was not to *identify* major news stories, however, as this could be achieved through simpler term volume change methods (Thelwall & Prabowo, 2007; Thelwall, Prabowo, & Fairclough, 2006).

Time series analyses of emotion have also been conducted for a range of offline and online texts to identify overall trends in the amount expressed (Dodds & Danforth, 2010). Although not focussing on spikes, this study confirmed that individual days significantly differed from the average volume when particular major news events or annual celebrations

(e.g., Valentine's Day) occurred. This conclusion agrees with a study of Twitter data from late 2008, which showed that average changes in Twitter mood levels correlate with social, political, cultural and economic events, although there is sometimes a delay between an event and an apparently associated mood change in Twitter (Bollen et al., 2009). A similar correlation has also been observed between a measure of average happiness for the USA based on Facebook status updates and significant events (Kramer, 2010). A different approach using Twitter data from 2008 and 2009 correlated the polarity of tweets relevant to topics (the US elections; customer confidence) with the results of relevant questions in published opinion polls. There was a strong correlation between the sentiment-based Twitter scores and the opinion poll results over time, suggesting that automatic sentiment detection of Twitter could monitor public opinion about popular topics (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). Separately, paid human coders via Amazon Mechanical Turk, have coded tweets about the first 2008 US presidential debate as negative, positive, mixed, or other. The time series method used on the resulting data was able to predict not only the outcome of the debate but also particular points of interest during the debate that triggered emotions (Diakopoulos & Shamma, 2010).

In summary, there are many sentiment-based time series analyses of online topics, and these have found that sentiment changes can be used to predict offline phenomena or associate with offline phenomena. No research has addressed the issue from the opposite direction however: are peaks of interest in online topics always associated with changes in sentiment?

Research question

The goal of assessing whether surges of interest in Twitter are associated with heightened emotions could be addressed in two ways: by measuring whether the average sentiment strength of popular Twitter events is higher than the Twitter average or by assessing whether an important event within a broad topic is associated with increased sentiment strength. Whilst the former may seem to be the more logical approach, it is not realistic because of the many trivial, commercial and informational uses of Twitter. These collectively make the concept of average Twitter sentiment strength unhelpful: hence the broad topic approach was adopted.

Motivated by Topic Detection and Tracking (TDT) task in the Text REtrieval Conferences (TREC), an event can be defined as something that happens at a particular place and time whereas a topic may be a collection of related events (Allan, Papka, & Lavrenko, 1998). In fact, these terms are recognised to be problematic to define: for instance, an event could also be more vaguely defined as a qualitatively significant change in something (Guralnik & Srivastava, 1999), and something that is spread out in time and space may still be regarded as an event, such as the O. J. Simpson incident (Allan et al., 1998). Here the term *event* is used with a narrow definition: something unrelated to Twitter that triggers an increase in the frequency of one or more words in Twitter. The *broad topic* for an event covers content that is related to the event, but not necessarily bounded in time, and is operationalised as a keyword search. The research question is therefore:

- Are the most tweeted events in Twitter usually associated with increases in positive or negative sentiment strength compared to other tweets about the same broad topic?

In order to operationalise the phrase, "associated with increased", in the research question, three pairs of Twitter categories were defined as follows for events and their broad topics. Here, the maximum volume hour is the hour on which the number of posts relevant to a broad topic is highest, and all volumes refer only to topic-relevant posts:

- Higher volume hours: Hours with at least 10% of the maximum volume.
- Lower volume hours: Hours with less than 10% of the maximum volume.
- Hours before maximum volume: All hours before the hour of the maximum volume.
- Hours after maximum volume: All hours after the hour of the maximum volume.
- Peak volume hours: The 5 hours before and after the maximum volume hour (11 hours in all).

- Hours before peak volume: All hours at least 6 hours before the maximum volume hour.

Based upon the above six categories, six hypotheses are addressed.

- H1n: For the top 30 events during a month, the average negative sentiment strength of tweets posted during higher volume hours will tend to be greater than the average negative sentiment strength of tweets posted during lower volume hours.
- H2n: For the top 30 events during a month, the average negative sentiment strength of tweets posted after the maximum volume hour will tend to be greater than the average negative sentiment strength of tweets posted before the maximum volume hour.
- H3n: For the top 30 events during a month, the average negative sentiment strength of tweets posted during peak volume hours will tend to be greater than the average negative sentiment strength of tweets posted before the peak volume hours.

Hypotheses H1p, H2p, and H3p are as H1n-H3n above but for positive sentiment strength.

Methods

The data used was a set of Twitter posts from February 9, 2010 to March 9, 2010 downloaded from data company Spinn3r as part of their (then) free access programme for researchers. The data consisted of 34,770,790 English-language tweets from 2,749,840 different accounts. The restriction to English was chosen to remove the complication of multiple languages. The data was indexed with the conversion of plural to singular words but no further stemming (c.f., Porter, 1980).

The top 30 events from the 29 selected days were identified using the time series scanning method (Thelwall & Prabowo, 2007; Thelwall et al., 2006) that separately calculates for each word in the entire corpus its hourly relative frequency (proportion of posts per hour containing the word) and then the largest increase in relative frequency during the time period. A similar approach has previously been applied to Twitter to detect “emerging topics” (Cataldi, Caro, & Schifanella, 2010). The increase in relative frequency is the relative frequency at any particular hour minus the average relative frequency for all previous hours. The "3 hour burst" method was used so that increases had to be sustained for three consecutive hours to count. The words were then listed in decreasing order of relative increase. This method thus creates a list of words with the biggest spikes of interest.

The top 30 words identified through the above method were not all chosen as the main events because in some cases multiple words referred to the same event (e.g., tiger, woods) and in other cases the words were not events but were hashtags presumably created by the mass media and being twitter-specific topics (e.g., #petpeeves, #imattractedto, #relationshiprules and #ff - Follow Friday, used on Fridays). All artificial hashtags of this kind were removed, multiple words referring to the same event were merged and the event "Olympics" was removed as this referred to multiple events over a period of weeks (the Winter Olympics 2010 in Canada).

For each of the selected topics, Boolean searches were generated to match relevant posts and to avoid irrelevant matches as far as possible (see Table 1). These searches were constructed by trial and error. During this process the broad topic containing the term was identified as well as the specific event that triggered the spike in the data. The searches were constructed to reflect narrow topics but not so narrow as to only refer to the specific event causing the spike so that comparisons could be conducted of relevant tweets before and after the event. In most cases this process was straightforward, as Table 1 shows. This led to one small keyword anomaly: several films are represented in the list due to the Oscars, but only the two with general names (Precious and Avatar) had "oscar" added to their search in order to remove spurious matches.

The next stage was to classify the sentiment strength of each tweet. Whilst many algorithms detect text subjectivity or sentiment polarity, a few detect sentiment strength (Pang & Lee, 2005; Strapparava & Mihalcea, 2008; Wilson et al., 2006). Nevertheless, the accurate detection of sentiment is domain-dependant. For instance, an algorithm that works well on movie reviews may perform badly on general blog posts. Twitter texts are short, because of

the 140 character limit on messages, and informal language and abbreviations may be common because of the shortness and the use of mobile phones to post messages. The SentiStrength algorithm (Thelwall et al., in press) is suitable because it is designed for short informal text with abbreviations and slang, having been developed for MySpace comments. It seems to be more appropriate than the most similar published algorithm (Neviarouskaya et al., 2007) because the latter has less features and has been less extensively tested for accuracy.

SentiStrength classifies for positive and negative sentiment on a scale of 1 (no sentiment) to 5 (very strong positive/negative sentiment). Each classified text is given both a positive and negative score, and texts may be simultaneously positive and negative. For instance, "Luv u miss u", would be rated as moderately positive (3) *and* slightly negative (-2). SentiStrength combines a lexicon - a lookup table of sentiment-bearing words with associated strengths on a scale of 2 to 5 - with a set of additional linguistic rules for spelling correction, negations, booster words (e.g., very), emoticons and other factors. The positive sentiment score for each tweet is the highest positive sentiment score of any constituent sentence. The positive sentiment score of each sentence is essentially the highest positive sentiment score of any constituent word, after any linguistic rule modifications. The same process applies to negative sentiment strength. The special informal text procedures used by SentiStrength include a lookup table of emoticons with associated sentiment polarities and strengths, and a rule that sentiment strength is increased by 1 for words with at least two additional letters (e.g., *haaaappy* scores one higher than *happy*). The algorithm has a higher accuracy rate than standard machine learning approaches for positive sentiment and a similar accuracy rate for negative sentiment strength (Thelwall et al., in press).

All posts were classified by SentiStrength for positive and negative sentiment strength and then hourly average positive and negative sentiment strength scores were calculated for each topic. Each score was calculated by adding the sentiment strength of each relevant post from each hour and dividing by the total number of posts for the hour. For instance if all posts had strength 1 then the result would be 1, whereas if all posts had strength 5 then the result would be 5. The averages varied between 1.3 and 3.0 (positive) and 1.1 and 2.9 (negative). The highest averages associated with topics with searches containing sentiment-bearing terms (hurt, care, killer and precious). The posts from each topic were split into three pairs of categories for the statistical tests, as defined in the research questions section above.

The non-parametric Wilcoxon signed ranks test was used to assess the six hypotheses. A Bonferroni correction for six tests was used to guard against false positives, so that p=0.008 is the critical value equivalent to the normal 5% level and p=0.002 is the critical value equivalent to 1%.

Results

Table 1 reports a summary of the topics chosen, the queries used and the average sentiment strength difference for each one. Wilcoxon signed ranks test were applied separately for the positive and negative sentiment data (n=30), giving the following results (with Bonferroni-corrected conclusions but original *p* values).

Negative sentiment

- H1n: There is strong evidence (p=0.001) that higher volume hours have stronger negative sentiment than lower volume hours.
- H2n: There is strong evidence (p=0.002) that hours after the peak volume have stronger negative sentiment than hours before the peak.
- H3n: There is strong evidence (p=0.002) that peak volume hours have stronger negative sentiment than hours before the peak.

Positive sentiment:

- H1p: There is no evidence (p=0.014) that higher volume hours have different positive sentiment strength than lower volume hours.
- H2p: There is no evidence (p=0.781) that hours after the peak volume have different positive sentiment strength than hours before the peak.

- H3p: There is some evidence (p=0.008) that peak volume hours have stronger positive sentiment than hours before the peak.

Table 1. List of topics (in decreasing order of spike size), searches used and sentiment strength differences**.

Topic (peak event)	**Search***	**Tweets**	**Neg. high – low**	**Neg. after – before max**	**Neg. peak – before peak**	**Pos. high – low**	**Pos. after – before max**	**Pos. peak – before peak**
Oscars (award ceremony)	oscar #oscar	90,473	0.100	0.125	0.122	0.127	0.016	0.147
Tsunami in Hawaii (warning issued)	tsunami hawaii #tsunami	43,216	0.075	0.067	0.088	**-0.142**	0.071	0.024
Chile or earthquakes (Chilean earthquake)	chile #chile earthquake quake	84,030	0.002	0.035	0.052	**-0.072**	0.039	**-0.019**
Tiger Woods (admits affairs)	tiger wood	62,205	0.226	0.164	0.269	0.014	0.071	**-0.012**
Alexander McQueen (death)	alexander AND mcqueen	7007	*0.593*	***-0.340***	*0.838*	0.158	0.175	0.136
The Hurt Locker (Oscar ceremony)	hurt AND locker	11,883	0.007	**-0.037**	0.002	0.141	**-0.012**	0.185
Sandra Bullock (Oscars and Razzies)	sandra AND bullock	7,063	**-0.083**	0.118	**-0.170**	0.180	**-0.034**	*0.324*
Shrove Tuesday pancakes (Shrove Tues.)	pancake	22,992	0.025	0.023	0.025	0.033	**-0.026**	0.003
Red carpet at the Oscars (Oscars arrivals)	red AND carpet AND oscar	1,883	0.042	0.073	0.089	0.040	**-0.004**	0.068
The Brit Awards (ceremony)	brit	15,031	0.059	0.166	0.089	0.053	0.044	0.020
Avatar and the Oscars (ceremony)	avatar AND oscar	1,391	0.178	0.259	*0.308*	0.131	0.005	0.093
Sachin Tendulkar (breaks international cricket record)	Sachin AND Tendulkar	1,134	0.002	0.009	**-0.044**	0.077	0.006	0.284
Google Buzz (launch)	google AND buzz	29,704	**-0.004**	0.051	0.134	0.093	**-0.057**	**-0.052**
Plane crash (Austin, Texas)	plane AND crash	2,895	**-0.027**	0.177	***-0.366***	**-0.006**	**-0.029**	0.173
Alice in Wonderland (Oscar ceremony)	alice AND wonderland	24,819	0.034	**-0.019**	0.028	0.135	0.086	0.089
Biggie Smalls (death anniversary)	biggie	6,438	0.012	0.054	**-0.001**	**-0.001**	0.032	**-0.013**
Rapper Guru (in coma)	guru	9,299	*0.435*	**-0.146**	0.066	0.066	**-0.060**	0.097
The Bachelor TV show (finale)	bachelor jake	18,734	*0.328*	0.144	*0.359*	0.070	0.022	0.040

Health care summit (meeting)	health AND care AND summit	2,846	0.012	0.202	0.033	0.046	0.061	0.052
Killer whale (attacks trainer)	killer AND whale	4,127	0.111	*0.474*	*0.566*	**-0.065**	**-0.045**	***-0.608***
IHOP restaurant (national pancake day 23 Feb.)	ihop	7,924	**-0.027**	0.03	0.020	**-0.041**	0.037	**-0.015**
Kathryn Bigelow (Oscar ceremony)	bigelow	4,044	**-0.017**	0.012	**-0.121**	0.167	**-0.074**	0.271
HTC (releases Google Android iPhone)	htc	9,052	0.052	0.000	**-0.036**	**-0.175**	**-0.116**	**-0.187**
Slam dunk competition (final)	dunk	9,932	0.269	*0.318*	*0.312*	**-0.025**	0.038	**-0.009**
James Taylor (singing at the Oscars)	jame AND taylor	559	0.274	**-0.022**	0.280	0.253	0.120	0.225
The Lakers basketball team (loose tight game)	laker	14,972	0.150	**-0.010**	0.080	0.093	**-0.107**	0.145
Lady Gaga (sings at the Brit Awards)	lady AND gaga	25,876	**-0.068**	0.015	0.006	0.053	0.055	0.035
Russia vs. Canada (Olympic hockey game)	russia AND canada	1,810	**-0.010**	0.039	0.050	0.128	**-0.077**	0.119
Precious at the Oscars (Oscar ceremony)	precious AND oscar	549	0.232	0.179	0.240	0.024	**-0.002**	0.019
Bill Clinton (hospitalised with chest pain)	bill AND clinton	2,574	*0.458*	0.104	*0.500*	**-0.084**	**-0.240**	**-0.069**

*Note that OR is default and plurals also match the terms given.
**Negative values are bold for emphasis and values above 0.3 are italic for emphasis

Discussion and limitations

The results give strong evidence that negative sentiment plays a role in the main spiking events in Twitter (H1n-H3n accepted) but only some evidence of a role for positive sentiment (H3p accepted; H1p, H2p rejected). Hence it is reasonable to regard spikes accompanied by negative sentiment strength increases as normal, and for associated positive sentiment strength increases as being at least unremarkable. There are several key limitations that affect the extent to which these findings can be generalised, however. First, the data cover only one month and include two special events (the Oscars and the Olympics). Other months may have a different sentiment pattern, particularly if dominated by an unambiguously positive or negative event. Second, the analysis covered the top 30 events during a month and it may be that the top 30 events for a significantly longer or shorter time period could give different results.

The sentiment strength algorithm used is also an issue, especially since sentiment-bearing terms are in some of the topics. This should not lead to spurious increases in sentiment associated with events, however, and should not undermine the results. The one possible exception is that The Hurt Locker (with the negative keyword *hurt*) was sometimes mentioned within other Oscar-related topics (Kathryn Bigelow, Precious, Alice in Wonderland, Sandra Bullock, Oscars) but these did not exhibit unusual increases in

negativity, with some recording decreases in some metrics, and so this does not seem to have had a major impact on the results. To confirm this, the six Wilcoxon tests were repeated with new sentiment values that ignored the main potentially problematic sentiment-bearing words (hurt, care, killer, precious), giving identical overall results and only one change in *p* value (from 0.014 to 0.013 for positive sentiment high vs. low volume hours).

The final major issue is that the events were calculated using hourly time series but time series based upon different time scales could have given different types of events. For instance, events that are broadcast live seem to have an advantage for hourly time series since viewers could tweet as soon as they finished, creating a sharp spike, whereas the spike for an event reported in the media would presumably be flatter because of the variety of media sources reporting it. To assess this, the full analysis was repeated using days rather than hours: 30 topics were selected based upon single day spikes in daily time series (19 of the 30 topics found were the same) and sentiment strengths were calculated on the basis of days rather than hours, and excluding hurt/cares/killer/precious from the sentiment dictionary. None of the results of the six new Wilcoxon tests were significant. The lack of significant results for day-based data counter-intuitively suggests that the time period length impacts on the role of sentiment in Twitter events. This may be due to the sparser data available for topics with little off-peak discussion because most discussion happens on the day of the event. It may also be partly due to the cruder slicing of time. For instance, some high volume hours are included as part of low volume days because their day is, on average, high volume. Topics may also be less emotionally-charged when selected by day because events causing an instantaneous reaction, such as sport event endings, may be less represented. For instance the Lakers' game ending was not selected from the daily data.

The importance of negative sentiment is surprising because many of the contexts are ostensibly positive: the Oscars, the Olympics, Shrove Tuesday, product launches. Nevertheless, even positive events trigger negative reactions. For the Oscars, posts during the ceremony include (slightly modified extracts): "guests look uncomfortable", "Fuck the Oscars", "excuse me while i scream", and "Vote for the Red Carpet Best and Worst Dressed". Similarly, for The Bachelor TV show finale, the comments included reactions against the show itself, such as "Hate that show" as well as disagreement with the outcome: "Jake is an idiot!" It seems that Twitter is used by people to express their opinions on events and that these posts tend to be more negative than average for the topic.

Events without negative sentiment strength increases are unusual, given the overall results. Whilst all 30 topics recorded an increase in average negative sentiment strength on at least one of the three measures, three recorded a decrease in average negative sentiment strength in two out of three measures. The case of Sandra Bullock is probably due to her attending a ceremony for the worst movie (the Razzies) just before the Oscars. This adds a second event to the main Oscars event, and one with a natural negative sentiment attached. Hence the main event is undermined by the second event (Figure 1). Infidelities by her husband were also revealed at this time, further complicating the data. The Austin plane crash is another case because the most reported crash was preceded by another two in the days beforehand. In contrast, Kathryn Bigelow was a big winner at the Oscars and seems to have escaped criticism (e.g., "Boring show but wow...Kathryn Bigelow made history!"), which illustrates that some positive events may have little or no negative backlash.

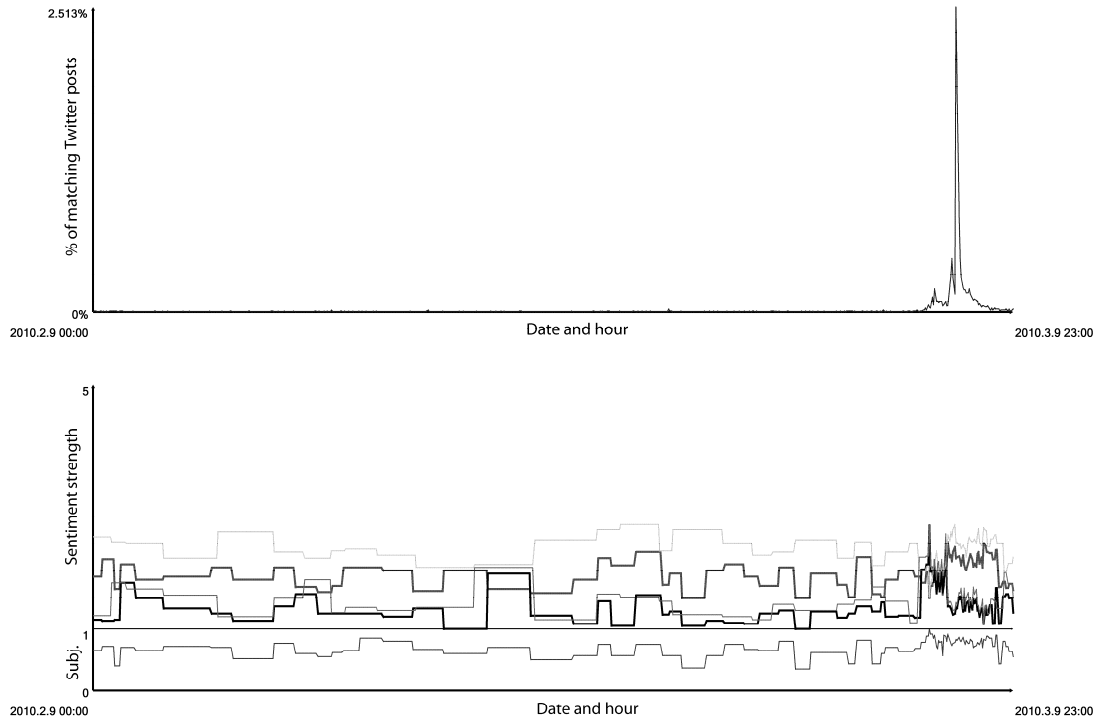

Figure 1. Twitter volume (top) and sentiment (bottom) time series for *sandra AND bullock*. In the lower graphs, the lowest line is the proportion of subjective texts. The thick black line is the average negative sentiment strength and the thick grey line is the average positive sentiment strength. The thinner lines are the same but just for the subjective texts (i.e., for which either positive or negative sentiment > 1). The sentiment data is bucketed into a minimum of 20 data points for smoothing – hence the city skyline appearance during periods with few matches. Note the two close events and the clear increase in negative sentiment during the peak.

Three events were associated with decreases in positive sentiment strength for all three measures. For Bill Clinton's heart attack and the killer whale attacking a trainer this seems to be natural. For HTC releasing a Google Android phone that was seen as directly competing with the popular Apple iPhone, this may be due to multiple events because a second event during the data occurred, Apple suing HTC over an iPhone patent. A more likely cause is the frequent topic of the HTC Hero mobile phone in other HTC-related posts, with *hero* being a positive sentiment term. This apparent decrease in positive sentiment is therefore probably an artefact of the term-based sentiment detection method.

A surprising pattern that the above analysis did not reveal is that the overall sentiment level was quite low in most cases. For instance, Figure 2 reports the Tiger Woods results surrounding his announcement that had a dramatic negative impact on his life and career and public perception of him, yet the overall negative sentiment strength is quite weak. The difference between negative sentiment strength in high and low volume times (0.226) is only 6% of the full range of sentiment strengths (i.e., 4) and the median difference for all 30 events is only 1%. To investigate why the sentiment strengths were not higher for Tiger Woods, as a case study, the first author conducted an informal content analysis on a random sample of 100 tweets from the peak hour. The types of tweet found were (together with a relevant tweet section, modified for anonymity): *humour* (21%, e.g., “show the 4 girls waiting for Tiger backstage!”); analysis (18%, e.g., “we will see which media has financial stakes in Tiger inc”); *cynicism* (13%, e.g., “statement was too monotone and coached”); *opinion or advice* (13%, e.g., “you shamed yourself and your family #idiotalert”, “meditate and find peace”); *information* (12%, e.g., “Read a recap of Tiger's apology here”); *sympathy* (12%, e.g., “Hard living in the pubic eye”); *uninterested* (11%, e.g., “stopped caring about this”). Clearly few tweets express an opinion about the event (less than 13%) and so it is unsurprising that there

was only a small increase in negative sentiment. The varied responses to the event are suggestive of the importance of the affordances perspective introduced above: few tweets seem to be simple reactions to this event and the majority seem to be using it to satisfy wider goals.

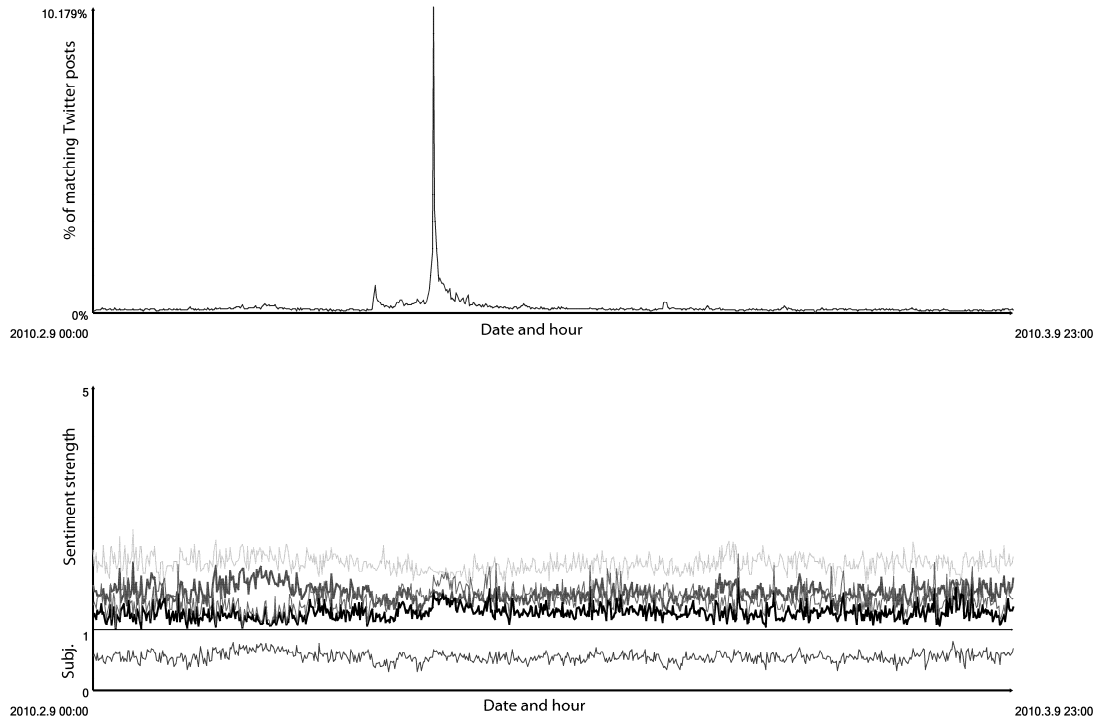

Figure 2. Twitter volume (top) and sentiment (bottom) time series for *tiger wood*. See the Figure 1 caption for additional information.

Finally, Figure 3 shows a classic event in the sense of a clear increase in negative sentiment strength around the announcement of Rapper Guru's coma. It is not surprising that the three events associated with the death or injury to well known people (also including Bill Clinton and Alexander McQueen) resulted in large increases in negative sentiment. Nevertheless there is not necessarily a decrease in positive sentiment strength in such negative events because of sympathy messages like, "hope GURU gets better soon. Peace." This illustrates that common sense ideas about likely changes in sentiment around particular types of events are not always correct and that increases in both sentiment polarities can potentially associate with events that are clearly positive or negative overall.

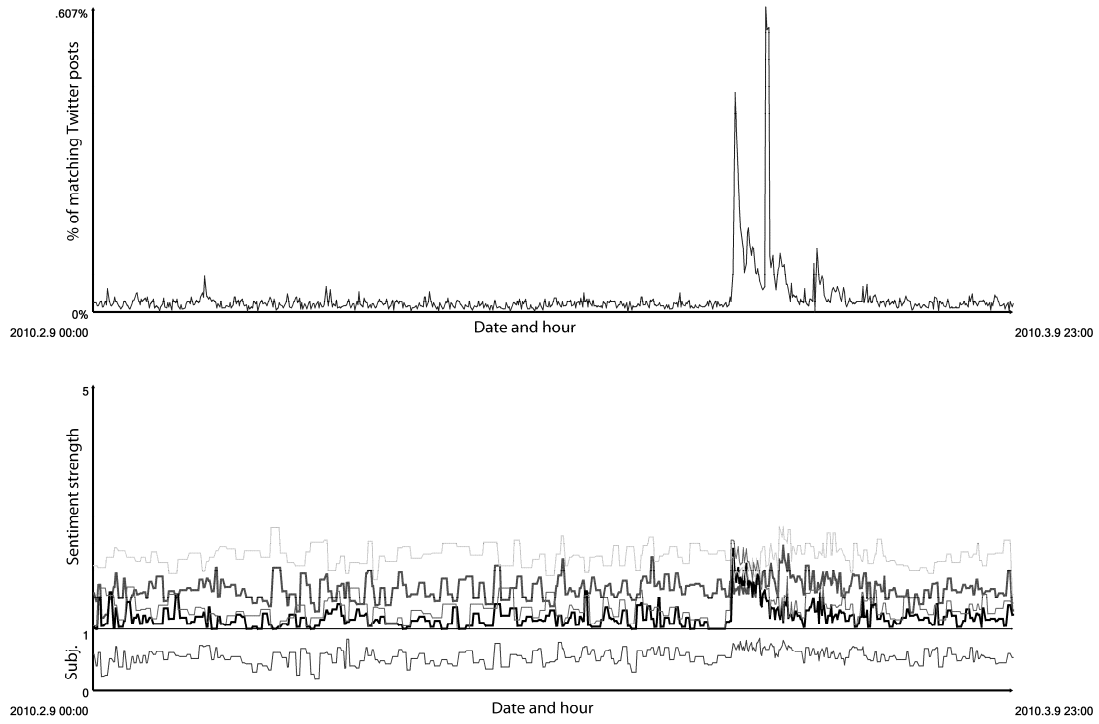

Figure 3. Twitter volume (top) and sentiment (bottom) time series for *guru*. See the Figure 1 caption for additional information.

Conclusions

The analysis of sentiment in the 30 largest spiking events in Twitter posts over a month gives strong evidence that important events in Twitter are associated with increases in average negative sentiment strength (H1n-H3n). Whilst there are exceptions and the hypotheses have only been tested on one time period, it seems that negative sentiment is often the key to popular events in Twitter. There is some evidence that the same is true for positive sentiment (H3p), but negative sentiment seems to be more central. Nevertheless, the overall level of sentiment in Twitter seems to be typically quite low and so the importance of sentiment should not be exaggerated.

The additional investigation did not find significant results for day-based statistics, suggesting that the fairly small changes in sentiment typically associated with significant events are hidden when averaged over days rather than hours. This underlines the fragile nature of the changes in average sentiment strength found.

From the perspective of analysing important events in Twitter, it is clear that it should be normal for such events to be associated with rises in negative sentiment. Even positive events should normally experience a rise in average negative sentiment strength as a reaction, although they would probably also experience stronger average positive sentiment. Despite this, it does not seem that more important events can be *identified* by the strength of emotion expressed. For example, the Bill Clinton topic had some of the largest increases in negative sentiment but was ranked only 30th for importance (as judged by Twitter spike size).

Despite the statistically significant findings, perhaps the most surprising aspect of the investigation was that the average changes in sentiment strength around popular events were small (typically only 1%) and were far from universal. Intuitively, it seems that events that are important enough to trigger a wave of Twitter usage would almost always be associated with some kind of emotional reaction. Even when reporting simple facts, such as the launch of important new products, it seems reasonable to expect a degree of excitement or at least scepticism. An informal content analysis of the Tiger Woods case revealed that only a minority of Tweets (under 13%) expressed a personal opinion, with the remainder exploiting the event for humour, expressing sympathy, cynicism or disinterest, analysing the event or giving information. This suggests that Twitter use is best conceived not as a reaction to

external events but as exploiting the affordances of these events for pre-existing personal goals, such as generating humour or applying analytical skills. This is the main theoretical implication of this study.

Finally, there are some practical implications of this research. The knowledge that big events typically associate with small increases in negative sentiment and sometimes with positive sentiment should aid Twitter Spam detection (e.g., Wang, 2010). For instance, crude attempts to Spam Twitter to promote an event may well result in artificially large increases in positive sentiment strength for the event. A non-constructive but useful practical implication is that it is unlikely that a system could be designed to detect major events from increases in sentiment strength rather than volume because the increases in sentiment strength are relatively small.

Acknowledgements

This work was supported by a European Union grant by the 7th Framework Programme, Theme 3: Science of complex systems for socially intelligent ICT. It is part of the CyberEmotions project (contract 231323).

References

Allan, J., Papka, R., & Lavrenko, V. (1998). On-line new event detection and tracking In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 37-45). New York, NY: ACM Press.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2007). Show me the money!: Deriving the pricing power of product features by mining consumer reviews. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 56-65). New York, NY: ACM Press.

Balog, K., Mishne, G., & Rijke, M. d. (2006). Why are they excited? Identifying and explaining spikes in blog mood levels. *11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, Retrieved July 8, 2010 from: http://staff.science.uva.nl/~mdr/Publications/Files/eacl2006-moodsignals.pdf.

Bansal, N., & Koudas, N. (2007). BlogScope: a system for online analysis of high volume text streams. In *Proceedings of the 33rd international conference on Very large data bases* (pp. 1410-1413). New York, NY: ACM Press.

Bassett, E. H., & O'Riordan, K. (2002). Ethics of Internet research: Contesting the human subjects research model *Ethics and Information Technology, 4*(3), 233-247.

Bifet, A., & Frank, E. (2010). Sentiment knowledge discovery in Twitter streaming data. In *Proc 13th International Conference on Discovery Science* (pp. 1-15). Berlin: Springer.

Blumler, J. G., & Katz, E. (1974). *The uses of mass communications: Current perspectives on gratifications research*. Beverly Hills, CA: Sage.

Bollen, J., Pepe, A., & Mao, H. (2009). Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena. *arXiv.org*, arXiv:0911.1583v0911 [cs.CY] 0919 Nov 2009.

boyd, d., Golder, S., & Lotan, G. (2009). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *Proceedings of HICSS-43*, Retrieved November 12, 2009 from: http://www.danah.org/papers/TweetTweetRetweet.pdf.

Bruza, P., & Weeber, M. (2008). *Literature-based discovery*. Berlin: Springer.

Case, D. O. (2002). *Looking for information: A survey of research on information seeking, needs, and behavior*. San Diego, CA: Academic Press.

Cataldi, M., Caro, L. D., & Schifanella, C. (2010). Emerging topic detection on Twitter based on temporal and social terms evaluation In *Proceedings of the Tenth International Workshop on Multimedia Data Mining table of contents* (pp. A4). New York, NY: ACM Press.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. In *Proceedings of the International AAAI*

Conference on Weblogs and Social Media (pp. Retrieved August 9, 2010 from: http://an.kaist.ac.kr/~mycha/docs/icwsm2010_cha.pdf).

Choudhury, M. D., Sundaram, H., John, A., & Seligmann, D. D. (2008). Can blog communication dynamics be correlated with stock market activity? In *Proceedings of the Nineteenth ACM conference on Hypertext and Hypermedia* (pp. 55-60). New York, NY: ACM Press.

Culotta, A. (2010). Detecting influenza outbreaks by analyzing Twitter messages. *arxiv.org*, Retrieved September 1, 2010 from: http://arxiv.org/PS_cache/arxiv/pdf/1007/1007.4748v2011.pdf.

Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 1195-1198). New York, NY: ACM Press.

Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies, 11*(4), 441-456.

Efron, M. (2010). Hashtag retrieval in a microblogging environment. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 787-788). New York, NY: ACM Press.

Enyon, R., Schroeder, R., & Fry, J. (2009). New techniques in online research: Challenges for research ethics. *21st Century Society, 4*(2), 187-199.

Fox, E. (2008). *Emotion science*. Basingstoke: Palgrave Macmillan.

Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. *Lecture Notes in Computer Science, 3646*, 121-132.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 79-84). New York: ACM.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 62-82). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum.

Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. *Fourth International AAAI Conference on Weblogs and Social Media*, Retrieved August 9, 2010 from: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM2010/paper/view/1513/1833.

Gruhl, D., Chavet, L., Gibson, D., Meyer, J., & Pattanayak, P. (2004). How to build a WebFountain: An architecture for very large-scale text analytics. *IBM Systems Journal, 43*(1), 64-77.

Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). The predictive power of online chatter. In R. L. Grossman, R. Bayardo, K. Bennett & J. Vaidya (Eds.), *KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 78-87). New York, NY, USA: ACM Press.

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through Blogspace, *WWW2004*. New York, Retrieved July 5, 2010 from: http://www.tackelberz.de/wp-content/uploads/2009/10/5-gruhl.pdf.

Guralnik, V., & Srivastava, J. (1999). Event detection from time series data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 33-42). New York, NY: ACM Press.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: University Press

Honeycutt, C., & Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. In *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1-10). Los Alamitos, CA: IEEE.

Hookway, N. (2008). Entering the 'blogosphere': some strategies for using blogs in social research. *Qualitative Research, 8*(1), 91-113.

Huang, J., Thornton, K. M., & Efthimiadis, E. N. (2010). Conversational tagging in twitter. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia* (pp. 173-178). New York, NY: ACM Press.

Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management, 6*(3-4), 248-260.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science & Technology, 60*(11), 2169-2188.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). New York, NY: ACM Press.

Johnson, S. (2009). How Twitter will change the way we live. *Time*, Retrieved August 9, 2010 from: http://www.time.com/time/business/article/2010,8599,1902604,1902600.html.

Katz, E. (1959). Mass communication research and the study of popular culture. *Studies in Public Communication, 2*(1), 1-6.

Kerbel, M. (2000). *If it bleeds, it leads An anatomy of television news*. Boulder, CO: Westview Press.

Kinsinger, E. A., & Schacter, D. L. (2008). Memory and emotion. In M. Lewis, J. A. Haviland-Jones & L. Feldman Barrett (Eds.), *Handbook of emotions (3 ed)* (pp. 601-617). New York, NY: The Guildford Press.

Kramer, A. D. I. (2010). An unobtrusive behavioral model of "Gross National Happiness". *Proceedings of CHI 2010*, 287-290.

Krishnamurthy, B., Gill, P., & Arlitt, M. (2008). A few chirps about twitter. In *Proceedings of the first workshop on Online social networks table of contents:* (pp. 19-24). New York, NY: ACM Press.

Kumar, R., Novak, J., Raghavan, P., & Tomkins, A. (2003). On the bursty evolution of blogspace. In *WWW2003*. Budapest, Hungary, http://www2003.org/cdrom/papers/refereed/p477/p477-kumar/p477-kumar.htm.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on world wide web* (pp. 591-600). New York, NY: ACM Press.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). Social media & mobile internet use among teens and young adults. *Pew Internet & American Life Project*, Retrieved February 5, 2010 from: http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults.aspx.

MacLean, A., Carter, K., Lövstrand, L., & Moran, T. (1990). User-tailorable systems: pressing the issues with buttons. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 175-182). New York, NY: ACM Press.

Mishne, G., & Glance, N. (2006). Predicting movie sales from blogger sentiment. *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, Retrieved August 8, 2010 from: http://www.aaai.org/Papers/Symposia/Spring/2006/SS-2006-2003/SS2006-2003-2030.pdf.

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. New York, NY: ACM Press.

Nahl, D. (2006). Affective load. In K. E. Fisher, S. Erdelez & L. E. F. McKenchnie (Eds.), *Theories of information behavior* (pp. 39-43). Medford, NJ: Information Today.

Nahl, D. (2007a). The centrality of the affective in information behavior. In D. Nahl & D. Bilal (Eds.), *Information and emotion: The emergent affective paradigm in information behavior* (pp. 3-37). Medford, NJ: Information Today.

Nahl, D. (2007b). Domain interaction discourse analysis: A technique for charting the flow of micro-information behavior. *Journal of Documentation, 63*(3), 323-339.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2007). Textual affect sensing for sociable and expressive online communication. *Lecture Notes in Computer Science, 4738*, 218-229.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (pp. 122-129). Menlo Park, CA: The AAAI Press

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of LREC 2010*, Retrieved October 5, 2010 from: http://www.lrec-conf.org/proceedings/lrec2010/pdf/2385_Paper.pdf.

Pang, B., & Lee, L. (2004). Sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of ACL 2004* (pp. 271-278). New York: ACL Press.

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd Annual Meeting of the ACL*, 115-124.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 1*(1-2), 1-135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 79-86). Morristown, NJ: Association for Computational Linguistics.

Perse, E. M. (1990). Involvement with local television news: Cognitive and emotional dimensions. *Human Communication Research, 16*(4), 556-581.

Porter, M. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130-137.

Riloff, E., Patwardhan, S., & Wiebe, J. (2006). Feature subsumption for opinion analysis. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 440-448.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*. New York, NY: ACM Press.

Savage, M., & Burrows, R. (2007). The coming crisis in empirical sociology. *Sociology, 41*(5), 885-899.

Seaton, J. (2005). *Carnage and the media: The making and breaking of news about violence*. London: Allen Lane.

Stafford, T. E., Stafford, M. R., & Schkade, L. L. (2004). Determining uses and gratifications for the Internet. *Decision Sciences, 35*(2), 259-288.

Strapparava, C., & Mihalcea, R. (2008). Learning to identify emotions in text, *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 1556-1560). New York, NY: ACM.

Swanson, D. R., Smalheiser, N. R., & Bookstein, A. (2001). Information discovery from complementary literatures: Categorizing viruses as potential weapons. *Journal of the American Society for Information Science and Technology, 52*(10), 797-812.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (in press). Lexicon-based methods for sentiment analysis. *Computational Linguistics*.

Tenopir, C., Nahl-Jakobovits, D., & Howard, D. L. (1991). Strategies and assessments online: Novices' experience. *Library and Information Science Research, 13*(3), 237-266.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (in press). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*.

Thelwall, M., & Prabowo, R. (2007). Identifying and characterising public science-related concerns from RSS feeds. *Journal of the American Society for Information Science & Technology, 58*(3), 379-390.

Thelwall, M., Prabowo, R., & Fairclough, R. (2006). Are raw RSS feeds suitable for broad issue scanning? A science concern case study. *Journal of the American Society for Information Science and Technology, 57*(12), 1644-1654.

Thelwall, M., Wouters, P., & Fry, J. (2008). Information-centred research for large-scale analysis of new information sources. *Journal of the American Society for Information Science and Technology, 59*(9), 1523-1527.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (pp. 178-185). Menlo Park, CA: The AAAI Press.

Wang, A. H. (2010). Don't follow me: Spam detection in Twitter. In *Proceedings of the International Conference on Security and Cryptography (SECRYPT)* (pp. Retrieved October 5, 2010 from: http://ceas.cc/2010/papers/Paper%2021.pdf).

White, M. (2002). Representations or people? *Ethics and Information Technology, 4*(3), 249-266.

Wigand, F. D. L. (2010). Twitter in government: building relationships one Tweet at a time. In *Proceedings of the Seventh International Conference on Information Technology* (pp. 563-567). Los Alamitos, CA: IEEE.

Wilson, T., Wiebe, J., & Hoffman, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics, 35*(3), 399-433.

Wilson, T., Wiebe, J., & Hwa, R. (2006). Recognizing strong and weak opinion clauses. *Computational Intelligence, 22*(2), 73-99.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann.