Executive Summary

Tweeting the Terror: Modelling the Social Media Reaction to the Woolwich Terrorist Attack

Forthcoming in

Social Network Analysis and Mining

Pete Burnap*[1], Matthew L. Williams[2,] Luke Sloan[2], Omer Rana[1], William Housley[2], Adam Edwards[2], Vincent Knight[3], Jeffrey Morgan[2], Rob Procter[4] & Alex Voss[5]

[1]School of Computer Science & Informatics, Cardiff University, UK

[2]School of Social Sciences, Cardiff University, UK

[3]School of Mathematics, Cardiff University, UK

[4]Department of Computer Science, University of Warwick

[5] Department of Computer Science, University of St. Andrews

ABSTRACT

Little is currently known about the factors that promote the propagation of information in online social networks following terrorist events. In this paper we took the case of the terrorist event in Woolwich, London in 2013 and built models to predict information flow *size* and *survival* using data derived from the popular social networking site Twitter. We define information flows as the propagation over time of information posted to Twitter via the action of retweeting. Following a comparison with different predictive methods, and due to the distribution exhibited by our dependent *size* measure, we used the zero-truncated negative binomial (ZTNB) regression method. To model *survival*, the Cox regression technique was used because it estimates proportional hazard rates for independent measures. Following a principal component analysis to reduce the dimensionality of the data, social, temporal and content factors of the tweet were used as predictors in both models. Given the likely emotive reaction caused by the event, we emphasize the influence of emotive content on propagation in the discussion section. From a sample of Twitter data collected following the event (N=427,330) we report novel findings that identify that the sentiment expressed in the tweet is statistically significantly predictive of both size and survival of information flows of this nature. Furthermore, the number of offline press reports relating to the event published on the day the tweet was made was a significant predictor of size, as was the tension expressed in a tweet in relation to survival. Finally, time lags between retweets and the co-occurrence of URLS and Hashtags also emerged as significant.

Key Findings:

Executive Summary

- Our first hypothesis was that tweets containing negative sentiment and high levels of racial tension would produce large and long-living information flows following the terrorist event. Contrary to this hypothesis tweets containing positive sentiment had an 11 percent higher chance of being retweeted after the event compared to tweets containing negative sentiment.
- Results suggest there was no significant association between tension and size of information flow, however significant associations did emerge in relation to duration of information flow. Contrary to the hypothesis, tweets containing high racial tension had shorter survival rates than those with lower levels of racial tension. Mirroring the sentiment finding, this suggests Twitter users were less likely to prolong the information flow where the content of a tweet was antagonistic.
- Our second hypothesis was that the number of newspaper stories relating to the event would influence both the size and survival of information flows on Twitter. The number of related newspaper stories published on the day a tweet was posted had a significant positive association with size of information flow. We found an increase in 100 news headlines about the event increased the rate of retweets by a factor of 5 percent. This provides evidence to suggest that people react to stories published in offline media through the mechanism of online social media in relation to this kind of event.
- Our third hypothesis was that temporal factors would influence both the size and survival of information flows surrounding the event. The results suggested that by the fifth time a tweet is retweeted, the timelag is highly significant and negatively associated with size of the potential information flow. For every additional second time lapse between the original tweet and the fifth retweet, the retweet rate reduces by 1 percent. This suggests the quicker a tweet is retweeted the larger it will become. Furthermore, time elapsed between the original tweet the fifth retweet was highly significant and positively associated with survival rate, indicating the slower a tweet is retweeted, the longer it will survive. Or put another way, tweets that receive their first five retweets faster also die out quicker following a burst of sudden interest. The association between rapid propagation of information and size could be explained through the theory of the issue attention cycle (Downs 1972), where the prominence of interest (i.e. large number of retweets) is a factor of enthusiasm in a current issue. This rapid burst of interest in a particular tweet could explain the size. Similarly, using the same theory, following a burst of enthusiasm, comes a rapid decline in interest, and a rapid decay in information propagation. Therefore, tweets that receive a large amount of interest early in their lifetime tend to survive for less time as interest wanes.
- Our fourth hypothesis was that social factors such as number of followers, potential reach of the tweet, and number of previous tweets will influence both the size and survival of information flows surrounding the event. We found that, for size, the number of followers of the tweeter was positively associated and highly significant. An increment of 100,000 followers increased the rate of retweets by a factor of 10 percent. Interestingly, the reach after the fifth retweet was not significantly correlated with size, while number of followers was, suggesting that the number of followers of the original tweeter is more significant as a predictor

than the number of people it might have reached upon propagation. Neither were significant for survival. In relation to number of previous tweets, there was a significant negative association with both size and survival. This new finding possibly suggests that by increasing the total number of tweets the user increases the freshness of their tweet-stream, decreasing the relevance of older information and possibly curtailing its propagation and survival in favour of the new content. Another possible explanation is that during these kinds of events, new information becomes available very frequently and given the user's propensity to tweet a lot, it is likely they could be relaying new information as it emerges.

- Finally we hypothesized that the presence, frequency and co-occurrence of content linking features such as Hashtags and URLs within the tweet would influence both the size and survival of information flows surrounding the event. Hashtag and URL frequency were negatively associated with size, with an increase of Hashtag and URL instances decreasing the rate of retweets by 15 percent and 35 percent respectively. This could be because including a number of Hashtags and URLs reduces the space available in the 140-character post for the inclusion of information that may be of interest and subsequently retweeted. However, of novel importance here is the strong positive statistical association between the co-occurrence of URLs and Hashtags in a tweet, and the information flow size and survival. The presence of a URL and a Hashtag increased the rate of retweets by a factor of 78 percent. This may suggest that tweets that are identified as relating to the event via the use of a Hashtag, and provide additional sources of (possibly evidential) information via a URL, could become large and long-lasting information flows due to their discoverability and enriched content.
- This study has provided evidence that the opinion/emotional factors of tweets are statistically significant predictors of both information flow size and survival following a terrorist event. Contrary to popular perception, while negative and racially tense content did emerge on Twitter following the event, it failed to propagate via retweeting. In short, negative and racially tense content remained small in number and lasted for short periods, dying out quickly. This indicates that on the whole, members of the public who used Twitter in this case chose to propagate positive and supporting content. Following on from this work we intend to test the predictive efficacy of our models on other cases that exhibit similar characteristics (e.g. the Boston terrorist event and the 2011 riots in England). The outcome of this ongoing research will be a general predictive model for Twitter information flow size and survival following major socially disruptive events.