

18th July 2014

Social Media Analytics are Evolving - From Twitter-based Crisis Mapping to Large-scale Real-Time Situation Assessment with Trust and Credibility Analysis

In recent years there has been a growing trend for the use of publically available social media content (e.g. Twitter, YouTube, Facebook, Instagram) for analytics in areas such as journalism, crisis management, political analysis and business intelligence. With social media content freely available and updated in real-time it is no wonder analysts are turning to it to discover trending topics, buzzing locations, emerging visionaries and so much more.

In politics social media sites such as Facebook and Twitter are used to track [5] [8] sentiment (i.e. positive or negative) in real-time during election campaigns, analysing major speeches, political debates and policy announcements as they happen [12]. For crisis management Twitter has been used to plot accurate real-time crisis maps [7] of incident reports during natural disasters such as hurricanes and tornados. Businesses are also making good use of social media [6] [11] to track public response during product launches, plan new store locations and target influential people [3] within market sectors.

Current analytic toolsets for social media can be broadly categorized as dashboards, crisis mapping toolkits and in-depth analysis frameworks. Dashboard applications (e.g. Tweetdeck, Sulia, Storyful, Flumes, WebLyzard) allow people such as Journalists to track news stories, alerting them to new and relevant content, trending topics and influential people. These dashboard applications allow click-through from content to author, providing contact details for a subsequent manual verification process (e.g. verification of content via a phone call). Tools for crisis mapping also follow a manual verification process (e.g. Ushahidi), using teams of volunteers to verify web and social media incident reports before adding them to a crisis map for display. For in depth analysis there are tools supporting sentiment analysis (e.g. Bing Elections, SocialMention), social network graph visualization (e.g. MentionMapp, Bottlenose) and topic tracking (e.g. Trackur).

Three challenges key to social media analytics are (a) scaling up approaches (i.e. big data techniques) to handle the content volumes we see today, (b) augmenting and improving the existing manual process for social media content verification, ultimately improving both trust and credibility in the final analysis and (c) increasing support for multi-lingual analytics.

The trust and credibility challenge is particularly important as fake news stories are becoming both increasingly common and increasingly difficult to spot. Examples of recent fake news (see figure 1) include sharks swimming in New York after the 2012 Hurricane Sandy [4], images of a Tsunami crashing into the statue of liberty and a 'FIFA corruption' tweet [1] sent out 'before' the 2014 world cup final with correct details of who won and scored in the game.

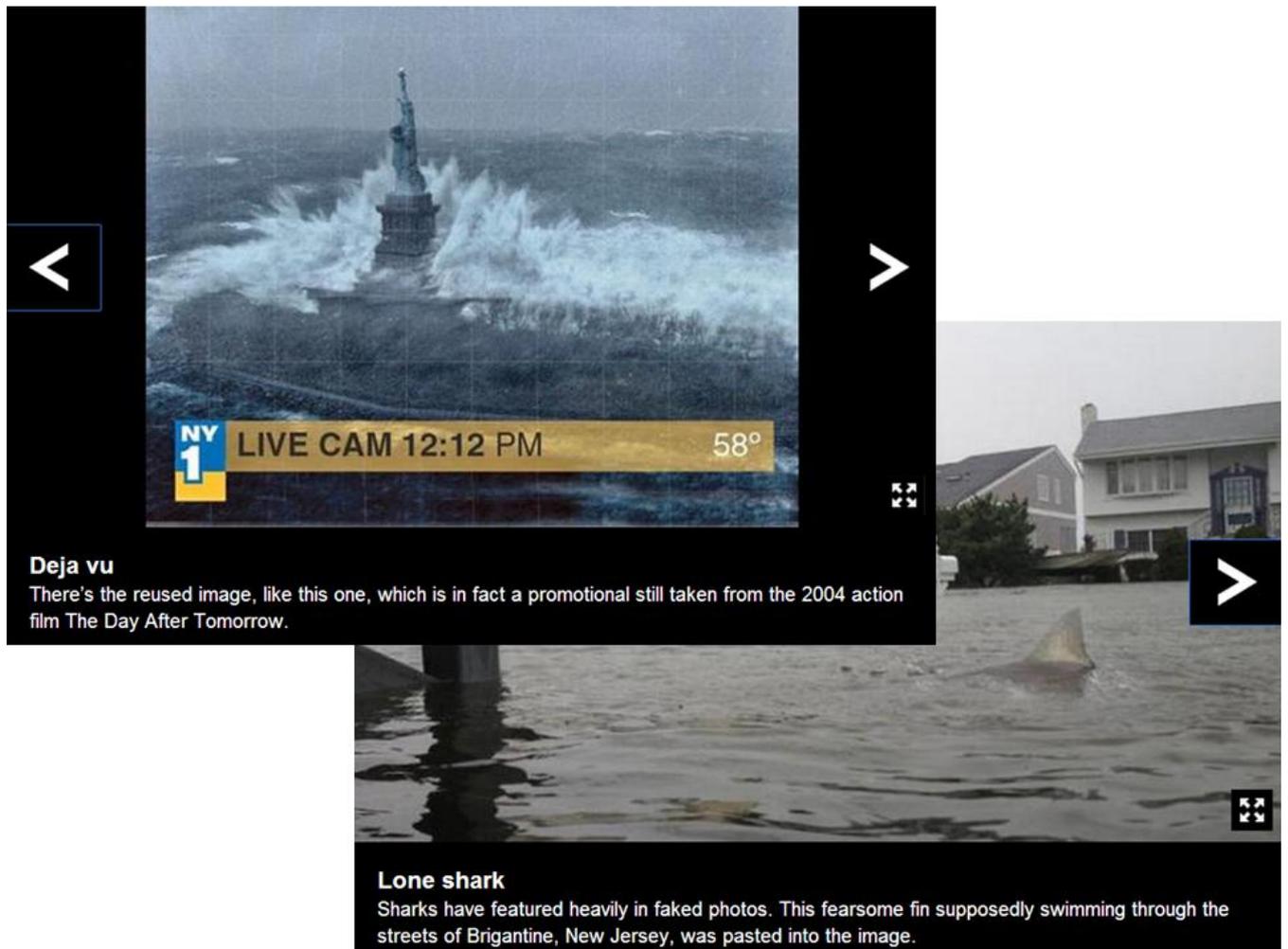


Figure 1. Examples [4] of fake news stories during Hurricane Sandy 2012, based on doctored images, which were widely propagated across social media. Source - BBC News - © 2012 BBC

Crisis Mapping of Natural Disasters

One example showing how social media analytics are evolving can be seen in the crisis mapping platform [7] developed by the TRIDEC project [10]. In TRIDEC social media analytics were used to allow Tsunami early warning centres to monitor Twitter for reports within 'at risk' coastal areas near known geological fault lines which have the potential to cause a Tsunami. Real-time monitoring is important as early wave impact assessments can be used to warn people on coastlines further away, allowing them to get to safety. Such crisis maps can help inform the general public during a crisis and can be used by emergency response teams to complement official information streams such as hardware-based sensor reports including satellite imaging and aerial photography.

Streams of twitter messages were geo-parsed in real-time and mentions of known locations detected. These sets of real-time location mentions were then aggregated, statistically analysed and plotted on a real-time crisis map. This approach produces high precision crisis maps as can be seen in figures 2 and 3. In the REVEAL project [9] these techniques are evolving, scaling up to run on a cluster of computers in a data centre using the Storm [2] distributed computation framework. This means large numbers of locations to be monitored simultaneously such as the entire coastline of a country.

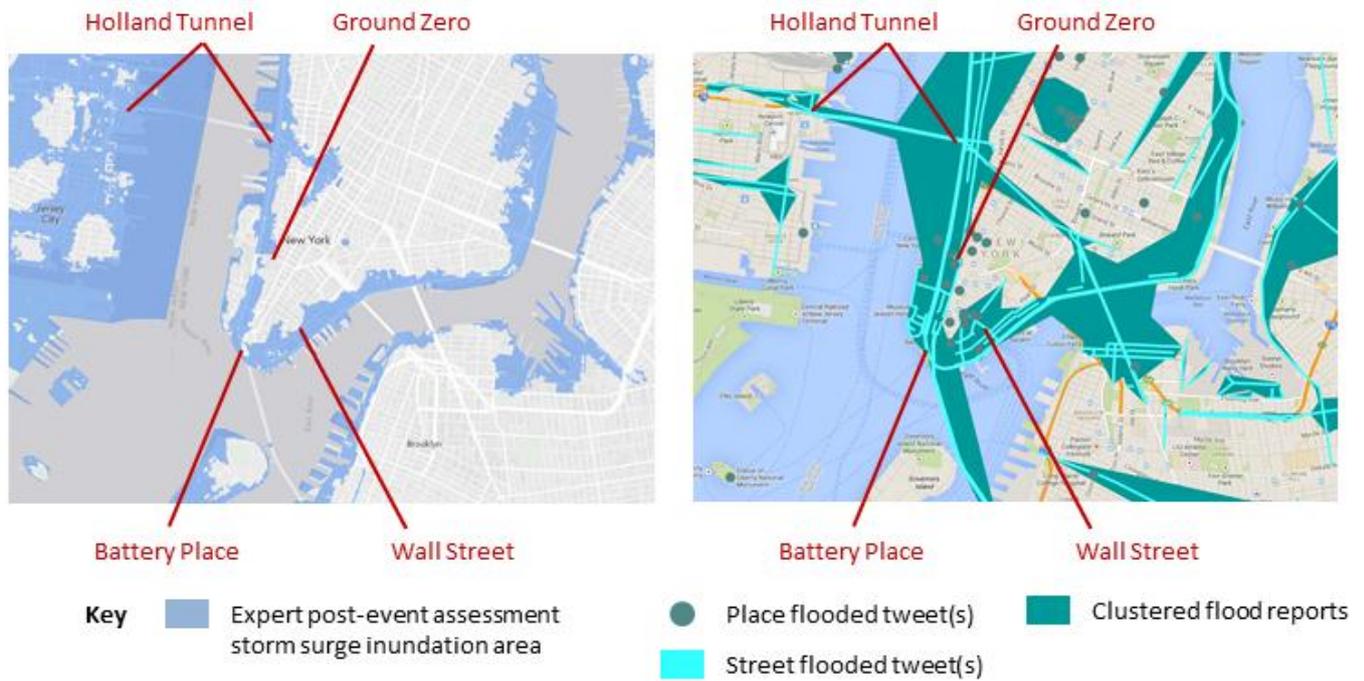


Figure 2. Crisis mapping of New York's 2012 flooding from Hurricane Sandy. The left hand image is the official post-event Storm surge assessment map from the US National Geospatial Agency. The right hand image is the incident map TRIDEC was able to create from tweets crawled during the event. Mapping courtesy of ArcGIS ESRI portal and Google Maps.

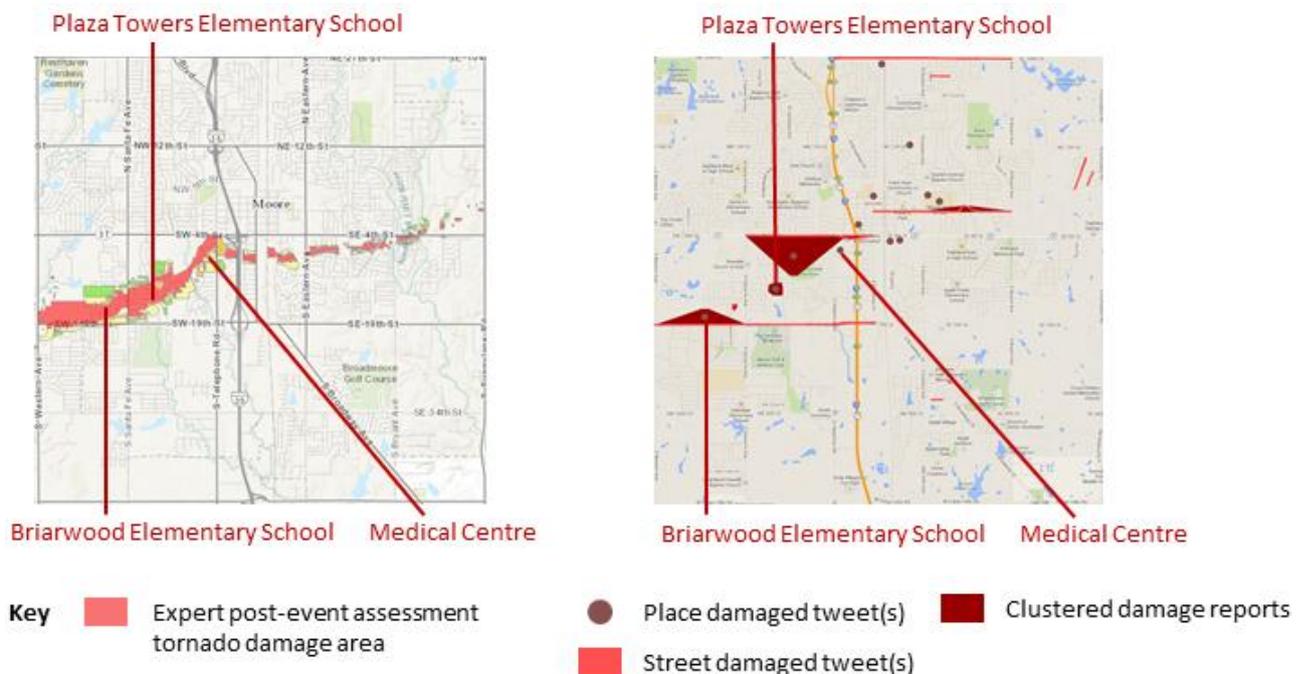


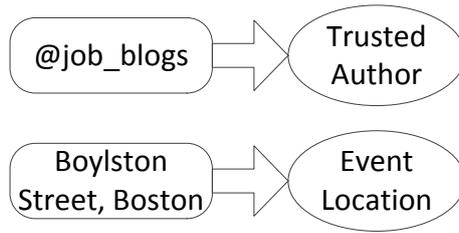
Figure 3. Crisis mapping of Oklahoma's 2013 tornado. The left hand image is the official post-event damage assessment map from the US National Geospatial Agency. The right hand image is the incident map TRIDEC was able to create from tweets crawled during the event. Mapping courtesy of ArcGIS ESRI portal and Google Maps.

Large-scale Situation Assessment with Trust and Credibility Analysis

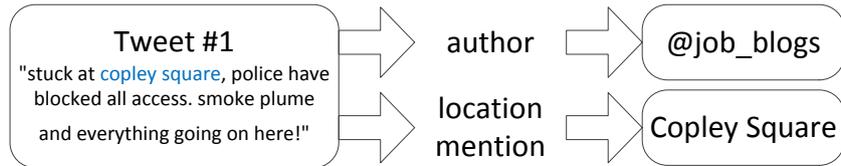
The REVEAL project is adding support for journalists and business analysts. In addition to real-time mapping visualizations, situation assessments will show trending topics (e.g. topics, hashtags, URI's) and emerging social networks (e.g. influential people, well connected people). These situation assessments will be personalized to the news story and/or business need of each analyst, providing a real-time interactive view across a range of social media sources (e.g. Twitter Facebook, Instagram, YouTube).

Analysts need more than just clever visualization from a situation assessment of course. These techniques are evolving to support an evidential approach for assessing the trust and credibility of large volumes of social media content. This knowledge-based approach allows automated classification of existing evidence, useful to provide a variety of filtered viewpoints. It also allows prior knowledge from the analyst to be utilized, such as lists of trusted 'news hounds' that a journalist has learnt to trust from past experience. Inference can be made of new information resulting from logical deductions from the evidence at hand. Figure 4 shows examples of the types of inference possible.

Prior Knowledge : Event Boston Bombing, 2013



Evidence Available : Twitter



Types of Inference

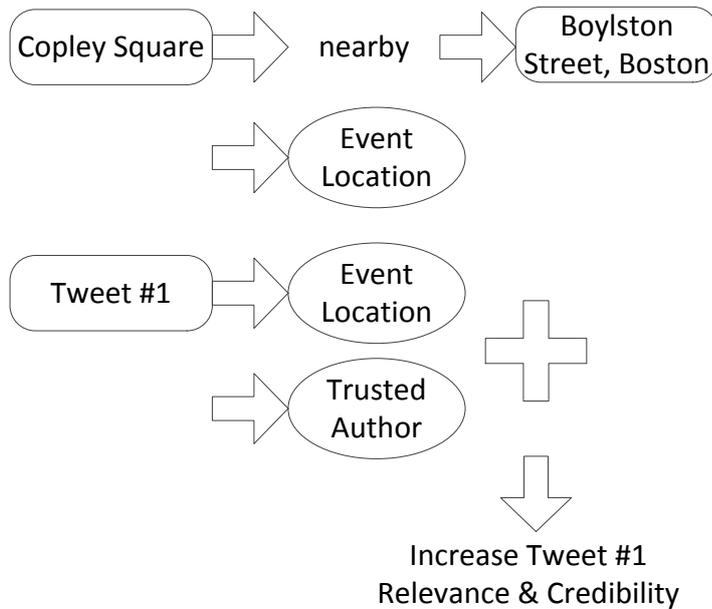


Figure 4. Example of evidential inference from the Boston bombing event. This simple conclusion is not very useful on its own, but becomes powerful when scaled up by combining with inference from 1000's of other facts extracted from social media reports. This type of technique could offer insights into trust and credibility not practical to achieve with manual approaches

Future Trends

The technology driving social media analytics is evolving. Applications are moving beyond dashboard type analytics toolkits and scaling things up. The volume of social media content that can be processed is increasing and there is an increasing focus on adding trust and credibility analysis to the situation assessment pictures compiled. This won't replace the need for manual verification, a phone call to the content author is always going to be an effective way to ascertain if someone is telling the

truth, but it will help analysts focus their attention on what matters in the sea of social media content out there.

Acknowledgement

The work presented in this article is part of the research and development in the REVEAL project (grant agreement 610928), supported by the 7th Framework Program of the European Commission.

About the author



Stuart E. Middleton is a senior research engineer at the University of Southampton IT Innovation Centre. His main research interests are social media, sensor systems, data fusion and ontologies. Stuart has a PhD in Computer Science from the University of Southampton.

Stuart E. Middleton, University of Southampton IT Innovation Centre, @stuart_e_middle

<http://www.it-innovation.soton.ac.uk> <http://users.ecs.soton.ac.uk/sem/>

REVEAL project, @RevealEU

<http://revealproject.eu/>

References

- [1] "'FIFA Corruption' Twitter account fraudulently claims to have predicted decisive moment in World Cup final", news.com.au, July 2014, <http://www.news.com.au/sport/football/fifa-corruption-twitter-account-fraudulently-claims-to-have-predicted-decisive-moment-in-world-cup-final/story-fnkjl6g2-1226987886361>
- [2] Apache Storm, Distributed and fault-tolerant real-time computation, APACHE incubator project, <https://storm.incubator.apache.org/>
- [3] Chishick, T. "the "Fry" effect", Blog, Agenda21, Mar 2011, <http://www.agenda21digital.com/the-fry-effect/>
- [4] Eveleth, R. "Hurricane Sandy: Five ways to spot a fake photograph", BBC Future, Oct 2012 <http://www.bbc.com/future/story/20121031-how-to-spot-a-fake-sandy-photo>
- [5] Feldman, R. "Techniques and Applications for Sentiment Analysis", Communications of the ACM, vol.56, issue 4, pp. 82-89, April 2013
- [6] Kent, P. "Viewpoint: Big data and big analytics means better business", BBC News, Oct 2012, <http://www.bbc.co.uk/news/business-19969588>
- [7] Middleton, Stuart E.; Middleton, Lee; Modafferi, Stefano, "Real-Time Crisis Mapping of Natural Disasters Using Social Media," Intelligent Systems, IEEE , vol.29, no.2, pp. 9-17, Mar.-Apr. 2014 <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6692841>

- [8] Moore, M.T. "Twitter index tracks sentiment on Obama, Romney", USA Today, Jan 2012, <http://usatoday30.usatoday.com/news/politics/story/2012-08-01/twitter-political-index/56649678/1>
- [9] REVEAL project, EC FP7 Grant Agreement 610928, <http://revealproject.eu/>
- [10] TRIDEC project, EC FP7 Grant Agreement 258723, <http://www.tridec-online.eu/home>
- [11] Wall, M. "Location tech and mobile map out way to better business", BBC News, June 2014, <http://www.bbc.co.uk/news/business-27781078>
- [12] WeGov project, EC FP7 Grant Agreement 248512, <http://www.wegov-project.eu/>