Crowds, Crisis, and Convergence: Crowdsourcing in the Context of Disasters

Kate Starbird
University of Washington

The Social Media Surge during Hurricane Sandy

On October 29, 2012, Hurricane Sandy slammed into the U.S. eastern seaboard, becoming one of the deadliest and costliest storms in U.S. history. “Superstorm Sandy” caused 72 direct fatalities in the U.S. and tens of billions of dollars of damage, due in large part to a catastrophic storm surge that flooded hundreds of thousands of homes and businesses (Blake et al., 2013). The aftermath of the storm brought major disruptions to transportation systems, long-term power outages, and gas rationing in parts of New York and New Jersey.

Like other disaster events in recent years, Sandy precipitated a huge surge in social media use. Twitter reported that they had hosted more than 20 million tweets with search terms related to the event during a six-day window around the U.S. impact. Instagram, a popular photo-sharing site, announced that users posted more than ten photos per second as Sandy came ashore.

Existing research suggests a large portion of this content would have come from users outside affected areas, and that much of it would have been “derivative”—i.e. re-posted and re-mixed content that echoes through social media (Starbird et al., 2010). However, these platforms did facilitate information-sharing behaviors that impacted response efforts. Residents of affected areas shared first-hand reports of actionable information, including photos of flooded streets, videos of trees falling and houses catching fire, and tweets reporting stranded people. Emergency responders turned to social media to broadcast storm warnings and to quell rumors. Problems with the propagation of misinformation drew widespread attention. One Twitter user reported, among other dubious claims, that the New York Stock Exchange had been flooded, and this misinformation spread rapidly before being called out by other Twitter users in what one writer called a “savage self-correction” (Herrman, 2012). Social media platforms and other online forums also played host to self-organized community response efforts and other forms of volunteerism, including the establishment of a Twitter hashtag to share information about open gas stations and a related project by a group of high school students to create and maintain a live “gas map”—an online map that displayed in real-time where gas was available.

The role of social media during Sandy’s lead-up, impact, and response generated considerable media attention, with some claiming the event marked a significant shift in the use of these services for emergency response, and at least one journalist suggesting (in personal communication) that Sandy was the first “social” disaster. Though we can appreciate the sentiment in the latter claim, it was one we had heard before, e.g. after the 2010 Haiti Earthquake and the 2011 Japan Tsunami. In fact, social media were already becoming an established feature of disaster events. And, disasters have always been inherently *social*.

Sociology of Disaster Meets Web 2.0

Sociologists of disaster have long known that people will “converge” onto the scene of disaster events (Fritz & Mathewson, 1957; Kendra & Wachtendorf, 2003). Fritz & Mathewson (1957) explained that though much of this convergence is physical, convergence can also be *informational*—i.e. people use available channels to share and seek information. Palen and colleagues (Hughes et al., 2008; Palen et al., 2010) connected this phenomenon to what now occurs online, whereby disaster events act as catalysts for massive “digital” convergence—of the kind that can generate 20 million tweets in six days.

This digital convergence carries considerable promise for improving disaster response. First-hand observations from citizen reporters on the ground of events have the potential to increase situational awareness for other affected people and for responders. Social media can also be used for formal crisis communications, and emergency responders are increasingly turning to these platforms for outgoing messaging during and between disaster events.

As the examples from Hurricane Sandy suggest, there remain several significant challenges in leveraging social media as a real-time information source. The first is volume. Clearly, it is difficult for an affected individual to make sense of tens of tweets and photos per second. Relatedly, if the focus is on finding actionable information coming from the ground of events, a vast majority of social media data can be considered *noise*—some portion is completely off-topic, and another large percentage contains repeated, *retweeted* or otherwise "derivative" information (Starbird et al, 2010). Another particularly vexing issue is the problem of lost context, whereby information loses the connection to its original author, time, or place. For example, a tweet sent at 4pm indicating a voluntary evacuation for a fire-affected neighborhood could become dangerous misinformation if reposted a few hours later, after the evacuation has become mandatory. Misinformation and intentional disinformation are major concerns. Additionally, the unstructured nature of social media content represents a challenge for those trying to make sense of it in aggregate form.

Automatically Filtering and Classifying the Flood of Data

Purely computational solutions for filtering and otherwise processing these social media streams show promise, but have some limitations. Though terms of service and protocols continually change, accessing social media data is often the easiest part of the problem, as many social media platforms provide Application Programming Interfaces (API) for collecting public data. Storing and searching these massive data sets presents a more complex challenge, one addressed in the broader conversations around dealing with "big data." Because the textual content of these social media streams is not quite the "natural language" for which traditional

Natural Language Processing (NLP) techniques have been designed and tested, new approaches for computational content analysis are needed. Additionally, accuracy is extremely important in time- and safety-critical environments like those of disaster response, and currently even the best automatic classification techniques along the more simple data dimensions (e.g. identifying situational awareness information) achieve only about 80% accuracy (Verma et al., 2011).

Harnessing the Power of the Crowd

Another solution for filtering the flood of data during disasters takes a *human computation* or *crowdsourcing* approach, using a large number of people, connected via the Internet, to manually process the data. In this research vein, we are very much following the crowd. During recent disaster events, members of the connected crowd have consistently appropriated social media platforms and other available tools to improvise response efforts, often in the form of informational assistance—e.g. the New Jersey students and their gas map. This new *digital volunteer* behavior aligns with another long-recognized disaster phenomenon, *spontaneous volunteerism*, whereby people make themselves available to help in various capacities, often by improvising to fill gaps in formal response efforts (Kendra & Wachtendorf, 2003).

During the 2009 Red River Floods, volunteer programmers created algorithms that automatically tweeted river heights at various locations (Starbird et al., 2010). After the Haiti earthquake, a group of self-named "voluntweeters" utilized the Twitter platform to help coordinate aid efforts, eventually connecting with each other to form an emergent organization (Starbird & Palen, 2011). In another highly publicized effort during that event, a group of students at Tufts University created and maintained a public map of humanitarian needs, translating and geo-locating thousands of reports arriving from Haitian people via an SMS short code (Meier & Munro, 2010). During the impact of Hurricane Irene in the Catskills in September 2011, a group of journalists performed as “crowdsourcerors,” organizing a community information-sharing and response effort through a combination of a Liveblog, Facebook, Twitter, and even radio broadcasts and phone calls from landline phones in more remote areas (Dailey & Starbird, in submission).

Though every event spawns new crowd-powered solutions to newly recognized needs, a number of ongoing, virtual volunteer organizations have been established, including the Standby Task Force, Humanity Road, Crisis Commons, and several Virtual Operation Support Teams connected to emergency responders. These groups use available online tools to respond to disaster events all over the world. However, questions remain regarding how they will sustain membership and how they can connect the products of their work and this new capacity more broadly to the established work practices of formal responders. One research opportunity lies in understanding the work of digital volunteers and designing to support this activity—for example, by developing crowdsourcing solutions that align with the motivations of disaster volunteers, i.e. initial altruism that soon becomes augmented by social and reputation “capital.”

Using the Noise to Find the Signal

Another approach seeks to leverage the collective behavior of the larger crowd to address information-processing challenges. Social media users, intentionally and not, work to shape the information space through their behavior within it. Instead of viewing crowd activity as simply noise, it is possible to consider every re-post, “like,” “follow,” and user mention as productive crowd work—and to use this “noise” to find the signal. For example, we can design algorithms to identify misinformation through features of crowd behavior—i.e. sensing the “savage self-correction” of dozens of voices publicly questioning false information. We may also be able to use the crowd as a sensor for other kinds of information, e.g. actionable information. We have made progress demonstrating that retweet and follow patterns on Twitter can be used to home in on users tweeting from the ground of event, but there is still work to be done in designing solutions that function in real-time—and questions remain in how best to communicate the products of these solutions to decision-makers.

Integrating Machine and Human Computation

The most powerful solutions in this space may indeed manifest as integrations of machine and human-powered solutions, where Machine Learning algorithms learn from volunteers' and other crowd members' online actions, and where these same algorithms feed processed data back to volunteers who can verify and synthesize the output before passing it along to responders and affected citizens. Along these lines, it will be important to design solutions that align with the values and motivations of digital volunteer communities, and to structure them to fit into formal emergency response processes.

Human-Centered Design in the Context of Disaster Events

Massive online convergence is now an established feature of crisis events, one that carries with it great potential for improving outcomes during response efforts, *if* we can manage to get the right information to the right people at the right time—and in the right form. The challenges manifesting at this intersection of crowds and crises are both technical and social. Solutions may benefit from a human-centered lens that seeks to understand and support the informational needs and goals of people—affected people, responders, volunteers, and the broader crowd—during disaster events. It is likely that the most effective solutions will find ways to integrate the work of the crowd with computational algorithms that can scale up with the ever-increasing size of the information-processing problem.

References

Blake, E. S., Kimberlain, T. B., Berg, R. J., Cangialosi, J. P., Beven II, J. L. (2013). Tropical Cyclone Report, National Hurricane Center (February 12, 2013). United States National Oceanic and Atmospheric Administration's National Weather Service.

Dailey, D. & Starbird, K. (In submission). Journalists as Crowdsourcerers: Responding to Crisis by Reporting with a Crowd.

Fritz, C. E., & Mathewson, J. H. (1957). *Convergence behavior in disasters: A problem in social control*. Washington, DC: National Academy of Sciences.

Herrman, J. (2012). Twitter is a Truth Machine. Blog on *Buzzfeed*. Posted October, 2012. Available at: http://gofwd.tumblr.com/post/34623466723/twitter-is-a-truth-machine

Hughes, A., L. Palen, J. Sutton, S. Liu, & S. Vieweg. (2008). "Site-Seeing" in Disaster: An Examination of On-Line Social Convergence. *Proc. of Information Systems for Crisis Response and Management Conference (ISCRAM) 2008*.

Kendra, J. M. & Wachtendorf, T. (2003). Reconsidering Convergence and Converger: Legitimacy in Response to the World Trade Center Disaster. *Terrorism and Disaster: New Threats, New Ideas: Research in Social Problems and Public Policy, 11*, 97-122.

Meier, P. & Munro, R. (2010). The Unprecedented Role of SMS in Disaster Response: Learning from Haiti. *SAIS Review*, *30*(2), 91-103.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010). A vision for technology-mediated support for public participation and assistance in mass emergencies and disasters. In *Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference* (Edinburgh, United Kingdom, April 14 – 16, 2010). Swinton, UK: ACM-BCS Visions of Computer Science, British Computer Society, 1-12.

Starbird, K., Palen, L., Hughes, A. & Vieweg, S. (2010). Chatter on The Red: What hazards threat reveals about the social life of microblogged information. *Proc. of the ACM 2010 Conference on Computer Supported Cooperative Work*. New York: ACM, 241-250.

Starbird, K., & Palen, L. (2011). 'Voluntweeters': Self-organizing by digital volunteers in times of crisis, *Proc of the 2011 ACM Conference on Human Factors in Computing Systems*. New York: ACM, 1071-1080.

Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A. & Anderson, K. NLP to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency. In the *Fifth International AAAI Conference on Weblogs and Social Media*, 17-21.