

#Londonsburning: Integrating Geographic Topical, and Social Information during Crisis

Kimberly Glasgow* Alison Ebaugh Clayton Fink

Johns Hopkins University Applied Physics Laboratory
11000 Johns Hopkins Rd
Laurel, MD 20723

kimberly.glasgow@jhuapl.edu (*primary author)
alison.ebaugh@jhuapl.edu
clayton.fink@jhuapl.edu

Abstract

Social media, such as microblogging, is a potentially powerful medium through which the city and its citizens can connect, particularly during times of crisis or extreme events. Yet neither the city nor its citizens are monolithic entities. We propose methods to integrate geographic, topical, and social information and behavior to improve situational understanding throughout the city, and enable more effective public and official response. We demonstrate this approach using Twitter data from the 2011 London riots.

Overview and Approach

This paper will examine the use of Twitter, a microblogging service, during one of the most tumultuous and violent periods in the recent history of one of the major capital cities of the world, London. In August 2011, London was rocked by a series of riots. The riots included mass looting and widespread arson. These events led to five deaths, hundreds of injuries, thousands of arrests, and over one hundred million pounds of damage in the city. The rioting spread to other major urban areas in England, including Manchester.

Data from social media collected during these events provides an unparalleled opportunity to examine the London riots from the perspective of the citizen in the midst of these events. We analyze Twitter data from a collection of over 14 million tweets from London for the time period of the riots and their immediate aftermath, as well as other publically available data. We hypothesize that although the

riots were a powerful, riveting event throughout the entire city, examining social media can illuminate finer distinctions in the behavior of its citizens that can aid in understanding patterns of response to these events. The integration of the information and methods discussed in this paper can help develop geographically based situational understanding of the riots within the city, and can provide insight into the networks and individual citizens who responded. Local authorities within an urban area could use these insights to engage with their citizenry more effectively and to more accurately assess how their citizens perceive them.

The citizens of a city may be well positioned to detect early indications of an event or provide updates, to assist with preparations for an impending event, to help prevent or minimize damage, and to support recovery (Starbird et al. 2010) (Vieweg et al. 2010). The role of the citizen, empowered by Information and Communication Technologies (ICTs) such as microblogs will continue to evolve as technologies develop, and both citizens and authorities become more sophisticated in using these technologies during disasters or other emergency events (Palen, et al. 2009), (Jaeger et al. 2007). London is known as a city with a relatively high rate of Twitter usage, and government entities such as the London Local Authorities (or Boroughs) have adopted Twitter as well (Panagiotopoulos and Sams 2011) to communicate with their residents. Thus, examining Twitter usage during the London riots is appropriate.

In order to better understand the city in a time of crisis or other disruptive event, social media activity relating to the event must be identified from within the larger stream. We describe an approach for doing so, leveraging the Twit-

ter convention of hashtags to identify relevant communications. Hashtags are words, phrases, acronyms, or abbreviations that can be explicitly incorporated in a tweet to signal association with the topic or meaning of the term. Hashtags are preceded by a “#”, such as *#prayforlondon*. Use of a hashtag makes a tweet easily discoverable by anyone interested in that tag. We identify a set of relevant hashtags relating to an aspect of the riots, which tell us *what* to attend to in the broader dataset.

In addition to identifying relevant social media within the larger stream, we locate communications specifically within the administrative districts of London, the London boroughs. This allows us to see *where* within London these communications are occurring. This data is examined in the context of the number of riot-related crimes or incidents that occurred within that borough.

To better understand *who* is communicating within the city, we construct a set of social networks from the relevant tweet data. As identifying central actors within a network is important (Faust & Wasserman, 1992), we measure actor prominence in these networks and assess actor similarity for the most prominent actors. Actor similarity may be an indication of roles that individuals play in networks (Marsden & Friedkin, 1993). These network roles may allow far more subtle understanding of actor importance than a simple examination of number of followers or number of tweets. We find network structure derived from actor similarity helps distinguish individuals who are active at a local grassroots level. We also situate these actors within their boroughs.

Data and Methods

We began collecting user content from Twitter after the riots started on August 6, 2011 using Twitter’s Application Program Interface (API) for search. The Search API is a RESTful interface that allows for various queries against recently posted tweets, including location based queries that will return a sample of tweets from a given location specified as a coordinate pair and a radius. We performed this query using the center coordinates of the Greater London administrative area of 51.502 latitude and -0.127 longitude, with a radius of 20 miles. Since the API’s location based queries only return a sample of the tweets from a location at a particular time, we also used the Search API to query all of the recent tweets of users whose tweets were returned for the given location. We continued the location and user-based queries from August 9 through September 30, 2011, giving us data for the later stages of the riots as well as the weeks afterward.

The location-based queries return metadata containing location information associated with each tweet. This metadata can come from two sources: the location field given in a user’s profile or a *geotag* which will only be

present for users using a mobile device with a Twitter client that has been enabled for geotagging. User-based queries will also return geotags – for tweets from mobile devices with geotagging enabled - but not the profile locations. The profile location may contain a user-entered location name (such as *London* or *Hackney*) or a latitude and longitude automatically populated from a user’s mobile client.

Nearly 400,000 unique hashtags appear in the 14 million tweet corpus. During the week of rioting, the most common hashtag was *#londonriots*. It was used over 25,000 times that first week, and appeared nearly 33,000 times in the data overall. Many other hashtags relating to the London riots or to recovering from the riots emerged spontaneously in the first days of the event. Some riot recovery hashtags continued to be used many months later.

We focus on Twitter communications relating to a specific aspect of recovery, cleaning up the damage from the riots. A review of the most frequently used hashtags in the first week of the disaster identified a seed set of hashtags for riot cleanup. That seed set was used to examine the remainder of the tag corpus, and a final set of 65 unique hashtags relevant to riot cleanup was identified. These include *#riotcleanup*, *#riotwombles*, *#londoncleanup*, and *#cleantottenham*. Tweets containing one or more riot cleanup hashtags form the basis for geographically based situational awareness. They also allow us to identify citizens within the city of London who played a significant role in riot cleanup communications through Twitter. Most of these citizens can be further located to a specific borough of London.

From a geographical perspective, in order to obtain a precise location of the tweeter at the time of the tweet, we only considered Twitter data from mobile devices with enabled geotagging. This set of tweets was further reduced to only include those that fell within the Greater London area. From a total of 32,977 tweets with relevant hashtags, 556 were used to paint the situational state of the urban environment geographically over the course of the riots. (The larger tweet set was used for determination of prominent actors.)

There are thirty-three London boroughs including the City of London that make up the Greater London Area. Each borough is a “local authority”, playing a major role in the school system and in the provision of social and public services. We use the boundaries of the boroughs to partition the city.

Rioting and destruction were not equally distributed across London during the riots. Some boroughs were more heavily affected, while others were untouched. We use data gathered by a news site, the Guardian, on incidents of rioting and related crimes reported to the authorities. This data includes approximate times and locations for each reported incident. A total of 245 crimes and incidents were reported by the Guardian data source over the week of the

riots. From that total, 131 crimes or incidents were reported within the Greater London boroughs. The incidents documented range from a peaceful protest consisting of local residents of Tottenham, Haringey borough (the incident widely considered a “spark” that helped trigger the riots), to missiles being thrown at police officers in Eltham, located in Greenwich borough. The distribution of crimes over the London Boroughs is reported in Figure (1) as well as the relative percentage of population of each London Borough.

London Borough	Twitter Activity (%)	Riot Incidents (%)	Population (%)
Barnet (BA)	1%	2%	2%
Barking & Daggenham (BD)	1%	2%	4%
Bexley (BE)	0%	0%	3%
Bromley (BO)	0%	5%	3%
Brent (BR)	0%	2%	4%
Camden (CA)	6%	2%	3%
City of London (CL)	2%	0%	0%
Croydon (CR)	0%	5%	4%
Ealing (EA)	1%	4%	4%
Enfield (EN)	11%	5%	4%
Greenwich (GR)	0%	7%	3%
Haringey (HA)	13%	20%	3%
Hammersmith & Fulham (HF)	1%	0%	2%
Hillingdon (HI)	1%	0%	3%
Hackney (HK)	19%	5%	3%
Hounslow (HO)	1%	0%	3%
Harrow (HR)	0%	0%	3%
Havering (HV)	1%	0%	3%
Islington (IS)	12%	2%	2%
Kensington & Chelsea (KC)	1%	1%	2%
Kingston Upon Thames (KT)	0%	0%	2%
Lambeth (LA)	2%	7%	4%
Lewisham (LE)	1%	2%	3%
Merton (ME)	0%	5%	3%
Newham (NE)	1%	0%	3%
Redbridge (RE)	2%	0%	3%
Richmond Upon Thames (RT)	0%	0%	2%
Southwark (SO)	1%	18%	4%
Sutton (SU)	0%	0%	2%
Tower Hamlets (TH)	2%	0%	3%
Wandsworth (WA)	7%	3%	3%
Westminster (WE)	8%	2%	4%
Waltham Forest (WF)	9%	3%	3%

Figure 1: Twitter Activity and Riot Incidents in London

Prominent Actor Characterization

Twitter functionality and conventions allow a user to express more nuanced communicative behaviors than the simple broadcast of a message. Even for a publicly visible message, users can specify an intended recipient toward whom the message is directed. Users can echo or repeat another user’s tweet by retweeting it, and credit the other user. Finally, a user can easily comment on another user by username. These behaviors can be considered analogous to *talking to* another, *quoting* another, and *talking about* another. These three distinct behaviors thus connect the

sender of a tweet to others in ways that are sociologically and semantically different. The patterns and networks that emerge may provide insight into roles played by individuals in these networks of crisis communication, how they are perceived by other network members, and the robustness of the network as a whole (Borgatti et al. 2009).

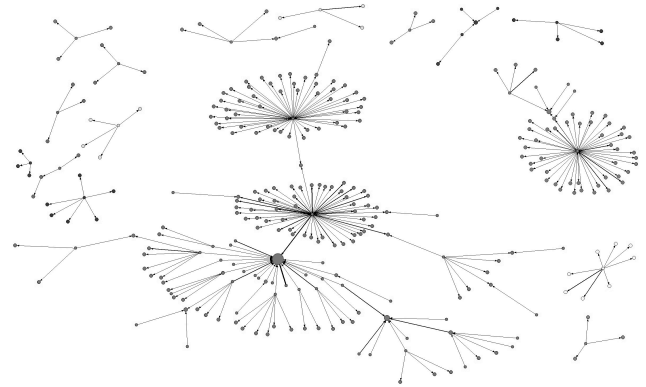


Figure 2: Social Network of Riot Cleanup Direct Addressals (component size > 3)

All tweets containing riot cleanup hashtags were used to construct a set of social networks reflecting riot cleanup communications for each of the three types of behaviors described above. These tweets generated network data encompassing roughly 4000 users and 5000 ties. The first network, pictured in Figure (2), captures only tweets that directly addressed (“talked to”) another twitter user. The smallest of the three networks, it contained 437 ties ranging in strength from 1 to 6, where tie strength is a count of the number of tweets from one member of the pair to the other. The network of retweeted messages (not shown) was much larger, containing 3035 ties (strength 1-7). The third network (also not shown) captures those who mentioned another user in their tweet. It was intermediate in size, containing 1526 ties, but had the largest range in tie strength (strength 1-50). All of the networks contained multiple unconnected components.

For each of these three networks, social network metrics that assess an actor’s centrality or prominence within the network were computed (in-degree, out-degree, in-2step-reach, and out-2step reach) using UCINET 6 (Borgatti et al., 2002). *Degree centrality* assesses how prominent, or “central” an actor is, based on how many ties the actor has to others in the network (Freeman, 1979). Indegree captures how many ties come in to an actor, while outdegree measures how many ties originate with the actor. *K-step reach* counts the number of nodes that a given actor can reach in k or fewer steps. K=2 finds both neighbors, and “neighbors of neighbors” who can be reached. These twelve metrics were summed to produce an aggregate representation of actor prominence across all three networks.

This aggregate measure was used to select the most central actors in the riot cleanup communications networks from the 4000 actors in the network. The metrics were also used to create an actor profile. These profiles were correlated and used to create an actor-by-actor similarity matrix. This matrix, rendered as a network (ties indicate similarity >.55) shown in Figure (3), shows the majority of prominent actors are linked in a single component with two main clusters. The remaining actors appear in two isolated dyads seen in the bottom left corner of Figure (3).

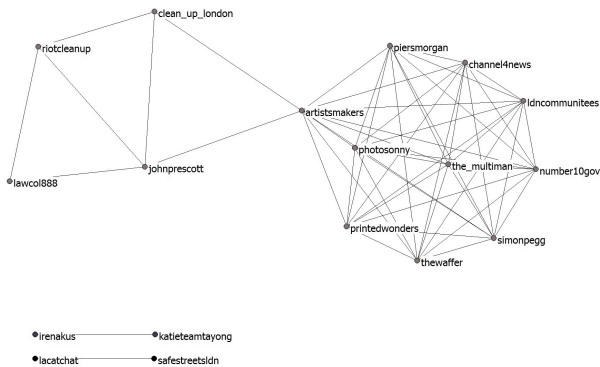


Figure 3: Actor Similarity Network

Geographic Characterization

The Twitter stream for the months of August to September 2011 is a rich source of information about the London riots and riot cleanup activity. It can be examined from the per-

spective of the location of riot cleanup communications within the city. It can also be used to illuminate the network position and connections of those most prominently involved in the riot cleanup network. This information can be further interpreted against a backdrop of riot crimes.

In order to gain an overall situational understanding of the city during the timeframe of the riots we have integrated the relative usage of the Twitter communications relating to riot cleanup with the geographically defined London boroughs. We examined each occurrence of the 65 riot cleanup hashtags for geotagged tweets. Once geographical coordinates were obtained, we were able to assign a London borough to every geographical coordinate and thus every occurrence of a hashtag. This aggregation technique paints the situational state of the urban environment of London during the riots as depicted in social media for the entire time period. (Comparable representations for individual time slices are possible, though not presented here.) Figure (4) represents the frequency of usage of hashtags related to riot cleanup per London Borough.

From the distribution of riot cleanup tweets across the London Boroughs, relatively high social media usage is clustered in the north central area of London, specifically the boroughs of Islington (65 occurrences), Haringey (72), Enfield (61), Waltham Forest (49) and Hackney (106).

In order to clarify the relationship of riot cleanup tweeting within a borough to the actual amount of local rioting or related behavior residents may have been responding to, we integrated the reported crimes and incidents pertaining

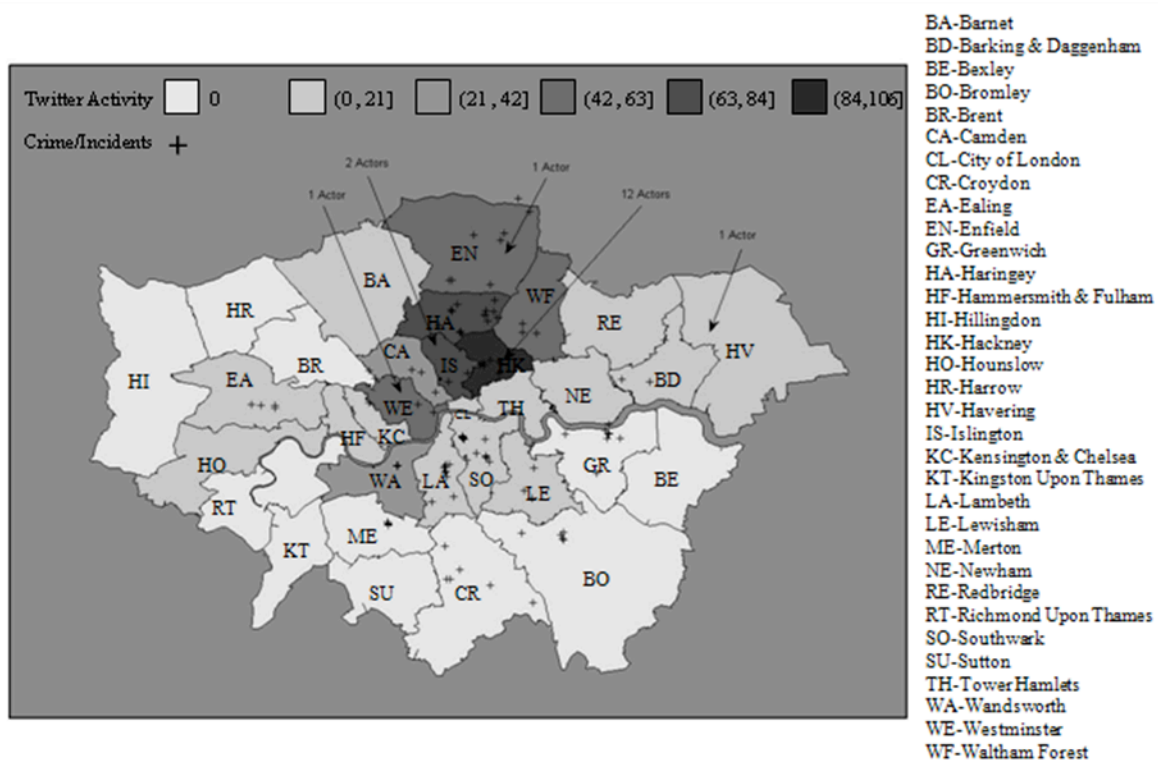


Figure 4: Twitter activity, Crime, and Actor locations

to the riot events within the London boroughs with the frequency of usage of the riot cleanup hashtags. Figure (4) also indicates the precise location for each crime or incident reported from within the Greater London area.

Discussion

Of all the crimes and incidents reported, 40% occurred in the 8 London boroughs that had the highest occurrence of geotagged riot cleanup hashtags (at least 21 occurrences). This suggests that integrating the social media based voices of citizens of Greater London area with geographic information may yield an early indication or detection of crisis, or ongoing situational awareness during recovery.

Currently a crime or incident represents an individual riot event and does not reflect the magnitude of the event nor the impact the event may have had on the community. For example, a single incident could represent a large mob looting a block, or a single burglary of a local establishment. Without a more meaningful indication of the level of impact that the crime or incident had on the urban environment it is difficult to state a more precise conclusion with regard to the relationship between the number of crimes/incidents in a borough and the frequency of riot cleanup hashtags.

As seen in Figure (4), the majority of the most prominent actors were also located in boroughs that reported more crimes, particularly in Hackney. Hackney with just 3% of London's population, seen in Figure (1), has two thirds of the prominent network actors.

Others have studied influential Twitter users, noting the dominance of celebrities and mass media in their ranks (Kwak et al. 2010). For riot cleanup communications, we find prominent actors who are neither. A number of these appear in the largest component of the network. Further, the structure of the large component of the actor similarity network, with its two clusters, provides an indication of whether an actor might be classified as celebrity/mass media, or whether that actor appears to play a different role.

Of the five actors in the smaller cluster of the large component in Figure (3), four - *artistsmakers*, *clean_up_london*, *lawcol888*, and *riotcleanup* - reflect grassroots citizen efforts to self organize to restore the city. One of these actors, *artistsmakers*, originally coined *#riotcleanup*. Given that there are over 900,000 Twitter users in the data, and 4,000 of them appeared in the network of riot cleanup communications, network information demonstrates value in discovering and clustering interesting actors. Through combining social network information directly based on actors' behavior in terms of *talking to*, *talking about*, and *quoting* others with information developed from a derived network computed from actor simi-

larity, we introduce a novel method to identify prominent actors.

From the perspective of the city, to include the local authorities responsible for restoring services and order, recognizing prominent actors in riot recovery offers promise to improve responsiveness and delivery of needed support. Understanding the larger network in which these actors are embedded may illuminate communication or coordination gaps that could be bridged. Interestingly, none of the Local Authorities, which had Twitter accounts and were actively tweeting during the period of the riots, were among the most prominent actors. Of the few who did appear in the network of riot cleanup communications, Lewisham was most prominent, yet ranked only 162nd based on the aggregate prominence measure while Enfield was 659th. Developing a better understanding of their positions in the networks of whom is *talked to*, *quoted*, or *talked about* during important events in their borough might empower Local Authorities to respond differently to their citizens.

Conclusion and Future Directions

Future directions for this research include exploring methods to increase the number of relevant tweets. In order to map activity to specific points within boroughs, we used only the subset of relevant tweets with geotags. While this ensures geographic precision, it could introduce other biases. Incorporating methods based on user profile location descriptions or tweet information to map users to geographic or administrative districts might provide a way to consider more tweets, despite increasing uncertainty about their location. While the presence of a meaningful hashtag is a clear signal of relevance, it does not perfectly capture all relevant tweets. Use of temporal correlation methods, topic models, or other methods may help improve tweet recall without introducing excessive noise.

Enriching our representation of geographically based events, such as riot-related crimes or incidents is desirable. For example, the amount of damage from the reported event or the number of arrests stemming from the incident could provide deeper insight into the magnitude or impact of incident than a count of crime incidents provides. In addition to crime data, we also plan to consider other socio-economic information, such as the Indices of Multiple Deprivation, which combine economic, social, and housing indicators to assess overall deprivation in small geographic regions, and to provide aggregate measures of deprivation at the local authority (borough) level (Government, 2011).

We are considering ways to extend actor profiles beyond the metrics currently employed. Including additional social network metrics, adding information about broadcast tweets, or incorporating connections between network members (tweets to each other) can be explored.

This paper focused specifically on riot cleanup communications. Other classes of riot related activity based on semantically related hashtags were identified in the data. Networks based on those communications can be explored and compared to create a richer understanding of crisis communications.

We consider data from the entire time period (the week of riots and the following weeks) as a single data set. However, the tweet timestamps provide the ability to explore this data as events and activities unfold over time. Network evolution, actor roles, and geographically based activity levels could be examined from a temporal perspective.

The growth of social media has dramatically changed both the landscape for social interaction, and the amount of data potentially available relating to social phenomena such as social networks. As massive amounts of online data reflecting human activity become available, new opportunities will emerge to develop understanding of major, even societal level phenomena (Watts 2007). Microblogging services, such as Twitter, can play an important role in illuminating response to disasters or other crisis events. But as MacEachren (MacEachren et al. 2011) points out, leveraging social media during crisis situations requires addressing challenges of volume, limited means to sort relevant from irrelevant information, and limited ability to extract meaning from the “generally ill-structured and cryptic” formats used.

We begin to address those challenges, integrating data reflecting geographic, topical, and socially based information to better relate the city and its citizens. While this paper focused on a subset of riot recovery behavior, this approach can be applied to any crisis or response related activity for which relevant tags emerge. We can flexibly identify *what* communications to focus on. Our approach addresses critical questions of *who* is important through measuring actor similarity and network position. It locates *where* communications are occurring within regions of a city, and enables association with sociodemographic or event-based variables. This approach holds promise to enhance timely understanding of citizen response in times of crisis, and to enable local authorities to engage more effectively in times of great need.

Acknowledgements

We would like to thank Dr. Mykola Hayvanovych, Aaron Billups, and Christopher Perkins for their contributions.

References

- Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, 323(5916), 892–895.
- Faust, K., & Wasserman, S. (1992). Centrality and prestige: A review and synthesis. *Journal of Quantitative Anthropology*, 4(1), 23–78.
- Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215–239.
- Government, D. for C. and L. (2011, March 23). The English Indices of Deprivation 2010: Technical Report. Publication (Reports and summaries). Retrieved February 12, 2012, from <http://www.communities.gov.uk/publications/corporate/statistics/indices2010technicalreport>
- Jaeger, P. T., Shneiderman, B., Fleischmann, K. R., Preece, J., Qu, Y., & Fei Wu, P. (2007). Community response grids: E-government, social networks, and effective emergency management. *Telecommunications Policy*, 31(10–11), 592–604.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web* (pp. 591–600).
- MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., et al. (n.d.). SensePlace2: GeoTwitter Analytics Support for Situational Awareness.
- Marsden, P. V., & Friedkin, N. E. (1993). Network Studies of Social Influence. *Sociological Methods & Research*, 22(1), 127–151.
- Palen, L., Vieweg, S., Liu, S. B., & Hughes, A. L. (2009). Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event. *Social Science Computer Review*, 27(4), 467–480.
- Panagiotopoulos, P., & Sams, S. (2011). Twitter in local government: A study of Greater London authorities. Conference Paper. Retrieved February 27, 2012, from <http://bura.brunel.ac.uk/handle/2438/6239>
- Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on the red: what hazards threat reveals about the social life of microblogged information. *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 241–250).
- Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness. *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 1079–1088).
- Watts, D. J. (2007). A twenty-first century science. *Nature*, 445(7127), 489.