Social Triangulation: A new method to identify local citizens using social media and their local information curation behaviors

Rob Grace
The Pennsylvania State University
wrg123@ist.psu.edu

Jess Kropczynski
The Pennsylvania State University
jkropczynski@ist.psu.edu

Scott Pezanowski
The Pennsylvania State University
Sap124@ist.psu.edu

Shane Halse
The Pennsylvania State University
seh297@ist.psu.edu

Prasanna Umar
The Pennsylvania State University
pxu3@ist.psu.edu

Andrea Tapia
The Pennsylvania State University
atapia@ist.psu.edu

ABSTRACT

Local citizens can use social media such as Twitter to share and receive critical information before, during, and after emergencies. However, standard methods of identifying local citizens on Twitter discover only a small proportion of local users in a geographic area. To better identify local citizens and their social media sources for local information, we explore the information infrastructure of a local community that is constituted prior to emergencies through the everyday social network curation of local citizens. We hypothesize that investigating social network ties among local organizations and their followers may be key to identifying local citizens and understanding their local information seeking behaviors. We describe Social Triangulation as a method to identify local citizens vis-à-vis the local organizations they follow on Twitter, and evaluate our hypothesis by analyzing users' profile location information. Lastly, we discuss how Social Triangulation might support community preparedness by informing emergency communications planning.

INTRODUCTION

Local citizens represent an important source and target for information during emergency situations. These citizens are located in a geographic vicinity and many are social media users, however, matching social media users to a particular geographic vicinity poses a challenge of discovery for researchers as well as local citizens and officials. As Landwehr and Carley (2014) write:

> Locals at the site of the disaster who are posting information about what they are witnessing are in many ways the gold of the social media world, providing new, actionable information to their followers. They are few in number, and while their messages are sometimes reposted they often don't circulate broadly. Locating their content is an ongoing challenge akin to finding a needle in a haystack. (p. 2)

Local citizens using social media platforms such as Twitter have been previously identified through content analyses that attempt to triangulate a user's location using geographic references included in a tweet (Olteanu, Vieweg and Castillo, 2015; Starbird and Palen, 2012), or through direct location information such as geotags (de Albuquerque, Herfort, Brenning and Zipf, 2015; Kogan, Palen and Anderson, 2015). While content analyses necessarily locate users post hoc, depend on unique location references, and encounter difficulties at scale, geotags are affixed to only 1.5-3.2% of all tweets (Morstatter, Pfeffer, Liu and Carley, 2013). Users who include location information in their profile present an alternative, however, some users do not disclose their location, or decide to include ambiguous or non-geographic information instead (Wang, Harper and Hecht, 2014, p. 74).

These methods of identifying local citizens set important conditions for localness: the users are necessarily active citizens posting tweets often in the midst of emergency and using keywords or hashtags that come to be queried by those interested in accessing locally generated information (Grace and Leskovich, 2015; Starbird and Palen, 2010). In this respect important methods for keyword-based tweet collection and subsequent location-based sub-sampling have been advanced (Olteanu, Castillo, Diaz and Vieweg, 2014). However, the successive subsets of users who actively post messages, include queried keywords, as well as location data of some kind, stand to leave a significant population of local citizens who use social media to *receive* information *before* an emergency invisible to others in the information space, as well as public officials seeking to provide warning notices or gain timely on-the-ground information.

Instead, we explore the utility of a novel method of Social Triangulation (ST) to identify local citizens as recipients of local information who can be identified before emergencies occur. ST is motivated by the basic assumption that people using social media and living in a geographic area curate their social networks to include local organizations both physically present in their geographic area and disseminating social media intended for that area. In the case of Twitter, the assumption suggests local people follow local organizations. Similar to methods of community asset mapping performed by urban planners (Kerka, 2003; McKnight and Kretzmann, 1997), ST involves first cataloguing a comprehensive list of local organizations maintaining Twitter accounts- local media, citizens' associations, businesses, civic and emergency services, etc., in order to discover users curating their social networks to receive information from these organizations, and thus the opportunity to identify these users as local citizens. Moreover, by understanding patterns of social network curation among local citizens, that is, where and in what variety they receive local information through social media, we can understand the *information infrastructure* of a local community.

In this paper, we first describe the method of ST and then evaluate the validity of its basic assumption. We do this through an exploratory application of ST to a small city in the Northeastern United States. We catalogue and categorize 195 community assets in this community, identify their collective 185,176 Twitter followers, and then, using available profile location information, evaluate the extent to which these followers are indeed "local." We find promising evidence supporting our basic assumption: among 79,998 users for whom we have location information, we find that fully 68% identify themselves as local citizens. Moreover, the more local organizations a user follows the more likely they are to identify as local citizens: For followers of only one organization, 67% identify as local citizens, while 71% of those following 3-9, 84% of those following 10-49, and 98% of those following 50 or more organizations, respectively, identify as locals. We discuss further efforts to validate our assumption and future work using ST as a localization method in our discussion.

Additionally, we explore the information infrastructure of the community to understand from what information sources, and in what variety, local citizens receive local information on Twitter. Our exploratory study points to *differences in the level and position in which citizens are embedded in the information infrastructure of a community.* Specifically, we find that a majority of loosely embedded citizens tend to curate and receive information from only local media sources, while a deeply embedded minority is positioned to receive more and various kinds of community information relative to the latter, especially information disseminated by citizens' associations as well as civic and emergency services.

We therefore find utility in ST as a method supporting community preparedness, capacity, and resilience-building. By systematically cataloguing and mapping community assets that act as important sources of local information on social media, as well as the people who receive information from them (and those who don't), ST can inform emergency communication planning by pointing to groups of local citizens not directly receiving emergency or risk-related information, as well as the local social media accounts that reach these segments of the community. Drawn from our particular community study, we offer general recommendations for local emergency communications strategies.

This paper is organized as follows: first, we review work in crisis informatics literature to point out that local citizens, as information recipients, tend to be excluded by current collection and localization methods. Second, we present the procedure of ST by deploying it to explore the information infrastructure of a small city in the Northeastern US. We then describe our findings related to local information curation behaviors we observe, as well the evaluation of the localness assumption behind ST. Lastly, we describe the merits of ST as a community preparedness tool and opportunities for future work.

WHY IS IDENTIFING LOCAL CITIZENS IMPORTANT?

Local citizens consist of people living together in a geographic region. Local citizens constitute communities, and are thus the primary producers and recipients of community information. In crisis informatics research, local citizens represent those closest to and most affected by natural and man-made disasters, as well as more common emergencies, and take part in emergency situations as eyewitnesses and citizen responders. As a result, local citizens produce actionable, and situational information to other locals, emergency response and relief officials, as well as a global crowd of interested observers, journalists, and digital volunteers (Starbird, Muzny and Palen, 2012; Starbird, Palen, Hughes and Vieweg, 2010). Moreover, local citizens live and work in towns, cities, and regions where they intimately know and shape the geographic, cultural and built infrastructure, and thus stand to provide unique and critical information during times of emergency (Starbird, Muzny and Palen, 2012).

As information producers, local citizens provide "generative information," primary accounts and interpretations of emergency-related events that Starbird and Palen (2010) take as the "the core of the information production cycle" (p. 246). Studies identify local citizens as individuals who are more likely post messages related to emergencies in their vicinity than social media users geographically-removed from events (de Albuquerque et al., 2015; Lachlan et al., 2014). Moreover, suggesting the importance of social networks for disseminating information, studies observe that local citizens are more likely retweet generative information created by other local citizens than the globally-distributed crowd (Kogan et al., 2015; Starbird and Palen, 2010). Whereas local citizens more often provide information about affected people and actionable information useful to other local citizens, the globally distributed crowd of social media users most often post messages of sympathy and support for those affected by a crisis (Olteanu et al., 2015). However, the distributed crowd of users- by retweet- can amplify and direct attention to local citizens' messages, or disseminate relevant URLs to other information sources online (Starbird and Palen, 2010; 2012).

As information recipients, local citizens represent the people to whom government and relief organizations attempt to communicate situational updates and directives during emergency situations. Studies find government and emergency response organizations adopt social media in emergency situations (Graham, Avery and Park, 2015; Hughes and Palen, 2009), and use social media to disseminate emergency warnings to local citizens before emergencies (Rice and Spence, 2016; Veil, Buehner and Palenchar, 2011). During and after emergencies, governments and NGOs post social media to convey important information to affected locals on the progress of relief operations, and the location and availability of emergency assistance (Olteanu, Vieweg and Castillo, 2015; Tim, Pan, Ractham and Kaewkitipong, 2016). In non-emergency contexts, municipal governments and public services, such as police departments, seek to communicate crime and traffic updates to the public, and also solicit information about missing persons and wanted suspects (Huang et al., 2016).

However, in the globally-expansive information space of social media, identifiable local citizens remain rarities (Starbird et al., 2010). Analyzing across many crises events, Olteanu et al. (2015) find local citizens' social media to typically account for an average of 9% of crisis-related messages on Twitter. In non-emergency contexts, social media might substantially lessen in volume, but lack the contextual data features- hashtags or crisis-related keywords- that assemble and make local information commonly accessible (Bruns and Liang, 2012). Yet while identifying local citizens within the voluminous and complexly-assembled information space of social media remains "akin to finding a needle in a haystack," their critical roles as information producers and information recipients require effective methods of localization to inform community information sharing in both emergency and non-emergency settings.

HOW HAVE LOCAL CITIZENS PREVIOUSLY BEEN IDENTIFIED?

Typical methods in crisis informatics collect Twitter data vis-à-vis the words people include in the content of their messages or, alternatively, geolocation information people append to tweets or descriptive locations entered in user profiles. These methods typically involve engaging Twitter's public API and can be referred to as keyword or location-based collection respectively (Olteanu, Castillo, Diaz and Vieweg, 2014, p. 376). Importantly, these collection methods set important conditions for identifying local citizens.

Location-based Collection and Localization

Local citizens can be identified on the basis of geographic coordinates (i.e. geotag) associated with a particular tweet, or location information included in a user's profile. Using Twitter's public API, queries can be constructed to retrieve tweets containing particular words or hashtags, or tweets with a location-identifier corresponding to a geographic area or bounding box specified by a set of geographic coordinates. Due to issues of relevance and comprehensiveness involved with using Twitter's REST and Streaming API, keyword-based data collection often precede location-based subsampling and subsequent identification of locals (Kogan et al., 2015).

For instance, to identify geographically vulnerable Twitter users during Hurricane Sandy, Kogan et al. (2015)

first selected eight "carefully chosen" keywords to collect tweets through Twitter's Streaming API. Among the users whose tweets were collected, users posting at least one geotagged tweet within a boundary box encompassing part of the United States Eastern Seaboard were selected as the basis for a subsequent round of data collection involving the REST API, by which the authors could collect up to 3200 of these users' most recent tweets (p. 983). This approach to identifying locals thus involves two important filters: those who use a specified set of keywords in their tweets and, the small minority among those users who also posted a geotagged tweet in the affected locale.

Using geotags to identify local citizens remains limited by the paucity of users who include geographic coordinates in their tweets- as a result only 1-3% of tweets carry geotags (Morstatter, Pfeffer, Liu and Carley, 2013). Consequently, local citizens posting geotagged tweets represent only a fraction of those actively posting content on Twitter in a geographic location, and none of the people who might use Twitter but do not actively tweet, or at least did not do so during the period of data collection. Among the latter, a PEW Research Center (2015) study found that 28% of Twitter users did not tweet during a one month period, and an additional 33% posted fewer than nine messages. Among those who did tweet, 30% of all tweets were retweets (for news-related tweets this rose to 49%) - a notable figure when considering that location-based queries to the Streaming API do not retrieve retweets (Twitter, 2016).

Content and Keyword-based Collection and Localization

Alternatively, content and keyword-based methods revolve around the relationship between the lexical scope and distribution of words curated by both locals and a global crowd using Twitter to discuss a crisis, and the selection of specific keywords held to represent and, in effect, constitute the crisis information space. The users whose tweets contain these keywords and hashtags can subsequently be identified as local citizens by using tweet geotags or profile location information, as described above, or coding methods to determine a user's location on the basis of the content of the tweets they post, for instance, if it is judged to be an "eyewitness" account (Olteanu et al., 2015).

Using the Streaming API to examine the 2011 Egyptian political protests, for example, Starbird and Palen (2012) relied on Arabic speakers in their research lab to select certain keywords "as the most popular during the early days of this event" (p. 9). Thus if a local protester, even if at the center of Tahrir Square, did not include one of the keywords "egypt," "#egypt," or "#jan25," they would never be identified by the authors and, more significantly, perhaps other locals and the global crowd. Starbird and Palen (2012) coded the retweets of active users on the basis of "assertions of or references to being in Cairo during the period of the protests." Accordingly, users were assessed as either in Cairo, in Egypt but not in Cairo, or outside Egypt. The authors thus identified local citizens as any Twitter user posting a tweet including at least one of the three queried keywords, and judged as being in Cairo at least once within the defined collection period (p. 10).

Keyword-based methods necessarily focus on highly-visible terms that ignore tweets missing the particular keywords queried and the users who post them or do not tweet at all (Bruns and Liang, 2012; Olteanu et al., 2014; Vieweg et al. 2010). Keyword curation can amplify the non-local, as it "assume[s] to establish a dataset of the most visible tweets relating to the event in question, since it is the purpose of topical hashtags to aid the visibility and discoverability of Twitter messages" (Bruns et al., 2012). While keywords and hashtags stand to benefit those affected in the vicinity of a crisis, the emphasis on visibility and discoverability especially aids the geographically-removed crowd who generally lack existing social network connections to local people and organizations.

Moreover, Vieweg et al. (2010) suggest that familiarity with a locale, such as possessed by local citizens, can lead user's to omit the very keywords or hashtags that would identify them to others as local. What they describe as "markedness" refers to:

> how certain places, landmarks or items become taken-for-granted and expected when referred to in more general terms. The RR data set was collected based on search terms "red river" and "redriver", and within this data set, if someone mentioned "the river" or "the flood level" it was commonly understood to be about the Red River, which makes the Red River "unmarked"— no detail is necessary when referring to it. (p. 1086)

Thus if someone only provided information on "flood level" such data would only become visible to the authors if, in another tweet, they had also mentioned one of the two keywords or hashtags queried for data collection. In the course of their retrospective examinations of three distinct crises events, a school shooting, tornado, and flood, Saleem, Xu and Ruths (2014) find that initial tweets concerning these events often fail to include references to places or the type of emergency, as well as visible hashtags associated with the event. The authors conclude that "the first tweets carrying situational information tended to lack the kind of identifying keywords and hashtags that would make them easy to discover in a full Twitter stream" (p. 156).

Together location and content/keyword-based methods set important conditions for localness. These methods risk omitting the 97-99% of users without geotags, those tweeting without identifiable keywords, or those who do not actively tweet but rely on social media for local information. As an alternative, we now introduce ST in order to i) identify local citizens in non-emergency contexts, ii) identify a greater proportion of local citizens than standard methods, and c) identify local citizens without relying on geotag and profile location data, or data from tweet content.

METHOD

We deploy ST by cataloguing community assets located in a small city in the Northeastern US, and analyze the information curation behaviors of local citizens who curate the Twitter accounts of these community assets within their social networks. Information curation "involves future oriented activities," consisting of a "set of practices that select, maintain, and manage information in ways that are intended to promote future consumption of that information" (Whittaker, 2011, p. 7). Here, curation involves practices of "following" community assets, specifically local organizations, by Twitter users who, as a result, enable ongoing access to local information disseminated by these organizations. Shared patterns of curation among people following local organizations reveals the information infrastructure of the local community, and reveals which groups of local citizens curate their social networks to directly access types of local information, as well as those who do not.

Our exploratory analysis of ST is guided by three questions:

1. How many local organizations do people typically curate within their social networks?
2. What patterns of information curation exist among people following local organizations?
3. What is the relationship between information curation behavior and a person's geographic location?

We describe the information gained using ST by characterizing descriptives of user curation behavior, comparing Twitter users identified using ST to the location mentioned in their profile, and finally, we describe the types of local organizations that socially triangulated Twitter users co-follow to better understand the information infrastructure of the local community.

Our methodology involves a four phase process of social triangulation (ST) to locate Twitter users in a geographic area: (1) categorizing of community assets, (2) cataloguing community assets, (3) collecting user information, and (4) analyzing geographic characteristics and information curation patterns.[1] To socially triangulate Twitter users, we began by identifying categories of public and private organizations to organize our search for community assets in one city in the Northeastern Unites States. Urban planners often assess community capacity and resources using community asset mapping methods which search for organizations in a community by type (Kerka, 2003; McKnight and Kretzmann, 1997). Similarly, we identified types or categories of organizations- public institutions, citizens' associations, local economy (e.g. businesses), and media- which we subsequently modified and expanded by searching for and cataloguing organizations maintaining a Twitter account in the city of interest. The resulting eight categories of community assets are listed in Table 1.

The initial categories functioned primarily to organize our search for local organizations. This involved a grounded process in which online directories and search engines were used to explore the diversity of local organizations in a particular community asset category and determine which types of organizations in that category maintain an identifiable social media presence. For example, beginning with the community asset category of local economy we used community directories (e.g. yellow pages) and general online searches (e.g. restaurants in...) to categorize and identify types of businesses maintaining Twitter accounts. We subsequently identified three categories- restaurants, bars, and entertainment. Importantly, the categories reflect local economic assets with a social media presence and not a comprehensive catalogue of community businesses. For example, beginning within the community asset category of citizens' associations, we initially speculated that the numerous churches and worship centers in the city might have a significant social media presence, however, upon investigation, we subsequently discovered that very few were on Twitter.

[1] ST represents a form of geolocation inference, techniques using social network relationships to infer a user's location (Backstrom, Sun, and Marlow, 2010; Jurgens, Finethy, McCorriston, Xu, and Ruths, 2015). However, ST differs by focusing on the process of establishing a single "ground truth" location by cataloguing important information sources located in a geographic area, and in relation to which all users' locations are inferred. For Twitter most geolocation inference methods use geotags or profile location information for ascertaining ground truth, and thus encounter similar constraints of other localization methods (Jurgens et al., 2015).

Table 1. Categories of Community Assets

Category	Description	Number
Citizens' Associations	Volunteer, social, and nonprofit organizations, e.g. habitat for humanity, environmental conservation groups, historical society	41
Civic Services	Civic government and public services, e.g. municipal government, public library, public transit	18
Emergency Services	Emergency management and response services, e.g. police and fire departments, EMS, university emergency alerts	13
Schools	Municipal school district, e.g. elementary, middle, and high schools, local vocational school	16
Bars	Establishments of good cheer, e.g. bars, saloons, taverns, wineries, and pubs	27
Entertainment	Recreational businesses, e.g. minor league baseball team, golf courses, movie theaters	5
Restaurants	Local (non- chain) food-serving establishments, e.g. cafes, diners, delis, pizzerias	43
Media	Local media, e.g. newspapers, newsletters, news websites, radio, television	32
Total		195

The second stage of ST involved developing a comprehensive catalogue of organizations for each category. The grounded process of identifying local organizations and arranging them into categories described above yielded an extensive catalogue of organizations, however, we sought to develop, to the best extent possible, a listing of all identifiable organizations both located in the local city and maintaining a Twitter account for each category selected. Additionally, we created a project Twitter account to utilize Twitter recommendation algorithms to suggest additional organizations and searched among the followers of these organizations, as well as the accounts each organization follows, to identify any organizations that did not appear through our online searches.

Third, using Twitter's REST API we collected the Twitter IDs of accounts following each local organization. For example, among the public services in the city we catalogued the city's bus system and collected the account IDs for the 1,011 people that follow the official bus system account on Twitter. In total, we collected 185,176 Twitter IDs, each following at least one of the 195 local organizations catalogued. The follower counts for each organization reflect the period of our data collection, which occurred during the week of December 5, 2016.

The fourth phase of ST involves evaluating the localization assumption behind ST, and analyzing the kinds and variety of local information users access through their social networks. For the latter we analyze 1) the curation patterns among those following local organizations and 2) the relationships among organizations with respect to users' curations patterns. The two analytical methods are further described below.

Geographic Analysis

We utilized profile location self-identified among the 185,176 Twitter users in order to understand their location relative to the number of organizations they follow. Using a Java program written to search the Twitter API, we found Twitter accounts available for 168,452 of these accounts. Of these, 79,980 accounts included profile information.

By importing this information into Google Fusion Tables we geocoded each of the self-described locations by relying on similar tools that are used to identify non-uniform information in Google Map search. Errors in geocoding can occur due to misspellings in the location, and ambiguous locations which match multiple, similarly named places in the world. During Google's geocoding process, between 12 – 15% of the locations were reported as ambiguous, and unavailable for geocoding. The identifiable locations were plotted on a Google Map to better understand the locations self-identified by Twitter users with respect to the number of local organizations they follow.

Social Network Analysis

Using the data collected, we constructed a 2-mode matrix where one mode was comprised of the Twitter accounts of each of the 195 local organizations and the second mode was comprised of the 185,176 Twitter users that followed those organizations. With an interest in uncovering patterns in curation behaviors among various data streams, we created a weighted 1-mode affiliation matrix that represents Twitter user co-followership of the 195 local organizations. In this network, each node is a local organization and each link among organizations is weighted by the number of Twitter users that co-follow both organizations.

Organizational categories were used as node attribute information in order to describe the relationships among types of local communities. Further, we developed a method to examine an attribute-based model of structuration of the local network. This was performed by first creating a table of homophily scores using a standard E-I index. Here, each “same category” tie is treated as an internal group tie and every “different category” tie is treated as an external group tie, in which the number of ties external to the groups (E) minus the number of ties that are internal to the group (I) are divided by the total number of ties in the network (Krackhardt and Stern, 1988). The homophily scores table was then treated as a 2-mode matrix to create a 1-mode affiliation matrix representing inter-category ties among various types of Twitter information streams resulting in a community asset network.

RESULTS

The following section describes the results of the descriptive and comparative analysis of Twitter followers identified using ST. We begin with a descriptive account of information gathered using our method and follow with analysis to better understand information curation behaviors among Twitter followers of the local organizations we catalogued.

Descriptive Analysis of ST

ST identifies people who curate local organizations within their social networks, and thereby become recipients of local information and embedded in the local information infrastructure. Among users curating at least one of the 195 citizen’s associations, public institutions, businesses, and media organizations in their social networks, how many and what categories of organizations do they follow? Figure 1 displays the following distribution of the 185,176 users, indicating different *levels* of embeddedness within the local information infrastructure. On average each user follows 1.82 organizations, with 75.3% (139,440) of all user following only one of the 195 organizations we catalogued in the city.

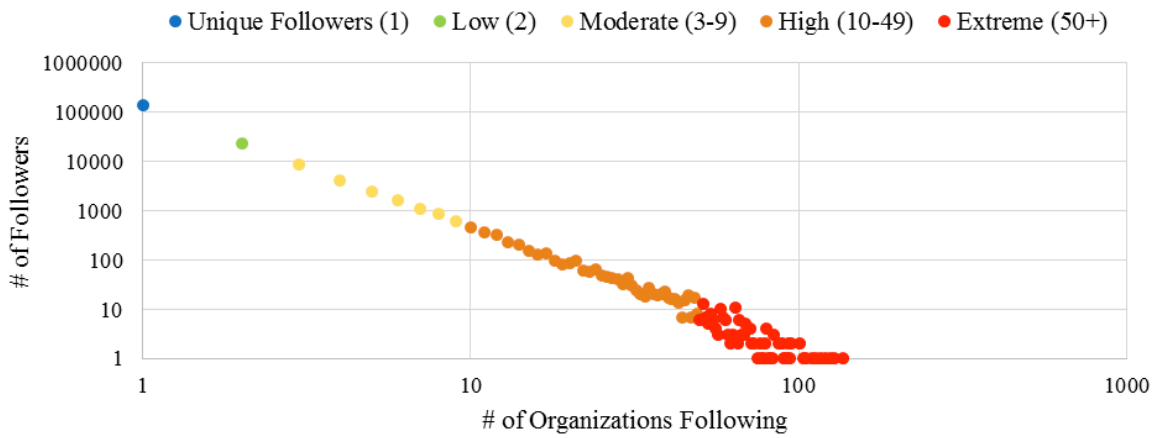


Figure 1. Distribution of users relative to the number of local organizations they follow (colored by levels of local information following).

Dividing the 185,176 users according to how many organizations they follow reveals five types of local information followers and information recipients we categorize as: unique, low, moderate, high, or extreme. As described, the majority of all users follow only one local organization and thus we consider them as unique recipients of local information. Low followers, 12.4% (23,027) of all users, follow only two organizations.

Medium, 10.5% (19,397), and high followers, 1.7% (3,147), follow between 3-9 and 10-49 organizations respectively. Accounting for only 0.1% or 165 of all users, extreme followers curate 50+ organizations within their social network, such as one user who follows a total of 139 local organizations. This following distribution indicates, on one hand, a highly embedded minority of users (medium-extreme followers) receiving multiple information streams within the community and, on the other, a weakly embedded majority (unique-low followers) that directly receive information from only one or two organizations.

Looking to the categories of organizations users follow points to important differences in what kind of organizations people choose to curate within their social networks, receive information from, and thereby different positions of embeddedness within the local information infrastructure (Figure 2). Comparing unique followers to high and extreme followers, for instance, reveals a stark contrast in their curation of media organizations (see Figure 2). Among the 139,440 people who only follow one local organization, over 106,000- 58% of all users- follow one of 32 local media organizations. In contrast, medium, high, and extreme followers curate multiple and more diverse organizations within their social networks. Among high and extreme followers, citizens' associations and public services account for approximately 40% and 50%, respectively, of the organizations curated within their social networks. These distributions indicate that local citizens are differently positioned within the local information infrastructure, that is, different segments of the community receive information from different sets of local sources. Thus a minority of highly embedded users follow more organizations and receive a different variety of local information from a balance of media and business organizations, and civic offices and volunteer groups. On the other hand, a majority of weakly embedded users follow few organizations and of less variety, with most receiving information only from local newspapers, radio stations, or telelvision sources.

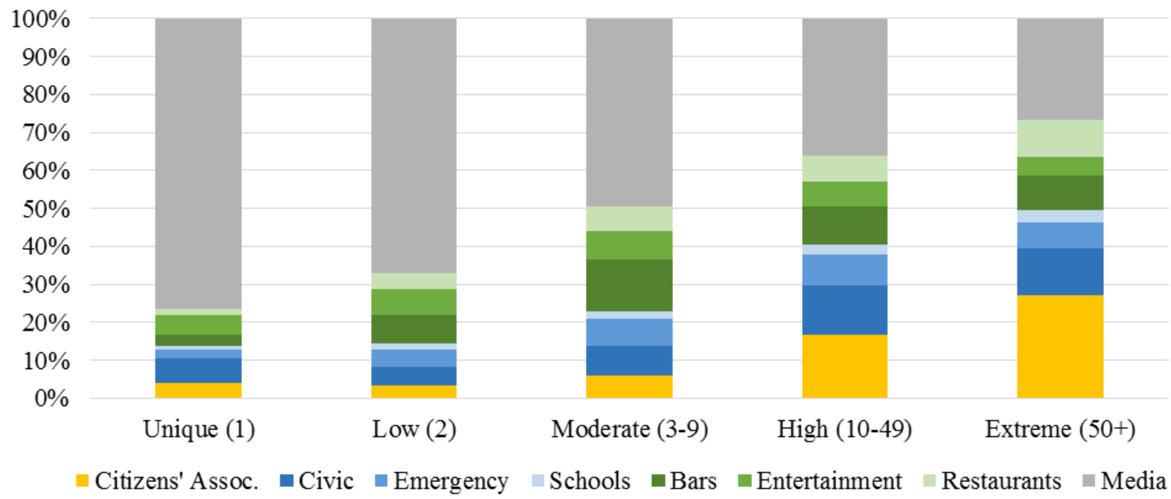


Figure 2. Proportion of Community Assets Followed by Each Type of Local Information Recipient.

Evaluation of Followers' Location

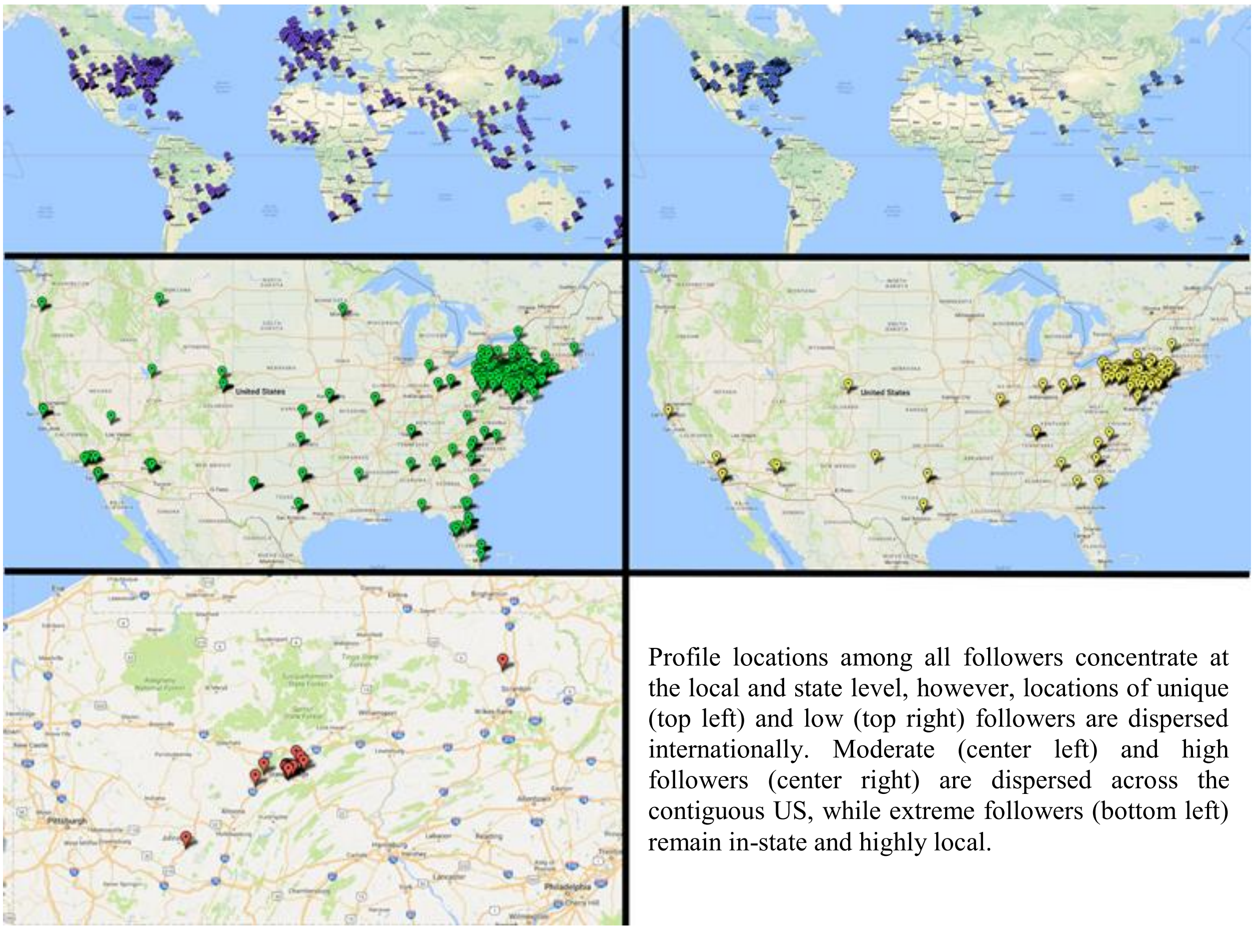
Next, we evaluate the location information of people who follow local organizations in order to understand the relationship between information curation behavior and a user's location. That local citizens will tend to follow and receive information from organizations in the geographic area in which they live motivates ST as a method for identifying local citizens. In order evaluate this assumption we use the location information users include in their Twitter profile to better understand the location of people following local organizations and how this might vary according to the number of organizations they follow (Table 2). Of the 185,176 users following a local organization, 43% include an identifiable location in their profile. Among those with location information, 91% are local to the state, and 68% identify their location within the municipality. Altogether, we find that 72,638 followers, or 29% of all users following a local organization, are local citizens of the city.

Differences, however, appear among followers at different levels of embeddedness within the local information infrastructure. That is, depending on how many local organizations a user follows, we observe differences in the availability of location information and the proportion of followers who identify as living in the local community. First, among unique followers we find only 41% to have included an identifiable location on their Twitter profile. In comparison, 68% of high followers and fully 88% of all extreme followers included profile locations. Significantly, in the locations identified among these users, we find that nearly all followers who include location information on their profile are located in state (91%), and 68% identify as local to the municipality itself. Moreover, among high and extreme followers who include profile locations, 84% and 98%, respectively, identify as local citizens.

Table 2. Evaluation of Follower Geographic Locations

Orgs. Following	Total Users	Users Located	%	In State	%	In City (25km)	%
Extreme (50+)	165	146	88%	146	100%	143	98%
High (10-49)	3147	2142	68%	2058	96%	1809	84%
Moderate (3-9)	19397	10404	54%	9755	94%	7361	71%
Low (2)	23027	10530	46%	9796	93%	7064	67%
Unique (1)	139440	56756	41%	50883	90%	37788	67%
Total	185176	79978	43%	72638	91%	54165	68%
Among all users				72638	39%	54165	29%

Lastly, we mapped the profile location information included by users using Google Fusion Tables and Google Maps (Figure 3). Separately mapping the self-identified locations of users at different levels of embeddedness in the local information infrastructure, dramatic differences in the geographic dispersion of users' locations become revealed. Among unique and low followers, users identify locations spanning around the globe. For instance, a follower of a local bar indicates in his profile that he is from Iraq, while a baseball-related twitter account in Cuba follows the local minor league baseball team.



Profile locations among all followers concentrate at the local and state level, however, locations of unique (top left) and low (top right) followers are dispersed internationally. Moderate (center left) and high followers (center right) are dispersed across the contiguous US, while extreme followers (bottom left) remain in-state and highly local.

Figure 3. Geographic dispersion of followers identified by ST

In contrast, the user locations of moderate and high followers are less dispersed and remain concentrated at the state and regional level, although many locations are indicated throughout the United States. Among extreme followers, all identify their location within the local municipality except two whose location remains in-state and within 200km. While the presence of a major public university with a large international student population likely accounts for many of the internationally-dispersed user locations, the locations of users outside the municipality become less dispersed and more regionally concentrated among those who follow multiple local organizations.

Description of Local Information Curation Behaviors

Figure 4 displays the co-following network structure of the 195 local organizations who share a link if they are followed by the same Twitter user. The network is very dense, wherein 96.6% of all possible ties are present, and has a low overall degree centralization at only 3% of the network being centralized around a few nodes. This network is very tightly bound with a network diameter of two and the average distance between any two nodes is 1.034. While many whole network measures of the network are not very descriptive of distinguishable measures in the network, a weighted measure of centrality shows that the whole network homophily E-I index among organizations links by same organizational type is 0.6892. In an exploration of the one-mode representation of the local information infrastructure divided by local information recipient types found similarly high densities with low diameter. This indicates that by the methods used in this analysis, this community appears to have very little fragmentation in its information infrastructure.

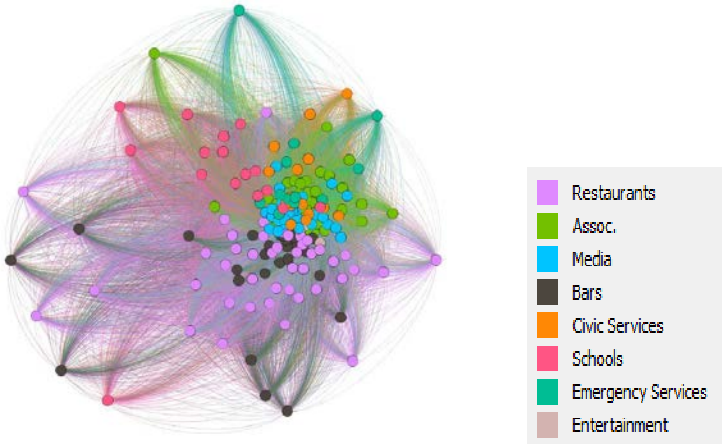


Figure 4. Sociogram of 1-mode affiliation matrix of local organizations (colored by organizational categories).

The next phase of this work was to utilize homophily scores for individual categories to develop an attribute-based model of structuration weighted by overlapping same category ties, which is displayed in Figure 5. As seen below, Civic Services (such as the local police department and bus) is the organization type with the most followers. Civic Services weakest links are to entertainment, followed by the media. Citizens' Associations is the organization type with most overlap in followership with other organizational types, in some ways placing it at the center of the information infrastructure of the community. Although Citizens' Associations display strong ties to most categories, its weakest link was to Entertainment, followed by Media. Comparatively, the Entertainment category had relatively weak ties to most categories. In a comparative analysis among the types of local information recipients, Twitter users in the extreme and high categories produce a similar information infrastructure, whole those in low and moderate categories produce an information infrastructure that displays strong ties among citizens associations and all other categories, as well as moderately strong tie among bars and emergency services. By virtue of this methodology, individuals in the unique category do not produce links among organizational types.

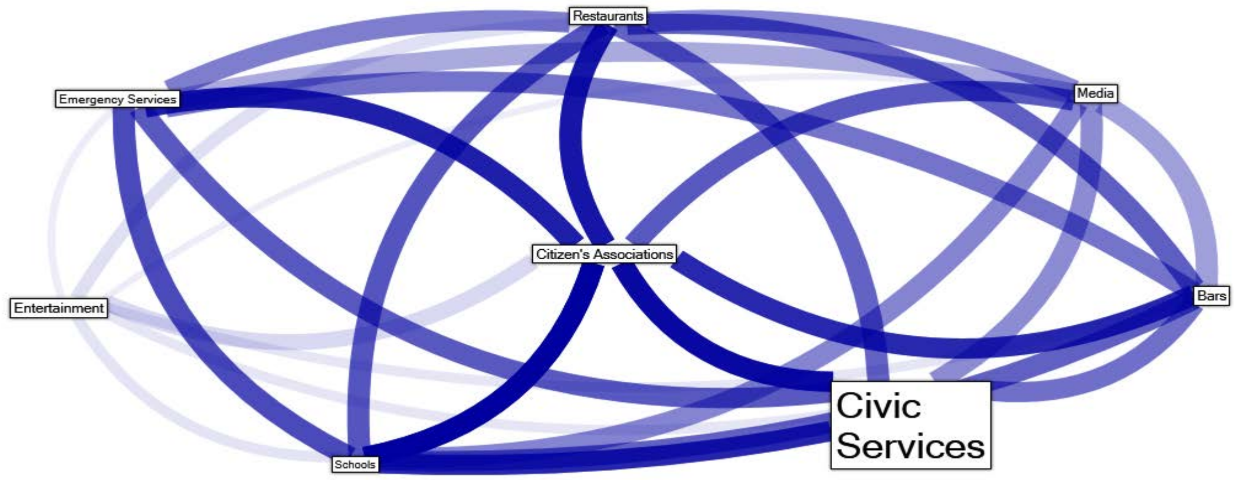


Figure 5. Attribute-based Model of Structuration of Community Assets (Labels sized by number of followers; ties weighted by co-followership).

DISCUSSION AND CONCLUSION

In the exploratory study we present, we find evidence to support the underlying assumption of ST: that local citizens tend to curate their social networks around organizations in their geographic area. Among the 79,998 users following a local organization in the particular city situating our study, and for whom we have location information, we find that fully 68% identify themselves as local citizens. Moreover, the more local organizations a user follows the more likely they are to identify as local citizens. For followers who have curated only one or two local organization in their social network, 44,885 or 67% identify as local citizens. This increases for those who follow 3-9 organizations (71%), 10-49 organizations (84%), and among those following 50 or more organizations, 98% identify as local.

Moreover, by identifying local citizens as information recipients, we find that users following local organizations on Twitter differ with respect to their level and position of embeddedness within the community information infrastructure. Local media accounts on Twitter, including television, radio, and news organizations, feature the most followers among the categories of local organizations we catalogued. However, the vast majority of users following only one or two local organization on Twitter (160k+) disproportionately curate media organizations within their social networks. Thus we see local citizens within this majority as information recipients weakly embedded within the local information infrastructure, and positioned so as to receive information predominantly from local media sources.

Additionally, we identify a sizeable minority of users (20k+) who curate their social networks to follow multiple and various kinds of local organizations on Twitter, to include not only media accounts, the most followed organizations in the community, but more citizens' associations and civic organizations relative to the majority of weakly embedded citizens. Thus we see this minority of moderate, high, and extreme follower as information recipients strongly embedded and distinctly positioned in the information infrastructure of the community.

We thus see value in developing ST for community preparedness and resilience-building to better map the information infrastructures existing in local communities in order to support local emergency communication planning. For the city in which we deployed ST, for instance, the Twitter accounts of emergency services collectively have a total of 16k followers. Among these, the municipal police department, which routinely posts public updates of crime and traffic incidents, has the most (8k). In comparison, the two local television stations and newspaper each have over 15k followers. Through the analysis of user curation behaviors, the followers of these organizations occupy different levels and positions within the information infrastructure of the community.

In the case of emergency situations, however, civic and emergency services will be called on to disseminate important public warnings and situational updates. Previous studies find that municipal governments and emergency managers often lack guidelines for emergency communications planning guidelines that might result in effective social media dissemination strategies (Huston et al., 2015; Rice and Spence, 2016). ST stands to inform the development of general and community-specific guidelines for emergency communications. Our exploratory deployment of ST leads us to three preliminary conclusions about our method and the particular information infrastructure that we have studied. First, in our sample, local information recipients that only followed one unique category are more commonly users that live outside of the geographic area of interest.

When using ST to identify information that local citizens are sharing, their inclusion may be problematic, however, if using ST to push information to local citizens, this group should certainly not be eliminated, as they constitute the majority of identified local citizens despite being the least connected. Second, while the media may attract a great many followers, these followers may only be loosely embedded in the information infrastructure of their community and may not be aware of information shared by emergency services directly. Lastly, in this community, the followers of citizens' associations appear to be highly embedded individuals and may be useful sources to seek information from during an emergency.

LIMITATIONS AND FUTURE WORK

These findings suggest ST holds promise for uncovering citizens in a geographic location and revealing the local organizations from which they receive information. However, important limitations must be considered that require future work. First, we acknowledge that additional parameters to refine the aggregated datasets of local organizations and followers will require further exploration. One limitation of this work is that despite our best efforts to identify local organizations using community directories, online search, and recommendation system suggestions, additional efforts must attempt to achieve more comprehensive catalogues of local organizations. The 195 local organizations used in this analysis are likely a non-comprehensive list of local organizations within the municipality on Twitter. Future research should focus on looking for organizations that may be fragmented from the primary information infrastructure.

Second, our analysis only recognizes organizations as streams of community information, however, alternative or multiple information streams might be more appropriate. Social networks among local citizens remain absolutely critical as information channels on social media platforms. Moreover, we recognize that prominent individuals can be more influential than organizations in the creation and dissemination of information through social media. We see the concept of ST as suitable for cataloguing both salient citizens and organizations as a geographic "ground truth" in relation to which the location of users following these local personal or organizational accounts can be evaluated and inferred.

Third, and lastly, our effort to ascertain users' locations through the use of available profile location information requires further methods of evaluation. We see as an immediate opportunity the collection and analysis of geotagged tweets posted by users following local organizations. In this regard, users with geotagged posts might be identified as local citizens according to three separate localness metrics defined by Johnson, Sengupta, Schöning, and Hecht (2016): if the user posts tweets in the municipal locale at least n-days apart (n-days metric); if the majority of the user's tweets occur in the locale (plurality metric);or if the geographic median of all the user's tweets falls within the locale. These metrics could be compared among and between users following the same number of organizations as well as categories of organizations. Such an analysis would further our understanding of the relationship between user information curation behavior and geographic location.

MY REFERENCES

Backstrom, L., Sun, E., and Marlow, C. (2010) Find me if you can: improving geographical prediction with social and spatial proximity, *Proceedings of the 19th international conference on World Wide Web*, Raleigh, NC.

Bruns, A., and Liang, Y. E. (2012) Tools and methods for capturing Twitter data during natural disasters, *First Monday*, *17*, 4.

de Albuquerque, J. P., Herfort, B., Brenning, A., and Zipf, A. (2015) A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management, *International Journal of Geographical Information Science*, 29, 4, 667-689.

Grace, R., and Leskovich, R. (2015) Challenges in the use of social media data for the next generation analyst, *Proceedings of SPIE Sensing Technology + Applications*, International Society for Optics and Photonics. Baltimore, MD.

Graham, M. W., Avery, E. J., and Park, S. (2015) The role of social media in local government crisis communications, *Public Relations Review*, 41, 3, 386-394.

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

Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., ... & Griffith, S. A. (2015) Social media and disasters: a functional framework for social media use in disaster planning, response, and research, *Disasters*, 39, 1, 1-22.

Johnson, I. L., Sengupta, S., Schöning, J., and Hecht, B. (2016) The Geography and Importance of Localness in Geotagged Social Media, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, Santa Clara, CA.

Jurgens, D., Finethy, T., McCorriston, J., Xu, Y. T., and Ruths, D. (2015) Geolocation prediction in Twitter using social networks: A critical analysis and review of current practice, *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media,* Palo Alto, CA.

Kerka, S. (2003) Community asset mapping, *Clearinghouse on adult, career, and vocational education: Trends and issues alert*, 47, 1-2.

Kitchin, R. (2014) The data revolution: Big data, open data, data infrastructures and their consequences, Sage, Thousand Oaks, CA.

Kogan, M., Palen, L., and Anderson, K. M. (2015) Think local, retweet global: Retweeting by the geographically-vulnerable during Hurricane Sandy, *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*, Vancouver, CA.

Krackhardt, David and Robert N. Stern (1988) Informal networks and organizational crises: an experimental simulation, *Social Psychology Quarterly*, 51, 2, 123-140.

Landwehr, P. M., and Carley, K. M. (2014) Social media in disaster relief, *Data mining and knowledge discovery for big data*, Springer, Berlin.

Mazer, J. P., Thompson, B., Cherry, J., Russell, M., Payne, H. J., Kirby, E. G., and Pfohl, W. (2015) Communication in the face of a school crisis: Examining the volume and content of social media mentions during active shooter incidents, *Computers in Human Behavior*, 53, 238-248.

McKnight, J. L., and Kretzmann, J. P. (1997) Mapping Community: Chapter 10 Capacity, *Community organizing and community building for health*, Rutgers University Press, New Brunswick, NJ.

Morstatter, F., Pfeffer, J., Liu, H., and Carley, K. M. (2013) Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose, *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, Cambridge, MA.

Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S. (2014) CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises, In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*,. Ann Arbor, MI.

Olteanu, A., Vieweg, S., and Castillo, C. (2015) What to expect when the unexpected happens: Social media communications across crises, *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*, Vancouver, Canada.

PEW Research Center. (2015) The evolving role of news on Twitter and Facebook, *Pew Research Center*, Retrieved from http://www.journalism.org/files/2015/07/Twitter-and-News-Survey-Report-FINAL2.pdf

Qu, Y., Huang, C., Zhang, P., and Zhang, J. (2011) Microblogging after a major disaster in China: A case study of the 2010 Yushu earthquake, *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, Hangzhou, China.

Rice, R. G., and Spence, P. R. (2016) Thor visits Lexington: Exploration of the knowledge-sharing gap and risk management learning in social media during multiple winter storms, *Computers in Human Behavior*.

Saleem, H. M., Xu, Y., and Ruths, D. (2014) Novel Situational Information in Mass Emergencies: What does Twitter Provide?, *Procedia Engineering*, 78, 155-164.

Starbird, K., and Palen, L. (2010) Pass it on?: Retweeting in mass emergency, *Proceedings of the 7th International ISCRAM conference,* Seattle, USA.

Starbird, K., and Palen, L. (2012) (How) will the revolution be retweeted?: information diffusion and the 2011 Egyptian uprising, *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, Seattle, WA.

Starbird, K., Muzny, G., and Palen, L. (2012, April) Learning from the crowd: collaborative filtering techniques for identifying on-the-ground Twitterers during mass disruptions, *Proceedings of the 9th International ISCRAM Conference*, Vancouver, Canada.

Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. (2010, February) 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*, Savannah, GA.

Tim, Y., Pan, S. L., Ractham, P., and Kaewkitipong, L. (2016) Digitally enabled disaster response: the emergence of social media as boundary objects in a flooding disaster, *Information Systems Journal*.

Twitter. (2016) Streaming API request parameters: Locations. Retrieved from https://dev.twitter.com/streaming/overview/request-parameters#locations

Veil, S. R., Buehner, T., and Palenchar, M. J. (2011) A work-in-process literature review: Incorporating social media in risk and crisis communication, *Journal of contingencies and crisis management*, 19, 2, 110-122.

Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. (2010, April) Microblogging during two natural hazards events: what twitter may contribute to situational awareness, *Proceedings of the CHI 2010 Conference on Human Factors in Computing Systems*, Atlanta, GA.

Wang, T. Y., Harper, F. M., and Hecht, B. (2014, November) Designing Better Location Fields in User Profiles, *Proceedings of the 18th International Conference on Supporting Group Work*, Sanibel Island, FL.

Whittaker, S. (2011) Personal information management: from information consumption to curation, *Annual Review of Information Science and Technology*, 45, 1, 1-62.