# Spatial, temporal, and content analysis of Twitter for wildfire hazards

**3 authors:**

Zheye Wang
Kent State University
**15** PUBLICATIONS   **58** CITATIONS

SEE PROFILE

Xinyue ye
Kent State University
**158** PUBLICATIONS   **884** CITATIONS

SEE PROFILE

Ming-Hsiang Tsou
San Diego State University
**84** PUBLICATIONS   **1,086** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   GeoPrivacy View project

Project   Computational Social Science View project

CrossMark

ORIGINAL PAPER

# Spatial, temporal, and content analysis of Twitter for wildfire hazards

Zheye Wang[1] · Xinyue Ye[1] · Ming-Hsiang Tsou[2]

**Abstract** Social media data are increasingly being used for enhancing situational awareness and assisting disaster management. We analyzed the wildfire-related Twitter activities in terms of their attributes pertinent to space, time, content, and network, so as to gain insights into the usefulness of social media data in revealing situational awareness. Findings show that social media data can characterize the wildfire across space and over time, and thus are applicable to provide useful information on disaster situations. Second, people have strong geographical awareness during wildfire hazards and are interested in communicating situational updates related to wildfire damage (e.g., containment percentage and burned acres), wildfire response (e.g., evacuation), and appreciation to firefighters. Third, news media and local authorities are opinion leaders and play a dominant role in the wildfire retweet network.

## 1 Introduction

As more and more fire-prone areas have been urbanized, people's livelihoods in the western USA have been severely influenced by the increasingly frequent wildfires (Pyne 2004; Collins 2008; Collins and Bolin 2009). In October 2003, the Cedar Fire—the largest fire in California history—caused a huge damage to San Diego County (Goldstein 2008). In May 2014, San Diego County witnessed several destructive wildfires within 1 month (see Table 2).

✉ Xinyue Ye
  xye5@kent.edu

[1] Department of Geography, Kent State University, Kent, OH, USA

[2] Department of Geography, San Diego State University, San Diego, CA, USA

The increasing wildfire activities, with the associated risks for nature and society, have attracted attention from researchers as well as emergency managers (Rodrigues and de la Riva 2014). In order to achieve a better understanding of the occurrences and patterns of spread of wildfires, efforts by domain scientists from various perspectives have included wildfire exposure modeling (Ager et al. 2014a, b; Thompson et al. 2015; Youssouf et al. 2014), wildfire risk assessment (Chuvieco et al. 2010, 2012; Martínez et al. 2009; Padilla and Vega-García 2011; Rodrigues et al. 2014), wildfire and wildland–urban interface (WUI) (Herrero-Corral et al. 2012; Massada et al. 2009; Schulte and Miller 2010), wildfire–climate interactions (Gillett et al. 2004; Liu et al. 2014; Westerling et al. 2006), and among others. To deal with the risks posed by wildfires at the early stages and reduce increased costs, wildfire management agencies have incorporated various wildfire detection systems, e.g., the general public, lookout towers, terrestrial mobile brigades, and aerial reconnaissance (Rego et al. 2013). The Wildland Fire Decision Support System (WFDSS) has also been developed to provide advanced tools for burn probability modeling and exposure analysis thus to improve the real-time wildfire suppression decision making (Calkin et al. 2011).

Although many studies and practices have been conducted in dealing with wildfire issues, most of them were not from a human-centric perspective and omitted the wildfire-related human behaviors (Slavkovikj et al. 2014). This might be due to lack of available data and the fact that the collection of survey data often required a large amount of time and budget. As an emerging human-centric sensing technology, social media outlets, such as Twitter, Facebook, or LinkedIn, have gone beyond a platform for sharing people's personal life and have become a data source for possible looks of people's behavioral patterns (Srivastava et al. 2012; Tsou and Leitner 2013; Tsou et al. 2013; Young 2014). Consequently, an increasing number of studies have started to characterize the way people becoming aware of, responding to, and recovering from disasters using social media data. According to Endsley (1995), situational awareness is "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." Disaster-relevant messages communicated by social media users have been used to analyze how people improve their situational awareness through the information contributed by others (Vieweg et al. 2010; De Albuquerque et al. 2015).

Space and time are strongly related to situational awareness in emergency events. Existing studies have revealed that social media users who are geographically proximate to the events are more likely to produce useful information for improving situational awareness. De Albuquerque et al. (2015) carried out a spatial analysis and found a strong spatial relationship between locational proximity to floods and the usefulness of the messages for crisis management. By analyzing the earthquake-related tweets in Japan, Acar and Muraki (2011) found that "people in directly affected areas tend to tweet about their unsafe and uncertain situation while people in remote areas post messages to let their followers know that they are safe." In addition, the temporal evolution of emergency events and the corresponding Twitter activities have been proved to be somewhat concurrent. Guan and Chen (2014) found that the ratio of tweets associated with Hurricane Sandy to general tweets increased gradually before this disaster, peaked when it landed, and then gradually decreased. Huang and Xiao (2015) indicated that messages posted by Twitter users varied with the temporal process of a disaster and thus could provide useful information for improving situational awareness at different stages of a disaster i.e., preparedness, response, impact, and recovery.

Besides analyzing spatial and temporal characteristics of disaster-related social media data, some studies focused on mining the actual content of social media messages to

improve knowledge about disaster situations. This was usually carried out by a data reduction process like classification, as these user-generated messages are extremely varied and some of them are not informative or relevant. Qu et al. (2011) divided the earthquake-related microblog messages with valuable information for improving situational awareness into four categories i.e., situation update, opinion expression, emotional support, and calling for action. Cameron et al. (2012) developed a platform for emergency situation awareness, which could detect emergent incidents and classify tweets as interesting or not. Imran et al. (2013a b) utilized machine learning methods to extract informative Twitter messages that can augment situational awareness and classify them into "fine-grained" classes, i.e., caution and advice, casualties and damage, donations, people, information sources and other. Imran et al. (2014) further designed an Artificial Intelligence for Disaster (AIDR) platform to automatically classify emergency-related Twitter messages into a set of user-defined situational awareness categories in a timely manner.

According to the two-step flow of communication theory, there are "gatekeepers" who filter and interpret information using their own perceptions before passing it on to the public (Xu et al. 2013). On social media, these "gatekeepers" are usually elite users or opinion leaders from whom the general public acquires information. In disaster situations, people may also tend to obtain situational updates and gain situational awareness from the informative messages shared by opinion leaders. However, to our knowledge, few studies have been conducted to investigate who the opinion leaders are and what roles they have played in the information exchange network related to disasters using social media data. Several exceptions to this include works by Cheong and Cheong (2011), Kogan et al. (2015), and Starbird and Palen (2010). Using social network analysis, Cheong and Cheong (2011) found that local authorities, traditional media reporters, and, etc. are important players in spreading situational information during 2010–2011 Australian floods. Kogan et al. (2015) indicated that local government authorities and the media are the most important nodes in the retweet network during the 2012 Hurricane Sandy. A similar phenomenon was also observed by Starbird and Palen (2010), "users are more likely to retweet information originally distributed through Twitter accounts run by media, especially the local media, and traditional service organizations."

Space, time, content, and network are all important attributes of emergency-related social media data and should be fully used to gain insights into emergency situational awareness. This paper presents the findings from examining the spatial and temporal variations of wildfire-related tweets and from our attempt to characterize wildfire by the discussion topics in the collected tweets, as well as from investigating the role of opinion leaders in people's acquisition of wildfire-related information. In the following sections, we first introduce our data and related methodology we used in the research. We then discuss the findings and their implications. Finally, we discuss what the findings suggest and what future pursuits on this topic can be.

## 2 Data and methodology

### 2.1 Data

In the present research, we used Twitter search API (https://search.twitter.com/) to collect wildfire-related Tweets. Our collection process included two phases. In the first phase, any tweet that contained either of the two keywords—"fire" and "wildfire"—was collected so

as to generate a holistic picture of online "wildfires." Those tweets showed that people usually tweeted about wildfires along with the places where the wildfires occurred. The examples listed below provide some such instances:

Flames jumping San Dieguito Rd, Camino del Sur, Rancho Bernardo evacuations in order.

The fires are right behind my parents' house in San Marcos, Please pray.

In the second phase, to solicit more information about specific wildfires, tweets associated with specific wildfires were gleaned based on keywords which are places where wildfires occurred. The keywords were randomly selected from a list of places (see Table 2), which are San Marcos and Bernardo. As the tweets collected using toponyms as keywords may contain noises which have nothing to do with wildfire, we filtered out these noises by checking whether a "fire" or "wildfire" also appeared in the collected tweets. In summary, the first phase focused on collecting general tweets related to wildfires, while the second phase centered upon the tweets pertinent to specific wildfires. Tweets collected in the first phase could be used in analysis of all dimensions (i.e., space, time, content, and network), whereas those tweets gleaned in the second phase are of particular importance for spatial analysis as only by identifying the accurate ignition locations of specific wildfires can we investigate the influence of geography (distance) on the spatial distribution of people's responses.

Temporally, our study period spans from May 13, 2014, when the first wildfire occurred, to May 22, 2014, when most of the destructive wildfires were 100 % contained. Spatially, a radius of 40 miles was set to specify a circular area (centered at downtown) to cover the majority of San Diego County.

There are several metadata fields in every tweet, such as the user's name, the timestamp when tweet was created. The spatial analysis was possible only with tweets that had geographic information (described in the metadata field "coordinates"). There were less than 5 percent of downloaded tweets that had coordinates associated with them. We collected 1334 geotagged tweets in the "fire" and "wildfire" pool. In terms of specific wildfires, after filtering out the noises, we retained 106 geotagged tweets with "Bernardo" as the keyword and 149 with "San Marcos" as the keyword (see Table 1 for a data summary). Other tweets did not have coordinates because the devices used to tweet messages did not have the built-in global positioning systems turned on.

Twitter users could either write their own words or re-post another's tweet (i.e., retweet) to generate a text. We focused on the texts of the "wildfire" tweets (including own generated tweets and retweets) to identify people's conversational topics on Twitter during wildfire hazards. After that, we built up a retweet network using "wildfire" retweets to gain knowledge about who the opinion leaders are and what role they have played.

| Table 1 Data summary for the collected tweets | Keyword | Total tweets | Tweets with geo-coordinates |
|---|---|---|---|
| | Fire | 32,410 | 3.9 % with geo-coordinates |
| | Wildfire | 2084 | 3.2 % with geo-coordinates |
| | Bernardo | 2049 | 5.2 % with geo-coordinates |
| | San Marcos | 5002 | 3.0 % with geo-coordinates |

## 2.2 Methodology

Several specific methods were used in our study: kernel density estimation (KDE) was performed to analyze the spatial pattern of wildfire-related tweets; text mining to identify conversational topics; social network analysis to detect the opinion leaders in wildfire hazards.

KDE is generally used to detect hot spots of spatial point data. Here, this technique was used to create a smoothed map of the wildfire-related tweets. KDE imported the coordinates of tweets and exported a raster formatted map where each cell was assigned a value to represent the intensity level (Han et al. 2015). Concentrated cells with intensive values are hotspots. To deal with the impact of population, a dual kernel density estimation (Dual KDE) was employed to map the spatial patterns of tweets associated with two specific wildfires (i.e., Bernardo fire and San Marcos fire). The population data were obtained at census block level. To transfer the areal data to point data, the population value at each census block is assigned to its centroid. After that, KDE was performed to generate the population map in a raster format. The formula of Dual KDE is given as:

Dual KDE Map = Each Cell Value of Tweets Map/Each Cell Value of Population Map

The content analysis is comprised of two facets: a text mining for identifying important terms and term clusters in wildfire-related tweets; a social network analysis for detecting users' structure and behavior in wildfire retweet network. The text mining of wildfire-related tweets was conducted using the "tm" package in R 3.1.2 (Feinerer et al. 2008; Feinerer and Hornik 2014). Since the raw tweets are unstructured and noisy, we need to clean them before calculating the term frequency and clustering terms. With reference to Ghosh and Guha (2013), we cleaned the raw tweets by removing URLs and stop words, converting a word's different morphological variants to the word's base form, and so on. In this process, some commonplace but meaningless words such as California, San Diego, county, and news were removed. Noteworthy, since our tweets contained many toponyms which are made up of more than one word (e.g., 4S Ranch), we combined those words to make sure that each toponym be represented by one word to avoid double counting. After cleaning the raw tweets, we obtained a term-document matrix, where a row stood for a term and a column for a tweet (Zhao 2012). We then calculated the frequency of terms to check their variation in importance. To identify the conversational topics related to wildfire, we employed k-means clustering method to identify clusters where terms appeared in the same group. With this method, terms which appeared frequently in the same document were grouped into one cluster, and terms which were grouped into one cluster were more likely to be seen in the same document than those separated into different clusters.

The social network analysis was also based on the textual content of tweets. Retweet (abbreviated as RT) is a function provided by the Twitter website, which allows users to tweet the content that has been posted by others. We can directly identify retweets, because the text of a retweet always starts out like "RT@Username." In the retweet network, the nodes are users who retweet other users' messages, as well as users who are retweeted by others. Another R package—"igraph" (Csardi and Nepusz 2006)—was implemented to conduct the social network analysis.

**Table 2** Overview of the major wildfires in May, 2014. *Source*: complied from http://www.fire.ca.gov/

| Major wildfires | Time of outbreak (UTC) | Time of 100 % contained (UTC) | Location | Long/lat | Acres |
|---|---|---|---|---|---|
| Bernardo Fire | May 13, 11:00 | May 17, 20:14 | Off Nighthawk Lane, southwest of Rancho Bernardo | −117.133/33.003 | 1548 |
| Tomahawk Fire | May 14, 9:45 | May 19, 9:20 | Traveled from Naval Weapons Station, Fallbrook to Camp Pendleton | −117.285/33.353 | 5367 |
| Poinsettia Fire (Carlsbad fire) | May 14, 10:30 | May 17, 12:00 | Off Poinsettia Ln & Alicante Rd in Carlsbad | −117.278/33.112 | 600 |
| Highway Fire | May 14, 13:00 | May 15, 18:30 | Off Old Hwy 395 and I-15 in the Deer Springs area | −117.162/33.312 | 380 |
| River Fire | May 14, 12:12 | May 19, 9:20 | North River Road and College Blvd., Oceanside | −117.747/33.251 | 105 |
| Cocos Fire (San Marcos Fire) | May 14, 16:00 | May 22, 18:15 | Village Drive and Twin Oaks Road, San Marcos | −117.160/33.114 | 1995 |
| Freeway Fire | May 14, 17:43 | May 20, 11:30 | Naval Weapons Station, Fallbrook | −117.260/33.370 | 56 |
| Pulgas Fire | May 15 14:45 | May 21, 17:00 | Off Interstate 5 at Las Pulgas Rd, north of Oceanside | −117.463/33.303 | 14,416 |
| San Mateo Fire | May 16, 11:24 | May 20, 23:30 | In the Talega area of Marine Corps Base Camp Pendleton | −117.300/33.286 | 1457 |

# 3 Spatial and temporal analysis of wildfire Twitter activities

In this part, we analyze the spatial and temporal relationship between social media activities and wildfire disruptions from the following two perspectives. First, we checked the temporal evolution of wildfire tweets and then compared it with the wildfire's temporal information (i.e., time of outbreak and time of 100 % contained) collected from authoritative source. Second, we examined whether the impact areas are clusters of wildfire tweets or not. Table 2 demonstrates some basic spatiotemporal information of the major wildfires occurred in our study period. The CAL FIRE only provided the longitudes and latitudes for several fires, so the geo-coordinates for other fires were inferred from their locations (see the fourth column in Table 2). The information in Table 2 provides a basis for our spatial and temporal analysis. We can see from this table that six of the nine wildfires occurred on May 14, which could explain why May 14 experienced a sudden increase in wildfire tweets (as shown by Fig. 1). A temporally concurrent evolution of wildfire and its related tweets could also be observed from Figs. 2 and 3. More specifically, the Bernardo fire and San Marcos fire both had their corresponding tweets peak on the day after the breakout day. This 1-day time lag is probably because it takes time to spread information.

The spatial information (i.e., ignition locations) associated with wildfires are used to identify the impact areas. Figure 3 shows that downtown area is the largest hot spot in terms of the number of "fire" and "wildfire" tweets. This may be due to the fact that, although San Diego is far away from the ignition locations, a large population could still generate numerous Twitter activities. Additionally, the digital divide between downtown and other areas in San Diego County might also explain this, as people in urbanized areas have more access to information and communication technologies (ICTs) than people in other areas do. Although non-spatial factors (e.g., population and digital divide) could explain to some extent the spatial pattern of wildfire tweets, geography still matters. To filter out the influence of population, dual KDE was performed to detect the clusters of tweets related to Bernardo fire and Cocos fire (see Figs. 4, 5 respectively). As shown by Figs. 4 and 5, the downtown area has become a low-value cluster, whereas clusters with values higher than medium are close to the wildfires' ignition locations. Considering that the number of geotagged tweets on Bernardo fire and Cocos fire is small, we included a population map of San Diego County
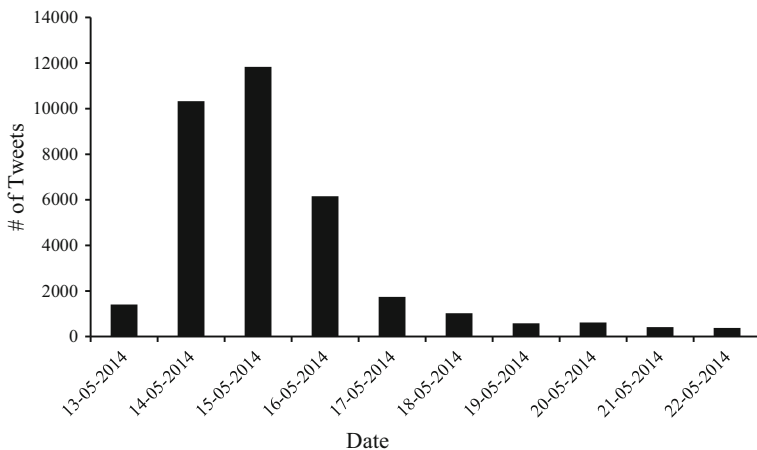


**Fig. 1** Temporal evolution of wildfire-related tweets with keywords of "fire" and "wildfire"
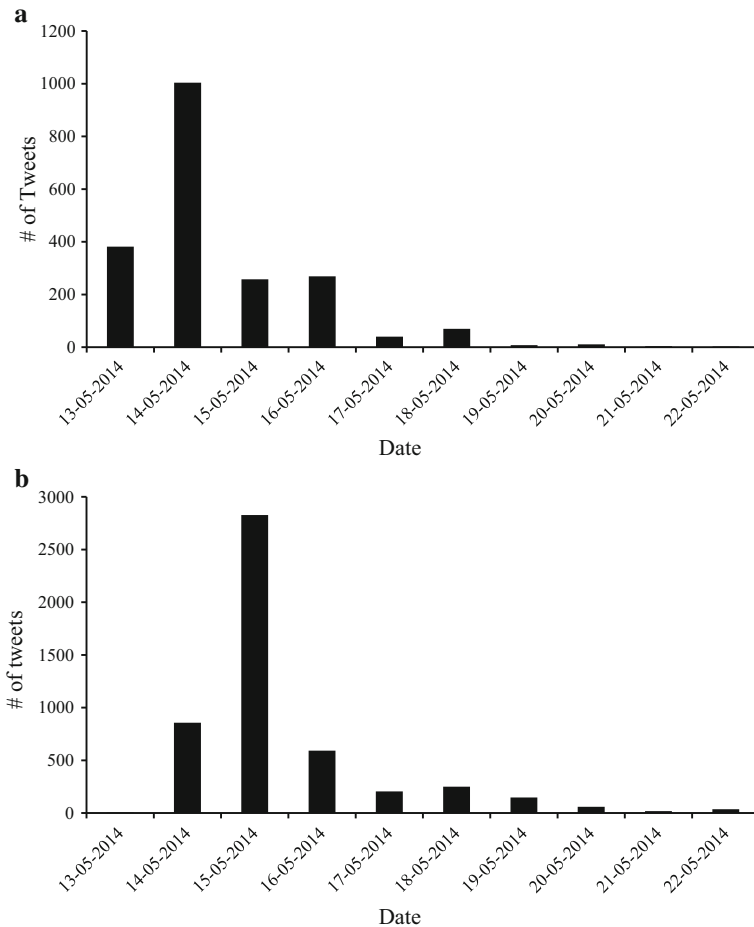
**a**



**b**



Fig. 2 Temporal evolutions of tweets with keywords including **a** Bernardo, **b** San Marcos

(Fig. 6) to exclude the impact of "small number problem." As seen from Fig. 6, areas around Bernardo are also heavily populated. Actually, Bernardo (Rancho Bernardo) is one of the populous areas in San Diego County (https://www.sandiego.gov). As such, our finding is consistent with those in previous studies. For example, Albuquerque et al. (2015) showed that Twitter messages which were geographically closed to flooded areas were more likely to be related to floods. Crooks et al. (2013) found that the majority of earthquake-related tweets originated from within the impact area. Our research provides new evidence on the relationship between geography and emergency-related social media activities.

## 4 Topics and network

We first look at the importance of a term in tweets. Figure 7 shows us the top 10 frequent words. If a term appears frequently in tweets, it is regarded as important. Hence, as shown by Fig. 7, the most important term is "evacuate." This is consistent with what we
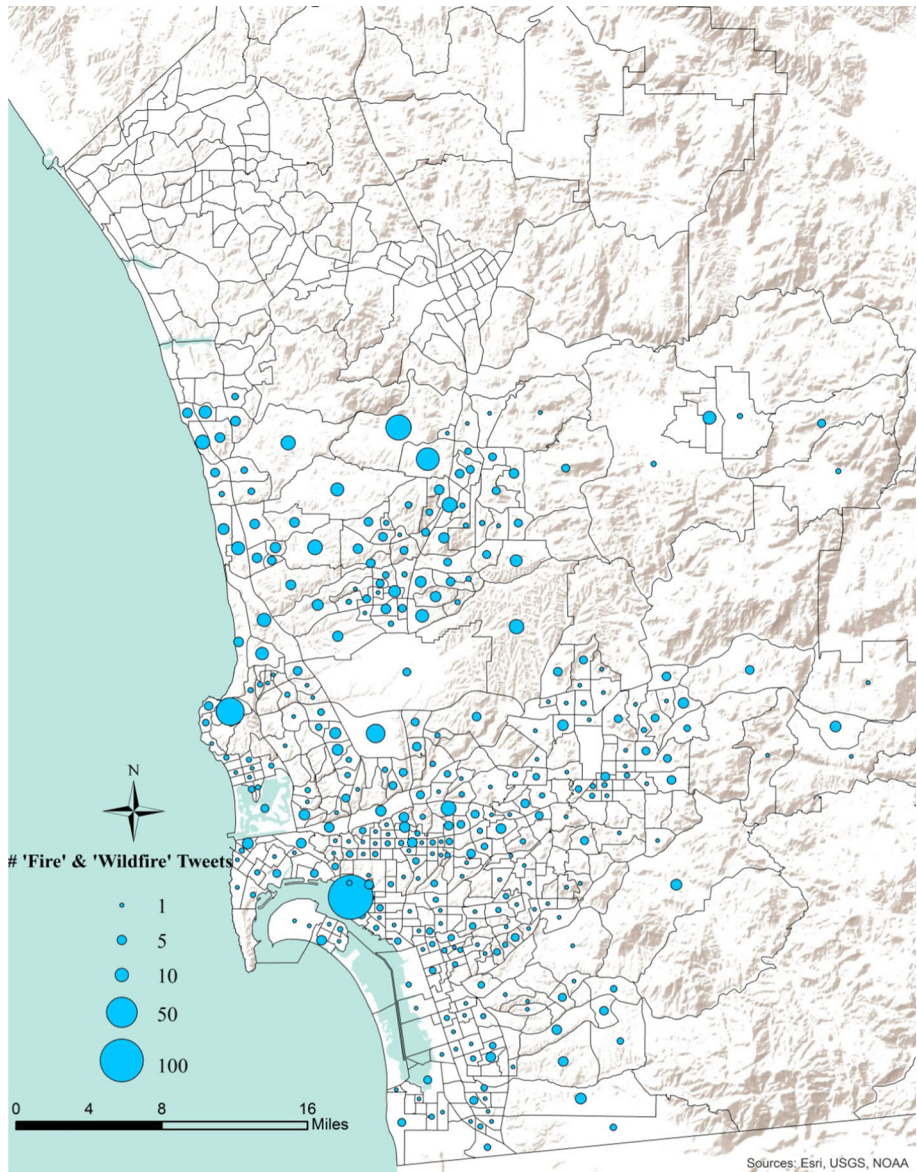
**Fig. 3** Spatial distribution of geotagged "fire" and "wildfire" tweets

expected, because the most urgent thing in wildfire situations is to evacuate and people always try to inform as many people as possible to evacuate. In these "evacuate" tweets, a large part talked about the evacuation of homes, resulting in a high frequency of "home." For instance, someone tweeted "Check @KPBSnews for updates on Bernardo fire in San Diego County. 700 acres burned, 20,000 homes being evacuated" and anther one posted a similar message saying that "California wildfire prompts evacuation of 20,000 homes." "Home" was also jointly tweeted with "burn" and "Carlsbad" when some Twitter users
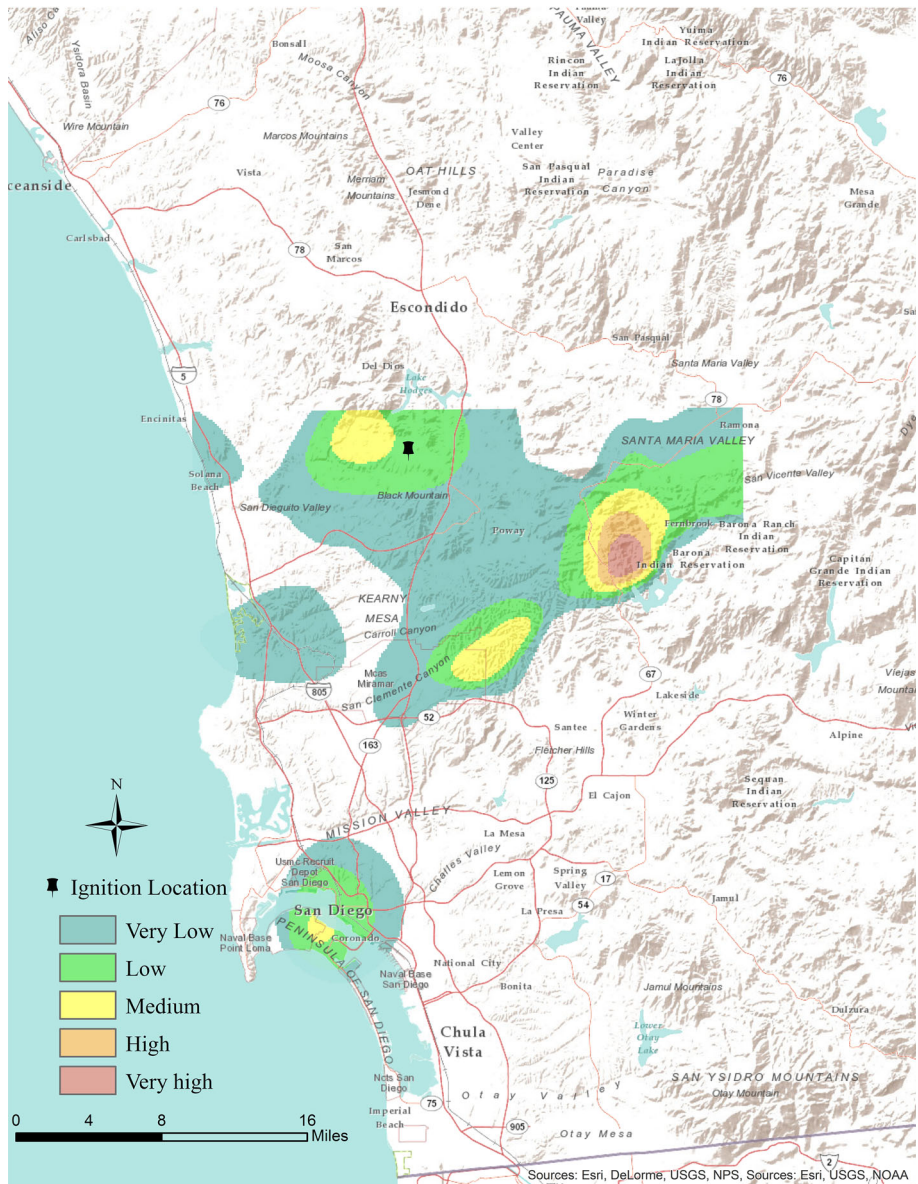
**Fig. 4** Dual kernel density estimation of geotagged tweets on Bernardo fire

reported the wildfire damage in Carlsbad such as "#CALFIRE official says wildfire has burned at least 30 homes in #Carlsbad. Homes all is same neighborhood. #PoinsettiaFire #CBS8." Similarly, "acre" was also used to indicate damage caused by wildfires like "California's wildfire season has ravaged nearly 10,000 acres." When people were not sure about the exact location of a wildfire, they tend to use "place name + area" to indicate a fuzzy place impacted by the wildfire. This could be evidenced by such tweets as
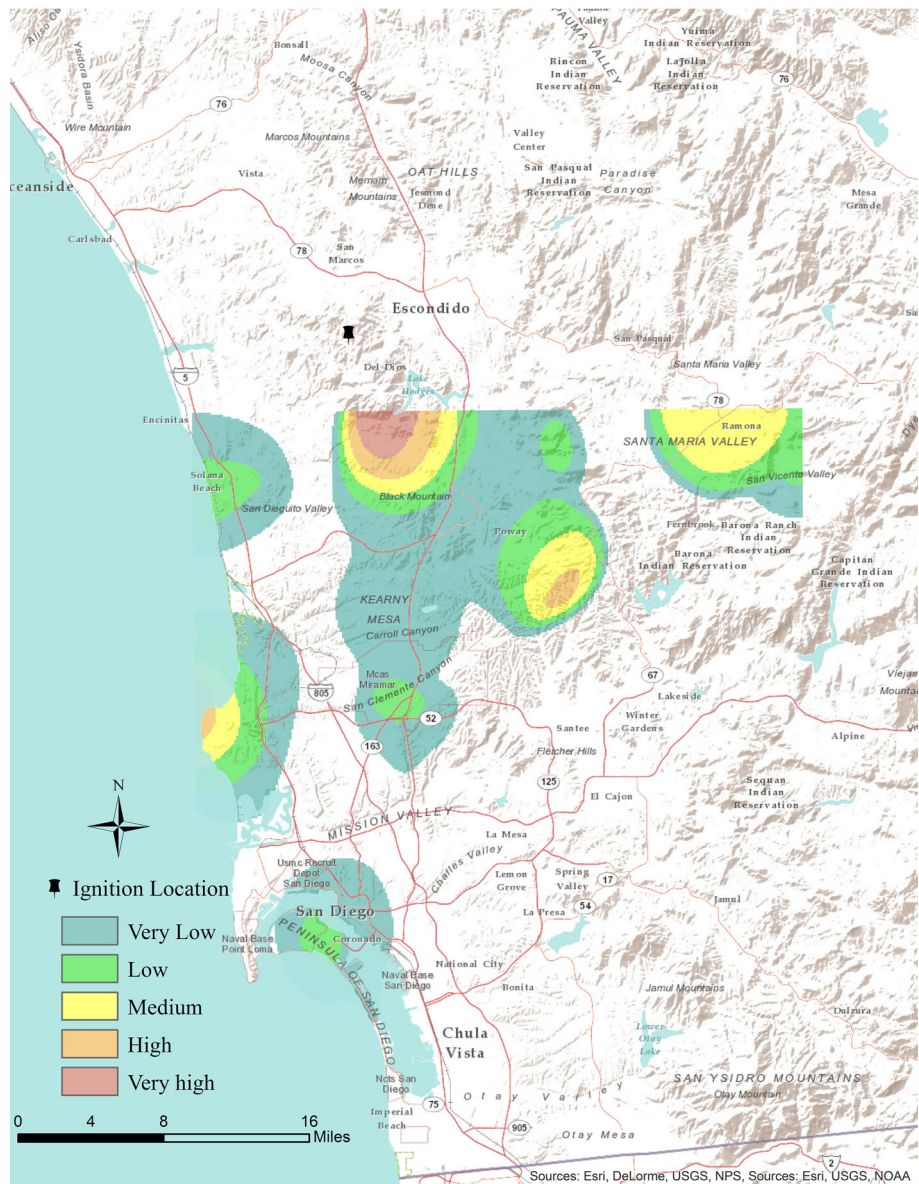
**Fig. 5** Dual kernel density estimation of geotagged tweets on Cocos fire

"evacuations ordered in 4S Ranch area due to brush fire." To explicitly label their tweets as situational updates, users added "now" or "update" in them. Two examples are shown here: "Now: San Diego County says evacuation orders over for all 20,000 homes in wildfire danger #BernardoFire" and "UPDATE: #BernardoFire is now 700 acres and 5 percent contained". Similar to "Carlsbad," "Bernardo" was tweeted because of the breakout of one wildfire in there. School kids expressed their emotions when they knew
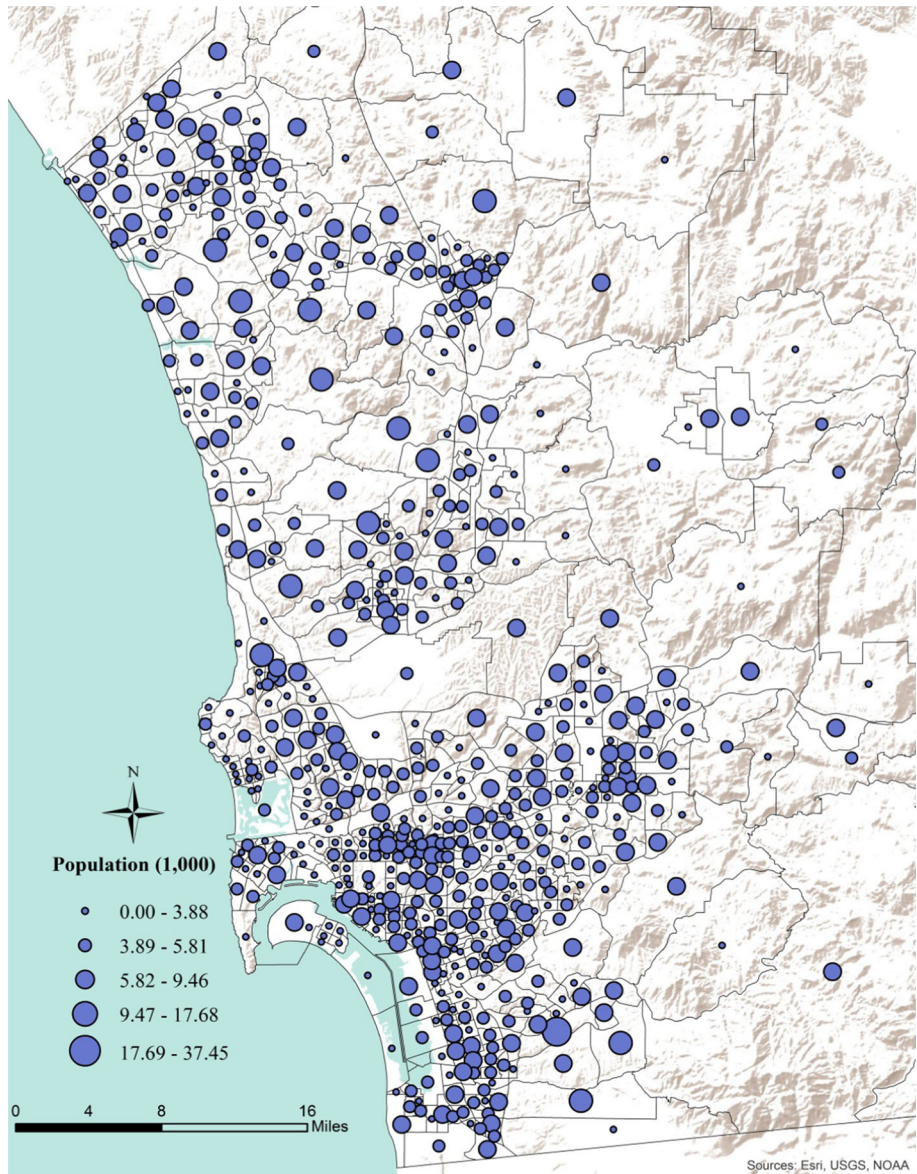
**Fig. 6** Spatial distribution of population in San Diego County

that their "school" would be closed. This could be exemplified by "Due to the wildfire in Carlsbad. School is canceled tomorrow. This is exciting. I know I'm not the only one."

Table 3 shows the seven clusters, and within each cluster, only top three terms are shown. The number of clusters specified here is to ensure that we get the most but differentiated topics. We can see from Table 3 that these clusters represent different topics. Specifically, cluster 1 stands for the topic related to people's thankfulness to firefighters; the topic revealed by cluster 2 is about the burned homes in Carlsbad; wind is a keyword in

**Fig. 7** Term frequency plot
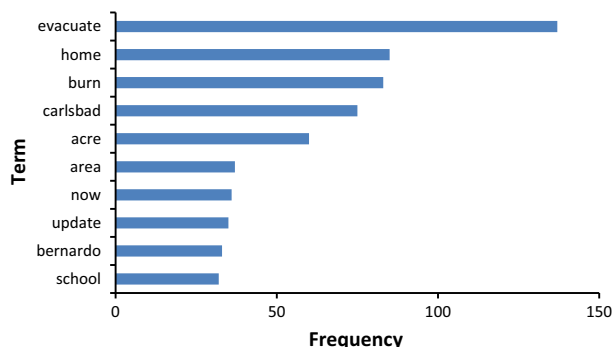


**Table 3** Term clusters in wild-fire tweets

| Number | Term clusters |
| --- | --- |
| Cluster 1 | know; thank; firefight |
| Cluster 2 | home; Carlsbad; burn |
| Cluster 3 | wind; Carlsbad; area |
| Cluster 4 | Carlsbad; contain; acre |
| Cluster 5 | burn; evacuate; 4S Ranch |
| Cluster 6 | acre; burn; contain |
| Cluster 7 | evacuate; home; Bernardo |

**Fig. 8** Indegree cumulative distribution of the retweet network
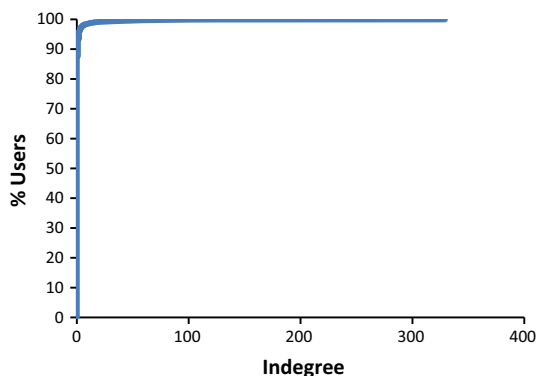


cluster 3, as it fanned the wildfire in Carlsbad area; a topic relevant to the containment percentage and impacted acres of Carlsbad wildfire is disclosed by cluster 4; cluster 5 represents the topic associated with the evacuation caused by a burning wildfire in 4S Ranch; cluster 6 is a topic on damage report, as it reveals the number of acres burned and the wildfire containment percentage; the last cluster talks about the evacuation of homes in Bernardo. These clusters reveal the main topics in the wildfire-related conversations on Twitter. People tweeted about wildfires together with the places where they occurred, as seen from cluster 2, cluster 3, cluster 4, cluster 5, and cluster 7. This reflects Twitter users' geographical awareness during wildfire events. People also communicate situational updates related to wildfire damage on Twitter, as seen from cluster 2, cluster 4, and cluster
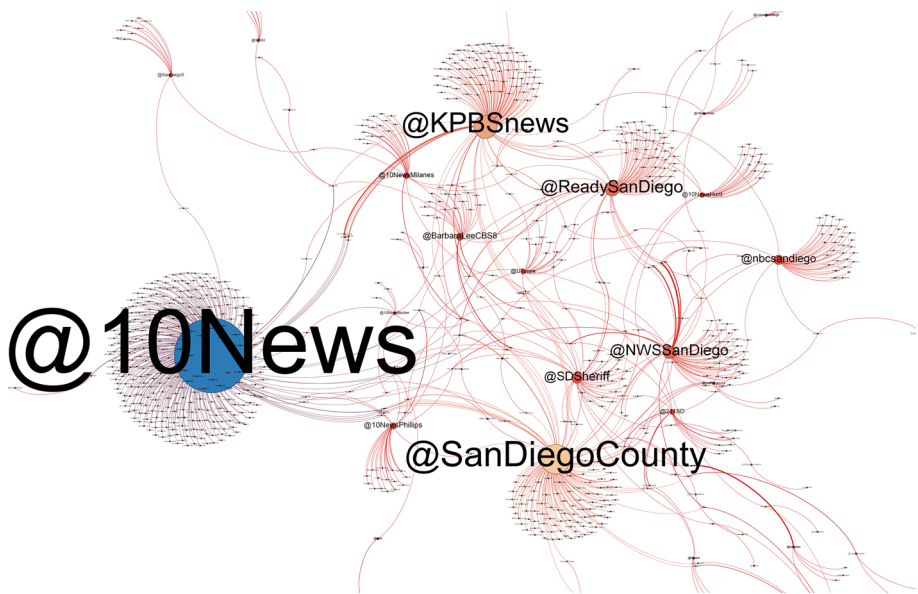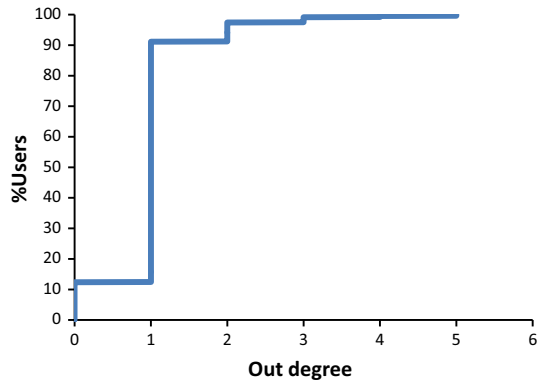
**Fig. 10** Major part of the retweet network

6. As shown by cluster 5 and 7, evacuations caused by wildfires are involved in tweets, representing that users are concerned about how to respond to wildfires. Different from other clusters, cluster 1 shows people's appreciation for firefighters.

The social network analysis was built based on the retweet relationship. In the retweet network, if user A retweets a message posted by user B, an edge runs from a node representing user B to another node representing user A, indicating that information has diffused from B to A. After building the network based on the retweet relationship, we calculated the indegree and outdegree for each node. The indegree of node A is represented by the times that user A has been retweeted by all other users. The outdegree of node A is the total times that user A has retweeted other users. Figure 8 shows the indegree cumulative distribution of the retweet network, from where we can see more than 85 % nodes had no users retweet their messages. Furthermore, according to Fig. 9, upward 90 % of

users retweeted only one user or none. The indegree and outdegree results show a polarized structure in the retweet network. That is, there are dominant users which act as hubs in the information exchange network during wildfire hazards. The major part of the retweet network was visualized to show its polarized structure.

As can be seen from Fig. 10, the node size is proportional to the times that it is retweeted by others. The nodes of @10news, @KPBSnews, and @nbcsandiego are Twitter accounts owned by three local news media in San Diego. Some accounts for the local government are also retweeted by numerous users. For example, @SanDieoCounty is an official Twitter account for the County of San Diego, @NWSSanDiego for the National Weather Service Office in San Diego, and @ReadySanDiego for the Office of Emergency Services in San Diego. We can see that people tend to acquire reliable information from either government or local news media during disasters. This provides implications for emergency management, since social media like Twitter could serve as a useful information propagation tool for emergency responders to improve the public's situational awareness.

# 5 Conclusion and discussion

Social media data are increasingly being used for enhancing situational awareness and assisting disaster management (Wang et al. 2015; Li et al. 2016; Yang et al. 2016). We analyzed the wildfire-related Twitter activities in terms of their inherent attributes, i.e., space, time, content, and network to gain insights into the usefulness of social media data in revealing situational awareness.

First, we analyzed the spatial and temporal patterns of wildfire-related tweets. Our analysis confirmed a temporally concurrent evolution of wildfire and wildfire-related Twitter activities. Meanwhile, a spatial coupling between wildfire disruptions and the related Twitter activities was also observed. Hence, social media data can characterize the disaster across space and over time, and thus are applicable to provide knowledge associated with disaster situations.

Second, people's conversations on social media tend to vary highly in terms of their subjects. Mining topics can reduce data chaos and extract useful information to enhance situational awareness and accelerate disaster response. We found that people's geographical awareness is strong during emergency events, and people are also interested in communicating wildfire damage (e.g., containment percentage and burned acres), wildfire response (e.g., evacuation), and appreciation to firefighters.

Third, opinion leaders play an important role in wildfire retweet network. We found that some elite users such as local authorities and traditional media reporters are dominant in the retweet network, which is consistent with the findings in previous studies. This polarized structure of retweet network has both advantages and disadvantages. On the one hand, situational announcements from authoritative sources are accurate and objective. On the other hand, eyewitness reports might not be able to attract sufficient attention.

There are drawbacks in our research which should be considered in our future pursuits on this topic. First, although the searching range could cover the majority of San Diego County, some places (e.g., Carlsbad) where wildfire occurred were not contained. Second, the 1 % sample limitation may lead to question that whether the sampled data are a valid representation of the overall wildfire Twitter activities. Third, the social network in our research is only based on the retweet relationship, while other types of network such as

followers network which is based on "who follows whom" could be used in future study. Fourth, the social network analysis centered on the investigation of opinion leaders in wildfire situation and thus overlooked the information diffusion process including its components, phases, and characteristics. Finally, since the four dimensions (i.e., space, time, content, and network) were separately analyzed, their simultaneous analysis might be able to provide some new insights into disaster management.

# References

Acar A, Muraki Y (2011) Twitter for crisis communication: lessons learned from Japan's tsunami disaster. Int J Web Based Communities 7(3):392–402

Ager AA, Day MA, Finney MA, Vance-Borland K, Vaillant NM (2014a) Analyzing the transmission of wildfire exposure on a fire-prone landscape in Oregon, USA. For Ecol Manag 334:377–390

Ager AA, Day MA, McHugh CW, Short K, Gilbertson-Day J, Finney MA, Calkin DE (2014b) Wildfire exposure and fuel management on western US national forests. J Environ Manag 145:54–70

Albuquerque JB, Herfort B, Brenning A, Zipf A (2015) A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. Int J Geogr Inf Sci. doi:10.1080/13658816.2014.996567

Calkin DE, Thompson MP, Finney MA, Hyde KD (2011) A real-time risk assessment tool supporting wildland fire decisionmaking. J For 109(5):274–280

Cameron MA, Power R, Robinson B, Yin J (2012) Emergency situation awareness from Twitter for crisis management. In: Proceedings of the 21st international conference companion on World Wide Web. ACM, pp 695–698, Apr 2012

Cheong F, Cheong C (2011) Social media data mining: a social network analysis of tweets during the Australian 2010–2011 floods. In: 15th Pacific Asia conference on information systems (PACIS). Queensland University of Technology, pp 1–16

Chuvieco E, Aguado I, Yebra M, Nieto H, Salas J, Martín MP et al (2010) Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. Ecol Model 221(1):46–58

Chuvieco E, Aguado I, Jurdao S, Pettinari ML, Yebra M, Salas J et al (2012) Integrating geospatial information into fire risk assessment. Int J Wildland Fire 23(5):606–619

Collins TW (2008) What influences hazard mitigation? Household decision making about wildfire risks in Arizona's White Mountains. Prof Geogr 60(4):508–526

Collins TW, Bolin B (2009) Situating hazard vulnerability: people's negotiations with wildfire environments in the US Southwest. Environ Manag 44(3):441–455

Crooks A, Croitoru A, Stefanidis A, Radzikowski J (2013) Earthquake: Twitter as a distributed sensor system. Trans GIS 17(1):124–147

Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJ Complex Syst 1695(5):1–9

Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. Hum Factors J Hum Factors Ergon Soc 37(1):32–64

Feinerer I, Hornik K (2014) tm: text mining package. R package version 0.6

Feinerer I, Hornik K, Meyer D (2008) Text mining infrastructure in R. J Stat Softw 25(5):1–54

Ghosh D, Guha R (2013) What are we 'tweeting' about obesity? Mapping tweets with topic modeling and geographic information system. Cartogr Geogr Inf Sci 40(2):90–102

Gillett NP, Weaver AJ, Zwiers FW, Flannigan MD (2004) Detecting the effect of climate change on Canadian forest fires. Geophys Res Lett 31(18). doi:10.1029/2004GL020876

Goldstein BE (2008) Skunkworks in the embers of the Cedar Fire: enhancing resilience in the aftermath of disaster. Hum Ecol 36(1):15–28

Guan X, Chen C (2014) Using social media data to understand and assess disasters. Nat Hazards 74(2):837–850

Han SY, Tsou MH, Clarke KC (2015) Do global cities enable global views? Using Twitter to quantify the level of geographical awareness of US cities. PLoS ONE 10(7):e0132464

Herrero-Corral G, Jappiot M, Bouillon C, Long-Fournel M (2012) Application of a geographical assessment method for the characterization of wildland–urban interfaces in the context of wildfire prevention: a case study in western Madrid. Appl Geogr 35(1):60–70

Huang Q, Xiao Y (2015) Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. ISPRS Int J Geo-Inf 4(3):1549–1568

Imran M, Elbassuoni SM, Castillo C, Diaz F, Meier P (2013a) Extracting information nuggets from disaster-related messages in social media. In: Proceedings of ISCRAM, Baden-Baden, Germany

Imran M, Elbassuoni S, Castillo C, Diaz F, Meier P (2013b) Practical extraction of disaster-relevant information from social media. In: Proceedings of the 22nd international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee, pp 1021–1024

Imran M, Castillo C, Lucas J, Meier P, Vieweg S (2014) Aidr: artificial intelligence for disaster response. In: Proceedings of the companion publication of the 23rd international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee, pp 159–162, Apr 2014

Kogan M, Palen L, Anderson KM (2015) Think local, retweet global: retweeting by the geographically-vulnerable during Hurricane Sandy. In: Proceedings of the 18th ACM conference on computer supported cooperative work and social computing. ACM, pp 981–993, Feb 2015

Li S, Ye X, Lee J, Gong J, Qin C (2016) Spatiotemporal analysis of housing prices in China: a big data perspective. Appl Sp An Policy. doi:10.1007/s12061-016-9185-3

Liu Y, Goodrick S, Heilman W (2014) Wildland fire emissions, carbon, and climate: wildfire–climate interactions. For Ecol Manag 317:80–96

Martínez J, Vega-Garcia C, Chuvieco E (2009) Human-caused wildfire risk rating for prevention planning in Spain. J Environ Manag 90(2):1241–1252

Massada AB, Radeloff VC, Stewart SI, Hawbaker TJ (2009) Wildfire risk in the wildland–urban interface: a simulation study in northwestern Wisconsin. For Ecol Manag 258(9):1990–1999

Padilla M, Vega-García C (2011) On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. Int J Wildland Fire 20(1):46–58

Pyne SJ (2004) Pyromancy: reading stories in the flames. Conserv Biol 18(4):874–877

Qu Y, Huang C, Zhang P, Zhang J (2011) Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In: Proceedings of the ACM 2011 conference on computer supported cooperative work. ACM, pp 25–34, March 2011

Rego F, Catry FX, Montiel C, Karlsson O (2013) Influence of territorial variables on the performance of wildfire detection systems in the Iberian Peninsula. For Policy Econ 29:26–35

Rodrigues M, de la Riva J (2014) An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ Model Softw 57:192–201

Rodrigues M, de la Riva J, Fotheringham S (2014) Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. Appl Geogr 48:52–63

Schulte S, Miller KA (2010) Wildfire risk and climate change: the influence on homeowner mitigation behavior in the wildland–urban interface. Soc Nat Resour 23(5):417–435

Slavkovikj V, Verstockt S, Van Hoecke S, Van de Walle R (2014) Review of wildfire detection using social media. Fire Saf J 68:109–118

Srivastava M, Abdelzaher T, Szymanski B (2012) Human-centric sensing. Philos Trans R Soc A Math Phys Eng Sci 370(1958):176–197

Starbird K, Palen L (2010) Pass it on? Retweeting in mass emergency. International Community on Information Systems for Crisis Response and Management

Thompson MP, Haas JR, Gilbertson-Day JW, Scott JH, Langowski P, Bowne E, Calkin DE (2015) Development and application of a geospatial wildfire exposure and risk calculation tool. Environ Model Softw 63:61–72

Tsou MH, Leitner M (2013) Visualization of social media: seeing a mirage or a message? Cartogr Geogr Inf Sci 40(2):55–60

Tsou MH, Yang JA, Lusher D, Han S, Spitzberg B, Gawron JM et al (2013) Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 US presidential election. Cartogr Geogr Inf Sci 40(4):337–348

Vieweg S, Amanda LH, Starbird K, Palen L (2010) Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1079–1088

Wang Y, Wang T, Ye X, Zhu J, Lee J (2015) Using social media for emergency response and urban sustainability: a case study of the 2012 Beijing rainstorm. Sustainability 8(1):25

Westerling AL, Hidalgo HG, Cayan DR, Swetnam TW (2006) Warming and earlier spring increase western US forest wildfire activity. Science 313(5789):940–943

Xu P, Wu Y, Wei E, Peng TQ, Liu S, Zhu JJ, Qu H (2013) Visual analysis of topic competition on social media. IEEE Trans Vis Comput Graph 19(12):2012–2021

Yang X, Ye X, Sui DZ (2016) We know where you are: in space and place-enriching the geographical context through social media. Int J Appl Geosp Res (IJAGR) 7(2):61–75

Young SD (2014) Behavioral insights on big data: using social media for predicting biomedical outcomes. Trends Microbiol 22(11):601–602

Youssouf H, Liousse C, Roblou L, Assamoi EM, Salonen RO, Maesano C et al (2014) Quantifying wildfires exposure for investigating health-related effects. Atmos Environ 97:239–251

Zhao Y (2012) R and data mining: examples and case studies. Academic Press, New York