Dr. Nikos Athanasis
University of Piraeus, Department of Digital Systems, Greece
athanasis@unipi.gr
Prof. Marinos Themistocleous
University of Piraeus, Department of Digital Systems, Greece
mthemist@unipi.gr
Prof. Kostas Kalabokidis
University of the Aegean, Department of Geography, Greece
kalabokidis@aegean.gr

The Convergence of Social 'Big Data' and Decision Support Systems for Natural Disasters Management

ABSTRACT

The influence of climate change trends and anthropogenic causes has radically increased the number of natural disasters worldwide over the last decades. A variety of decision support systems have been developed in different regions that provide crucial functionalities for natural disaster management based on models, simulations, and algorithms. Nevertheless, there are unpredictable factors during an emergency which are too complex to be evaluated in the current decision support systems. Facts about disasters collected from the small number of individuals located near the scene of a disaster have been proved to be the most useful when dealing with specific disaster situations. Unfortunately, this data is rare and difficult to locate within the greater sea of social media postings related to the disaster. Social media has emerged recently as a potential resource to improve the management of crisis situations such as disasters triggered by natural hazards. They enable the affected population to timely publicize an amount of disaster-related information and is considered as an additional information source for coping with natural disaster management. From the perspective of public authorities, the main challenge to using social media is the lack of control over the lineage of the information, and thereby an unknown reliability and trustworthiness. Since social media come from autonomous users, many of them may be highly noisy and lack authenticity. As a result, most public agencies have been more reluctant in adopting social media information than nongovernmental or volunteer organizations. This article presents a state-of-the-art approach towards the effective and accurate enhancement of decision support tools for natural disaster management with social media. A 'big data' - oriented approach is proposed, in order to cope with challenges of huge amounts of data, in different formats and varying quality that must be processed quickly. The novelty of the approach lies in the enrichment of geospatial content retrieved prediction simulations with real-time disaster-related information from tweet messages during an emergency. The article aims to highlight the role of geospatial big data such as geo-located tweet messages in the effective confrontation of natural disasters.

KEY WORDS

Big Data, Social media, Twitter, real-time disaster management, civil protection

Introduction

Social media has been used to disseminate a wide range of public safety information before, during and after natural disasters by providing assistance towards the establishment of situational awareness (Herfort et al., 2014). At the same time, the growing use of electronic devices equipped with Global Positioning System (GPS) receivers has increased the amount of geoinformation available in social media platforms (e.g. blogs, chat rooms, discussion forums, wikis, YouTube Channels, LinkedIn, Facebook and Twitter) and transformed them into location-based social networks (Beigi et al., 2016). Social media messages with a geographic reference can be described by the terms of Volunteered Geographic Information (VGI) (Stefanidis et al., 2013), Neogeography (Goodchild and Glennon, 2010) and Crowdsourcing (Gao et al., 2011).
Despite the intensive research activities regarding the contribution of social media in natural disasters management, many challenges are still open. The form of geo-social media is highly unstructured and thematically diverse, while valuable knowledge is often implicit and cannot be easily processed through automation (Sahito et al., 2011). Such data structure heterogeneity has a direct impact on the ability to store, manage or process effectively. Apart from data diversity, there is an enormous volume of social geospatial data – especially during emergencies – that must be analyzed as soon as possible. With all the high volume, high speed and varied structure of social media content, significant challenges emerge on how to deal with this 'big data' problem (Athanasis et al., 2017). Even though the amount of the available data is huge, reliable information is rare to find and difficult to locate within the enormous pool of social media postings (Landwehr and Carley, 2014). It is easier than ever to search the web for information, but filtering out falsehood and off-topic discussions from the huge online content still remains difficult (Herfort et al., 2014). Thus, it is still an open research question how emergency management agencies and the public can capitalize on the abundance of geo-social media by reducing the volume to credible and relevant content (Spinsati and Osterman, 2013).
This article presents a state-of-the-art approach towards an enhancement of decision support tools for natural disaster management with social media. Its novelty lies in the enrichment of geospatial content retrieved from Geographic Information Systems (GIS) modeling outcomes with real-time disaster-related information from social media during an emergency incident. Instead of solely relying on social media sources or 'a posteriori' analysis through classification (Lofi et al., 2012), or machine learning approaches (Terpstra et al., 2012), the applied methodology is based on the combination of spatial danger rating models with geo-social tweet messages. As a result, the volume of the underlying tweet messages that have to be processed is significantly reduced and the possibility to include erroneous messages is minimized. By using existing and well-studied geographical models for danger rating, the open problem of handling social media information during the occurrence of natural disasters can be tackled.

Related Work

A wide range of studies highlights the contribution of social media to natural disasters management. A survey about how individuals and organizations use social media in disaster events is described in Landwehr and Carley (2014). Social media visualization in location-based knowledge discovery has been analyzed in MacEachren et al. (2011) and Terpstra and deVries (2012).

For wildfires, De Longueville et al. (2010) use location-based social networks as a reliable source of spatiotemporal information, by analyzing the temporal, spatial and social dynamics of Twitter activity during a major forest fire event. Sinnappan et al. (2010) categorize tweets during the 2009 Black Saturday bushfires in Australia, while Sutton et al. (2008) describe the analysis of tweets in California during the wildfires of 2007. Vieweg et al. (2010) focus on the analysis of Twitter data during the Spring 2009 Red River Floods and Oklahoma grass fires events; they identified features of information generated during emergencies and described how Twitter can contribute to enhancing situational awareness. Spinsanti and Ostermann (2013) enrich VGI with geographic context found in Spatial Data Infrastructures (SDI) or other databases with a geographic component; they present a system designed to retrieve, process, analyze and evaluate social media content on forest fires by integrating authoritative data sources with VGI.
Regarding the contribution of social media in flooding management, Vieweg et al. (2010) and Starbird et al. (2010) analyze Twitter messages during the flooding of the Red River Valley in the United States and Canada in 2009, seeking to discern activity patterns and extract useful information. A similar approach is followed in Herfort et al. (2014) and De Albuquerque et al. (2015) who explore the relations between spatial information from social media messages and geographic information retrieved from hydrological data and official sensor data. Triglav-ˇCekada and Radovan (2013) show how volunteered geographical information has been used to map serious floods in Slovenia in 2012. Fuchs et al. (2013) followed a visual exploration and analysis methodology for a set of geolocated tweets from Germany regarding the severe flooding throughout Germany in the summer of 2013.
Regarding the analysis of social media in earthquakes response, Sakaki et al. (2010) and Crooks et al. (2013) investigated the use of Twitter for detecting and estimating the trajectory of earthquakes in real time. Acar and Muraki (2011) applied open-ended questionnaires to selected Twitter users and also analyzed the tweets sent in response to the Tohoku earthquake and the consequent tsunami in Japan. Earle et al. (2010) describe the contribution of Twitter in earthquake response.
A large number of related approaches focus on the contribution of VGI in disaster-related information but examine the social media as a stand-alone information source (e.g. in Acar and Muraki, 2011). In contrary, Tomaszewski et al. (2014) combine Federal Emergency Management Agency reports with related terms found in tweet messages. Spinsanti and Ostermann (2013) follow a ranking and clustering methodology of tweets by enriching the tweet context with different geospatial characteristics found in related SDIs. Albuquerque et al. (2015) follow a similar approach to enhance the identification of relevant messages from geo-social media as VGI and geographic features of flood phenomena derived from authoritative data (sensor data, hydrological data, and digital elevation models).
Even in the cases where the geo-social content is combined with GIS data from diverse sources, this accomplishment requires 'a posteriori' analysis of the messages mostly though classification (Albuquerque et al., 2015), machine learning (Sakaki et al., 2010) or natural language processing methods (Corvey et al., 2010). This analysis, however, adds crucial time overheads that hinder the timely and effective response to an emergency.

Methodology

The first pillar of the proposed methodology (Fig. 1) consists of the GIS modeling component that utilizes the Minimum Travel Time (MTT) algorithm (Finney, 2006), as a fire behavior prediction algorithm inside a web-based GIS system. By running fire simulations through the MTT algorithm, parameters such as major flow paths, spread rate, time of arrival and fireline intensity can be calculated. The perimeter polygon of the simulated fire restricts the area where the tweet messages will be filtered before shown in the web-based GIS system.

By the time a new fire is ignited and the perimeter has been calculated, tweet messages that consist of up to 140 Unicode characters are steadily analyzed by the geo-social component. The geographic location of any tweet message is described in the metadata field 'coordinates', which is also known as geo-tag. In general, users can geo-reference messages in Twitter either manually (e.g. by entering the name of a city in the field 'location') or automatically when a client application has access to the coordinates of a GPS receiver. Because in most situations only a small fraction of tweets are geo-referenced by users, an external gazetteer component of the geocoding API of ESRI1 is used. The component searches the tweet messages for place names (toponyms) and assigns coordinates if a toponym is found.

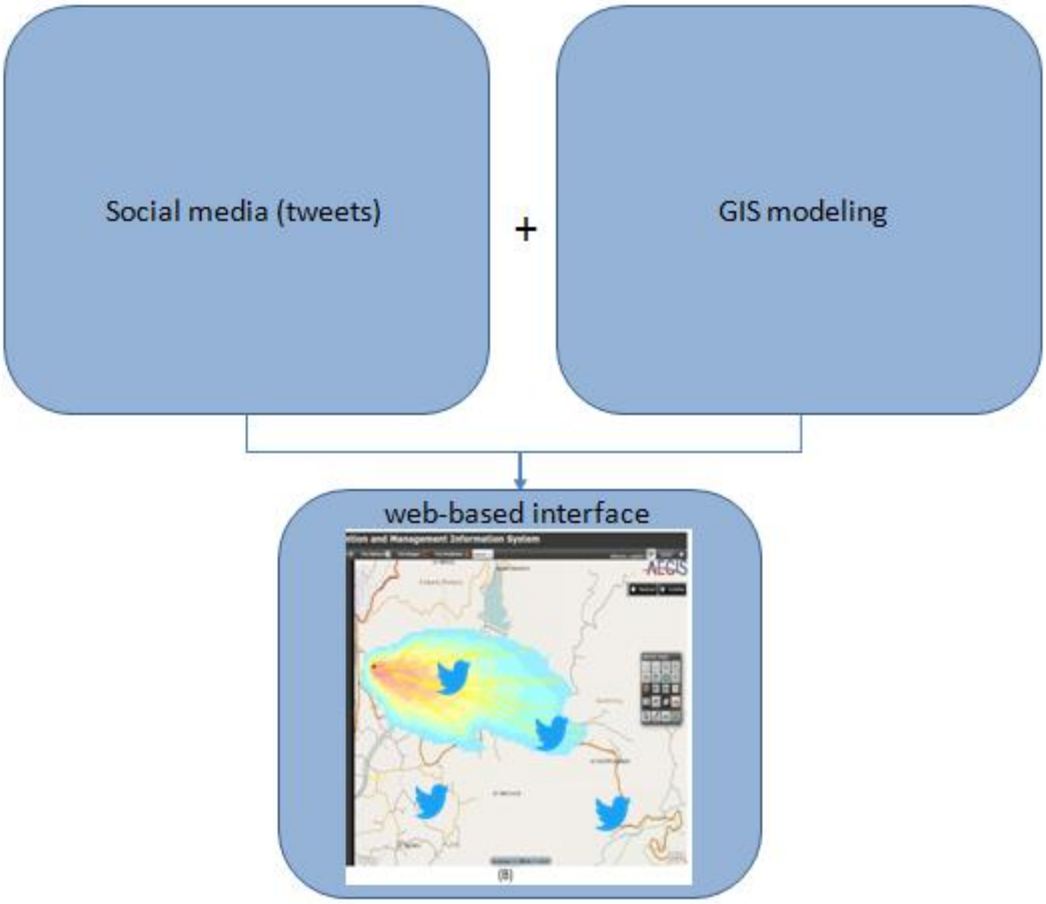

Fig. 1 – Conceptual design of the proposed methodology

For identifying messages containing relevant to the incident information, Twitter messages are filtered based on specific keywords that are common practice in the analysis of Twitter messages (Vieweg et al., 2010; Graham et al., 2012; Kongthon et al., 2012). Tweets containing the Greek keywords 'photia' or 'pyrkaya (meaning 'fire'), 'sismos' (meaning 'earthquake') are retained. By

[1] https://geocode.arcgis.com/arcgis

following the aforementioned approach, tweet messages can be visualized on top of a Web-GIS system.
In the core of the proposed architecture lies the Apache Kafka component, an open source stream processing platform (Fig. 2). It is a high scalable message queue storage, capable of process streaming data such as tweet messages. Apache Kafka works together with the Apache Hadoop2 framework. The Hadoop open-source work framework provides tools for organizing, managing and transforming large-scale data. On top of Hadoop runs the Hadoop Distributed File System (HDFS) that is a distributed file system designed to run on commodity hardware. Inside the big data cluster, virtual machines called workers receive the twitter messages and distribute them to the brokers, which are responsible for replicating the messages.

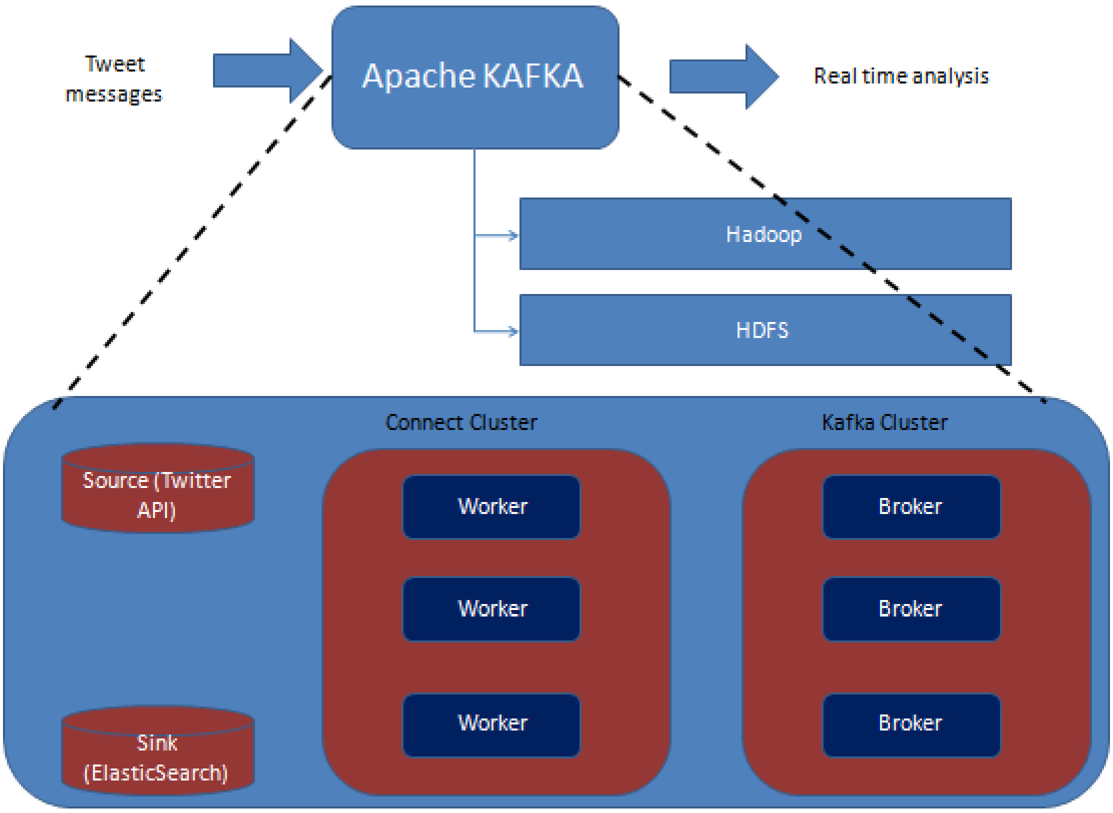

Fig. 2 – Architectural components of the proposed methodology

Tweet messages are retrieved from the Twitter Source by utilizing the Twitter API and stored in Kafka topics. The Kafka Connect API is utilized that receives messages from any sources (such as Twitter) and redirects them into related sinks (i.e. Cassandra[3], PostgreSQL[4], ElasticSearch[5]). In the proposed methodology, tweet messages are retrieved from the Twitter API (that is used as a source) and stored in a Kafka Topic. From the topic, the Producer API is used to connect the source (i.e. Twitter) to any Kafka topic as a stream of records for a specific category (i.e. a specific natural disaster event). From there, the consumer API is used to get out the tweeter messages from the Twitter topics into ElasticSearch, a distributed big data search and analytics engine capable of near real-time use cases.

[2] http://hadoop.apache.org/
[3] http://cassandra.apache.org
[4] https://www.postgresql.org/
[5] https://www.elastic.com

Fig. 3 describes how ElasticSearch is used for filtering out the off-topic tweets. The area of interest for the specific wildfire (i.e. the arrival time based on the MTT fire behavior modeling) is retrieved in a JSON format from the ArcGIS Server that is used to store all output results of the fire simulations. This polygon about the area of interest is used as a GeoPolygon Query inside the ElasticSearch big data store, to exclude the off-topic messages and visualize the meaningful messages through the web-based GIS visualization platform.

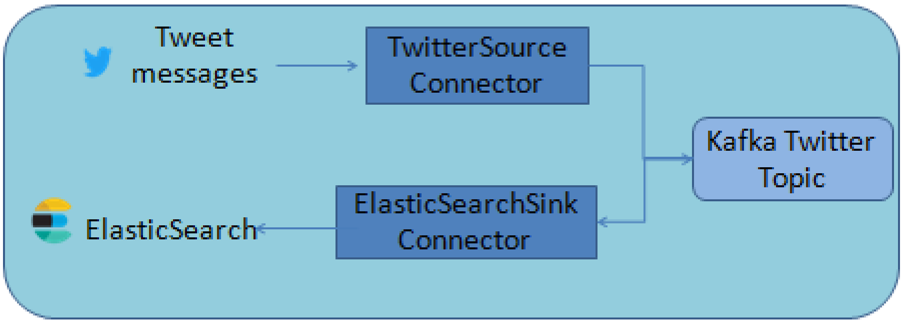

Fig. 3 – Architectural components of the proposed methodology

Results and Conclusion

The present work aims to highlight the role of the social big data, towards a more sophisticated transfer of knowledge among the civil protection authorities, emergency response crews, and the affected population. The results from our case studies show that social media content encloses potentially useful information and can act as an additional communication channel for citizens who have been affected by a disaster. Our approach follows a 'big data' architecture to cope with challenges of huge amounts of data, in different formats and varying quality that must be processed quickly. It is based on the enrichment of geospatial modeling results with real-time disaster-related information from social media during an emergency incident. Compared to similar studies, the added value is by combining wildfire behavior modeling outputs with tweet messages in order to increase the accuracy and efficiency of the tweets during an emergency.

Big Data technology emerges as a technology capable of successfully addressing contemporary digital challenges. The Big Data ecosystem provides high-volume, high-velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization.

The proposed highly scalable architecture relies exclusively on big data components. Thus, it can be applied to different geographical areas, to different types of social media and to a variety of natural disasters. Even though the Big Data ecosystem integrates many platforms and software components, it is mainly based on distributed storage and processing of very large datasets on computer clusters.

Our goal is to extend the proposed methodology and apply it first in a web-based system which will be tested by local authorities in Lesvos Island, Greece, throughout the next wildfire season. The objective is to avoid situations similar as to a forest fire in Lesvos that burned approximately 550 ha on 30 August 2015; a clear wind direction shift occurred when a fast-moving front inside a gully surpassed the mountain ridge after the first hours of fire spread (Kalabokidis et al., 2016). The fire prediction simulations of MTT were conducted for wildfire propagation, but the model did not take into consideration unpredictable factors such as the wind shift. The utilization of this system may potentially not only support civil protection and fire control services in the organization of effective wildfire management and control but also contribute to the immediate and massive alert of firefighters and/or people who are at risk during a fire outbreak.

REFERENCES

Acar, A. and Muraki, Y., 2011. Twitter for crisis communication: lessons learned from Japan's tsunami disaster. International Journal of Web Based Communities, 7(3), pp.392-402.

Athanasis, N., Themistocleous, M. and Kalabokidis, K., 2017. Wildfire Prevention in the Era of Big Data. In European, Mediterranean, and Middle Eastern Conference on Information Systems (pp. 111-118). Springer, Cham.

Beigi, G., Hu, X., Maciejewski, R. and Liu, H., 2016. An overview of sentiment analysis in social media and its applications in disaster relief. In Sentiment analysis and ontology engineering (pp. 313-340). Springer, Cham.

Corvey, W.J., Vieweg, S., Rood, T. and Palmer, M., 2010, June. Twitter in mass emergency: what NLP techniques can contribute. In Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media (pp. 23-24). Association for Computational Linguistics.

Crooks, A., Croitoru, A., Stefanidis, A. and Radzikowski, J., 2013. # Earthquake: Twitter as a distributed sensor system. Transactions in GIS, 17(1), pp.124-147.

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), pp.667-689.

De Longueville, B., Annoni, A., Schade, S., Ostlaender, N. and Whitmore, C., 2010. Digital earth's nervous system for crisis events: real-time sensor web enablement of volunteered geographic information. International Journal of Digital Earth, 3(3), pp.242-259.

Earle, P., Guy, M., Buckmaster, R., Ostrum, C., Horvath, S. and Vaughan, A., 2010. OMG earthquake! Can Twitter improve earthquake response?. Seismological Research Letters, 81(2), pp.246-251.

Finney, M.A., 2006, March. An overview of FlamMap fire modeling capabilities. In Fuels management—how to measure success: conference proceedings (pp. 28-30). USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S. and Stange, H., 2013, November. Tracing the German centennial flood in the stream of tweets: first lessons learned. In Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information (pp. 31-38). ACM.

Gao, H., Barbier, G. and Goolsby, R., 2011. Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intelligent Systems, 26(3), pp.10-14.

Goodchild, M.F. and Glennon, J.A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth, 3(3), pp.231-241.

Herfort, B., de Albuquerque, J.P., Schelhorn, S.J. and Zipf, A., 2014. Exploring the geographical relations between social media and flood phenomena to improve situational awareness. In Connecting a digital Europe through location and place (pp. 55-71). Springer, Cham.

Kalabokidis K., Ager A., Finney M., Athanasis N., Palaiologou P., Vasilakos C. 2016. AEGIS: a Wildfire Prevention and Management Information System. Natural Hazards and Earth System Sciences 16(3):643-661.
Landwehr, P.M. and Carley, K.M., 2014. Social media in disaster relief. In Data mining and knowledge discovery for big data (pp. 225-257). Springer, Berlin, Heidelberg.
Lofi, C., Selke, J. and Balke, W.T., 2012. Information extraction meets crowdsourcing: a promising couple. Datenbank-Spektrum, 12(2), pp.109-120.
MacEachren, A.M., Robinson, A.C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J. and Mitra, P., 2011. Geo-twitter analytics: Applications in crisis management. In 25th International Cartographic Conference (pp. 3-8).
Sahito, F., Latif, A., and Slany, W., 2011. Weaving Twitter stream into linked data: a proof of concept framework. In: Proceedings of the 7th international conference on emerging technologies, Islamabad, Pakistan, 1–6. doi:10.1109/ICET.2011.6048497.
Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, April. Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web (pp. 851-860). ACM.
Sinnappan, S., Farrell, C. and Stewart, E., 2010. Priceless tweets! A study on Twitter messages posted during crisis: Black Saturday. ACIS 2010 Proceedings, 39.
Spinsanti, L. and Ostermann, F., 2013. Automated geographic context analysis for volunteered information. Applied Geography, 43, pp.36-44.
Starbird, K., et al., 2010. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In: Proceedings of the 2010 ACM conference on computer supported cooperative work, 6–10 February, Savannah, GA. New York: ACM Press, 241–250.
Sutton, J.N., Palen, L. and Shklovski, I., 2008. Backchannels on the front lines: Emergency uses of social media in the 2007 Southern California Wildfires (pp. 624-632). University of Colorado.
Terpstra, T., De Vries, A., Stronkman, R. and Paradies, G.L., 2012. Towards a realtime Twitter analysis during crises for operational crisis management (pp. 1-9). Burnaby: Simon Fraser University.
Tomaszewski, B., Dickens, K., Sawant, A.A., Ogireddy, P.K., Yu, Q. and Holden, E., 2014, September. Visually contextualizing social media within spatial, temporal and thematic constraints for disaster situation awareness. In GeoVisual analytics: interactivity, dynamics, and scale, workshop@ GIScience (Vol. 23, pp. 1-4).
Triglav-ˇCekada, M. and Radovan, D., 2013. Using volunteered geographical information to map the November 2012 floods in Slovenia. Natural Hazards and Earth System Sciences, 13(11), p.2753.
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. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1079-1088). ACM.