Exploring the Geographical Relations Between Social Media and Flood Phenomena to Improve Situational Awareness

A Study About the River Elbe Flood in June 2013

Benjamin Herfort, João Porto de Albuquerque, Svend-Jonas Schelhorn and Alexander Zipf

Abstract Recent research has shown that social media platforms like twitter can provide relevant information to improve situation awareness during emergencies. Previous work is mostly concentrated on the classification and analysis of tweets utilizing crowdsourcing or machine learning techniques. However, managing the high volume and velocity of social media messages still remains challenging. In order to enhance information extraction from social media, this chapter presents a new approach that relies upon the geographical relations between twitter data and flood phenomena. Our approach uses specific geographical features like hydrological data and digital elevation models to prioritize crisis-relevant twitter messages. We apply this approach to examine the River Elbe Flood in Germany in June 2013. The results show that our approach based on geographical relations can enhance information extraction from volunteered geographic information, thus being valuable for both crisis response and preventive flood monitoring.

Keywords Social media · Volunteered geographic information · Disaster management · Flood · Situation awareness · Emergency management

1 Introduction

Managing an emergency puts high demands on authorities and crisis management organizations. Collecting as much information as possible about the unfolding crisis and making sense of that information in a timely manner is critical to subsidise relief efforts. One of the main challenges in emergency management is thus achieving

B. Herfort · J. P. de Albuquerque · S.-J. Schelhorn · A. Zipf
Chair of GIScience, Heidelberg University, Heidelberg, Germany

J. P. de Albuquerque (✉)
Department of Computer Systems, ICMC, University of Sao Paulo, Sao Carlos, Brazil
e-mail: jporto@icmc.usp.br

J. Huerta et al. (eds.), *Connecting a Digital Europe Through Location and Place*,
Lecture Notes in Geoinformation and Cartography, DOI: 10.1007/978-3-319-03611-3_4,

situation awareness, which can be defined as “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” (Endsley 1995).

Social media platforms like Twitter, Flicker or Instagram are growingly used by crisis-affected individuals. Hence, they are used to share local knowledge that can be a vital source of crisis-relevant information. Although this topic has been a subject of research in the recent years (Vieweg et al. 2010; Kongthon et al. 2012; Sakaki et al. 2010; Terpstra et al. 2012; Imran et al. 2013), the process of collecting and analysing social media information has to be further improved and evaluated to offer better insights and information that really contributes to situation awareness.

Scientific research on crisis management and social media has concentrated on filtering and classifying microblog posts, e.g. tweets, applying crowdsourcing (i.e. manual message classification by volunteers) (Gao et al. 2011; Rogstadius et al. 2011; Lofi et al. 2012) or natural language processing and machine learning (Sakaki et al. 2010; Terpstra et al. 2012; Imran et al. 2013). Nevertheless, a crucial problem remains unsolved. During a crisis, the volume and the velocity of posted tweets are extremely high. Distinguishing messages that contain critical information from off-topic messages in an efficient and reliable way is the basic requirement for any feasible approach for dealing with the information overload. However, existing approaches are only partially successful in this regard.

The use of existing and well-studied geographical models about natural hazards hold a non-exploited potential to tackle the open problem of handling social media information during a crisis. In the end, this could lead to crisis-relevant and actionable information, thus contributing to situation awareness and better decision-making. There is initial work in this field (e.g. Triglav-Čekada and Radovan 2013), but this is not as comprehensively studied as other approaches. This chapter thus seeks to contribute with a new approach for leveraging existing geographical knowledge of the flood phenomena in order to improve the usefulness of VGI in crisis management.

In the pursuit of this goal, we apply a geographical approach to prioritize crisis-relevant information from social media. Our methodology is based on specific geographical relations of flood phenomena, for example hydrological features and models of terrain and affected areas. Furthermore, we conduct a case study for the River Elbe Flood in Germany in June 2013 to validate our approach. Combining information from tweets, water level measurements and digital elevation models, we thus seek to answer the three research questions as follows.

- RQ1. Does the spatiotemporal distribution of flood-related tweets match the spatiotemporal distribution of the flood phenomenon?
- RQ2. Does the spatial distribution of flood-related tweets differ dependent on their content?
- RQ3. Is distance to flood phenomena a useful parameter to prioritize social media messages in order to improve situation awareness?

The remainder of this chapter is organized as follows. In Sect. 2, related work on social media and volunteered geographic information in the context of crisis management is presented. In Sect. 3 we present our research approach. Information about the case study and the different datasets we employed is given in Sect. 4. In Sects. 5 and 6 we present the methodology and the results of our study. Finally, Sect. 7 concludes the chapter by discussing our findings and possibilities for future work.

2 State of the Art

Social media is significantly influencing social interactions. Social media is a “disruptive technology” (Hiltz and Plotnick 2013), especially because it provides an alternative to traditional authoritative information from governmental institutions like civil protection or mapping agencies (Goodchild and Glennon 2010). The ability to communicate and share geographical data through simple, freely-available tools that can be quickly learnt without demanding professional or scientific background is key to this development. This is described by the terms Neography, Volunteered Geographic Information (VGI) and Crowdsourcing (Goodchild and Glennon 2010; Gao et al. 2011; Hudson-Smith et al. 2009).

Social media has also become a potentially useful tool during crises. Citizens adapted social media applications like social sites, document management, multimedia sharing, microblogging and geo-location systems to suit their crisis management needs. Twitter, for instance, enables victims to quickly connect with the rest of the world, and so can help to minimize the effects of catastrophes and supports disaster relief (Kaewkitipong et al. 2012). On the one hand, social media offers a new communication channel for government agencies to reach the media and informing affected citizens (Chatfield and Brajawidagda 2013). Therefore, the pervasive use of social media causes significant implications for emergency management practice and policy (Palen 2008).

Many studies have examined the way of extracting crisis-relevant information from social media messages (e.g. Yin et al. 2012; MacEachren et al. 2011; Imran et al. 2013). In particular, scientific research focused on twitter messages, so called tweets, using either crowdsourcing or machine learning techniques. Sakaki et al. (2010) investigated the real-time nature of twitter and were able to detect crisis-related twitter messages using a support vector machine (SVM). Kongthon et al. (2012) analysed twitter messages about the flood that affected Thailand in 2011, concluding that, due to its up-to-the-minute character, analysis and classification of twitter messages can be useful in coordinating resources and efforts and in preparing and planning for disaster relief. Imran et al. (2013) tested an automatic method for filtering crisis-relevant social media messages vis-à-vis a crowdsourcing approach, i.e. based on manual classification by volunteers. Their results show that machine learning can be utilized to extract structured information from unstructured text-based twitter messages.

Vieweg et al. (2010) analysed twitter messages referring to the Red River Floods in spring 2009. Graham et al. (2012) analysed the use of Twitter during the UK floods in November 2012, by mapping geo-referenced tweets mentioning the words "flood" and visually checking whether the distribution of tweets corresponds to rainfall data and official flood alerts. The authors conclude that the digital trails of twitter messages are mostly matched to official data on floods and metereological precipitation. Triglav-Čekada and Radovan (2013) gathered information about the November 2012 floods in Slovenia from VGI. Their research shows that volunteered image gathering is a comparable alternative to satellite imagery.

Furthermore, there are several studies that examine the use of social media as a tool for improving situation awareness during crises (Yin et al. 2012). The so-called "crisis maps", as exemplified by the Ushahidi platform (Okolloh 2009), are the most recent entrants to the social media field (Goolsby 2010). Meier (2012) compares the value that live crisis maps can provide for situation awareness with a bird's-eye view of an unfolding event. MacEachren et al. (2011) develop and implement tools for visually-enabled information foraging and sense-making.

Most of the extant research is focused on analysing data from social media as a stand-alone information source, although situation awareness should arise from the combination of different data sources. Gao et al. (2011) state that scientific data could augment VGI to provide more detailed insights on information requirements and needs during a disaster. The integration and fusion with official and scientific data sources could lead to progress in validating and verifying information gathered from social media and thus improve the fitness-for-use of VGI as a source for crisis relevant information. This is the direction pursued in the present study, which is applied to the case of the floods in the river Elbe basin in Germany in 2013.

3 Research Approach

This chapter addresses the problem of enhancing the extraction of useful information from VGI and social media for improving situation awareness during emergencies. In contrast to the approaches reviewed in the previous section, which resort to either crowdsourcing or machine learning, our approach is based on the geographical relations between flood phenomena and social media messages.

Inspired by Tobler's first law of geography (Tobler 1970), we assume that near things are more related than distant things. Regarding crisis events, this implies that the spatiotemporal characteristics of the catastrophe affect the spatiotemporal characteristics of VGI and social media messages. As such, our approach seeks to leverage existing knowledge about the spatiotemporal characteristics of flood phenomena to improve information extraction from social media. In doing this, the hypothesis posed here is that social media messages which are closer to the flooded areas are 'more related' to the unfolding event, thus being more useful for improving situation awareness.

Our approach explores the relations between spatial information from twitter messages and the knowledge about flood phenomena both from hydrology and official

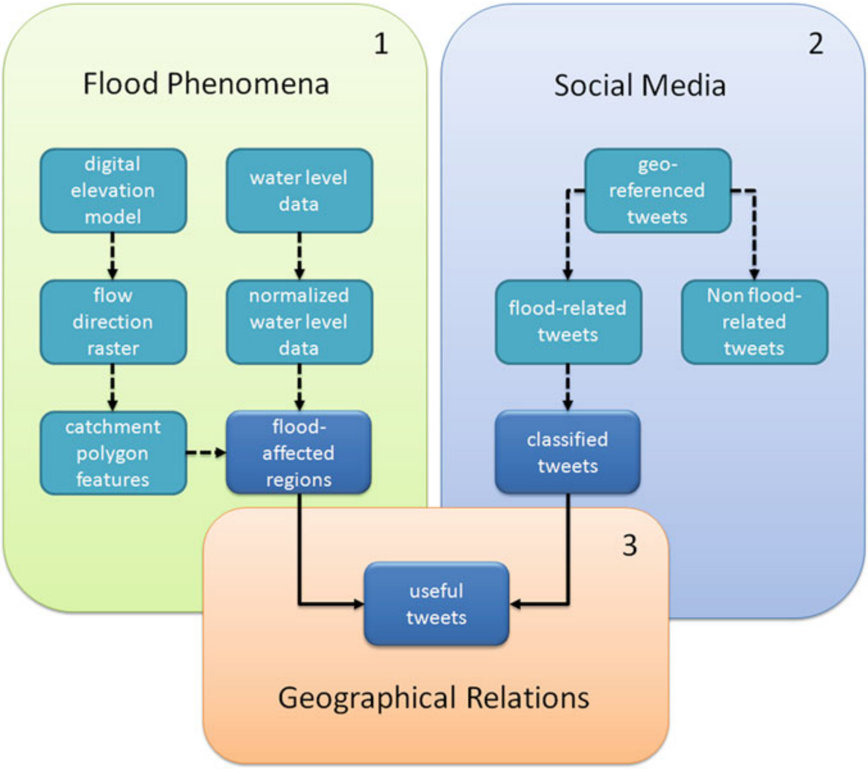

Fig. 1 Research approach

sensor data. The goal is to test our hypothesis that the distance to flood phenomena is a useful resource to prioritize messages for improving situation awareness.

Figure 1 schematically depicts our approach. It is divided into three main components: (1) gathering information on flood phenomena, i.e. flood-affected regions; (2) gathering information from social media, i.e. georeferenced twitter messages; (3) analysing the geographical relations between the information on flood phenomena (1) and social media messages (2) to assess the usefulness of tweets.

In this chapter, this approach is applied to analyse the use of Twitter during the River Elbe flood in 2013.

4 Description of the Case Study and Datasets

This section provides a description of our case study followed by an explanation of the datasets we employed.

4.1 River Elbe Flood

In the period from 30th May to 3rd June 2013 extreme heavy rain affected large parts of eastern and central Europe. The distribution of precipitation in the basin of the rivers Elbe, Moldau and Saale reached values two to three times higher than that for an average June. This is equivalent to a centennial probability of occurrence. The soil was already highly saturated at this time due to a wet climate in May 2013. Therefore, the heavy rain rapidly resulted in surface runoff causing the severe flood situation. The monthly average flow was three to four times higher than the longstanding average and in some places even higher than the higher value ever recorded.

The same finding follows from the examination of the water level data. Some gauging stations measured values that were never recorded before. For instance, at "Magdeburg-Strombrücke" the water level reached 7.46 m. That is an increase by more than 70 cm compared to the former maximum. Another characteristic of the flood was the huge stretch of the flood wave. The alert phase 4 (the highest in Germany) that was announced by the government lasted for 6 days along the rivers Elbe, Mulde, Elster and Neiße in Saxony and Saxony-Anhalt. This implies that dikes and dams were at risk of destruction for almost a whole week. The water levels in general did not return to their normal state until 16th of June 2013 (Sächsisches Landesamt für Umwelt and Landwirtschaft und Geologie 2013).

4.2 Datasets

The Twitter dataset contains of 60.524 geo-referenced short text messages ("tweets") within the territory of Germany. Each message consists of 140 Unicode characters at a maximum. Besides the actual text message string every tweet contains several additional fields representing metadata, such as a UTC time when the tweet was created, entities like *hashtags* (i.e. keywords preceded by #) and URLs, as well as an integer representation of the unique ID or information about the user who posted the tweet. The geographic location of a tweet is described in the metadata field "coordinates". The inner coordinates array is formatted as geoJSON.[1]

Users can geo-reference messages in Twitter either manually (e.g. by entering the name of a city in the field "location") or automatically via a client application that access the coordinates of a GPS receiver. Unfortunately, only a small fraction (3 % is the estimated average) of tweets are currently georeferenced by users, and this consists of a limitation for analysis approaches based on the location like the current study.

Twitter offers a number of Application Programming Interfaces (APIs), which can be used for automatically recovering data. For this study, we queried the Twitter streaming API using the 1 % garden hose access, during the period from 08th June 2013 1.30 pm to 10th June 2013 midnight, and collected every geo-referenced tweet

[1] https://dev.twitter.com/docs/platform-objects/tweets

within a bounding box covering Germany. Afterwards we further filtered tweets by their location and excluded those outside the territory of Germany.

In addition to the twitter dataset, we also gathered official water level data from 54 monitoring stations along the rivers Elbe and Saale provided by the German Federal Waterways and Shipping Administration and the German Federal Institute for Hydrology through the German online gauge system "Pegel Online" via *web feature service*.[2] In this manner, our second dataset includes information about the location of each measurement station, the current water level, the average flood water level over a time period from 1st November 2000 to 31st October 2010, and the highest water level ever recorded. The water level measurements were provided in a 15-minute resolution for the whole period analysed.

As a third dataset, we used HydroSHEDS drainage direction information derived from elevation data of the Shuttle Radar Topography Mission (SRTM) at 3 arc-second resolution in order to compute hydrographical features of the river Elbe basin. This includes information about flow accumulation, stream network and catchment boundaries (Lehner et al. 2008).

5 Methodology

This section describes the detailed methodology used in this chapter, by further elaborating the procedures used to apply the approach described in Sect. 3 and schematically depicted in Fig. 1. The next section explains the steps conducted in preparing the datasets employed (Sect. 4.2), followed in Sect. 5.2 by the description of the analytical procedures used.

5.1 Data Preparation

The first step of our data preparation consisted of defining the flood-affected regions based on the digital elevation model (for catchment areas) and on official data (river water levels). Starting with the HydroSHEDS flow direction raster, based on SRTM elevation data, we computed catchment polygon features for each location where two streams flow together using the ArcHydro Toolset for ArcGIS. The detailed workflow is shown in Fig. 2. This way of proceeding guarantees that any cell within a catchment drains into the same stream. Catchments therefore contain no more than one stream by definition.

In the next step, we analysed the water level data collected from 54 water level measurement stations along the rivers Elbe and Saale. To assess the severity of the flood at the gauge station, we computed the difference between the daily maximum

---

[2] http://www.pegelonline.wsv.de/webservice/wfsAktuell

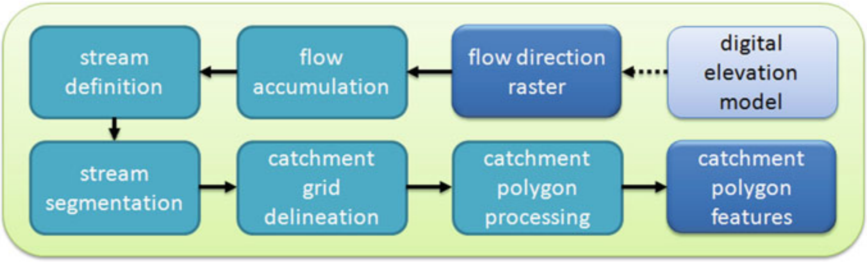

Fig. 2 Catchment processing workflow

water level and the average flood water level for the time period from 1st November 2000 to 31st October 2010.

In the third step, we combined both information on catchments and water levels based on the location of the monitoring stations. The normalized water level values were then matched to the corresponding catchment regions. If more than one water level measurement station was found to be within one given catchment region, we calculated the arithmetic mean of the values measured by those corresponding stations. If the computed flood level for a catchment exceeded the average flood water level by more than 100 cm it was considered "flood-affected".

The fourth step involved performing a content analysis of the twitter messages to identify messages that contain useful information. For doing so, we first filtered the twitter messages, sorting them out into the categories "flood-related" and "non-related". This was accomplished using keyword filtering as common practice in the analysis of twitter messages (e.g. Graham et al. 2012; Kongthon et al. 2012; Vieweg et al. 2010). Tweets containing the keywords in German "Hochwasser", "Flut", "Überschwemmung" ("Hochwasser", "Flut" and "Überschwemmung" are the German words meaning "flood") and the English word "flood", regardless of case-sensitivity, were considered "flood-related". The selection of these keywords was based on the definition of the German dictionary "Duden" for the word "Hochwasser". Furthermore, we included the additional words "Deich" (dike) and "Sandsack" (sandbag), which were found to be common in reports in the media.

Finally, we classified the flood-related tweets into thematic categories, based on a manual content analysis. The content-based classification of messages requires a well-defined set of categories, which heavily depends on the crisis context analysed, i.e. it varies for each crisis phenomenon and event. We analysed the categories proposed by Imran et al. (2013) ("caution and advice", "information source", "donation", "causalities and damages", "unknown") and Vieweg et al. (2010) (warning, preparatory activity, fire line/hazard location, flood level, weather, wind, visibility, road conditions, advice, evacuation information, volunteer information, animal management, and damage/injury reports). However, neither of the previous sets of categories was well suited for our case study, the River Elbe flood. We thus used these previous works as a guideline and adapted them to derive an own set of categories that we considered necessary.

Table 1 Thematic categories based on content analysis

Category	Description
"Volunteer actions" (VA)	Tweets related to flood combating Example: "Keine Ahnung, bin auf der Sandsackfüllstation in der Listemannstr. #Magdeburg #Hochwasser" ("I have no clue. I'm at the sandbag filling point at Listemann-street. #Magdeburg #flood")
"Media" (M)	Tweets related to media coverage, politicians and political events Example: "schaue mir das Hochwasser am Fernseher an. schrecklich. und dann gibt es auch noch Plünderer. unglaublich. #SpiegelTV" ("Watching the flood on TV, horrible, there are even looters, unbelievable. 'SpiegelTV")
"Traffic conditions" (TC)	Tweets related to road or rail traffic, traffic jams or other restraints Example: "Der #ice644 von Berlin nach Köln/Bonn soll übrigens fahren—aktuell aber zehn Minuten Verspätung. #hochwasser" ("By the way, the ice644 will go from Berlin to Cologne/Bonn, current delay 10 min #flood")
"Flood level" (FL)	Tweets related to hydrological or physical measurements, not only quantitative ("719 cm") but also qualitative information ("water level sinks") Example: "aktuelles Foto aus #Lostau: 08.30 Uhr Pegel MD Strombrücke: 719 cm #hochwasser http://t.co/uv3NkMMcIw" ("Latest photos from #Lostau: 08.30 am water level MD Strombrücke 719 cm #flood http://t.co/uv3NkMMcIw")
"Other" (O)	Tweets not related to any of the previous categories Example: "Ich wünsche den #Hochwasser betroffenen weiterhin alles Gute, und trotz alledem allen einen schönen #Sonntag" (To all #flood-affected people: Let's hope for the best. Despite all that, have a nice #Sunday")

We grouped flood-related twitter messages into five categories: "volunteer actions" (VA), "media" (M), "traffic conditions" (TC), "flood level" (FL) and "other" (O). Table 1 presents a detailed description of the categories and their characteristics.

5.2 Data Analysis Procedure

The analysis of our data was guided by our three research questions (see Sect. 1). For answering the first research question, we sought to determine whether the spatiotemporal distribution of flood-related tweets matches the spatiotemporal distribution of the flood phenomenon. For doing this, we first generated a density map by executing a kernel density function using ArcGIS software, in order to allow a visual analysis of the spatial distribution of tweets for the time period analysed (8th–10th June 2013). In the following step, we calculated the distance between each tweet and the nearest flood-affected catchment. In order to test if flood-related tweets are closer to the flooded areas than non-related tweets, we computed and compared the average distance of the two groups (flood-related tweets *versus* non-related tweets) using an independent sample t-test.

For answering our second research question, i.e. whether the spatial distribution of flood-related tweets differ depending on their content, we again firstly performed a visual analysis by producing density maps using the kernel density function. Next, we calculated the average distance for each of the categories of Table 1 and performed a post-hoc analysis (LSD) to test the mean distances depending on the categorization for statistical significance.

The final step of our study consisted of answering our third and overall research question by assessing to what extend the distance of messages to flood phenomena is a useful parameter for prioritizing social media messages in order to improve situation awareness. For doing this, we verified if the categories whose messages are closer to flood-affected areas are more useful to improve situation awareness than the categories with more distant messages. In this analysis, we considered social media messages that contain information which is not available through other sources as being more useful to improve situation awareness. As such, the criteria for defining the usefulness of social media messages that we adopt consists of the capacity to enrich and complement other information sources.

6 Results

The results of our study are presented in the following sections. The next section provides an exploratory description of the data collated, serving as a basis for the detailed analysis based on our research questions (Sect. 5.2).

6.1 Data Description

Figure 3 shows flood-affected catchments and the severity of the flooding calculated from digital elevation data and water level data for the time period from 8th to 10th June 2013. Comparing the three maps, one can visualize the shift of the flood peak from the upper reaches (southeast) in the map of 8th June to the lower reaches (north) in the map of 10th June. As such, on 8th June 2013 the catchments along the river Elbe in the federal state of Saxony were most affected, whilst the lower reaches of the river Elbe were not affected until 10th June 2013.

The results of the first classification of twitter messages based on keywords are listed in Table 2. Overall we examined 60,524 tweets within the territory of Germany. The majority (99.34 %) of them do not contain the keywords. These tweets were marked as “non-related”. For the period from 8th to 10th June 2013 we selected 398 tweets containing the keywords and marked these tweets as “flood-related”.

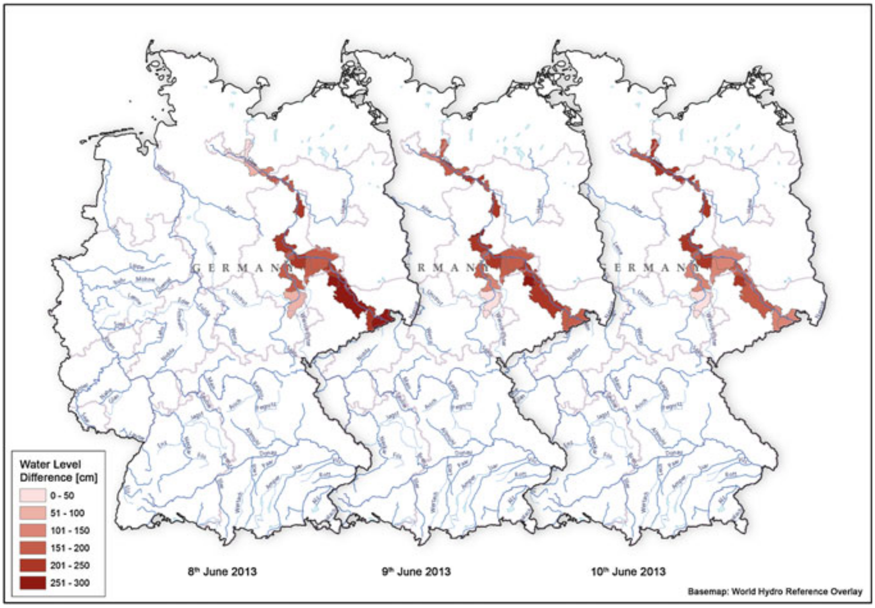

Fig. 3 Spatiotemporal distribution of flood-affected catchments based on official water level information

Table 2 Classification of twitter messages using keyword-based filtering

Period	8th–10th June 2013	8th June 2013	9th June 2013	10th June 2013
# all tweets	60,524 (100 %)	14,286 (100 %)	23,093 (100 %)	23,145 (100 %)
# flood-related tweets	398 (0.66 %)	75 (0.52 %)	197 (0.85 %)	126 (0.54 %)
# non-related tweets	60,126 (99.34 %)	14,211 (99.55 %)	22,296 (99.15%)	23,019 (99.46 %)

Table 3 Classification of twitter messages based on content analysis

Period	8th–10th June 2013	8th June 2013	9th June 2013	10th June 2013
# all tweets	398 (100 %)	75 (100 %)	197 (100 %)	126 (100 %)
# VA	113 (28.39 %)	24 (32.00 %)	59 (29.95 %)	30 (23.81 %)
# M	57 (14.32 %)	8 (10.67 %)	30 (15.23 %)	19 (15.08 %)
# TC	30 (7.53 %)	4 (5.33 %)	8 (4.06 %)	18 (14.29 %)
# FL	32 (8.04 %)	5 (6.67 %)	15 (7.61 %)	12 (9.52 %)
# O	126 (31.66 %)	34 (45.33 %)	85 (43.15 %)	47 (37.30 %)

Table 3 shows the distribution of flood-related tweets based on content classification. More than a quarter (28.39 %) of all flood-related tweets contain information referring to volunteer actions, whereas flood-related tweets referring to media, traffic conditions or flood level reach a much less share. About 30 % of the flood-related tweets were classified as "other" and therefore do not contain any viable information.

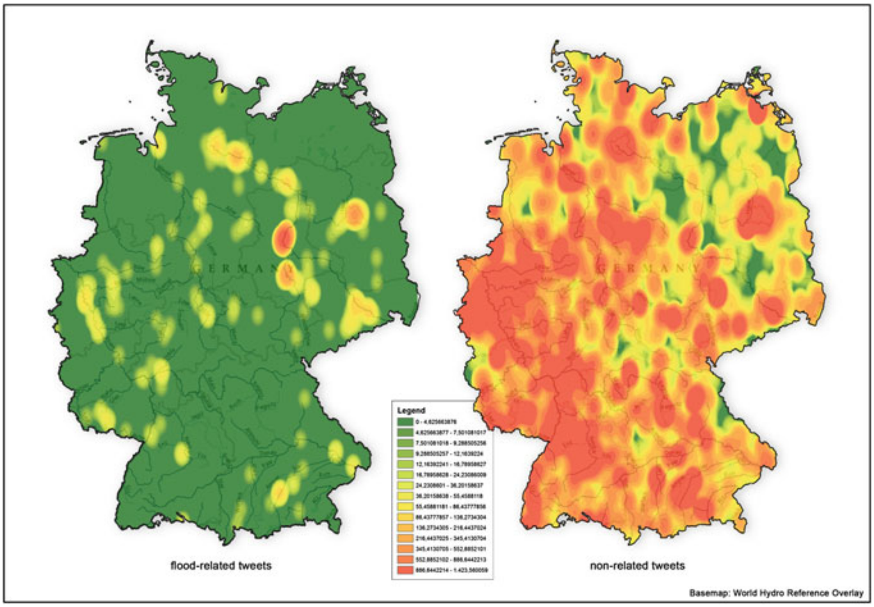

Fig. 4 Spatial distribution of flood-related and non-related tweets

6.2 Analysis

The analysis of the results achieved is presented as follows by addressing each of our three research questions in turn.

6.2.1 RQ1: Does the Spatiotemporal Distribution of Flood-Related Tweets Match the Spatiotemporal Distribution of the Flood Phenomenon?

Firstly, we examined the spatial distribution of flood-related and non-related twitter messages to review whether they follow the spatiotemporal distribution of the flood phenomenon. Figure 4 shows the density of tweets depending on keyword classification. Flood related tweets (on the left side) show peaks in the regions of Magdeburg, Berlin and Halle. Overall flood-related tweets appear only in a few parts of Germany. Non-related tweets (on the right side) concentrate in dense populated regions, e.g. urban areas like Berlin, Hamburg, Munich and the Ruhr area. The tweets cover almost all of Germany, except for some regions in the federal states of Brandenburg and Mecklenburg-Hither Pomerania.

Comparing the spatial distribution of flood-related tweets to the spatial distribution of flood-affected catchments (see Figs. 3 and 4) one can notice similarities the first look. Not the location of all flood-related tweets, but at least of a considerable amount of them does correspond to the location of flood-affected catchments. To further

Table 4 Average distances to flood-affected catchments

	# tweets	Average distance [km]	Standard deviation
Non-related	60,126	221	125
Flood-related	398	78	121

examine this relationship we statistically analysed the distance of all tweets to flood-affected catchments (Table 4).

Using the t-test, we found out that the distance to flood-affected catchments for flood-related twitter messages was statistically significantly lower (78 $\pm$ 121 km) compared to non-related twitter messages (221 $\pm$ 125 km), $t(60522) = 22.674$, $p = 0.000$.

This implies that the locations of flood-related twitter messages and flood-affected catchments match to a certain extent. In particular this means that mostly people in regions affected by the flooding or people close to these regions posted twitter messages referring to the flood.

That is remarkable as there are for instance far more tweets posted in greater distance to flood-affected regions compared to the number of tweets posted in the proximity to flood-affected regions and as such as that media coverage about the River Elbe Flood was enormous since it was one of the most severe floods ever recorded in Germany. Regarding these circumstances one would have expected a great amount of tweets referring to the flood posted in the urban areas like Munich, Hamburg or the Ruhr area. However, that was not the case. The majority of tweet referring to the flooding were posted by locals.

6.2.2 RQ2: Does the Spatial Distribution of Flood-Related Tweets Differ Depending on Their Content?

Considering that most flood-related tweets were posted by locals it seems probable that these messages contain local knowledge only available to people on site. To review this assumption we analysed the spatial distribution of flood-related tweets depending on their content.

Figure 5 visualizes the spatial distribution of tweets referring to the flooding according to the content based classification. At a first look, one can notice that tweets classified as "other" do not follow the spatial distribution of flood-affected catchments in a particular way. Tweets containing information about "volunteer actions" and "flood level" show a spatial distribution that is similar to the spatial distribution of flood-affected catchments. On the contrary tweets containing information about "media" or "traffic conditions" do not show such a match. To review our observations we statistically analysed the distance of all flood-related tweets to flood-affected catchments (Table 5).

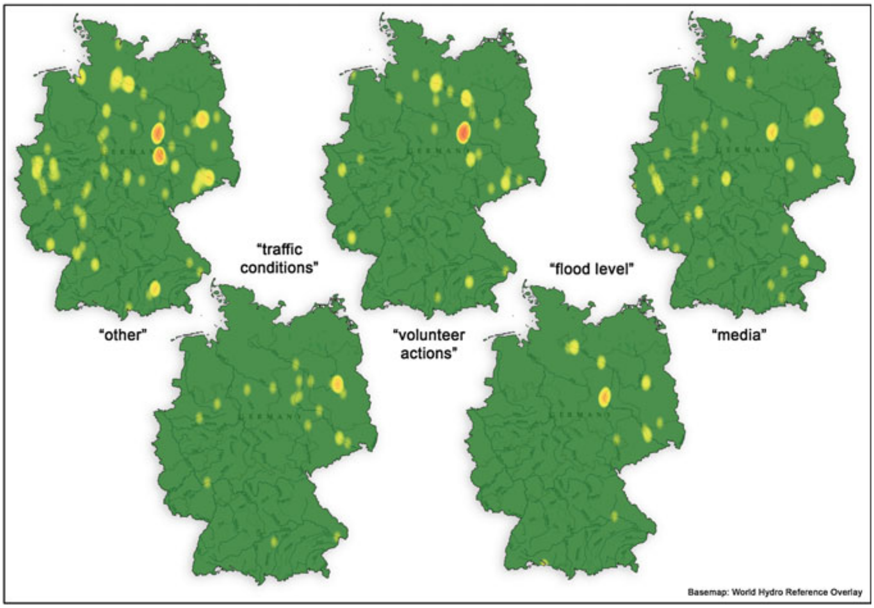

Fig. 5 Spatial distribution of twitter messages depending on content classification

Table 5 Average distances to flood-affected catchments depending on categories

Category	# tweets	Average distance [km]	Standard deviation
VA	113	34	110
FL	32	44	113
TC	30	78	125
O	166	90	114
M	57	150	108

By analysing flood-related tweets based on their content we find that the spatial distribution of tweets differs between the various categories. Especially tweets containing information about VA and FL tend to be concentrated in proximity to flood-affected catchments. Tweets containing information about "media" and "other" show the opposite characteristics.

The application of the post-hoc analysis (LSD) confirms that the differences between categories are statistically significant (Table 6). The average distance of the messages in the categories VA and FL do not differ significantly. In contrast, messages in VA and FL do have significantly different average distances from the messages in O and M.

Table 6 Average differences in distance to flood areas between the categories

	VA	FL	TC	O	M
VA	X	−10 (0.677)	−44 (0.065)	−56 (0.000)*	−116 (0.000)*
FL	10 (0.677)	X	−34 (0.244)	−46 (0.041)*	−107 (0.000)*
TC	44 (0.065)	34 (0.244)	X	−12 (0.617)	−72 (0.006)*
O	56 (0.000)*	46 (0.041)*	12 (0.617)	X	−61 (0.001)*
M	116 (0.000)*	107 (0.000)*	72 (0.006)*	61 (0.001)*	X

6.2.3 RQ3: Is Distance to Flood Phenomena a Useful Parameter to Prioritize Social Media Messages in Order to Improve Situation Awareness?

Analysing the results of the previous research question, we observe that the categories of twitter messages can be divided into three groups as regards to the distance to flood-affected areas:

- Group A: messages in FL and VA are the closest messages to the flooded areas.
- Group B: messages in TC have average distance between the other groups.
- Group C: messages in O and M are more distant to flooded areas.

Applying the criteria we defined for usefulness of social media messages for improving situation awareness (Sect. 5.2), we can conclude that messages in Group A are the most useful ones. Indeed, information about current flood levels is crucial for situation awareness and can complement existing water level measurements, which are only available for determined geographical points where gauging stations are located. Since volunteer actions are increasingly organised via social media, this is a type of information which is very valuable and completely missing from other sources. Hence, our results show that the twitter messages that are closest to the flood-affected areas (Group A) are also the most useful ones. Therefore, we can answer positively our research question, concluding that the distance to flood phenomena is indeed a useful parameter to prioritize twitter messages towards improving situation awareness.

7 Discussion and Conclusion

In this chapter we present a new approach to extract crisis-relevant information from social media platforms like Twitter. Our results show that the spatial distribution of twitter messages referring to the flooding of the river Elbe in Germany in June 2013 is significantly different from the spatial distribution of off-topic messages. We further found that flood-related tweets that contain more useful information for situation awareness (e.g. volunteer actions and flood level) are significantly closer to flood-affected regions than others. This implies that distance to flood phenomena is a useful parameter to prioritize social media messages.

This approach to leverage geographical relations to prioritize social media messages can make a contribution for both research and practice. One potential use of our approach is for enhancing other approaches to classification of social media messages. This could be accomplished by using the prioritization according to geographical relations produced by our approach as weights in the algorithms of existing machine learning techniques. Moreover, the proximity to disaster hotspots could be used for ranking messages to be processed by volunteers in crowdsourcing deployments.

The generality of the results presented here should be investigated by applying our approach in similar analyses of other flooding events. Future work should also concentrate on refining the approach by including information from other social media platforms (e.g. Instagram or Flickr). The integration of other official datasets, e.g. precipitation data or satellite images, is another avenue for future work towards better understanding the relations between social media and crisis phenomena from a geographical perspective. Implementing more detailed hydrological models will additionally extend the validity of our method regarding flood phenomena. Furthermore, our results could be generalised by investigating the value of exploring geographical relations for prioritizing social media messages in other disasters besides floods.

Acknowledgments The authors would like to thank Matthias Stork for the ideas and discussions on hydrological models, as well as the German Federal Waterways and Shipping Administration and the German Federal Institute for Hydrology for providing the water level data. João Porto de Albuquerque would like to express gratitude to FAPESP (grant 2012/18675-1) and CAPES (grant 12065-13-7) for funding.

References

Chatfield AT, Brajawidagda U (2013) Twitter early tsunami warning system: a case study in Indonesia's natural disaster management. In: System sciences (HICSS), 2013 46th Hawaii international conference on (pp. 2050–2060). IEEE

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

Gao H, Barbier G, Goolsby R (2011) Harnessing the crowdsourcing power of social media for disaster relief. Intell Syst IEEE 26(3):10–14

Goodchild MF, Glennon JA (2010) Crowdsourcing geographic information for disaster response: a research frontier. Int J Digital Earth 3(3):231–241

Goolsby R (2010) Social media as crisis platform: the future of community maps/crisis maps. ACM Trans Intell Syst Technol (TIST) 1(1):7

Graham M, Poorthuis A, Zook M (2012) Digital trails of the UK floods—how well do tweets match observations? The Guardian Datablog. http://www.guardian.co.uk/news/datablog/2012/nov/28/data-shadows-twitter-uk-floods-mapped. Accessed 20 June 2013

Hiltz SR, Plotnick L (2013) Dealing with information overload when using social media for emergency management: emerging solutions. In: Proceedings of the 10th international ISCRAM conference, pp 823–827

Hudson-Smith A, Crooks A, Gibin M, Milton R, Batty M (2009) NeoGeography and web 2.0: concepts, tools and applications. J Location Based Serv 3(2):118–145

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

Kaewkitipong L, Chen C, Ractham P (2012) Lessons learned from the use of social media in combating a crisis: a case study of 2011 thailand flooding disaster

Kongthon A, Haruechaiyasak C, Pailai J, Kongyoung S (2012) The role of Twitter during a natural disaster: case study of 2011 Thai flood. In: Technology management for emerging technologies (PICMET), 2012 proceedings of PICMET'12: 2227–2232

Lehner B, Verdin K, Jarvis A (2008) New global hydrography derived from spaceborne elevation data. EOS, Trans Am Geophys Union 89(10):93–94

Lofi C, Selke J, Balke WT (2012) Information extraction meets crowdsourcing: a promising couple. Datenbank-Spektrum 12(2):109–120

MacEachren AM, Robinson AC, Jaiswal A, Pezanowski S, Savelyev A, Blanford J, Mitra P (2011) Geo-twitter analytics: applications in crisis management. In: Proceedings of the 25th international cartographic conference, Paris, France

Meier P (2012) Crisis mapping in action: how open source software and global volunteer networks are changing the world, one map at a time. J Map Geogr Libr 8(2):89–100

Okolloh O (2009) Ushahidi or "testimony": web 2.0 tools for crowdsourcing crisis information. Participatory Learn Action 59(1): 65–70

Palen L (2008) Online social media in crisis events. Educause Q 31(3):12

Rogstadius J, Kostakos V, Laredo J, Vukovic M (2011) Towards real-time emergency response using crowd supported analysis of social media. In: Proceedings of CHI workshop on crowdsourcing and human computation, systems, studies and platforms

Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie (2013): Gewässerkundlicher Monatsbericht mit vorläufiger Auswertung des Hochwassers Juni 2013. http://www.umwelt.sachsen.de/umwelt/wasser/707.htm. Accessed 1 September 2013

Sakaki T, Okazaki M, Matsuo Y (2010) 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

Terpstra T, de Vries A, Stronkman R, Paradies GL (2012) Towards a realtime Twitter analysis during crises for operational crisis management. In: Proceedings of the 9th international ISCRAM conference, Vancouver, Canada

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240

Triglav-Čekada M, Radovan D (2013) Using volunteered geographical information to map the November 2012 floods in Slovenia. Nat Hazards Earth Syst Sci Discuss 1(3):2859–2881

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

Yin J, Lampert A, Cameron M, Robinson B, Power R (2012) Using social media to enhance emergency situation awareness. IEEE Intell Syst 27(6):52–59