

Information Propagation in Social Networks during Crises: A structural Framework

Daniela Pohl and Abdelhamid Bouchachia

Abstract In crisis situations like riots, earthquakes, storms, etc. information plays a central role in the process of organizing the intervention and decision making. Due to their increasing use during crises, social media (SM) represents a valuable source of information that could help obtain a full picture of people needs and concerns. In this chapter, we highlight the importance of SM networks in crisis management to show how information is propagated through. The chapter also summarizes the current state of research related to information propagation in SM networks. In particular three classes of information propagation modes are identified: network analysis and community detection, structural (role-oriented) information propagation, and infrastructure-oriented information propagation. The chapter describes an analysis framework that deals with structural information propagation for crisis management purposes. Structural propagation is about broadcasting specific information obtained from social networks to targeted sinks/receivers/hubs like emergency agencies, police department, environment department, etc. Specifically, the framework aims to identify the discussion topics, known as sub-events, related to a crisis (event) from SM contents. A brief description of techniques used to detect topics and the way those topics are used in structural information propagation are presented.

1 Introduction

The last decades brought several technical innovations which changed peoples' communication behavior due to light-weight, powerful mobile phones, mobile In-

Daniela Pohl
Institute of Information Technology, Alpen-Adria-Universität Klagenfurt, Universitätsstr. 65-67,
Austria, e-mail: daniela@itec.aau.at

Abdelhamid Bouchachia
Smart Technology Research Center, Bournemouth University, Fern Barrow Poole BH12 5BB, UK,
e-mail: abouchachia@bournemouth.ac.uk

ternet, social media and mobile applications. With these innovations in recent years, social media networks (e.g., Twitter, Flickr, YouTube, Facebook etc.) form new communication and news channels. People have now the possibility to document every situation they are involved in. They can propagate updates, requests, opinions, and information of general public interest.

In other terms, social media users are actually producer and consumer of information - all at the same time. SM networks facilitate the dissemination of information along social links connecting communities formed in these networks. These connections are of two types: strong ties which are usually formed by frequent contact (family and friends) and weak ties which are formed by infrequent contact.

Over the past SM networks have been studied by different communities focussing on various aspects related to the identification of communities, the propagation of information through the network, modeling the relationship between information producer and information consumer, analysis of sentiment, topic detection. Due to these advances, SM have emerged as research avenue in the area of crisis management studies. As stated in Sec. 2 below, several studies highlight the importance of social media in crisis management. People tend to use SM particularly when communicating with emergency services is difficult due to overload or emergency services were not on site sufficiently early when the crisis happened.

SM networks can be analysed to uncover topics discussed during a crisis, to identify the people needs, and to discover rumor so that it can systematically be countered. Therefore, efficient methods for analysing SM need to be developed in order to:

- detect the important information related to people needs, human casualties and infrastructure damages from SM contents
- understand the community and network structure and
- find communication hubs to propagate useful information that can efficiently be exploited to stabilize the situation, give hints, publish shelter information, and counter against rumors.

This chapter summarizes the research on information propagation in SM networks for crisis management. A framework for dealing with structural information propagation is described. Methods for analysis of SM content, developed by the authors¹, are discussed to show how *structural information propagation* can be used in crisis management.

The rest of the chapter is organized as follows. Section 2 highlights the relationship between social media and crisis management by showing illustrative recent studies. Section 3 summarizes research that focuses on information propagation in social media during a crisis. Section 4 proposes an analysis framework for disseminating and propagating valuable information to be used by targeted people, like first responders, in order to organize their intervention. In particular, the framework de-

¹ We dont claim to provide any original results in this chapter. We simply summarize the different research studies relevant to information propagation in the context of social media and crisis management

scribes how topics discussed in SM networks during crises are detected and tracked to be propagated later on to targeted people. Section 5 concludes the chapter.

2 Crisis Management and Social Media

Crisis Management referring to disaster, emergency or catastrophe management indicate all actions taken before, during, and after a disaster. In general, CM tasks are divided into four phases [15]: *mitigation*, *preparedness*, *response* and *recovery*.

- *Mitigation* deals with challenges regarding risk assessment and crisis prevention to minimize the possibility of disaster occurrence.
- *Preparedness* includes all steps to increase readiness in case of disaster (e.g., training, public awareness, preventive actions, etc.).
- *Response* covers actions to reduce casualties and to stabilize the situation during the disaster.
- *Recovery* focuses on re-establishing the damaged infrastructure and facilitating normal course of life.

In all these phases, information about the situation at hand is important for situational awareness, planning, and decision making. Research shows that SM networks are increasingly used during crisis situations, e.g., to publish information on the current situation. There are several studies highlighting the importance of different SM networks. For instance, Palen [26] described the usage of SM platforms during the Virginia Tech shooting and the southern California wildfire. Vieweg et al. [47] showed how Twitter was used during two emergencies: Red river floods and the Oklahoma grassfire in 2009. They analyzed tweets and identified different categories of tweets (i.e., warning, information on impact, weather, evacuation, etc.). Choudhary et al. [9] studied tweets obtained during the Egyptian Uprising. In particular, they analyzed the sentiment and the topics discussed during the event. Reuters et al. [39] analyzed SM in the context of emergent groups during a disaster. They identified several roles of users in such groups, e.g., reporters who generate contents, retweeter who forward and propagate the information, etc. They also studied SM in the context of European incidents, e.g., the eruption of Eyjafjallajkull and the mass-panic during Loveparade in Duisburg [40]. Terpstra et al. [45] analyzed the data collected during a festival in Belgium (Pukkelpop) which was hit by a heavy storm. They used keyword-based filters to identify tweets with specific content, like damages or casualties. Their analysis showed that tweets can be seen as realtime information sources for situational awareness in crisis management.

Beside Twitter and Facebook, photo sharing tools, like Flickr, have been used to exchange information about events of different/incidents. Liu et al. [21] described the importance of photo sharing via SM during emergency cases. The work depicted the usage of Flickr through six different crisis between 2004 and 2007, e.g., Indian Ocean earthquake, Hurricane Katrina, Minneapolis bridge collapse, etc. They showed that there were eyewitness photos which can be used for a formal emergency

response [21]. Perng et al. [27] analyzed Twitter for the Norway attacks in 2011. Flickr was analyzed by Fontugne et al. [14] in the context of the Japan Tsunami. The study showed that it is possible to identify crisis situations and their impact using the content of Flickr, in particular the metadata of the crisis pictures. Dashti et al. [11] studied Twitter data for disaster recovery and engineering reconnaissance. They argued that for improving situational awareness during an emergency, the most important messages are those including pictures and location information. Yates and Paquette [50] described the use of SM during Haiti earthquake. They highlighted the importance of SM for communication and collaboration between aid units to share information in an easy and efficient way. Likewise, Dougle et al. [13] analyzed the use of SM during Haiti earthquake. The results show that SM is important for situational awareness stating the positive attitude of first responders to use social networks. Hughes et al. [16] collected data related to Hurricane Sandy from Twitter, Facebook, agencies websites and Nixle which is a notification service. They studied online communication of local fire and police services. The study showed that efficient browsing techniques for analysis of SM data are important.

Agencies make use of social networks to monitor the reaction of the public on different events and to communicate with the public during such events. For example, Deneff et al. [12] analyzed the tweeting behavior of two major police forces in UK during the 2011 riots. They highlighted the different strategies of the police forces in communicating with the public. They also presented the different pro and cons of these strategies. Although, relief units use SM in CM and daily work, development of new practice and supporting technologies (e.g., identify active users, understand the current information propagation trend, detect topics and events, etc.) are needed to connect both communities, the public and emergency services [7].

3 Information Propagation in Crisis Situations

SM networks offer simple communication channels to propagate information in a short time. Many research studies illustrate how information can be disseminated in real-time which enable just-in-time analysis of SM data upon the occurrence of events. For example, Stollberg and de Groeve [43] showed that it is possible to detect events by monitoring tweets that contain specific keywords (e.g., earthquake).

In the following we address the important research topics in relation to SM and CM:

- *Network analysis and community detection:* This is about the analysis of SM networks in static or dynamic fashion such as topology and changes of the network. The aim is to detect communities and uncover the communication channels for information spreading. It facilitates the analysis of ties within the communities. Details follow in Section 3.1.
- *Role and topic-oriented Information Propagation:* Analysis of which kind of information is propagated through SM networks and which users will receive

information. The question about which information to trust is another area that falls into this category. Details follow in Section 3.2.

- *Infrastructure-oriented information propagation*: It is about structural forwarding of information to responsible persons or facilitating an infrastructure to forward information. Details follow in Section 3.3.
- *Prediction and forecasting*: This is related to forecasting of several aspects like evolution, impact, behavior of users, etc. Usually the reliability of forecasting depends on the characteristics of the propagation itself [8]. The present chapter will not further discuss this topic as it is out of the intended scope.

The first two points are related to each other, because if the structure of the network is known, the mechanisms/ways the information propagates can also be modeled.

Information published on SM networks can emanate from two sources:

- *General public*: un-structural information can take different forms text, images, videos.
- *Crisis management*: structural information exchange and propagation within CM services.

Most of the research studies investigate data emanating from the public in order to enhance situation awareness, decision making, and resource management.

3.1 Network Analysis and Community-Detection

Tyshchuk et al. [46] examine Twitter data related to Japan tsunami in 2011 by considering social network analysis and natural language processing to understand and uncover communication patterns. The social network analysis was used to find and analyze communities using a random walk algorithm on time slices. Twitter texts were investigated to understand the role of community members [46].

Ren et al. [38] described a visual analytic tool called WeiboEvents². It comprises two parts: (1) a web-based analysis tool and (2) an expert analytic system. The tools are used to analyze and explore a retweet tree and to identify clusters of users. To get rid of the huge amount of data in SM networks, the web-based tool is used as crowd-sourcing tool, where the user can solve mini-tasks asked by the system (i.e., annotating tweets or the importance of users). The results are stored and shared with other users, allowing also an expert in front of the expert analytic system (in case of an emergency) to reuse already examined data. The system is used in China since 2012.

Kumar et al. [20] and Morstatter et al. [23] described TweetXplorer, a tool to identify: who is important, where do relevant tweets originate from, and when do different trends occur. The number of retweets is used to identify important users

² Weibo is a very popular Chinese microblogging platform.

and tweets. Additionally, time and location are used in visual analytics. TweetXplorer visualizes the obtained retweet-network in a graph, map and keyword-cloud representation. The graph represents the retweet network, the map shows the heat-map of tweets and the most important keywords of a time period are visualized in a word cloud. They evaluated their tool using data about Hurricane Sandy. The experiments showed that clusters of users discussing specific topics (e.g., power outage, hospital) can be identified [23].

Sutton et al. [44] examined the amplification of tweets during Boston bombing in 2013. They collected data originating from organizations responsible for emergency management during the incident from Twitter. The data was used to identify users or other organization (i.e., user profiles) in the network which amplify information by retweeting the original tweet [44]. These amplifying user profiles must be considered in a standardized communication process (helping to support data propagation). To identify such users, they analyzed the tweets through creating a retweet-network, where the nodes represent the different organizations and the edges show interaction (i.e., if one node retweets messages from the other node) [44]. The influence of a node is given by the amount of retweets the node performed. It shows that local organization has the most influence in creating new tweets/information and non-local organization acts as supporter for the propagation of the information through retweeting.

Klein et al. [19] suggested a framework based on graph analytic to analyze the Twitter network during an incident in real-time. They tested a framework based on tweets gathered from Twitter with keywords related to emergencies (e.g., earthquake, flood, etc.). It tracks the communication network behind the tweets to identify the leading members in the network. In addition, the content of tweets itself is also analyzed through lexical analysis.

Most of the works focus on Twitter, due to the easy access and the fact that Twitter is a platform with real-time characteristic during a crisis [41]. The presented research work above showed the identification and analysis of the networks based on activities/retweets between users.

Network analysis and community detection are important for CM, especially in the phases *preparedness* and *response*, because they help us (1) to understand how information is propagated, (2) to identify sources of rumors and (3) to create more efficient communication plans.

3.2 Role and topic-oriented Information Propagation

Starbird and Palen [42] analyzed the spread of information, especially, the retweet behavior during the Egyptian uprising. They collected a huge number of tweets, approximately 2.2 million tweets. 956 most tweeting user accounts were analyzed and labeled manually using the following coding: *in Cairo*, *in Egypt but not in Cairo*, and *not in Egypt*. Tweets that are propagated are related to solidarity, detained friends/relatives and violence. The most retweeted information was sent from

local accounts in Cairo, for example, from journalists and bloggers. The authors also stated that the retweeting behavior of other users served as “crowd-based approval” of the content propagated.

Purohit et al. [36] showed that it is possible to identify influential users (i.e., virtual responders). These users can be regarded as disseminators to communicate information to the general public. The identification of influential users is based on a graph built from tweets lying in a specific time window. Nodes represent users and directed edges represent the interaction between users. Additionally, the profession of users can be identified by comparing the user profile with a lexicon representing the most important roles (i.e., journalism, police, etc.) during an emergency. Visualization was used to browse through the graph and identify propagation paths of false information.

Mendoza et al. [22] analyzed tweets from the Chilean earthquake in 2010. They examined the correlation between followers and number of tweets. The correlation showed that the most active users have a high amount of followers in the network. The topology of the network also uncovers related communities. The authors examined the propagation of tweets containing the word “earthquake”. They found different patterns of propagation (i.e., tree-based and cyclic). Tree-based patterns produce direct information propagation, while cyclic patterns represent comments or replies. Very interesting was the difference in propagation of rumors and true information. Rumors tended to cause more questions by other persons in the network and therefore it is possible to detect rumors easily.

Ireson et al. [18] analyzed local communities in the case of Sheffield flooding in 2007. They used local forums and blog posts to analyze the frequency of posts from individuals. They analyzed groups related to topics. It turned out that users with high-frequency posts provided important information to the situation at hand. Reuter et al. [40] [39] analyzed incidents in Europe and the USA. They distinguished between virtual and real emergent groups. They identified several user groups by roles: *helper*, *repeater*, *retweeter*, *reporter*, and *reader*. Hughes et al. [17] identified Twitter as a major broadcast medium for information dissemination. They showed that some users act as “information hubs” in collecting and disseminating information. In [16], Hughes et al. also analyzed SM from another perspective: how is SM used by Police & Fire Services to propagate information to the public. They examined data collected from Facebook, Twitter, Nixle and the website of the police/firefighters department related to the Hurricane Sandy in 2012. The results showed that the most important communication channel was Facebook followed by Twitter. They can be used to communicate closure of transportation or areas and to publish information from third parties like weather conditions.

The identification of roles allows to refine communication plans for disseminating information in a fast way during *response*.

3.3 *Infrastructure-oriented Information Propagation*

Propagation of information is not always straightforward, especially when the communication infrastructure has (partially) been destroyed. This is very important before or during *recovery* phase of CM. Therefore, research focuses on ad-hoc networks for forwarding messages (i.e., tweets). Al Akkad et al. [1] discussed the physical perspective of information spreading focusing on a store-and-forward mechanism to overcome the limitation of the damaged infrastructure. They developed an application for Android mobile phones which takes care of the publication of tweets. It uses WiFi technology to join and leave islands of connectivity. If a mobile device reaches an island with Internet access (i.e., another mobile phone in the ad-hoc network with Internet connection), messages stored on the mobile phone get published. Similar systems can be found in [1]. Moreover, there are also different routing algorithms to transmit information within ad-hoc networks in an efficient way. Rafflesberger and Hellwagner [37] described a comparison of routing algorithms for emergency response operations. Zhou et al. [51] collected microblogs from Weibo to build a Naïve Bayes classifier that allows routing information to emergency departments (e.g., police, fire department, ambulances, etc.).

Routing of information was also examined in the context of first responders (i.e., not only for social media data). In Netten et al. [25, 24], for example, important information is forwarded to people also interested in that information or topic due to their role. They recorded conversation between relief units which was then processed to detect conversation topics.

Wei et al. [48] discussed the difference in disseminating information through strong and weak ties during disaster situations. They created and ran a simulation environment emulating these strong and weak ties. The results showed that weak ties have a high influence in propagating information, as they often have a bridging function between communities with strong ties. This has also been shown by Bakshy et al. [3].

Infrastructure-oriented propagation is important particularly during the *response* and *recovery* phases when the infrastructure is not completely restored. Efficient and intelligent routing mechanism of information helps propagating important facts about the crisis.

4 From SM to Structural Information Propagation

In crisis management information must be first uncovered, analyzed and then directed to people in particular *structures* such as emergency services who will make decision. In the following we describe a framework for structural information propagation during emergency management. Figure 1 sketches the general processing chain for analyzing SM in CM. First, in the *information propagation* step, the people (e.g. including victims, involved people, bystanders, etc.) propagate information about the current situation. The information is analyzed using various techniques

stemming from *Information Retrieval* and *Network Analysis* by examining the network structure, role and discussion topics. In [31], we investigated social media analysis approach to detect topics related to a crisis. The outcome of the investigations are then used for *structural propagation*, by redirecting the topics to the crisis services.

Figure 2 provides details of the last two steps. Relying on specific interfaces such as the Twitter Streaming API, relevant data is obtained. Results of the *analysis* are then used for decision making. Information can also be published for the *public* and redirected to the *first responders* or other *agencies* such as police and fire departments.

Information propagated through the network may call for some correcting actions. The reason for corrective actions could be rumors or false information, but also additional information for prevention. Social media can be used to convey information to the public. Results of the network analysis can be used to propagate information in a focused and efficient way.

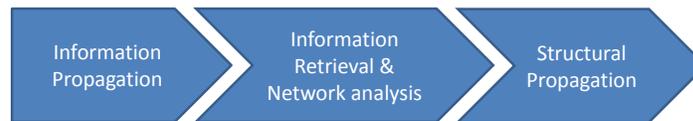


Fig. 1 General Processing Chain

To track structural information, it is important to understand which topic is currently being discussed on social networks, to interpret that topic, and to develop the appropriate action strategy based on what is currently discussed and propagated.

We have developed several topic detection algorithms for social media. Topics can be hotspots or incidents like flooding, collapsing of bridges, etc. in a specific area. These topics are called *sub-events* which can be identified by accumulating/aggregating all messages posted in relation to the same topic. Retweets/re-posts amplify the importance of topics as the amount of messages discussing the topic increases. The advantages of the framework are (1) to uncover the important sub-events during a crisis, and (2) to propagate the analysis results to agencies or other involved parties. Our work on detecting sub-events from SM data is summarized in the following.

4.1 Topic Detection

To uncover the discussion topics in SM during a crisis, we developed clustering approaches. Two strategies were investigated: offline where a corpus is readily available and online where data is obtained and sequentially processed in real time .

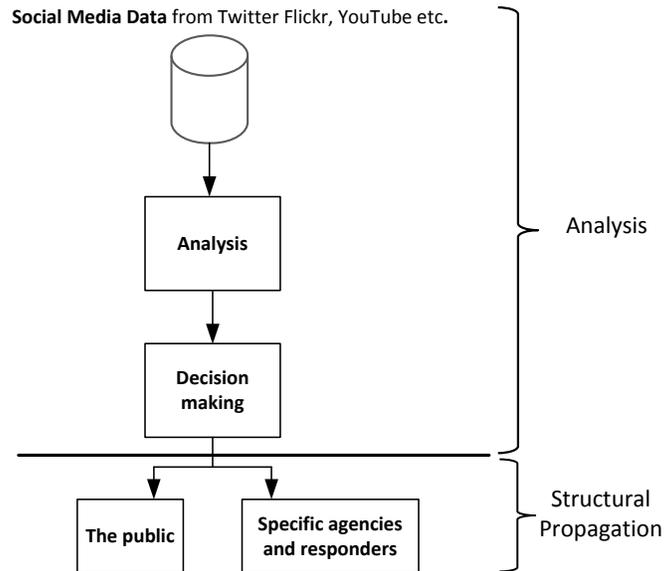


Fig. 2 Information Processing and Propagation

Offline clustering methods were used to detect sub-events which help understand how and what people communicated about an event. Moreover, the offline strategy is used afterward and for training purposes by the agencies. In parallel we developed an online method which was integrated into a system tested during an exercise in the BRIDGE project³.

The different processing steps for the offline and online approaches are summarized in Fig. 3. First, data is gathered via standardized APIs from the SM networks. Data is then geo-tagged before text analysis is performed. Important terms/words are selected through appropriate features selection methods. Afterwards, the offline or online (topic) sub-event detection processes are executed (see Section 4.3 and 4.4). The resulting topics are summarized, labeled with human-readable keywords and visualized in a user-friendly interface.

Using the visualization tool, a person (or team) decides what to do. The following scenarios can be envisaged:

- They could be ignored if the identified issues are already handled by first responders or are out-dated
- The outcome is transmitted to specific persons on-site to take actions (response, rescue)
- Steps must be taken to overcome rumors or clarify the situation in the public

³ http://www.bridgeproject.eu/en/news/BRIDGE-Conducts-Its-Third-Successful-Demonstration-in-Stavanger-Norway_126, last visit: July 2014

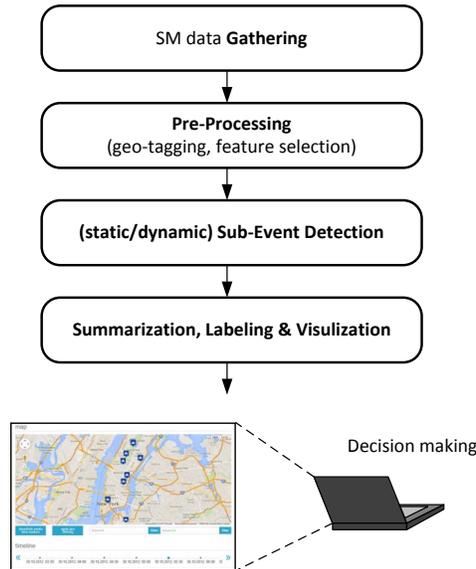


Fig. 3 Topic detection as analytic instrument for structural propagation (Markers by MapIcon-Collection mapicons.nicolasmollet.com)

The propagation could be supported by results of network analysis to find information hubs to disseminate and spread information in a more efficient way.

4.2 Datasets

In our studies we used different datasets as shown in Table 1. The datasets were collected from several social media networks (i.e, YouTube, Flickr and Twitter) based on indicators related to the emergency situation at hand. Search indicators are usually keywords that reflect on the targeted situation. The datasets are related to crises of different magnitudes. For example, the Mississippi Flood crisis could consist of many sub-events describing flood in cities like Vicksburg and New Orleans. The same is true for the Hurricane Irene which affected South Carolina, Virginia and New York. As to UK riots, different cities are affected like London, Manchester and Birmingham.

The first four datasets in Table 1 were used in studies for offline clustering. For the evaluation, the UK riots dataset was labeled based on the sub-events that could be identified by human using textual features and metadata [32]. Labeling requires the location, time and content to identify sub-events (see Section 4.3).

Table 1 Datasets [34]

Dataset Name	Duration	#Pictures	#Videos	#Tweets	Σ Items
Mississippi Flood (MF)	04-19 May 2011	2039	442	0	2481
Oslo Bombing (OB)	22 July 2011	31	222	0	253
UK Riots (UK)	06-10 Aug. 2011	178	274	0	452
Hurricane Irene (HI)	23-29 Aug. 2011	455	700	0	1155
Hurricane Sandy (HS)	29-30 October 2012	286	167	1003	1456

Hurricane Sandy was used to analyze the online clustering strategy (see Table 1). A large amount of media items for a shorter time period has been used to track the dynamics of social media networks. We used Twitter as real-time broadcast medium to collect relevant messages.

In the following, we present a summary of examined offline and online approaches.

4.3 Offline Approaches

In the offline clustering study, we have proposed several methods for sub-event detection:

- *Self-organizing maps (SOM)* in [29]: We used the MF, OB, UK, and HI datasets shown in Table 1. We compared our results with the ground truth (i.e., describing the timeline of the different incidents: MF: [35], OB: [5], UK: [4], and HI: [2]).
- *Agglomerative clustering (AC)* in [30]: Agglomerative clustering is a clustering approach which allows to visualize intermediate result-steps for the user in an intuitive representation, called dendrogram. We introduced a weighting mechanism for keywords/features based on the location of their appearance (title, body, tags). We also used the MF, OB, UK, and HI datasets and compared them to the ground truth.
- *2 Phase-Geo (2PG)* in [28]: We also introduced a 2 Phase-Geo (2PG) clustering approach based on self-organizing maps. It allows to take the (often) sparse geo-information into account when performing the clustering.
- *2 Phase-Geo-Time (2PGT)* in [28]: In another variation of the 2PG algorithm, we additionally considered the time-information of media items published. We evaluated both algorithms, 2PG and 2PGT, using MF, OB, UK and HI datasets to show the impact of geo-information.

Figure. 3 summarizes the steps performed during the analysis. The geo-tagging step is used only for approaches which need that tag in order to run the detection procedure.

In [32] we compared the SOM, AC, 2PG, and 2PGT algorithms using different criteria: **scalability**, **metadata quality**, **ground truth** and **clustering quality**. *Scalability* compares the runtime complexity, the needed parameters and the visualization of the clustering results for their clarity. *Metadata quality* examines the

sensitivity of the clustering approaches regarding the completeness of the metadata. We compared the clustering results against the corresponding ground truth.

The clustering quality is measured based on the **topic** and **item level**. For the *topic level* we check if the different approaches were able to identify the topics. We made use of clustering quality metrics (Dunn, Davies-Bouldin and Silhouette Index [32]) and the Normalized Mutual Information (NMI) to compare the similarity between the clustering results [32]. 2PG and 2PGT show an intuitive interface to visualize the data (i.e., map-based representation) compared to the other algorithms focusing on text-based visualization.

In addition, we compared different clustering approaches based on *item-level*. We labeled UK riots dataset with different granularity. First, labeling was performed with the focus on the City-District-Incident-Date (CDID) format. This means, items in the dataset with the same city, district, and describing the same incident (on the same date) are labeled similarly. We considered also a City-District (i.e, topics from the same city and district are summarized) and City level (i.e., topics related to the same city are combined). For comparison purposes, we used the purity metric to identify the best algorithm and we showed that AC and the 2PGT perform best for the different granularity levels.

The results are summarized in Table 2 based on the different criteria. For further details please refer to [32] and [34].

Table 2 Short summary of the evaluation, ($n \hat{=}$ number of media items) [32]

	SOM	AC	2PG	2PGT
Performance	$O(n^2)$	$O(n^3)$	$O(n^2)$	
Parameter	Number of clusters and terms used in the aggregation			
Representation	Table-representation with text rows		Map-based representation	
Metadata	Text		Text+Geo	Text+Geo+Time
Ground Truth (Topic-level)	Identifies most important sub-events 2PG and 2PGT easier to read			
Cluster Quality (Topic-level)	Similar in assigning data items			Future extension with streaming
Cluster Quality (Item-level)	AC and 2PGT performs best on item level evaluation			

It is easy to communicate those sub-events that are the focus of discussion in SM networks to the interested parties (i.e., relief units, incident commander, etc.). In an aftermath analysis, the obtained insights can be used to simulate or model the propagation of information for an exercise.

4.4 Online Approaches

Based on the findings of the offline approaches, we developed an online approach to detect and track sub-events in real-time during a crisis. We developed an online feature selection mechanism combined with an online clustering algorithm [31]. The feature selection approach extends the term-frequency-inverse document frequency (tf-idf) by introducing an additional weighting mechanism for trending keywords. This means, keywords with a high peak in incoming messages are highly weighted than keywords with less frequency. So, the most important features are identified for further consideration in the clustering stage.

Growing Gaussian Mixture Models [6] have been used to cluster the items with the feature sets resulting from the processing steps. Each cluster is represented as a multivariate Gaussian. The algorithm allows to split and merge existing clusters, to optimize the number of clusters. In the best case, each cluster represents one sub-event (topic). Outdated features are removed from the clusters by deleting the corresponding entries from the multivariate Gaussian's center and covariance matrix.

We evaluated the clustering algorithm using the Hurricane Sandy (HS) dataset [31]. Compared to the ground truth (HS: [49]), the algorithm identifies topics related to flood, power outage and different damages. Additionally, we compared the algorithm against a baseline offline algorithm, which calculates the features in advance from the entire dataset. Using Normalized Mutual Information (NMI) the online clustering showed a similar behavior as the baseline algorithm, although features are dynamically examined. We used the Silhouette metric to judge the quality of the clusterings: high Silhouette values indicate well separated clusters. The values for the online and offline approach are very similar (e.g., in average a difference of 0.16) [31].

We incorporated the developed online features selection and clustering approach in a real-time SM analysis tool called "Information Aggregator"⁴ for supporting crisis management. The tool allows the transmission of important outcome to first responders on-site of the incident [34] as a typical case of structural information propagation. That is, identified sub-events are used during crisis management as additional source for decision making and for defining new communication strategies. First responders have the benefit of using knowledge gained from SM analysis to execute management tasks [33].

5 Conclusion and Future Work

This chapter highlights the importance of social media in crisis management. It describes research on how information is propagated through social networks during crises. We categorized information propagation into three classes: network analysis and community detection, role and topic-oriented information propaga-

⁴ http://www.bridgeproject.eu/content/bridge_information_intelligence_flyer.pdf

tion, infrastructure-oriented information propagation. We suggested a framework for structural information propagation, that is about how hot topics discussed by people can be detected and forwarded to first responders or used to communicate information to the public. We briefly summarized our work in relation to topic detection using various offline and online algorithms.

In future, we plan to extend the topic detection framework by considering active learning to enhance the quality and efficiency of topic detection. We also plan to use the network structure to enable efficient information propagation.

Acknowledgements The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°261817 and was partly performed in the Lakeside Labs research cluster at Alpen-Adria-Universität Klagenfurt.

References

1. Al-Akkad, A., Raffelsberger, C., Boden, A., Ramirez, L., Zimmermann, A.: Tweeting When Online is Off? Opportunistically Creating Mobile Ad-hoc Networks in Response to Disrupted Infrastructure. In: Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management. University Park, Pennsylvania, USA (2014)
2. Avila, L.A., Cangialosi, J.: National Hurricane Center (2012). URL http://www.nhc.noaa.gov/data/tcr/AL092011_Irene.pdf
3. Bakshy, E., Rosenn, I., Marlow, C., Adamic, L.: The Role of Social Networks in Information Diffusion. In: Proceedings of the 21st International Conference on World Wide Web, WWW '12, pp. 519–528. ACM, New York, NY, USA (2012)
4. BBC News Europe: England Riots: Maps and Timeline (2012). URL <http://www.bbc.co.uk/news/uk-14436499>
5. BBC News Europe: Timeline: How Norway's Terror Attacks Unfolded (2012). URL <http://www.bbc.co.uk/news/world-europe-14260297>
6. Bouchachia, A., Vanaret, C.: Incremental Learning Based on Growing Gaussian Mixture Models. In: Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on, vol. 2, pp. 47–52 (2011)
7. Büscher, M., Liegl, M.: Connected Communities in Crises. In: H. Hellwagner, D. Pohl, R. Kaiser (eds.) Social Media Analysis for Crisis Management, no. 1 in 2. IEEE Computer Society Special Technical Community on Social Networking E-Letter (2014)
8. Castillo, C., Mendoza, M., Poblete, B.: Predicting Information Credibility in Time-Sensitive Social Media. *Internet Research* **23**(5), 560–588 (2013)
9. Choudhary, A., Hendrix, W., Lee, K., Palsetia, D., Liao, W.K.: Social Media Evolution of the Egyptian Revolution. *Commun. ACM* **55**(5), 74–80 (2012). DOI 10.1145/2160718.2160736. URL <http://doi.acm.org/10.1145/2160718.2160736>
10. Cui, K., Zheng, X., Zeng, D., Zhang, Z., Luo, C., He, S.: An Empirical Study of Information Diffusion in Micro-blogging Systems during Emergency Events. In: Y. Gao, K. Shim, Z. Ding, P. Jin, Z. Ren, Y. Xiao, A. Liu, S. Qiao (eds.) Web-Age Information Management, *Lecture Notes in Computer Science*, vol. 7901, pp. 140–151. Springer Berlin Heidelberg (2013)
11. Dashti, S., Palen, L., Heris, M.P., Anderson, K.M., Anderson, S., Anderson, S.: Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods. In: Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management. University Park, Pennsylvania, USA (2014)
12. Deneff, S., Bayerl, P.S., Kaptein, N.: Social Media and the Police - Tweeting Practices of British Police Forces during the August 2011 Riots. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI13). Paris, France (2013)

13. Dugdale, J., Van de Walle, B., Koepfinghoff, C.: Social Media and SMS in the Haiti Earthquake. In: Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion, pp. 713–714. ACM, New York, NY, USA (2012). DOI 10.1145/2187980.2188189. URL <http://doi.acm.org/10.1145/2187980.2188189>
14. Fontugne, R., Cho, K., Won, Y., Fukuda, K.: Disasters seen through Flickr Cameras. In: Proceedings of the Special Workshop on Internet and Disasters, SWID '11, pp. 5:1–5:10. ACM, New York, NY, USA (2011). DOI 10.1145/2079360.2079365. URL <http://doi.acm.org/10.1145/2079360.2079365>
15. Hiltz, S.R., van de Walle, B., Turoff, M.: The Domain of Emergency Management Information. In: Information Systems for Emergency Management, vol. 16, pp. 3–19. B. van de Walle, M. Turoff and S. R. Hiltz, Armonk, New York (2010)
16. Hughes, A., Denis, L.S., Palen, L., Anderson, K.: Online Public Communications by Police & Fire Services during the 2012 Hurricane Sandy. In: Proceedings of the ACM 2014 Conference on Human Factors in Computing Systems (CHI). Toronto (2014)
17. Hughes, A.L., Palen, L.: Twitter Adoption and Use in Mass Convergence and Emergency Events. ISCRAM Conference **6**(3/4), 248 (2009)
18. Ireson, N.: Local Community Situational Awareness during an Emergency. In: Digital Ecosystems and Technologies, 2009. DEST '09. 3rd IEEE International Conference on, pp. 49–54 (2009). DOI 10.1109/DEST.2009.5276763
19. Klein, B., Laiseca, X., Casado-Mansilla, D., Lopez-de Ipiña, D., Nespral, A.: Detection and Extracting of Emergency Knowledge from Twitter Streams. In: J. Bravo, D. López-de Ipiña, F. Moya (eds.) Ubiquitous Computing and Ambient Intelligence, Lecture Notes in Computer Science, pp. 462–469. Springer Berlin Heidelberg (2012)
20. Kumar, S., Morstatter, F., Liu, H.: Monitoring Social Media for Humanitarian Assistance and Disaster Relief. In: H. Hellwagner, D. Pohl, R. Kaiser (eds.) Social Media Analysis for Crisis Management. IEEE Computer Society Special Technical Community on Social Networking E-Letter (2014)
21. Liu, S., Palen, L., Sutton, J., Hughes, A., Vieweg, S.: In Search of the Bigger Picture: The Emergent Role of On-Line Photo-Sharing in Times of Disaster. In: Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM 2008) (2008). URL <http://www.cs.colorado.edu/~palen/Papers/isgram08/OnlinePhotoSharingISCRAM08.pdf>
22. Mendoza, M., Poblete, B., Castillo, C.: Twitter under Crisis: Can We Trust What We RT? In: Proceedings of the First Workshop on Social Media Analytics, SOMA '10, pp. 71–79. ACM, New York, NY, USA (2010)
23. Morstatter, F., Kumar, S., Liu, H., Maciejewski, R.: Understanding Twitter Data with Tweet-Explorer. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13, pp. 1482–1485. ACM, New York, NY, USA (2013). DOI 10.1145/2487575.2487703. URL <http://doi.acm.org/10.1145/2487575.2487703>
24. Netten, N., van Someren, M.: Identifying Segments for Routing Emergency Response Dialogues. In: 5th International Community on Information Systems for Crisis Response and Management (ISCRAM) Conference (2008)
25. Netten, N., van Someren, M.: Improving Communication in Crisis Management by Evaluating the Relevance of Messages. *Journal of Contingencies and Crisis Management* **19**(2), 75–85 (2011)
26. Palen, L.: Online Social Media in Crisis Events. *EDUCAUSE Quarterly (EQ)* **31**(3), 76–78 (2008). URL <http://www.educause.edu/>
27. Perng, S.Y., Büscher, M., Wood, L., Halvorsrud, R., Stiso, M., Ramirez, L., Al-Akka, A.: Peripheral Response: Microblogging During the 22/7/2011 Norway Attacks. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)* **5**(1), 41–57 (2013)
28. Pohl, D., Bouchachia, A., Hellwagner, H.: Automatic Identification of Crisis-Related Sub-events Using Clustering. In: 11th International Conference on Machine Learning and Applications (ICMLA), vol. 2, pp. 333–338 (2012). DOI 10.1109/ICMLA.2012.170

29. Pohl, D., Bouchachia, A., Hellwagner, H.: Automatic Sub-Event Detection in Emergency Management Using Social Media. In: In First Inter. Workshop on Social Web for Disaster Management (SWDM), In conjunction with WWW'12. Lyon, France (2012)
30. Pohl, D., Bouchachia, A., Hellwagner, H.: Supporting Crisis Management via Sub-Event Detection in Social Networks. In: Inter. Conf. on Collaboration Technologies and Infrastructures. Toulouse, France (2012)
31. Pohl, D., Bouchachia, A., Hellwagner, H.: Online Processing of Social Media Data for Emergency Management. In: International Conference on Machine Learning and Applications (ICMLA), vol. 2, pp. 408–413 (2013). DOI 10.1109/ICMLA.2013.83
32. Pohl, D., Bouchachia, A., Hellwagner, H.: Social Media for Crisis Management: Clustering Approaches for Sub-Event Detection. *Multimedia Tools and Applications* pp. 1–32 (2013). DOI 10.1007/s11042-013-1804-2
33. Pohl, D., Bouchachia, A., Hellwagner, H.: Supporting Crisis Management via Detection of Sub-Events in Social Networks. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)* 5(3), 20–36 (2013)
34. Pohl, D., Bouchachia, A., Hellwagner, H.: Crisis-related Sub-Event Detection Based on Clustering. In: H. Hellwagner, D. Pohl, R. Kaiser (eds.) *Social Media Analysis for Crisis Management*, 1. IEEE Computer Society Special Technical Community on Social Networking E-Letter (2014)
35. Public Health Emergency: 2011 Mississippi River Flooding (2012). URL <http://www.phe.gov/emergency/news/sitreps/Pages/2011mississippi-flooding.aspx>
36. Purohit, H., Bhatt, S., Hampton, A., Shalin, V., Sheth, A., Flach, J.: With Whom to Coordinate, Why and How in the Ad-hoc Social Media Communities During Crisis Response. In: Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management. University Park, Pennsylvania, USA (2014)
37. Raffelsberger, C., Hellwagner, H.: Combined Mobile Ad-Hoc and Delay/Disruption-Tolerant Routing. In: Proceedings of the 13th International Conference on Ad-hoc, Mobile, and Wireless Networks (ADHOC-NOW), *Lecture Notes in Computer Science (LNCS 8487)*, vol. 8487, pp. 1–14. Springer, Berlin Heidelberg (2014)
38. Ren, D., Zhang, X., Wang, Z., Li, J., Yuan, X.: WeiboEvents: A Crowd Sourcing Weibo Visual Analytic System. In: IEEE Pacific Visualization Symposium (PacificVis), pp. 330–334 (2014). DOI 10.1109/PacificVis.2014.38
39. Reuter, C., Heger, O., Pipek, V.: Social Media for Supporting Emergent Groups in Crisis Management. In: Proceedings of the CSCW Workshop on Collaboration and Crisis Informatics, no. 2 in International Reports on Socio Informatics, pp. 84–92 (2012)
40. Reuter, C., Marx, A., Pipek, V.: Crisis Management 2.0: Towards a Systematization of Social Software Use in Crisis Situations. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)* 4(1), 1–16 (2012)
41. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW '10, pp. 851–860. ACM, New York, NY, USA (2010). DOI 10.1145/1772690.1772777. URL <http://doi.acm.org/10.1145/1772690.1772777>
42. Starbird, K., Palen, L.: (How) Will the Revolution Be Retweeted?: Information Diffusion and the 2011 Egyptian Uprising. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12, pp. 7–16. ACM, New York, NY, USA (2012)
43. Stollberg, B., de Groeve, T.: The Use of Social Media within the Global Disaster Alert and Coordination System (GDACS). In: Proceedings of the 21st international conference companion on World Wide Web, WWW '12 Companion, pp. 703–706. ACM, New York, NY, USA (2012)
44. Sutton, J., Spiro, E., Fitzhugh, S., Johnson, B., Gibson, B., Butts, C.T.: Online Message Amplification in the Boston Bombing Response. In: Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management. University Park, Pennsylvania, USA (2014)

45. Terpstra, T., de Vries, A., Stronkman, R., Paradies, G.L.: Towards a realtime Twitter Analysis during Crises for operational Crisis Management. In: Proceedings of the 9th International ISCRAM Conference. Vancouver, Canada (2012)
46. Tyshchuk, Y., Li, H., Ji, H., Wallace, W.A.: Evolution of Communities on Twitter and the Role of Their Leaders During Emergencies. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '13, pp. 727–733. ACM, New York, NY, USA (2013)
47. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In: Proceedings of the 28th International Conference on Human factors in Computing Systems, CHI '10, pp. 1079–1088. ACM, New York, NY, USA (2010)
48. Wei, J., Bu, B., Guo, X., Gollagher, M.: The Process of Crisis Information Dissemination: Impacts of the Strength of Ties in Social Networks. In: *Kybernetes*, no. 2 in 43, pp. 178 – 191. Emerald Group Publishing Limited (2014)
49. Wikipedia Article: Effects of Hurricane Sandy in New York (2012). URL http://en.wikipedia.org/wiki/Effects_of_Hurricane_Sandy_in_New_York
50. Yates, D., Paquette, S.: Emergency Knowledge Management and Social Media Technologies: A Case Study of the 2010 Haitian Earthquake. *International Journal of Information Management* **31**(1), 6 – 13 (2011). DOI 10.1016/j.ijinfomgt.2010.10.001. URL <http://www.sciencedirect.com/science/article/pii/S0268401210001453>
51. Zhou, Y., Yang, L., Van de Walle, B., Han, C.: Classification of Microblogs for Support Emergency Responses: Case Study Yushu Earthquake in China. In: 2013 46th Hawaii International Conference on System Sciences (HICSS), pp. 1553–1562 (2013). DOI 10.1109/HICSS.2013.129