IT University
of Copenhagen

Social Media Analytics for Disaster Management

PhD Dissertation
by

Alivelu Mukkamala

Supervisor:
Roman Beck, IT University of Copenhagen, Denmark

Committee:
Andrea Carugati, Aarhus University, Denmark
Claudia Buengeler, Kiel University, Germany
Hanne Westh Nicolajsen, IT University of Copenhagen, Denmark

Acknowledgements

To have a PhD abbreviation besides one’s name looks nice but to acquire it one has to experience the PhD journey that is neither a cup of a tea nor a cake walk. But people who surround us make this journey attainable by giving their support. Whether at workplace, home or outside, I am thankful to everyone who ever encouraged me and for their support. Without acknowledging them my dissertation would be an incomplete piece of work.

To Roman Beck
Thanks for being patient whenever I was frustrated or stressful
Thanks for sharing your knowledge and guiding me
Thanks for encouraging me to pursue my ideas

To my colleagues
Thanks for all the fun

To my husband
Thanks for your technical, moral and emotional support
Thanks for bearing with me these three years
Thanks for always encouraging me

To my friends
Thanks for being stress relievers

Last but not the least. Thank you so much my dear wonderful sons, Siddharth and Rishi, for making this entire journey a memorable moment. I never forget those tiny little discussions we had and made me stronger in the weakest moments.

To Siddharth
Thanks for helping me these past 3 years
Thanks for proofreading my papers

To Rishi
Thanks for understanding me through this entire journey
Thanks for Cooking for me ☺

Abstract

In times of crisis involving disasters or other extreme events, victims of these events use social media to share information about their situation. The user-generated content contains vast amounts of valuable information, albeit mostly hidden, regarding the victims' needs, the urgency of supplies, and their situation following the disaster. Especially when drawn from adversely affected areas, these insights are useful for coordinating relief and rescue activities among communities and organizations devoted to improving conditions and saving lives. The insights are quite valuable for humanitarian operations to develop a real time understanding of the situation even before they arrive at the ground. Hence, it is imperative that we develop innovative methods for harnessing the potential of the user-generated content, in order to make disaster relief efforts more effective. Time is an important dimension in humanitarian relief activities, since providing assistance and supplies to victims in a timely manner helps save lives and mitigate the effects of the disaster. However, while extracting time-related information from user-generated social media content is highly important, the relevant scholarly literature has not yet explored this problem. With this in mind, in addition to exploring methodological aspects related to identifying and extracting real-time information and analyzing disaster-related social media data through a theoretical lens, this dissertation contributes to the understanding of the importance of extracting real-time information in the context of a disaster.

Our literature review revealed that there has been very little attention devoted to the use of social media in disasters in information systems (IS) research. For the empirical part of this dissertation, we focused on two different natural catastrophes: For Study 1, we applied a manual content analysis method to analyze the social media data related to Hurricane Sandy through the theoretical lens of situation awareness. For Study 2, we applied a supervised machine learning approach to analyze the media data related to the Chennai floods through the theoretical lens of social presence. For Study 3, we applied the unsupervised topic modeling method to analyze social media data related to the Chennai floods for the purpose of understanding the emerging phenomena during disasters. Furthermore, inductively derived reflections and findings based on the empirical studies (Study 1 and 2) further helped clarify the challenges and opportunities associated the methods point of view and social media data point of view.

The research we conducted offered support for the idea of a method of automatically extracting information from social media content in real time and served as a basis for the exploration of this idea. We developed a time-indicating dictionary—a time wordlist (T-wordlist) to automatically process and extract time-relevant information of needs and urgencies during disasters from social media content. This research extends a social media analytics method—the dictionary-based approach—by developing the T-wordlist to extract the time-relevant information from the social media content.

Overall this dissertation contributes to current research in two ways. First, research findings confirm the theoretical explanations of situation awareness and social presence and shed light on emerging phenomena. Most importantly, this dissertation extends the methodological aspects by analyzing the theoretical concept of social

presence, since thus far this concept has only been analyzed through survey strategy or by conducting interviews. The research at hand extends this concept further by analyzing Twitter messages and examining why people choose to help one another in times of crisis despite the lack of a personal relationship. Second, the contributions of this dissertation have important implications for practitioners. Since disaster management agencies must handle massive volumes of data during disasters, the T-wordlist could be embedded into their systems, resulting in an automatic extraction of time-relevant information concerning needs and urgencies.

Resume

I krisesituationer som involverer katastrofer eller andre ekstreme begivenheder, bruger ofrene for disse begivenheder sociale medier til at dele information om deres situation. Det brugergenererede indhold omfatter enorme mængder værdifuld information, om end mest skjult, om ofrenes behov, det akutte behov for forsyninger, og deres situation umiddelbart efter katastrofen. Særligt når de udtrækkes fra negativt påvirkede områder, er disse informationer nyttige med henblik på at koordinere nødhjælps- og redningsaktiviteter mellem lokalsamfund og organisationer som er dedikerede til at forbedre forholdene og redde liv. Disse informationer er meget værdifulde for humanitære indsatser fordi de er medvirkende til at der kan opnås en forståelse for situationen i realtid allerede før de ankommer på landjorden. Derfor er det essentielt, at vi udvikler innovative metoder til at fastholde potentialet i det brugergenererede indhold for at gøre nødhjælpsindsatser i forbindelse med katastrofer mere effektive. Tid er en afgørende faktor i humanitære nødhjælpsindsatser, eftersom det at stille hjælp og forsyninger til rådighed for ofrene i tide er medvirkende til at redde liv og minimere konsekvenserne af katastrofen. Omend det er yderst vigtigt at udtrække tidsrelateret information fra brugergenereret indhold på sociale medier, har relevant akademisk litteratur endnu ikke belyst dette område. Med dette for øje, bidrager denne afhandling, foruden at udforske metodologiske aspekter i forhold til at identificere og udtrække information i realtid og analysere katastrofe-relaterede data fra sociale medier gennem en teoretisk tilgang, til forståelsen af vigtigheden af at udtrække information i realtid under en katastrofe.

Vores litteraturgennemgang afslørede, at der er blevet givet meget lidt opmærksomhed til brugen af sociale medier i informationssystems- / IS-forskning. Til den empiriske del af denne afhandling har vi fokuseret på to forskellige naturkatastrofer: Til studie 1 anvendte vi en manuel indholdsanalysemetode for at analysere de data fra sociale medier, som omhandlede orkanen Sandy gennem en teoretisk anskuelse af *situation awareness*. Til studie 2 anvendte vi en superviseret maskinlæringstilgang til at analysere de mediedata der omhandlede Chennai oversvømmelserne gennem en teoretisk anskuelse af *social presence*. Til studie 3 anvendte vi den ikke-superviserede *topic modeling* metode for at analysere data fra sociale medier relateret til Chennai oversvømmelserne med henblik på at forstå de fænomener der opstår under katastrofer. Desuden hjalp induktivt udledte refleksioner og resultater baseret på empiriske studier (studie 1 og 2) med at blotlægge udfordringerne og mulighederne associeret med metodetilgangen og med sociale mediedata-tilgangen.

Forskningen vi gennemførte understøttede ideen om en metode til automatisk udtræk af information fra indhold på sociale medier i realtid og dannede et udgangspunkt for udforskningen af denne idé. Vi udviklede en tidsindikeret ordbog - en tidsordliste (T-ordliste) til automatisk at processere og udtrække tids-relevant information om mere og mindre akutte behov under katastrofer fra indhold på sociale medier. Denne

forskning udbygger en social medie-analysemetode - en ordbogsbaseret metode - ved at udvikle T-ordlisten til at udtrække tids-relevant information fra indholdet på sociale medier.

Overordnet bidrager denne afhandling til den nuværende forskning på to måder. For det første bekræfter forskningsresultater de teoretiske forklaringer af *situation awareness* og *social presence* og den kaster lys over fænomener der er under udvikling. Endnu vigtigere er det, at denne afhandling udbygger de metodologiske aspekter ved at analysere det teoretiske koncept om *social presence*, eftersom dette koncept indtil nu kun har været analyseret gennem undersøgelsesstrategi eller ved at gennemføre interviews. Den her foreliggende forskning udbygger dette koncept yderligere ved at analysere Twitter beskeder og undersøge hvorfor mennesker vælger at hjælpe hinanden i katastrofesituationer selvom de ikke har noget personligt forhold. For det andet har denne afhandlings bidrag vigtige implikationer for praktikere. Eftersom katastrofehåndteringsaktører skal håndtere massive datavolumen under katastrofer, kunne T-ordlisten indlejres i deres systemer og bevirke et automatisk udtræk af tids-relevant information om behov og akutte situationer.

Papers of the Dissertation

Paper 1
Mukkamala, AM & Beck, R (2016)
Disaster management and social media use for decision making by humanitarian organizations
In: Proceedings of the 49th Hawaii International Conference on System Sciences **(HICSS 2016)**; Kauai, Hawaii, USA

Paper 2
Mukkamala, AM & Beck, R (2016)
Enhancing Disaster Management Through Social Media Analytics to Develop Situation Awareness: What Can Be Learned from Twitter Messages About Hurricane Sandy?
In: Proceedings of the Pacific Asia Conference on Information Systems **(PACIS 2016)** Chiayi, Taiwan. (Completed research accepted as a short paper)

Paper 3
Mukkamala, AM & Beck, R (2017)
Presence of Social Presence during Disasters
In: Proceedings of the Pacific Asia Conference on Information Systems **(PACIS 2017)** Langkawi, Malaysia

Paper 4
Mukkamala, AM & Beck, R (2018)
THE ROLE OF SOCIAL MEDIA FOR COLLECTIVE BEHAVIOR DEVELOPMENT IN RESPONSE TO NATURAL DISASTERS
In: Proceedings of the European Conference on information systems **(ECIS 2018)**
(Under review)

Paper 5
Mukkamala, AM & Beck, R (2017)
Social media for disaster situations: Methods, opportunities and challenges
In: Proceedings of IEEE Global Humanitarian Technology Conference **(GHTC 2017)** San Jose, California, USA

Paper 6
Mukkamala, AM & Beck, R (2017)
The Development of a Temporal Information Dictionary for Social Media Analytics
In: Proceedings of International Conference on Information Systems **(ICIS 2017)** Seoul, South Korea. (Research-in-Progress)

Table of Contents – Cumulative Dissertation

Introductory Chapter…………………………………………………………... **-10-**

Paper 1…………………………………………………………………………... **-60-**

Paper 2…………………………………………………………………………... **-72-**

Paper 3…………………………………………………………………………... **-85-**

Paper 4…………………………………………………………………………... **-103-**

Paper 5…………………………………………………………………………... **-124-**

Paper 6…………………………………………………………………………... **-141-**

Table of Contents for Introductory Chapter

1 Introduction………………………………………………………………………….. -10-
1.1 Motivation and Research Questions ……………………………………………. -10-
1.2 Structure of the Dissertation…………………………………………………….. -13-

2 Theoretical Background……………………………………………………………….. -17-
2.1 Situation Awareness……………………………………………………………….. -17-
2.2 Social Presence……………………………………………………………………….. -17-
2.3 Emerging Phenomena: From Collective Awareness to Collective Support…….. -19-

3 Research Background……………………………………………………………… -20-
3.1 Disasters……………………………………………………………………………. -21-
3.2 Role of Social Media during Disasters……………………………………………….. -22-
3.3 Sentiment Analysis and Dictionaries……………………………………………. -26-
3.4 Temporal Information……………………………………………………………… -27-

4 Research Design……………………………………………………………………….. -28-
4.1 Research Methodology……………………………………………………………….. -29-
4.2 Data Collection……………………………………………………………………….. -30-
4.3 Empirical Analysis…………………………………………………………………. -31-
4.3.1 Manual Content Analysis…………………………………………………. -31-
4.3.2 Supervised Machine Learning Approach…………………………………. -32-
4.3.3 Unsupervised Machine Learning Approach………………………………. -33-

5 Main Research Results………………………………………………………………... -34-
5.1 Paper 1: Social Media Use for Disaster Management…………………………... -34-
5.2 Paper 2: Social Media Analytics to Develop Situation Awareness……………... -35-
5.3 Paper 3: Presence of Social Presence during Disasters…………………………. -37-
5.4 Paper 4: Emerging Phenomena on Social Media during Natural Disasters…….. -38-
5.5 Paper 5: Social Media for disasters: Methods,Opportunities and Challenges….. -40-
5.6 Paper 6: A Temporal Information Dictionary for Social Media Analytics……... -42-

6 Contributions to Research and Practical Implications…………………………... -44-
6.1 Contributions to Research…………………………………………………………. -44-
6.2 Practical Implications…………………………………………………………… -46-

7 Limitations and Future Research…………………………………………………….. -47-

8 References………………………………………………………………………… -48-

Introductory Chapter

1. Introduction

With the advent of the Internet and contemporary forms of information communication technologies (ICT) such as social media, contemporary economic, social, and political life is being increasingly affected by digital communication. Ironically, in tandem with these technological developments, the number of natural disasters is also increasing (Debarati Guha-Sapir 2016; Nan and Lu 2014). Especially during disaster situations, people depend on digital technologies, and on social media in particular, to share and receive information and to help those in need. As such, there has been an evolution of digital volunteerism (Cobb et al. 2014) bringing positive societal change during disasters by applying technology for humanitarian purposes. For example, the crowdsourcing platform MicroMappers encourages online volunteers to tag tweets and rate pictures posted during disasters in order to assess the impact of the damage faster, allowing humanitarian organizations to more effectively stage their operations (Meier 2016a).

As mentioned above, in recent years there has been a noteworthy rise in the number of natural disasters affecting communities throughout the world (Debarati Guha-Sapir 2016; Huang and Cervone 2016; Stiegler et al. 2011). For example, more than 9,000 people lost their lives in the Nepal earthquake in April 2015 (Economist 2015). Along with natural disasters, man-made disasters are also disruptive and make communities vulnerable.

Moreover, the availability of real-time information is essential for both disaster management agencies (DMAs) and humanitarian organizations for both effective decision-making and coordination of immediate response activities. The quicker the organizations react and respond, the more effective the relief efforts will be. Rescue teams can offer assistance to the affected population, assess infrastructure damage, and gauge the degree of economic loss and number of fatalities. In reality, the ability to evaluate who needs help, where to reach them, and what kind of help is needed, is limited by the constraints imposed by the difficulty of tracing and tracking information (Comfort et al. 2004). In this regard, until recently, the ICTs in disaster situations were brought in by nongovernmental organizations (NGOs) and intergovernmental organizations (IGOs). Several information systems have been implemented to improve the flow of information in an ad hoc manner but the adequacy of their performance is still questionable. The reason for poor performance can be attributed to the unsuitability of applying these systems, configured for routine environments, and to nonroutine and uncertain environments (Day et al. 2009). However, compared to the 1990s, the time it currently takes to receive information from disaster zones has decreased tremendously due to technological developments such as the Internet, mobile communications, and, social media (Meier 2016b).

1.1 Motivation and Research Questions

Social media are now commonly used during emergencies (Kaufhold and Reuter 2016) and have changed traditional communication practices. In the past, communication of disaster-related information had been unidirectional; information flowed from emergency management organizations to the public (Bunker et al. 2017; Latonero and Shklovski 2011). Now, with the emergence of social media it is possible to get real-time information about crises from affected people sharing their situation from within disaster zones. During and immediately after disasters, when conventional means of communication fail, technologies like social media act as an alternative means of sharing and gathering information (Huang and Cervone 2016; Simon et al. 2015). Rather than waiting for nongovernmental organizations (NGO) and governmental organizations to implement the technology necessary for carrying out disaster relief efforts, these activities can be performed by the people affected who still quite often have access to social media. As such, when other channels face difficulties and fail to work properly in disaster situations (Qu et al. 2009), people use available communication channels for information sharing and gathering. In disaster situations, people reorient (Sutton et al. 2008) their personal social media networks in order to propagate information and support relief efforts. As disasters unfold, the majority of social media users employ their social media access as a tool for providing and locating disaster-relevant information, inquiring about the well-being of friends and families, and reporting about the damages caused by the disaster (Al-Saggaf and Simmons 2015; Chong et al. 2014; Kaufhold and Reuter 2016; Shklovski et al. 2008). Moreover, recent disasters have illustrated how people use online forms to demonstrate community solidarity and offer support and assistance for victims (Torrey et al. 2007), and recent terror attacks in Brussels and Munich have shown how people cope with such man-made disasters by sharing information on Twitter (Bunker et al. 2017; Mirbabaie et al. 2014).

Social media is a valuable source of real-time information (Acar and Muraki 2011), and has paved the way for the development of different applications and tools to identify (Fuchs et al. 2013) real-time event information (Yin et al. 2012). Social media also acts as a preliminary rapid damage assessment immediately after a disaster (Kryvasheyeu et al. 2016). While sharing information, social media users play different roles during disasters (Lee et al. 2013; Reuter et al. 2013) and combine resources to make sense of the situation collectively (Bunker et al. 2017; Mirbabaie and Zapatka 2017; Subba and Bui 2010). Beyond the individual level, government organizations, media, and NGOs have also begun using social media platforms to share disaster-related information during disaster events. For example, during recent disasters, such as the Alberta floods in 2013, Boston bombing in 2013, Bp deep water oil spill in 2010, and West Texas explosion in 2013, Colorado wild fires in 2012 affected people and others used Twitter to gather and share information about the disaster situation (Olteanu et al. 2014; Olteanu. 2017; Starbird et al. 2015). In general, there has been a noteworthy rise in the number of disasters throughout the world in recent years, and researchers expect that disaster management agencies (DMAs) will increase social media use (Kaewkitipong et al. 2016; Kryvasheyeu et al. 2016), because of its potential utility in disaster situations. In this context, it is important to understand the opportunities and challenges that are associated with using social media as an information source for coordinating disaster relief activities. Moreover, it is argued that disaster-related social media research is still at a preliminary stage and is also an under researched phenomenon in the information systems (IS) discipline (Leong et al. 2015; Tim et al. 2017). Researchers should pay more attention to the role of the community during disasters, because during disasters “the community is a ‘victim’ that can only play a ‘reactive’ role in crisis response” (Leong et al. 2015). Thus, this dissertation’s main focus is to analyze and understand the user-generated content of affected communities and its usefulness for disaster management agencies. As such, this dissertation adopts a generally holistic perspective to understand the use of social media in real-time for disaster management agencies by addressing the following question:

> **RQ 1**: *How can social media be applied during disasters as a source of information for disaster management agencies?*
>
> ***(Papers 1** and **5)***

Conversations on social media will eventually generate an unprecedented amount of content in the form of individual messages. This user-generated content contains extensive valuable information, though often hidden, regarding the victims' needs, urgency of supplies, and so forth. Information from affected areas plays an important role during disaster situations and helps emergency responders make better disaster-response decisions. In such scenarios, timely information is as essential as access to food, water, shelter, etc. Concerning disaster response, the value of information depreciates significantly (e.g. by 60% to 80%) as time passes. More recently social media has reduced the amount of time it takes to get real-time information from affected areas (Meier 2016b). Since individual use of social media is increasing, , they might also expect DMAs to adopt social media (Lindsay 2011). Moreover, most emergency management personnel believe that social media is useful for sharing information and keeping in touch with the public, and believe that it provides an effective overview of disaster situations (Krstajic et al. 2012; Reuter et al. 2016; van Gorp et al. 2015). While DMAs are using social media as an additional means of communicating with the public during disasters (Chatfield and Brajawidagda 2013; Subba and Bui 2017), thus far, they have been unable to exploit social media for two-way communication (Muralidharan et al. 2011), nor have they successfully integrated user-generated information into their processes to assist with disaster response. DMAs still struggle to effectively incorporate social media into their management processes (Reuter et al. 2016; Wukich 2015). One reason for this could be the overwhelming amount of data generated on social media platforms during a disaster. Emergency agencies are faced with the challenge of aggregating and processing rescue requests from social media in real-time (Abbasi et al. 2012). Capturing data involves gathering meaningful information from huge volumes of social media messages. Message filtering would be needed to reduce the amount of superfluous information and retain only messages with high priority information (Avvenuti et al. 2016). Though disaster-related social media data contains a wealth of information, using traditional research methods to analyze it poses substantial challenges and difficulties (Sivarajah et al. 2017). Therefore, using analytics for this purpose in the field of information systems (Abbasi et al., 2016; Baesens et al., 2016; Chen et al., 2012) will be a useful endeavor. Therefore, this dissertation's second research question is specifically guided by a methods perspective, and focuses, in general, on applying different theoretical lenses to answer the following question:

> **RQ 2**: *How useful are the different content analysis / text analytics methods for identifying relevant information from social media channels during disasters?*
>
> ***(Papers 2, 3, 4)***

Relevant and timely information enhances the planning and mobilization of inter- and intra-organizational disaster relief efforts, which perform in conditions of extreme uncertainty. Since timely information flow obviously improves the flow of resources in interorganizational coordination, we argue that time is an important dimension in humanitarian relief activities, since providing assistance to victims in a timely manner and making supplies available on time are crucial elements of disaster response. Therefore, the possibility of extracting time-related information from user-generated content in social media is an important topic that has not yet been addressed in existing IS research. Thus far, researchers have used supervised (Abbasi et al. 2008; Gonçalves et al. 2013; Pang and Lee 2008) and unsupervised dictionary based approaches to conduct sentiment analysis (Abdulla et al. 2016; Taboada et al. 2011; Thelwall et al. 2011). Dictionaries have been used to conduct sentiment analysis in order to measure human emotions and sentiments (Dodds et al. 2011; Nielsen 2011), but none of the research has focused on time-related information concerning the urgencies and needs of disaster situations. Moreover, none of the studies have developed a dictionary with time-indicating words using a systematic method. Therefore, by addressing the following research question, this dissertation extends the social media analytics by developing a time-wordlist, which is an easy-to-use approach to extract temporal information from social media data.

> **RQ 3**: *How can time-indicating expressions be captured from social media data in close to real time?*

(Paper 6)

1.2 Structure of the Dissertation

This cumulative dissertation project consists of one literature review paper, four empirical study papers, and a case study paper that advance knowledge on the role of social media analytics for disaster management by addressing the previously mentioned research questions. As depicted in **Figure 1**, the overall structure is aligned with the previously formulated research questions based on the disaster-related social media data analysis and reflections of the empirical studies. The different papers are each allocated to one of the different methods of analyzing the disaster-related social media data, depending on the degree to which the theories are applicable, except **Paper 1** and **Paper 5,** which reflect the social media uses, opportunities, and challenges. **Paper 1** and **Paper 5** address the first research question. In particular, **Paper 1** implicitly addresses the research question by analyzing the extant literature to shed light on the use of social media in real time for disaster management phases. Particularly, the focus of **Papers 2-4** is to addresses the second research question. The papers mainly explore different methods such as manual content analysis, supervised machine learning approach, and the unsupervised topic modeling technique. Moreover, we applied two different theoretical lenses: situation awareness and social presence, and we inductively identified emerging phenomena of collective behavior. Along with theories, these methods contribute to an understanding of what type of information is shared as disasters unfold, who shares most of the situation updates, why people go to extra lengths to help each other, and why and how collective behavior phenomena emerge on social media. **Paper 5** answers the first research question and presents a holistic understanding of opportunities and challenges regarding social media data in general, and sheds lights on methods like content analysis and supervised machine learning approach based on empirical works (**Paper 2 and 3**). With respect to the third major research question, **Paper 6** develops and validates the time-wordlist to advance social media analytics, and contributes especially to the topic of dictionaries and the dictionary-based approach. Thereby, this paper provides a new way of extracting time-relevant information from user-generated content and offers support in gaining new insights that have not been explored thus far.

Paper 1 examines the extant research and its applicability to different disaster management phases. In order to lay a theoretical foundation for this research, we mapped the social media uses during disasters (Houston et al. 2015) to disaster management phases (Cozzolino 2012). Later, we coded the literature along the two dimensions (disaster management phase and social media use during disaster) and whenever we believed an article could be categorized into one or more of the categories we did so. The mapping of social media use during disasters to disaster management phases demonstrated how the social media in disaster classification actually resonates with the more commonly used disaster management phases. Our results revealed that social media is a relatively new technology that has not been specifically designed to support humanitarian aid during disasters. There is neither a holistic methodology/approach nor sufficient theoretical foundations concerning using social media for disaster management across the different response phases. Moreover, we argue that there is also a need to encourage IS research to engage more, in general, with humanitarian organizations to help them to develop better-aligned solutions.

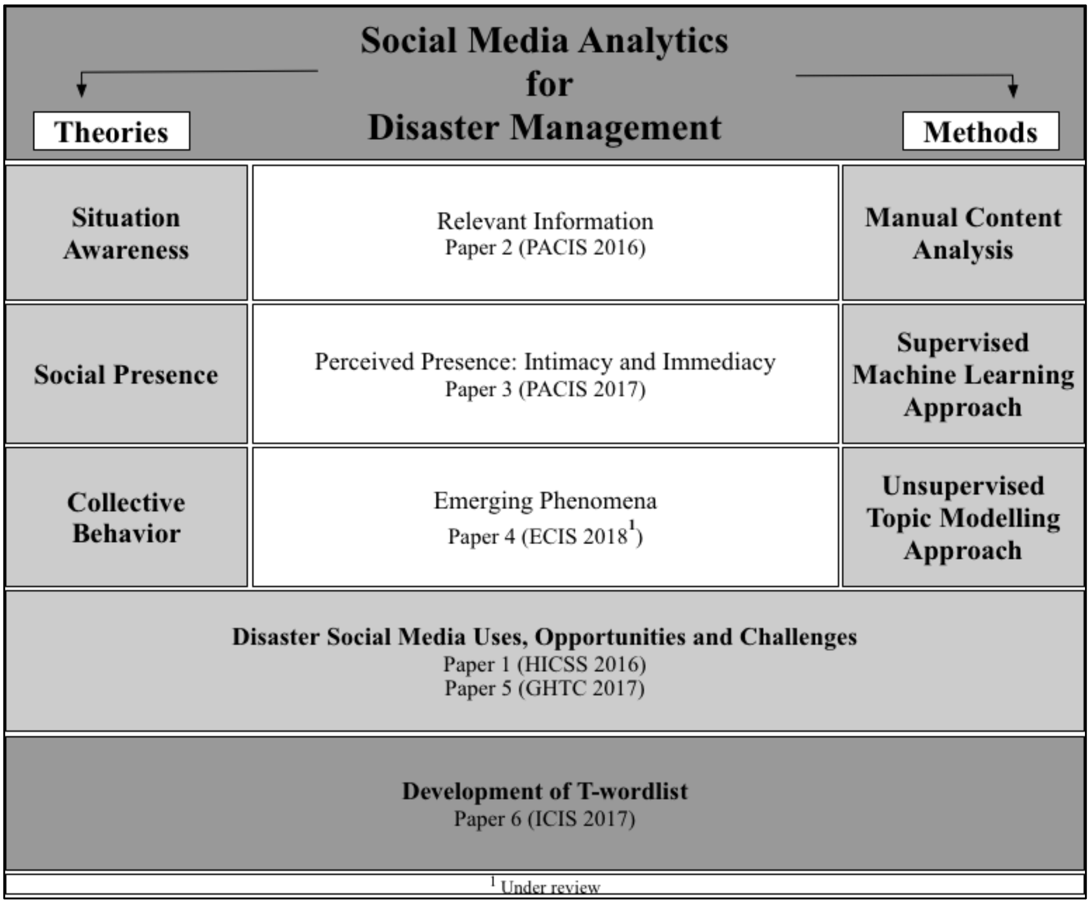

Figure 1. Structure of the Dissertation

Paper 2 analyzes how different information categories shared on social media can be identified and used to provide disaster response agencies with better intelligence and situation awareness. Situational awareness is important for emergency management officials to facilitate better decision-making. Therefore, in our first empirical research, we analyze 11 million Twitter messages related to Hurricane Sandy (Zubiaga 2015). The messages were generated from around the world, but we focused on Twitter messages emerging from hurricane-affected areas. Thus, we relied on the geolocation information of the tweets to filter the messages. To analyze the messages, we adapted two different coding classifications from prior research, which we developed for analysis from disaster-related Twitter data from an earthquake and floods, respectively (Qu et al. 2011; Starbird et al. 2010). We developed our coding framework based on the above-mentioned coding schemes and applied them to manually identify information sources and information types. Our results show that most of the informational messages were situation updates, whether produced by the original source or a secondary source. The percentage of original tweets (original source) was larger, indicating that affected individuals share real-time information while also asking for real-time help. This contributes to situation awareness for disaster management agencies.

Paper 3 takes a comprehensive approach toward illustrating how machine learning can be applied to analyze large volumes of textual content for exploring theoretical concepts such as social presence (Short et al. 1976). We assume that especially during disasters, along with people who seek help and support on social media, there are other online users who also feel intimacy and immediacy for the victims and provide support by sharing relevant information online and actively participating in relief activities. A tweet can also create a feeling of intimacy and immediacy based on the textual content. In order to understand how people express intimacy and immediacy as forms of social presence in times of disasters, we analyzed approximately 1.65 million tweets from a devastating flood in Chennai, India, which took place in December 2015. We operationalized intimacy and immediacy concepts and applied a supervised machine learning algorithm (naive Bayes classifier) to categorize the tweets. Our empirical results define intimacy tweets as more focused on expressing or showing moral support and offering online help, whereas immediacy tweets seek to demonstrate the vulnerability of the situation and motivate people to actively take part in relief activities. Moreover, information in immediacy tweets, which reflect the needs and urgencies of affected people, is important and valuable to disaster management agencies who are engaged in helping people and saving lives.

Paper 4 aims at understanding a process, an emerging phenomenon, comprising temporal and logical relationships between the topics (derived from messages) and the corresponding response relief activities during disasters. To identify how collective behavior develops, we have chosen an unsupervised text analytics method—topic modeling. Topic modeling is a data-driven approach, that can be applied directly onto the textual content to extract the topics (Debortoli et al. 2016). Therefore, we applied topic modeling as the computational grounded theory approach with a focus on analyzing the Twitter data of the Chennai floods. Subsequently we interpreted and coded the identified topics in order to conceptualize a collective behavioral process model. We argue that conditions/activities, such as collective awareness, collective concern, collective empathy, and collective support are necessary conditions for people to feel, respond, and act in forms of collective behavior. During disasters, collective behavior evolves as an emerging phenomenon, hence disaster management officials can make use of community-level intelligence for their tactical decision-making and for organizing emergency relief activities.

Paper 5 builds on **Paper 2** and **Paper 3** by inductively deriving the opportunities and challenges from a data and methods (manual content analysis and machine learning approaches) perspective. The main aim is to discuss the applied methods and their opportunities, as well as the challenges involving capturing and analyzing the social media disaster data, in order to facilitate the extraction of important information that will be useful to disaster management agencies. Based on two case studies and the extant literature, we conclude that social media provides information: social media offers firsthand observations, situation updates, information about the needs and urgencies of affected individuals, supports existing early warning systems, helps disseminate early warnings, acts as an additional communication channel, and so on. However, the challenges to be considered are: data volume and velocity, relevance and integration, and rumors and fake news. Geolocation information is quite useful to extract information from affected areas but the challenge is that the percentage of tweets that contains geoinformation is rather low. So DMAs should begin to create awareness among the public about the importance of enabling geoinformation, especially during disasters. During disasters, new structures in the form of digital emergent groups evolve and provide an opportunity for DMAs. But to start a dialogue and collaborate with them in relief activities, DMAs must move to the next level, rather than just adopt social media. Moreover, manual content analysis is effective and feasible only when the dataset is rather “small.” However, manually analyzing millions of tweets in real time is challenging and impractical. Therefore, using well-defined training datasets and supervised algorithms is necessary in order to classify “huge” amounts of data automatically. We also argue that automated text analysis methods must be further developed in order to understand and extract meaningful insights from user-generated content.

Paper 6 has a predominant social media analytics focus and contributes to a dictionary-based approach through developing a dictionary—or wordlist—comprising time-relevant words and phrases. Disaster-related social media data contain valuable information about the needs and urgencies of affected people. We argue that time plays a role in another dimension—namely, regarding the ability to prioritize emergency response forces along the specific urgencies of a population affected by a disaster. For example, tweets containing words like “just,” “urgent,” or “immediately” might indicate specific urgencies; if disaster relief forces can analyze the data quickly enough—meaning in an automated way—the data can help organize response priorities. Moreover, while there are many dictionaries available to conduct sentiment analysis, there are none devoted to extracting temporal information. Therefore, in this paper, we manually develop a dictionary, and to achieve our research objectives, we followed a methodology that consisted of three phases: First, we defined the criteria for the collection of time-indicating words and then collected the words; second, we collected the data and further applied the wordlist to the tweets that were related to crises and disasters; in the final step, we conducted the content analysis to validate our preliminary results.

2. Theoretical Background

From the information systems perspective, theories are useful either to test concepts or to extend an existing theoretical concept. Moreover, theories are helpful to understand and observe a phenomenon of real-world situations. In this dissertation, we applied two theories in a deductive way: situational awareness and social presence. Furthermore, we also inductively explored an emerging phenomenon—collective behavior process—which will be presented below.

2.1 Situation Awareness

The concept of situation awareness first appeared in the military aviation where environment is dynamic, complex, and risk is involved (Salmon et al. 2006). The widely used definition of situation awareness is “the perception of the 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 1988). The elements of situation awareness vary widely between domains and they depend on different set of goals and decision-making tasks from the respective domain. For example, the elements a pilot needs to know about will be quite different from a surgeon (Endsley 1995; Endsley and Garland 2000) and so are the needs of emergency managers to take decisions and perform tasks under uncertainties. In order to support good decision-making, it is important to gather information to develop a certain understanding about what is happening in a certain situation (Reilly et al. 2007).Especially when disasters take place suddenly, immediate response is necessary to help affected people. A few empirical research (Verma et al. 2011; Vieweg et al. 2010a; Vieweg et al. 2010b) used Twitter data in times of disasters with the aim shedding light on the role of information creation and sharing on social media platforms for creating situation awareness (SA). Situation awareness refers to the way that human beings extract meaning from information about their surroundings to develop mental models of situation by integrating the extracted information with their own knowledge to explore and anticipate further action (Seebach et al. 2011; Vidulich et al. 1994). Often, SA helps in emergency situations to implement response strategies and to derive decisions for combatting the crisis (Vieweg et al. 2010a). Even though SA has often been analyzed from an individual point of view, it can also be aggregated at the group level (Seebach et al. 2011) and Twitter facilitates group-level SA that can be subsequently used by disaster management agencies to develop an understanding about the real-world situation on the ground. Since Twitter is a social network platform, people experiencing the same situations often tend to exchange information and mental models. Situation updates that start at an individual level lead to engagement at a collective level. This collective-level knowledge sharing enhances the SA in a broader scope and helps emergency organizations make appropriate decisions and stage quick responses to disasters.

2.2 Social Presence

Social presence can be traced back to telecommunications research in the 1970s (Lowenthal 2009; Short et al. 1976) where it was viewed as a media characteristic. In the communication context, the degree of salience indicates the perceived feeling or significance of the other person being present in the interaction (Kehrwald 2008). The quality of a communication medium plays an important role as it can determine the way people interact and communicate. Hence, from the perspective of degree of response, social presence of a communication medium is assessed according to how well a medium can transmit information of nonverbal cues, facial expressions, posture, and attire (Gunawardena 1995).

Social presence is typically associated with two concepts: intimacy (Argyle and Dean 1965) and immediacy (Wiener and Mehrabian 1968). Intimacy describes how people act and become close during social interactions while immediacy focuses on interpersonal communication and communicative behavior (Short et al. 1976). Equilibrium of intimacy develops between any pair of individuals when the joint function based on the mutual exchange of a smile, conversation, eye contact, or physical distance occurs. People alter their behavior to maintain the intimacy whenever one of the functions changes (Argyle and Dean 1965). In contrast, immediacy is a measure of the psychological distance an individual places between the self and a target audience (Wiener and Mehrabian 1968). The selected communicative behavior of an individual leads to physical or psychological closeness in interpersonal communication (Wiener and Mehrabian 1968; Woods and Baker 2004). As an example, the use of television enhances intimacy to a greater degree than radio (Short et al. 1976). Along the same lines, it can be argued that a person can create an impression of formal or informal attitude while speaking with someone on the phone. In other words, a person can convey immediacy or non-immediacy through verbal and nonverbal communication while speaking with someone on the phone (Aragon 2003; Cobb 2009; Gunawardena and Zittle 1997). Therefore, social presence can be viewed as an attribute of the media in question as well as an attribute of the communicators and their presence in a sequence of interactions (Biocca et al. 2003; Gunawardena and Zittle 1997; Tu 2002).

Thus, researchers have often applied the theory of social presence originating from media studies to examine interactions between students and teachers in the context of online learning (Tu and McIsaac 2002). Most importantly, in online learning, online participants convey social presence through the messages they send and the interpretation of those messages by others. Visible activities, such as posting messages, replying and responding to others, and participating in group activities contain the social presence cues of the individuals who send and receive them (Kehrwald 2008). This confirms that despite the lack of existence of nonverbal cues in online environments, individuals grasp cues through language, style, and other details in order to build relationships (Walther et al. 2005). Thus, we can use the theory of social presence to explain and understand how people interact in online learning environments; however, there is still a lack of understanding concerning how to properly detect and measure social presence in social media environments. It is a known fact that social networking sites fall into the “medium” category because of the richness of different attributes such as textual content, videos, pictures, and other forms of information (URLs and website links) (Kaplan and Haenlein 2010). Previous studies have shown that online users’ interactions and engagement with other users is directly related to social presence (Lim et al. 2015) and have demonstrated that social presence plays an important role in fulfilling social connection needs in online environments (Han et al. 2015). While social presence has been previously used in research on social networks, it is unclear which role social presence plays during emergencies, e.g., to stimulate relief activities. **Paper 3** focuses on analyzing the data through social presence concepts, which clarify why people go to great lengths to help strangers during disasters despite a lack of face-to-face interaction.

2.3 Emerging Phenomena: From Collective Awareness to Collective Support

Several studies have observed collective behavior during disasters (Starbird and Palen 2011; Starbird et al. 2010); however, there is no research that explicitly analyzes collective behavior process from a social media point of view. As such, there is a need for research that fills this gap and examines the underlying causal mechanisms surrounding the collective behavior processes in the context of social media during disasters (Eismann et al. 2016). In response, our research on collective behavior during disasters primarily focuses on exploring emerging concepts **(Paper 4)** that lead to collective behavior during disasters. In general, a process model explains an emerging phenomenon with a set of sequential activities and its outcomes. Although a process typically describes the necessary conditions—which may not necessarily be sufficient conditions—it nevertheless provides an explanation that links conditions and activities (Markus and Robey 1988). Process observations help enhance understanding of the sequence of activities that lead to the outcomes (Crowston 2000). Through process modeling, we determined a sequence of activities and concepts that generated the collective behavior during a flooding situation. The concepts we discovered are: 1) collective awareness, 2) collective concern, 3) collective empathy, and 4) collective support. Our model illustrates how different information-sharing activities using social media become collaborative activities as the disaster unfolds. As mentioned above, we argue that these conditions/activities (collective awareness, collective concern, collective empathy, and collective support) are necessary in order for people to feel, respond, and act in forms of collective behavior. The process provides an explanation for people's actions and the altruistic behavior that emerges during disasters and other extreme situations.

The figure below (**Figure 2**) depicts the overall phenomena that occur during disasters from a natural disaster point of view. During disasters, people use social media because they feel the presence of others in the online environment. To fulfill needs of social connection, people engage with social media, sharing and receiving relevant information in order to decrease anxiety and panic (Han et al. 2015). In general, whenever a disaster strikes, affected individuals are exposed to personal experiences and observations in uncertain situations created by extreme events. Curious people in the immediate area or relatives of affected individuals also share or forward important information through social media. Along with concerns, opinions, frustrations and rumors, situation updates are posted on social media sites. According to the theory of situation awareness, these situation updates are very important for helping disaster management agencies to make decisions regarding how to react and respond. Individual-level situation awareness transforms itself into a group-level activity on social media sites through collaborative communicative activities. These ongoing collaborative interactions create collective awareness within the online community through sharing and receiving different information types. Subsequently, people feel sympathetic toward the affected population and start to feel collective concern, offering support, for example, in an attempt to combat the difficult situation. When affected individuals begin asking for help and making rescue requests, the vulnerability of the affected individuals evokes a feeling of empathy among the online social media users that ultimately leads to collective empathy. In the final stage of this process, online community members organize themselves into a body of collective support, which leads to collective behavior. Most importantly, the situation updates are useful to humanitarian relief organizations for making efficient and effective decisions facilitating efforts to support victims and save lives during disasters.

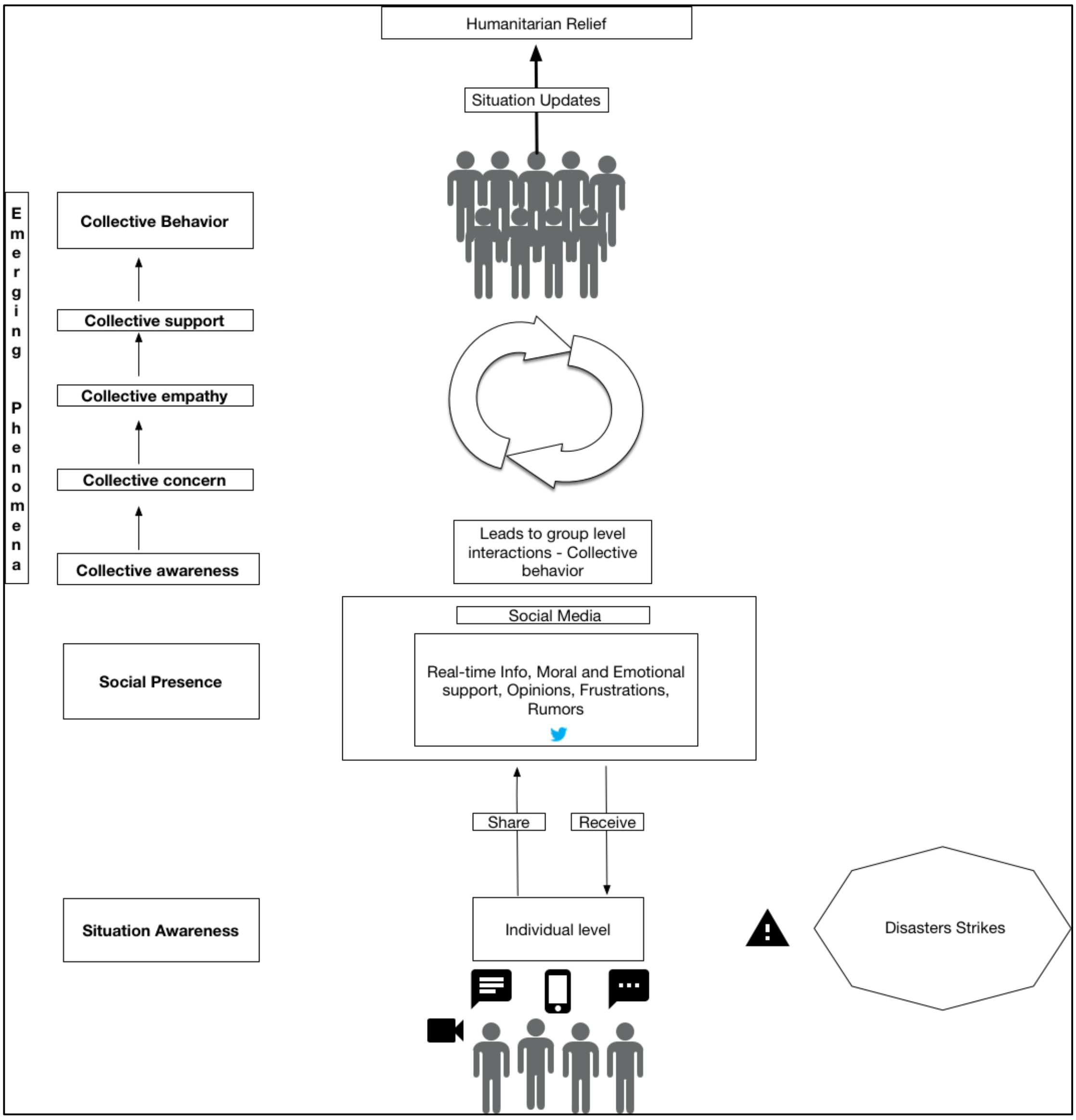

Figure 2. Theoretical Concepts— Phenomena during Disasters

3. Research Background

The first part of the present section provides an overview of the literature streams that form the foundation for this dissertation. The related literature on the role of social media during disasters is divided into three parts based on what methods were applied to analyze the disaster-related content. This section continues by outlining the current state-of-the-art research and demonstrates how the papers included in this dissertation advance the relevant research.

3.1 Disasters

While humankind is prone to different kinds of disasters, the number of natural disasters is increasing globally, and, of course, these disasters come with devastating consequences (Debarati Guha-Sapir 2016). A disaster is an unexpected event that disrupts and threatens the normal conditions of a community, thus exceeding the community's normal capacity to respond to adverse events. The disaster has an impact on people's lives, economy, and environment and demands urgent action despite difficulties. Disasters tend to be unpredictable and difficult to anticipate. Disasters include natural calamities, such as earthquakes, tornadoes, or hurricanes and also man-made destructive activities, such as terrorist attacks or industrial accidents. These types of disasters occur suddenly, and also demand an immediate response. Other kinds of natural crises, such as epidemic diseases or economic crises, also have an impact on society, but the impact is spread over a period of time. Clearly, each type of disaster must be managed differently (Cozzolino 2012).

Disasters threaten not only the technical infrastructure, such as power lines, roads, and communication, but also the social, organizational, and economic structures that support the daily operations of the community (Comfort et al. 2004). Even though emergency relief teams are trained to deal with disaster situations, they still require status updates and situational information about the environment. To combat a disaster, gathering important information is essential for emergency officials to mobilize resources (Comfort et al. 2004). Disasters are typically managed through four disaster management phases—mitigation, preparedness, response and recovery/reconstruction—which researchers model as a cycle featuring different activities (Lee et al. 2017). Researchers also describe these four activities as pre-disaster, response (during the disaster), and post-disaster phases (Houston et al. 2015). Disaster management is mainly concerned with predicting and understanding consequences of disasters in order to mitigate undesired consequences (Zagorecki et al. 2013). Having a good strategy for each phase is essential for efficient disaster management; therefore, disaster responders need appropriate information about the different activities within each of the four disaster management phases (Jayaraman et al. 1997).

The four main phases of disaster management cycle (Cozzolino 2012) are explained as follows:

Mitigation: The mitigation phase focuses on measures that either attempt to reduce the chance of a disaster occurring or attempt to reduce the impact of the disaster (Altay and Green 2006). The focus is on having policies in place based on previous experiences and taking measures to minimize the effects of potential damage (Tierney 1993). Hence, in order to assess the risks and potential consequences of an emergency, it is necessary to obtain knowledge about the issues surrounding that emergency (Vivacqua and Borges 2010).

Preparedness: The preparedness process involves attempting to anticipate emerging problems in future disaster situations and to planning ways to address those problems.

This phase involves improving response capacities and planning in advance to deal with the disaster in order to reduce vulnerability when it occurs. The main activities are personnel training, making emergency plans, and creating awareness about the level of a disaster's impact among the general public and government officials. Both mitigation and preparedness phases are considered pre-disaster (Tierney 1993; Vivacqua and Borges 2010).

Response: This is the phase that occurs during the disaster; it focuses primarily on physical safety and secondary hazards. The response phase is the most important phase because it focuses on preserving the community, preserving the environment, and saving lives (Altay and Green 2006). A few of the activities associated with this phase are warnings, evacuation, safely moving victims to shelters, medical treatment, and so on (Tierney 1993).

Recovery/reconstruction: In the aftermath of a disaster, the recovery phase involves both long-term and short-term activities intended to stabilize and bring the community back to normal conditions (Altay and Green 2006). Examples of activities include damage assessment and replacement, repairing public places, and tending to the recovery needs of victims at the community level (Tierney 1993).

Even though the four phases seem to be cyclical, they often overlap and merge based on the type and nature of a specific disaster. The commonly mentioned recurring problems that hinder organizational performance in disaster operations include lack of resources, lack of coordination, and poor communication (Vivacqua and Borges 2010). In general, disaster management demands a lot of on-site information in order to create the situation awareness necessary to appropriately handle the disaster. Social media has become a bottom-up information provider of real-world, real-time disaster information (Acar and Muraki 2011; Krstajic et al. 2012); however, the transition between focusing on dedicated technologies for disaster relief and the existing use of social media for such purposes has not yet been specifically investigated. Researchers have argued that work on disaster-related social media is still underdeveloped (Tim et al. 2017). Thus, based on extant literature**, Paper 1** maps different social media uses to disaster management phases in order to enhance the potential use of social media for disaster management agencies. Moreover, the empirical research focus of **Papers 2, 3,** and **4** concerns the social media data generated during natural disasters—i.e., hurricane and floods.

3.2 Role of Social Media during Disasters

More recently, the stakes have changed; previously, disaster information communication went in one direction, from governmental and relief organizations to the public. Now, ordinary people also use digital technologies to send and receive relevant information during disaster situations (Palen and Liu 2007). Following disasters, people want to know what happened and how they can help affected individuals (Vieweg et al. 2008; Vieweg 2012); not only are affected people involved in information sharing, but also individuals who are outside the affected regions. Whether it be an earthquake in Chile (Ahmed and Sargent 2014), floods in the Philippines (Lee et al. 2013), or Hurricane Sandy in the US (Shelton et al. 2014), recent disasters feature people involved in intense communications and collaborations on social media as a means of dealing with the adversity. Much rumor mongering was also identified on social media during the 2013 Boston marathon bombings in the US (Starbird et al. 2014) and 2008 terrorist attacks in Mumbai, India (Oh et al. 2011; Oh et al. 2013). Even during other man-made disasters, for example, a bombing in Brussels in 2016 (Mirbabaie and Zapatka 2017) or shootings in Munich in 2016 (Bunker et al. 2017) people used social media to make sense of these sudden and unexpected atrocities.

In this context, Twitter has attracted much research attention as a medium for sharing information, asking questions, and organizing response activities (Mirbabaie and Zapatka 2017; Starbird et al. 2010; Takahashi et al. 2015; Vieweg et al. 2010b; Vieweg 2012). However, thus far, none of this research has focused specifically on India and investigated the use of social media during disasters (natural disaster) there. Thus, the empirical studies of **Paper 3** and **Paper 4** focus on data related to the 2015 Chennai floods to explore the extent of social media use in India and what type of information individuals shared during the flooding situation. Moreover, in line with the focus of existing research, **Paper 1** provides insights into role of social media during disasters and for disaster management phases. The empirical studies of **Papers 2**, **3**, **4** focus on natural disaster-related datasets comprised of Twitter messages. **Paper 5** uses insights from previous work (**Paper 2** and **3**) to present a case study that illuminates the opportunities and challenges associated with analyzing social media data in general. For **Paper 6**, we used a disaster-related Twitter dataset to validate the preliminary work related to our dictionary.

We now turn to a categorization of the extant literature on the methods applied to analyze the social media data related to disasters.

Social Media Disaster Data: Manual Content Analysis

The technique of content analysis systematically categorizes text documents/messages into fewer categories based on explicit rules of coding (Krippendorff 1989; Morris 1994; Stemler 2001). By applying manual content analysis (qualitative data coding), different information types were inductively identified from different disaster-related social media data (Starbird et al. 2010; Vieweg et al. 2010b). For example, researchers applied (manual) content analysis to understand the usage patterns of Twitter communications during flooding in Red River valley in the US and Canada. They coded message contents, as well as author characteristics, to determine whether a message was the author's original content or shared from multiple secondary sources. Further coded the messages from original sources and secondary sources to identify the themes in order to evaluate information production patterns. The three important aspects of the information production lifecycle are generative, synthetic, and derivative aspects. Although 80% of the messages were contributed by local and peripheral individuals, less than 10% of the messages were from original sources (Starbird et al. 2010). Moreover, Twitter messages related to Oklahoma grass fires of 2009 and the 2009 Red River flooding were analyzed manually to understand the situation awareness. Researchers identified situation updates that contained useful insights and labeled them as: warning, preparatory activity, fire line/hazard location, flood level, weather, wind, visibility, road conditions, evacuation information, damage/injury reports, and so on. The most significant finding was that the two disaster events (floods and fire) were of a different quality, as were the situation updates reflecting the real-time conditions of the disaster situations. For example, during the grass fires, a higher percentage of information concerned "evacuation information and damage/injury, fire line/emergency location, and wind directions," while a significant number of situation updates related to the flooding concerned "preparatory activity, flood level, weather and volunteer information" (Vieweg et al. 2010b). Related to another disaster, the 2009 Black Saturday fires in Australia, Twitter messages from the Melbourne area were extracted in order to determine the amount of information that was meaningful and useful and to identify whether or not Twitter might be a useful source in future disasters. Researchers argue that people share information extensively on Twitter and share different types of messages in order deal with the adverse situation; they draw conclusions by clustering the messages according to predefined categories (Sinnappan et al. 2010). Using manual content analysis on four different Twitter disaster datasets, (Vieweg 2012) clustered tweets around three overarching categories: social, built, and physical environment. Furthermore, based on message content, Vieweg's study further divided each main category tweet into subcategories. Altogether, she identified thirty-five subcategories with different information types, and this shed light on the potential of social media during disasters and on the nature of information sharing behavior on social media sites. In analyzing the data of 2010 Chile earthquake, (Mendoza et al. 2010) found that along with useful information, people also spread rumors during chaotic situations. Finally, in their earthquake-related social media study (Ahmed and Sargent 2014) identified traces of participation by various organizations, such as agencies, business organization, and government agencies on social media during disasters.

The research discussed above focused on using manual content analysis to analyze information shared by people during disasters. (Gupta et al. 2013) have analyzed the spread of fake images on Twitter to evaluate Twitter activity during Hurricane Sandy, and (Hughes et al. 2014) have identified and analyzed online Twitter communications of police and fire services were identified and analyzed, but, significantly, there has thus far been no research studying the online communications of regular people. Therefore, in order to understand the online communications of people, in general, particularly concerning the source of messages and nature and type of messages, **Paper 2** focuses on manual content analysis from a situation awareness point of view. In order to understand whether affected individuals enable their geocoordinates during disasters, as part of our first empirical study, **Paper 2** focuses on the geolocation coordinates of individual tweets in order to locate relevant tweets from the area affected by Hurricane Sandy. We used various tools to process the geolocation information of Twitter messages.

Social media disaster data: Supervised machine learning approach

Social media data characteristics, such as volume and velocity, and limitation in manual content analysis make automatic computational methods necessary. To automate and process real-time disaster-related Twitter data a few systems and tools based on supervised machine learning approaches have been developed. For example, Tweet Tracker provides situation awareness immediately after a disaster (Kumar et al. 2011), and Emergency Situation Awareness (ESA) detects an earthquake using Twitter data, sends alerts, and assesses the impact of disasters (Yin et al. 2012). Classifiers were built based on tweet content to identify the signs and to further categorize the type of disaster (Karimi et al. 2013). Other systems like CrisisTracker

(Rogstadius et al. 2013) and Artificial Intelligence for Disaster Response (AIDR) (Imran et al. 2014) require human coders to annotate the data.

Apart from the systems mentioned above, classifiers are trained to extract and categorize situation awareness information from four emergency situations (floods, fires and earthquake) (Verma et al. 2011). The results reveal that situation awareness information is often objective, impersonal, and formal. There is evidence that more than 75% of the tweets from four emergency situations contain information indicating situation awareness (Verma et al. 2011). Moreover, classifiers were trained to automatically sort messages into different classes, such as, caution and advice, information source, donation, and causalities and damage, in order to more easily extract relevant information (Imran et al. 2013). Thus far, studies have focused on extracting situation awareness information using classifiers with supervised machine learning approaches; no studies have yet focused on other theoretical concepts such as social presence. Therefore, in **Paper 3,** we applied a supervised machine learning approach and also used the concept of social presence to classify the Twitter data related to the Chennai floods.

Social Media Disaster Data: Unsupervised Approach (Topic Modeling)

In the category of unsupervised approaches for textual analysis, topic modeling using Latent Dirichlet Allocation (LDA) is a prominent technique (Debortoli et al. 2016). Thus far, this approach has very rarely been applied to analyze the social media disaster data. In contrast to supervised learning methods, techniques like topic modeling do not require predefined categories to cluster messages into groups, but are able to detect categories indicatively. (Lee et al. 2013) used the topic modeling approach to extract eight prevalent topics (including traffic updates, weather agency updates, relief goods and rescue, and so forth) from the Twitter data related to the Philippine floods of 2012.

Extant literature has illuminated the fact that most of the studies mentioned above applied supervised classifiers to build systems, compared to the very limited attention devoted to unsupervised topic modeling techniques. Moreover, researchers have called for an investigation of the causal paths from individual-level actions to collective-level outcomes during disasters in order to understand self-organizing at the community level, from different theoretical and analytical perspectives (Nan and Lu 2014). Researchers have also invoked the need to determine the underlying causal mechanisms for the collective behavior process (Eismann et al. 2016). Therefore, to close these research gaps, in **Paper 4** we applied a data-driven analytics approach (topic modeling) and identified emerging phenomena along with individual actions and collective level outcomes during floods in a developing country—in this case, India.

Overall, however, there are few systems available using machine learning approaches—in comparison to compared to manual content analysis—and the application of machine learning techniques, in general, is rather uncommon in the extant research. One of the challenges of manual content analysis is that it is not feasible to use in real-time with huge data volumes. This could be a contributing factor explaining why the research on social media during disasters is considered to be still in its infancy and why it is, in any case, an underresearched topic (Tim et al.

2017). Disaster management agencies (DMAs) and humanitarian officials recognize the value of user-generated content and have begun to formally incorporate social media for the purposes of monitoring and communicating during emergency situations (Ehnis and Bunker 2012; Latonero and Shklovski 2011; Muralidharan et al. 2011; Panagiotopoulos et al. 2012; Vieweg et al. 2014). However, DMAs still struggle to incorporate social media into their routine processes; they are still experimenting and remain skeptical (Sheombar et al. 2015). DMAs need simple tools that are easy to use in order to integrate the social media data because it is difficult for DMA officials to turn "to powerful but less intuitive or less rapid tools such as programming languages" (Imran et al. 2015). Hence it is argued that, "disaster research needs to be able to use and create new methods" (Landwehr and Carley 2014). In this regard, **Paper 6** focuses on developing an easy-to-use method, a dictionary (wordlist) instead of a complex computational technique to extract temporal information concerning the needs and urgencies of individuals in emergency situations.

Analyzing content and preparing wordlists are not new research strategies. The area of sentiment analysis, or opinion mining, has cultivated considerable research activities in recent years (Pang and Lee 2008). Most of the research in sentiment analysis or opinion mining is based either on previous wordlists or involves adding more words to existing lists (Bradley and Lang 1999; Nielsen 2011) in order to further analyze data through different methods. Researchers have applied sentiment analysis during disasters to evaluate the feelings, concerns, or even panic of affected people (Beigi et al. 2016; Verma et al. 2011). Focusing on disasters and DMA perspectives, in particular, researchers have developed a crisis lexicon (wordlist) based on six different disaster-related datasets. With the help of crowdsourcing, researchers extracted frequently used crisis-relevant terms from tweet messages and compiled a lexicon (Olteanu et al. 2014). Moreover, the wordlist VerbNet, a collection of verbs, was developed to categorize disaster-related tweets. It consists of nine VerbNet classes that were routinely present in a dataset of situationally aware tweets (Vieweg 2012). It is argued that "while not perfect, systems incorporating these VerbNet is one step toward correctly validating data without human intervention"(Landwehr and Carley 2014). Even though there are a few domain-specific dictionaries available, there is no dictionary available to extract time-relevant information from social media disaster data. Thus, this dissertation contributes to the dictionary-based approach by developing a "T-wordlist" to categorize and extract time-relevant information from data, as exemplified in **Paper 6**. However, the next subsection will provide a brief overview of sentiment analysis and dictionaries.

3.3 Sentiment Analysis and Dictionaries

In general, for performing sentiment analysis, research focuses on two types of approaches: unsupervised, dictionary-based approaches (Abdulla et al. 2016; Backfried and Shalunts 2016; Taboada et al. 2011; Thelwall et al. 2011; Thelwall et al. 2010), and supervised, machine learning approaches (Abbasi et al. 2008; Gonçalves et al. 2013; Pang and Lee 2008; Read 2005; Yang et al. 2010). The dictionary-based approach assigns sentiment scores to a list of words to measure either semantic orientation or polarity (positive or negative), or the strength (valence) of a given text (Nielsen 2011; Taboada et al. 2011).

Constructing a sentiment dictionary manually is labor intensive and time consuming; hence, most of the sentiment analysis research depends on preexisting, manually constructed dictionaries. For example, the Linguistic Inquiry and Word Count (LIWC) application uses an internal dictionary that was compiled manually (Pennebaker et al. 2001). It consists of 4,500 words and word stems; words were collected from different sources, and more words were added to it over time. Altogether, there are around 66 categories with varying numbers of words that are assigned positive or negative values. Another lexicon, the Affective Norms for English Words (ANEW) contains 2,477 unique words. The words are scored for valency ranging from -5 (very negative) to +5 (very positive) (Bradley and Lang 1999). A few other dictionaries that are widely used in sentiment analysis include the Opinion lexicon (Wilson et al. 2005), SentiWordNet (Esuli and Sebastiani 2007), WordNet (Miller (Miller 1995), WordNet-Affect (Esuli and Sebastiani 2006), and AFINN-96 (Nielsen 2011). There are also certain domain- specific dictionaries (Olteanu et al. 2014; Temnikova et al. 2015) and also language- specific dictionaries (Madhoushi et al. 2015). Even though there are many dictionaries, it is still unclear how to build a dictionary in the best possible way (Nielsen 2011) and there are no standardized procedures or commonly accepted methods in place concerning how to build them (Deng et al. 2017). As mentioned above, while many sentiment analysis dictionaries exist, no dictionaries have been developed specifically for time-related information. **Paper 6** fills this research gap by developing our "T-wordlist" dictionary for identifying text messages that contain temporal information. Though there are no time-word dictionaries available, there is a considerable research focus on temporal information in the computer science discipline, which will be discussed below.

3.4 Temporal Information

It is of utmost important to identify the temporal information about *events* that is embedded in different types of text (Alonso et al. 2007) or to answer the questions in a given news article regarding events (Pustejovsky et al. 2003a). In this regard, since time plays an important role, researchers consider extracting and normalizing temporal information to be an essential task. Furthermore, according to (Schilder and Habel 2001), time-denoting expressions in a document are of three different types: explicit reference, index reference, and vague reference. Date expressions such as "18.08.71" provide an explicit reference and point to a precise moment, and can thus be easily normalized. As part of indexical reference, temporal expressions (such as "today," "by last week," etc.) can only be evaluated via the presence of a time stamp in the document. Other types of temporal expressions (such as "in several weeks," "in the evening," etc.) express vague temporal information that is difficult to place on a timeline. "The normalization task of a temporal tagger is to assign the same value to all expressions carrying the same semantics or referring to the same point in time" (Strötgen and Gertz 2013).

One of the first research initiatives—TIMEX2 (Ferro et al. 2001), a standard annotation of temporal expressions—was initially developed as part of the TIDES (Translingual Information Detection, Extraction, and Summarization) program. Based on the TIDES TIMEX2 annotation effort, TimeML (Pustejovsky et al. 2003a), a temporal markup language and Timebank corpus (Pustejovsky et al. 2003b) containing annotated events, times, and temporal relations was developed to identify events and temporal expressions in natural language texts. To some extent, researchers used informal discussions of online communities to tag, retrieve, and normalize temporal information (Wen et al. 2013). Most importantly, researchers explored and discussed the challenges in extracting temporal orientation from social media messages such as Facebook messages (Schwartz et al. 2015). In contrast to previous research, a temporal ontology, TempoWordNet (Ga et al. 2014; Hasanuzzaman et al. 2016), was automatically constructed by adding temporal information to the words from WordNet (Miller 1995) using a two-step classification approach. However, we argue that if time-related words were automatically extracted from WordNet, these words would not necessarily be representative of the words ordinary people use in their daily communications. Thus, in our preliminary study (**Paper 6**), we developed our T-wordlist by manually collecting temporal words.

Paper 6 was inspired by previous research on temporal information but it differs in two aspects when compared with the methods mentioned above. First, most of these methods use advanced natural language processing (NLP) techniques, such as parsing, classification, etc., to identify and extract temporal events and thereby to find temporal relationships between the events. In contrast, our approach uses a simple, easy-to-use lexicon-based approach using manually collected time words to identify and filter texts containing time-indicating information. Second, most of the above-mentioned temporal work primarily targets news media and articles (such as Wikipedia) where language styles are generally formal; hence, as also indicated in (Wen et al. 2013), applying these techniques to more informally written texts such as social media posts is challenging. Since our primary focus is identifying time-indicating expressions in social media texts, we chose a lexicon-based approach suitable for processing social media texts in close to real-time. There are many dictionaries available that are oriented toward sentiment analysis. A glimpse of extant literature will be presented below.

4. Research Design

Overall, our research design guides the research process from data collection to appropriate selection of theories, to choosing methods to analyze the data for exploring answers to the research questions. In particular, our research design justifies the methods we selected for this research and also discusses the inclusion of theory. The positioning of theory often leads to two different reasoning approaches: deductive and inductive (Blumberg et al. 2005). We present the theories applied as part of this dissertation in the theoretical background section. However, the main framework of this dissertation consists of multiple theories and empirical methods depending on the focus of the research activity: literature review versus empirical analysis. In order to expand the depth and breadth of research, this dissertation includes a mixed-methods research design, which will be further explained below.

In the following subsection, we first present research methodology; and then present various empirical analyzes in the second subsection.

4.1 Research Methodology

Social science research increasingly applies both quantitative and qualitative methods for studying research problems. However, with "the diffusion of the Internet, the proliferation of numerous nonwork related systems and **social media** have now made IT an integral part of individuals' lives." When IS researchers encounter such situations they face certain theoretical (Venkatesh et al. 2013) and methodological challenges (Sivarajah et al. 2017) concerning how to handle new forms of data such as social media. In this regard, researchers argue a mixed-methods (qualitative and quantitative) approach best allows IS researchers to conduct research in such situations and subsequently make contributions to theory and practice (Venkatesh et al. 2013). Moreover, it is also argued that mixed methods play an important role in the analysis of social media data; while quantitative and highly automated analyzes might be useful to for examining frequencies of interactions and structures of social networks, qualitative analysis facilitates understanding the meaning of social interactions (Ågerfalk 2013). Admittedly, since we have collected only one set of data—social media data—for the studies used here, this dissertation does not represent a mixed method according to a traditional mixed-methods point of view. In any case, we combined both qualitative and quantitative analyses within and across studies.

The empirical studies of this dissertation employ both quantitative and qualitative approaches. The applied descriptive statistical methods and computational techniques fall under the quantitative approach, while the manual content analysis of messages and manual coding of topics (topic modeling) fall under the qualitative approach. Manual content analysis (CA) is a suitable approach for analyzing digital texts such as online forums and social media conversations (Hamad et al. 2016). A primary objective of this dissertation is the analysis of social media disaster data, particularly textual messages, therefore **Papers 2**, **3**, **4**, **6** employed qualitative content analysis (manual coding) along with quantitative analysis, such as descriptive statistics and computational techniques and tools in order to derive answers to research questions. Using different coding schemes, we applied content analysis in **Paper 2** to analyze Hurricane Sandy data. In order to analyze the whole dataset of 1.65 Million tweets, in **Paper 3,** we developed a text classifier using a naive Bayes algorithm using manually coded training sets. Similarly **Paper 4** focused on applying an unsupervised computational and probabilistic technique—topic modeling, which facilitates clustering of words from Twitter messages using a tool (www.MineMyText.com). The clustering and labeling of topics is a qualitative approach because it involves the manual coding of topics. Following the described procedure, we combined both qualitative and quantitative analyses within and across studies. However, **Paper 1** (literature review) set the platform/foundation for the dissertation, while **Paper 5** offers reflections of empirical studies (**Paper 2, 3**) to further extend the research regarding the application of other techniques such as topic modeling (**Paper 4**) and to contribute to the development a dictionary-based approach through our creation of a wordlist (**Paper 6**).

In order to infer meaning from the textual content, the process of manual content analysis consists of "deductive" and "inductive" approaches (Elo and Kyngäs 2008; Hsieh and Shannon 2005). In general, the use of an approach (inductive or deductive) "partly" directs the required sequence of methods (Kondracki et al. 2002). The deductive approach begins with predefined categories or concepts and sifts the data to test the theory or extend the theory. The inductive approach is data driven and grounded in the data. The emerging results provide new insights to the researchers (Kondracki et al. 2002). **Paper 2** and **Paper 3** follow a deductive approach, whereas **Paper 2** builds on predefined categories to extract situation awareness information from the affected individuals (original source) of Hurricane Sandy. **Paper 3** investigates the perceived presence based on message content, and for this purpose we applied predefined theoretical concepts (social presence) to conduct the empirical analysis. However, for **Paper 1** we applied different theoretical lenses to analyze and categorize the extant literature but not to test the theory or extend it. As for **Paper 4** the appropriate approach was inductive, as topics were derived using a data-driven computational approach. We inductively derived further categories and themes by interpretation and by inferring the meanings of the topics. The discussions of methods, challenges and opportunities for **Paper 5** were derived inductively based on **Paper 2** and **Paper 3.**

4.2 Data Collection

In general, Twitter is a rich source of information and it disseminates information quickly in the form of short messages known as tweets (Java et al. 2007; Velichety and Ram 2013). Users can exchange messages by adding links, pictures, hashtags, and video clips, thereby empowering users to exchange rich information they regard as important. Twitter has garnered scholarly attention as the most effective channel during extreme events (Acar and Muraki 2011; Ahmed and Sargent 2014; Chatfield et al. 2014; Mirbabaie et al. 2014; Sinnappan et al. 2010). The entire focus of this dissertation is on disaster-related user-generated content, particularly on Twitter. We chose Twitter as a platform for the data because of its large-scale public data access. Moreover, it provides an opportunity for users to directly share their reality. Specifically, in one way or another, the dissertation papers all focus on Twitter messages. Especially for the empirical studies, we also considered additional metainformation concerning the relevant messages, along with the messages themselves. The details of the disaster cases and descriptions of their datasets are discussed below.

Hurricane Sandy: For the first empirical study, **Paper 2**, we extracted Twitter data related to Hurricane Sandy using publicly accessible tweet IDs (Zubiaga 2015). Hurricane Sandy, a natural disaster, hit the East Coast of the US in 2012 and made landfall on October 29th with devastating impact. Using the tweet IDs, we retrieved messages between May and June 2015, using the Twitter Rest APIs. The originally shared dataset comprised nearly 15 million tweets with IDs, but only approximately 11 million tweets were downloaded and retrieved. Altogether, 3,983,288 unique Twitter users sent messages in 61 different languages about Hurricane Sandy during the disaster. As explained in **Paper 2**, only 0.93% of the tweets contain geolocation information posted in the English language. We discuss the detailed data preprocessing in **Paper 2**.

Chennai Floods: Chennai, a southern Indian city, received a devastatingly high amount of rainfall in December 2015—especially during the first few days of December. The rainfall intensified and Chennai received 34 times the amount of

normal rain in a day (Misra 2015), which created a natural disaster of massive flooding in the city.
Social media monitoring tool Radian 6 was used to collect the Twitter messages using the hashtags #TNflood, #chennaiRains, #chennafloods, #chennaiRainsHelp, #IndiaWithChennai, and #chennaiMicro to extract related tweets. The timeline for the collected data was from November 30th through December 16th, 2015 with a total dataset consisting of 1.65 million tweets posted by 209,644 unique users. The Radian 6 tool provides few Twitter attributes such as tweet ID, author, content, and followers count, but it does not provide full metadata such as retweet status, retweet count, or original tweetId for retweets. Hence, we again downloaded the whole dataset via the open Twitter API using tweet-Ids with the help of a custom tool and reanalyzed the dataset to segregate the original tweets from the retweets based on the retweet status information of the tweet. **Paper 3** and **Paper 4** focus on analysing the perceived presence of social media users on the Twitter platform and understanding the emerging phenomena that evolved during Chennai floods. These papers also present the descriptive statistics and data preprocessing.

Disparate disaster datasets: For **Paper 6,** we chosen different datasets to build our wordlist. For that purpose, we collected 12 different publicly available disaster datasets from CrisisLex.org (Olteanu. 2017), which is a repository of crisis-related social media data. Each of the datasets consists of approximately 1,000 to 1,200 tweet messages from both natural and man-made disasters, such as wild fires, floods, shootings, and so forth. Together, the datasets consist of approximately 13,000 tweet messages.

4.3 Empirical Analysis

We organize the presentation of the various content analysis/text classification techniques applied in this dissertation based on the addressed research objectives respectively. We divide the empirical analysis into content analysis, a supervised machine learning approach, and an unsupervised (topic modeling) approach.

4.3.1 Manual Content Analysis

Content analysis helps researchers create their own context of inquiry and constructs to make the texts more meaningful. It is a qualitative research technique for achieving replicable, reliable, and valid inferences from data on an aggregate level that opens an avenue for understanding trends, patterns, and differences (Krippendorff 1989; Lombard et al. 2002). Even though researchers often criticize this approach for potential subjectivity biases because of the researcher’s assessment (Harwood and Garry, 2003), through a structured iterative process researchers can draw objective, replicable results and valid inferences. In general, a structured process (Morris 1994) describes a systematic approach for content analysis. This approach not only guided us through a step-wise iterative research process, but also made the whole analysis process more transparent. The first step in content analysis is to define a unit of analysis and then decide on predefined labels based on theoretical concepts. Thorough discussions are necessary between coders to get a better understanding of the coding scheme concerning what constitutes the categories and what does not. To eliminate subjective bias and discrepancies, text messages must be coded independently by at least two coders. Finally, results must be compared in order to resolve the discrepancies concerning the concepts. This iterative process helps increase the validity and reliability of the data by confirming intercoder agreement (Krippendorff 1989). We adopted the same procedure in all the empirical studies to decrease the discrepancies and also to achieve high values of intercoder agreement. In general, the unit of analysis can be a few words to a few sentences, and, therefore, a tweet message is a unit of analysis in **Paper 2**, **3**, **6**, since a tweet can be objectively identified by coders (Rourke et al. 2001). Even though **Papers 2**, **3**, **4**, and **6**, build on content analysis, the aim of applying content analysis is guided by different research objectives and for different purposes. In **Paper 2**, we derived relevant information from the messages by applying the content analysis (deductive approach), and later we visualized the information spread on a temporal axis with the help of graphs. For **Paper 3**, we applied content analysis (deductive approach) by manually labeling the messages for the training set to train the classifier. With the help of classifiers, we categorized the whole dataset containing 1.65 million tweets. For **Paper 4** (inductive approach), we applied topic modeling on the dataset, and defined the unit of analysis as a "topic." Furthermore, we individually analyzed topics, along with tweet content, to cluster the topics into meaningful categories and further derive the themes from the categories. The aim of **Paper 6** is different, as we filtered the tweet messages by applying the T-wordlist. Then we applied content analysis to validate the data in order to determine whether or not a tweet message contained temporal information.

4.3.2 Supervised Machine Learning Approach

Text classification can be defined as a process that comprises assigning a predefined category of labels to new texts or documents based on a probabilistic measure of likelihood using a training set of labeled texts (Yang and Liu 1999). A naive Bayes classifier is a probabilistic classifier that estimates the probability of a given text based on the joint probabilities of words and categories using a bag-of-words approach. The naive part of the classifier is that it assumes that the conditional probability of a word over a category is independent from the conditional probabilities of other words over that category. The naive Bayes assumption makes the classifier far more efficient and practical than the exponential complexity of other classifiers and also it works quite well for the text classification with a fair amount of accuracy; therefore, it is one of the most used techniques for text classification (Yang and Liu 1999; Zhang and Li 2007). With support from our colleagues, we built a naive Bayes classifier for classification of tweets. For this purpose we manually coded the tweets to get the training datasets, as explained in the previous section.

In general, a naive Bayes classifier requires a training set containing manually coded messages; 80% of the training set is used to train the classifier and the other 20% is used to test the accuracy of the classifier. The text classification is performed in several iterations with each iteration adopting different strategies (Narayanan et al. 2013) to enhance the accuracy of the classifier. We employed feature selection to yield the best accuracy for naive Bayes classifier strategies, such as stop words removal and bigram association measures.

In general, performance of a machine learning algorithm can be described using four measures: precision, recall, F-measure, and accuracy (Powers 2011; Yang 1999). These measures are super imposed over the statistical variables: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). True and false positives refer to the number of predicted values that are correctly identified and incorrectly identified, whereas true and false negatives refer to the predicted values that are correctly rejected and incorrectly rejected respectively. Building on these variables, precision is defined as the ratio of predicted positive values that are correctly identified as real positive values, i.e., TP/(TP+FP). Similarly, recall (also known as true positive rate) is defined as a ratio of correctly predicted positive values over all positive values, which is TP/(TP+FN). Moreover, the F-measure or F-score is a trade-off between precision and recall and it is defined as a single measure comprised of the harmonic mean of the precision and recall. These three measures—precision, recall and F-measure—provide performance information at the level of labels, whereas accuracy of the classifier provides information about overall performance of the classifier, which can be defined as (TP + TN) / (TP+FP+TN+FN).

Paper 3 uses supervised machine learning approach to classify 1.65 million Twitter messages. The accuracy of our classifier is 0.805 (80.50%) and out of 1.65 million tweets, 37% of the tweets were classified as social presence, containing intimacy and immediacy categories.

4.3.3 Unsupervised Machine Leaning Approach

According to text analytics, unsupervised methods can identify the underlying text features in a text corpus by using clustering methods without explicitly imposing the need for specifying the categories of interest before performing the textual analysis (Grimmer and Stewart 2013). Topic modeling is a popular unsupervised clustering method for text analysis that provides a quantitative technique for the analysis of qualitative data. Although the automated computational analysis of textual data is constrained by a computers' limited ability to process the meaning of human language, researchers have demonstrated that it is a valid and reliable tool when provided with sufficiently large datasets (Halevy et al. 2009). Hence, statistical techniques like topic modeling are emerging as a novel and complimentary strategy for researchers interested in analyzing large collections of qualitative data in a scalable and reproducible manner.

The latent Dirichlet allocation (LDA) algorithm is capable of inductively identifying topics running through a large collection of documents and assigning individual documents to these topics (Blei 2012; Blei et al. 2003). The idea behind LDA is rooted in the distributional hypothesis of linguistics (Firth 1957; Harris 1954), which posits that words that repeatedly co-occur in similar contexts (e.g., documents, paragraphs, sentences) tend to share meaning and, hence, can be used as proxies for describing the content of a text. For example, the co-occurrence of words like "temperature," "wind," "rain," and "sunshine" in a set of tweets can be interpreted as a marker for a common topic of these tweets, namely "weather." In contrast to hard classification or clustering methods, which assign each document to exactly one category, probabilistic topic modeling algorithms like LDA allow that documents belong to multiple categories (topics) with a varying degree of membership. So statistically speaking, LDA represents documents through a probability distribution over a fixed set of topics, and each topic, in turn, through a probability distribution over a fixed vocabulary of words. For example, a tweet may be 60% about the topic "weather"—which, in turn, is represented by words such as the ones mentioned above—and 40% about the topic “New York City,” which might contain words like “nyc,” “manhatten,” and “big apple.” Grouping and aggregating the topic distributions of a large number by metadata (e.g., author, time, geography) facilitates quantitatively summarizing content, detecting differences in content between subgroups of documents, or tracking and tracing the development of topics over time. By applying the topic model, **Paper 4** identifies latent concepts in order to explore the topics we derived from the data to investigate emerging phenomena.

5. Main Research Results

The following subsections provide a condensed overview of the research objectives, the research background, the empirical analyses, and the key findings of the six papers that are part of this dissertation.

5.1 Paper 1: Disaster Management and Social Media Use for Decision Making by Humanitarian Organizations

The emergence of social media makes it possible to acquire real-time information about the situation in a crisis or disaster zone from the affected people themselves. Rather than waiting for nongovernmental organizations (NGO) or governmental organizations to introduce technology for managing disaster relief efforts, individuals in affected areas often still have access to social media and are able to provide information electronically. Currently, the general public tends to use social media sites in a disaster situation to either obtain or contribute information (Palen and Liu 2007). People often switch to personal communication channels when traditional communication channels fail or face operating difficulties during a disaster (Qu et al. 2009).

While researchers claim that disaster-related social media research continues to be an underresearched area (Tim et al. 2017), it is still unclear how user-generated content can assist humanitarian organizations in real time in managing decisions and organizing disaster management and operations. Thus, **Paper 1** addresses this research gap by conducting a systematic literature analysis to answer essentially two research questions: *To what degree does existing research reflect the real-time use of information based on social media in disasters* and *How can existing research findings be used to help humanitarian organization respond to disasters?*

To answer these research questions, we conducted a systematic literature review (Webster and Watson 2002) to collect the relevant articles. In all, we analyzed 45 articles and clustered them in order to understand how many research articles focus on the analysis of real-time social media information of users affected by a disaster. We clustered the 45 articles into four different categories based on their research focus. In order to answer our research questions, we focused on the first and third categories, comprised of 15 articles. In order to set a theoretical basis for this research, we also adopted two guiding frameworks that deal with the functions of social media use during disasters (Houston et al. 2015) and disaster management phases (Cozzolino 2012). We coded the literature along the two dimensions: disaster management phase and social media use during disasters; we also categorized the articles based on their research focus along these dimensions. The mapping of articles allowed us to understand how social media usage in disasters actually resonates with commonly employed disaster management phases.

We found that most of the uses of social media during disasters in the literature overlap across different disaster management phases. Hence, some of the articles are related and mapped to more than one phase. The following are some of the important findings from **Paper 1**.

Mitigation: We were not able to identify any articles that fit into this category. One explanation could be that there are only limited application areas imaginable concerning how social media can be used before a sudden and unexpected disaster occurs. In cases of flooding or wildfires, it is nevertheless thinkable that social media might be used to mitigate the most severe effects of such catastrophes given enough time to prepare the general public.

Preparedness: Twitter-based monitoring applications and tools help event detection and location extraction (Kumar et al. 2011; Sakaki et al. 2010) and Twitter-based data analysis describes the process of information production and distribution by the general public or identifies the information shared by local residents to enhance situation awareness(Starbird et al. 2010).

Response: The key practices we observed in social media communication during disasters include sharing of real-time information, extending moral support to communities, and proposing relief activities. Twitter also acts as a system for sharing different types of messages for different purposes, such as providing situational updates, asking for help, expressing opinions, offering emotional support (Qu et al. 2011) and providing crisis communication (Heverin and Zach 2010). Monitoring applications based on tweet content helps to visualize the disaster-affected area and to provide geolocation information (Lingad et al. 2013; Yin et al. 2012), and crowdsourcing open software applications collect data and provide the visual information of affected areas as well as needs and urgencies of the victims (Currion et al. 2007; Gao et al. 2011).

Recovery/Reconstruction: In this phase, affected people attempt to reconnect to their community through web forums and social media platforms such as Twitter (Qu et al. 2011; Qu et al. 2009; Shklovski et al. 2008; Sutton et al. 2008). However, crowdsourcing applications also become useful in the reconstruction phase to locate missing friends and family members.

Finally, according to **Paper 1**, mapping and applying the research findings of the collected articles to disaster management can illuminate the true potential of social media for emergency management. Our findings revealed that social media has the capacity to assist with a decisions support system that NGOs and emergency relief organizations could use to improve the efficiency of decision-making activities.

5.2 Paper 2: Enhancing Disaster Management through Social Media Analytics to Develop Situation Awareness: What Can Be Learned from Twitter Messages about Hurricane Sandy?

Among the different social media platforms, Twitter is one of the largest and fastest communication media for disseminating news and information in the form of short messages known as tweets (Java et al. 2007). Human nature dictates a desire to communicate and create a feeling of closeness with one another in uncertain times. People share and read news and messages more often during disaster events such as floods, wildfires (Starbird et al. 2010; Vieweg et al. 2010b), and earthquakes (Qu et al. 2011). Therefore, on an aggregated level, real-time information about the situation on the ground contains useful and, for disaster response forces, important intelligence information, which relief organizations, such as the Red Cross or the US Federal Emergency Management Agency, can use to more accurately coordinate their activities. For our research, we were specifically interested in the value of Twitter messages for contributing to situation awareness in areas affected by a disaster.

Despite an increasing interest in social media analytics in information systems research, only a limited number of publications focus on the uses of social media during disasters—e.g., the use of Twitter messages for disaster management. **Paper 2** addresses this gap and strives to answer the research questions; namely, *Which different types of information relevant to disaster management can be identified and what are the nature and characteristics of this shared information?* In order to illustrate what a disaster management agency can learn from social media data, we analyzed approximately 11 million tweets sent between October 25th and November 5th, 2012, during the period of Hurricane Sandy, which hit the US East Coast and made landfall at New York City on October 29, 2012.

Data preprocessing was an important aspect of the study because we were most interested in tweets shared from the affected area, and thus focused on extracting messages shared by Twitter users in the affected areas of Hurricane Sandy. The longitude and latitude coordinates of a tweet indicate the exact location of a Twitter user at the time the information was shared (Graham et al. 2014). In order to extract the tweets from the hurricane path, filtering the tweets based on their geographical location information was an important task. Using the Cosmos tool (Burnap et al. 2015), we processed the whole dataset of 11 million tweets and extracted the tweets that contained longitudinal and latitudinal coordinates. We also used the CartoDB (CartoDB 2015) tool to 1) filter the English language tweets, 2) visualize the data, and 3) narrow down the tweets geographically to the path Hurricane Sandy took along the coastline. As a result, we ended up with 68,800 tweets originating from the hurricane-affected area between October 25 and November 5, 2012. We also manually screened the tweets to exclude commercial content such as advertisements and spam, and also conducted manual content analysis. Our first level of the coding scheme consisted of two information source categories: original and secondary (Starbird et al. 2010). Our secondary coding scheme consisted of informational messages and action-related, opinion-related and emotion-related messages (Qu et al. 2011). We combined both coding schemes to obtain an integrated coding scheme and, with the help of our new coding scheme, we analyzed 677 geolocated tweets.

Paper 2 evidenced only 1.05% of 11 million tweets actually contained geolocation information. Moreover, out of 677 tweets, 75% of the tweets represented original source material, while 25% were categorized as secondary sources. In the pre-disaster phase, the number of tweets from secondary sources exceeded the number of tweets from original sources. As the disaster unfolded, people slowly started sharing their own personal experiences, observations, and also discussed the situation based on common knowledge or by using material from other sources (Starbird et al. 2010). Hence, from October 27th onwards, the tweets in the original category increased until October 31st, since the hurricane hit New York City on October 29th. During the disaster phase, people's desire to understand, share, and estimate the impact of the current situation increased considerably. Approximately 73% of the tweets were informational messages, 15% tweets were emotion related, and opinion- and action-related tweets comprised 8% and 4% of tweets respectively.

The affected individuals shared real-time information either to provide situation updates or because they were in need of real-time help. This bottom-up generated information would be useful for emergency responders and disaster management agencies to keep track of real-time situations in the affected areas. This type of community-level situation awareness would help disaster management organizations make better decisions and would facilitate more effective handling of the disaster response.

5.3 Paper 3: Presence of Social Presence during Disasters

Twitter is changing traditional communication practices during emergencies due to real-time user-generated content, which is enhancing collective collaboration (Vieweg et al. 2008). Considering the media-related component of social media, based on the degree of social presence, social networking sites fall into the "medium" category because users can connect and communicate with others through text messages and can also upload videos, pictures, and other forms of information (URLs and website links) (Kaplan and Haenlein 2010). During disasters, despite the lack of face-to-face interactions, people ask for help, provide moral support, and even help each other on social media platforms. It is important to explore what makes people perceive others' needs, feel empathy, and lend support on the basis of short text messages. Thus far, research has used survey strategies or interviews to understand the social presence of online learning (Gunawardena and Zittle 1997; Tu and McIsaac 2002) or social media platforms (Al-Ghaith 2015; Xu et al. 2012). Especially during disasters, when people seek help and support on social media sites, other online users feel intimacy with and immediacy for those affected and provide support by sharing relevant information online and actively participating in relief activities. Without the feeling of social presence (intimacy and immediacy) people who are in need do not ask for help nor do people come forward to lend support. There is no research as yet that has explored social presence based on tweets during disasters. In this regard, using the theory of social presence as a methodological framework, this paper demonstrates how seemingly ephemeral and hastily written text messages on Twitter can create a feeling of intimacy and immediacy that can stimulate information sharing at a community level. This community-level information sharing can serve as a useful source of information for disaster management agencies. **Paper 3** addresses the research question: *How can social presence be detected through a content analysis of tweets, and what role does it play for building relationships, cooperation, and collaboration during times of emergency?*

In order to understand how people express intimacy and immediacy as forms of social presence in times of disasters, we analyzed approximately 1.65 million tweets from a devastating flood in Chennai, India, which took place in December 2015. We applied a supervised machine learning approach to analyze the entire dataset. Our methodology was three-fold: 1) operationalize social presence concepts: intimacy and immediacy on social media, 2) conduct manual content analyses to develop a training dataset for message classification, 3) train and use a naive Bayes classifier to classify the dataset.

The results in **Paper 3** explain that the overall accuracy of the classifier is fairly high (i.e., around 80% of the prediction is considered accurate for correct predictions). In terms of precision and recall, the immediacy category received good values that were better than those of the intimacy category. For the F-measure, a value of around 0.6-0.7 indicates a good performance and the F-measure values for all categories indicated reasonably good performance other than for the intimacy categories. Out of 1.65 million tweets, 37% of tweets were classified as social presence and contained intimacy and immediacy categories, whereas the remaining 63% of tweets were classified as belonging to no predefined category. This is consistent with previous research that states that during disasters people post and share different types of messages containing suggestions, comments, criticism, etc., (Qu et al. 2011; Vieweg 2012). People also discuss media and government and vent their frustrations or anger; such tweets could be not categorized as social presence tweets. However, when it comes to retweets, which constitute 80% of the dataset, the proportion of social presence retweets was higher; around 42% of retweets could be categorized as related to intimacy or immediacy. Immediacy-related messages address the needs and urgencies of affected individuals, while intimacy-related messages aim primarily at providing moral support and creating awareness.

Emergent support groups evolve immediately after a disaster unfolds, and research has demonstrated that that these groups work together collectively to cope with the situation (Drabek and McEntire 2003). Our research illustrates how online social presence can create a sense of responsibility and common identity during disasters among social media users. One of the reasons underlying active participation on social media sites during emergencies may be the perceived online social presence. Moreover, people, who feel higher levels of social presence continue to use and interact more on Twitter (Han et al. 2015). Thus, social presence plays a significant role; despite the lack of face-to-face communication, social media sites facilitate a sense of intimacy and immediacy among individuals, leading to actions of assistance, if even through the act of simply reading and retweeting a tweet.

5.4 Paper 4: The Role of Social Media for Collective Behavior Development in Response to Natural Disasters

Twitter has garnered scholarly attention (Chatfield et al. 2014) for being the most effective channel used during extreme events (Mirbabaie et al. 2014). While communicating with others on Twitter, users perceive the presence of others (Han et al. 2015) and play different roles (Lee et al. 2013; Reuter et al. 2013) while coordinating their behaviors in order to make sense and cope with the situation at hand (Bunker et al. 2017; Mirbabaie and Zapatka 2017; Stieglitz et al. 2017).

Prior research has focused on identifying the categories and taxonomies of the types of information being shared among users during disasters (Olteanu et al. 2015; Qu et al. 2011; Vieweg 2012). However, extant research lacks theoretical attention devoted to the dynamics of and relationships between the identified categories. Against this background, a primary objective of our research is to identify the topics that are shared via social media during disasters to stimulate collective behavior, and to investigate how users' information-sharing behavior changes as the event unfolds. In other words, we aim at understanding a process that is based on the temporal and logical relationships between the topics that are shared on social media and the corresponding activities that disasters stimulate. We built a process model to explore an emerging phenomenon on social media during a disaster. To identify how collective behavior develops, we chose an unsupervised method—topic modeling—for which manual content analysis is not necessary to train the system, as it can be applied directly to extract the topics (Debortoli et al. 2016). For this research, we applied topic modeling as the computational grounded theory approach and analyzed the topics that were extracted from the data. Since the topics are data driven, we manually interpreted and further clustered these topics into generic categories, and then further into themes, and traced their development over the period of the disaster. We focused on analyzing the Twitter data of the Chennai flooding using topic modeling and subsequently coded the results in order to conceptualize a collective behavioral process model.

To interpret and code the identified topics, two researchers independently inspected all 50 topics generated for each day along with the relevant tweet content. Subsequently, to decrease subjective bias we compared the results and synthesized them in order to minimize errors. For most of the topics (around 80%) the researchers generated very similar labels and identical labels. The researchers discussed disagreements in interpretation until consensus was reached. In order to reach a higher level of abstraction about the topics, we organized all 50 topics that were identified for a given day in descending order based on their probabilities. We then focused on the top 10 most prevalent topics for each day and tried to manually cluster them into higher-order categories in order to get an overview of the most prevalent categories for each day. For example, we clustered Topic 5 ("updates on routes"), Topic 28 ("water level"), and Topic 42 ("public and private organizations opened/closed") into the category situation updates, since all of these topics were related to information-sharing about the condition in real time. Similarly, we followed the same procedure to cluster all the topics into different categories.

Later we built a process model to explore an emerging phenomenon on social media during a disaster. For this model we plotted topics along the timeline of the disaster and further clustered the topics into categories for two reasons: First, to recognize which categories emerge at which point in time, thereby signaling relevance while the disaster unfolds. Second, to develop a process model that explains collective engagement of social media users in times of disasters. After aligning each day's categories on a more abstract level, we derived the themes and defined their names. For example, the categories 1) initial information about disaster, 2) situation updates, and 3) criticism about insufficient attention, explicitly provided information on rain updates, road updates, etc., but implicitly this information dissemination created awareness among social media users. Hence, we grouped the three categories under the theme "collective awareness." In the same manner, we created the rest of the themes, based on the prominent categories arising each day. The main findings of **Paper 4** are along four main themes: collective awareness, collective concern, collective empathy, and collective support.

The collective behavior process has enhanced our understanding of the emerging phenomena (emerged concepts) on social media sites during a disaster situation. The process or sequence of categories explains first, what types of information people share on social media over the period of a disaster, and second, why and how peoples' information-sharing behavior changes and further leads to collaborative and cooperative activities as the event unfolds. In addition, the discovered process revealed situations and activities performed by victims as well as by witnessing individuals. Moreover, the activities that individuals perform or perceive on social media sites during a disaster have the capacity to reveal latent behavioral patterns. To some extent, the exploration of themes through categories enhanced our understanding of those latent behavioral patterns that make up the sequence that begins with collective awareness, concern, and empathy, and ultimately leads to collective support. In this way, social media use has the capacity to enhance and influence community resilience in a positive way (Kaufhold and Reuter 2016).

5.5 Paper 5: Social Media for Disaster Situations: Methods, Opportunities and Challenges

Humankind is prone to different types of disasters, such as earthquakes, floods, epidemics, etc., and has dealt with the devastating consequences of disasters throughout history. Disasters are events that disrupt the normal functioning of a community impact people's lives, economy and environment. When compared to epidemic diseases or economic crises, both natural and man-made disasters occur relatively suddenly and require immediate relief activities (Cozzolino 2012; Debarati Guha-Sapir 2016). During disasters, the availability of real-time information is essential for both disaster management agencies and humanitarian organizations to make effective decisions and coordinate immediate response activities. The quicker the organizations can react and respond, the more effective the relief efforts will be. However, the volume and the velocity of social media data make it challenging to monitor and extract valuable information.

In this regard, by using two case studies that we explored as examples, **Paper 5** discusses the applied methods and their opportunities and challenges regarding the analysis of social media disaster data for extracting important information that can assist disaster management agencies. The two case studies we used are Hurricane Sandy (US, 2012) and the Chennai floods (India, 2015). We believe the following observations will be beneficial to practitioners and academics who are interested in social media analytics in general. In particular, these insights could be useful for disaster management agencies who seek to integrate social media into routines and processes.

Social Media: Provider Of Information: People use different social media platforms, such as Twitter (Eismann et al. 2016), Facebook (Eismann et al. 2016; Kaewkitipong et al. 2016), Flickr (Liu et al. 2008), etc. during disasters. Our first empirical study (**Paper 2**) evidences that during disasters people share firsthand observations and experiences on social media. Moreover, people who witness an unfolding event or live in the vicinity of an event immediately share information on social media (Krstajic et al. 2012). Social media complements the existing early warning systems because it facilitates rapid information dissemination (Chatfield and Brajawidagda 2013), and because early warnings can also be detected on social media by monitoring message content (Sakaki et al. 2010). Situation awareness is important for disaster management agencies to make appropriate decisions concerning disaster response activities. Similarly, creating public awareness by disseminating disaster preparedness and warnings using social media is also important.

However, we did notice certain challenges associated with using social media disaster data. Data volume and velocity are major issues complicating the handling and processing of social media data in real time, since ongoing social media activity explodes during disasters. For example, 11 million tweets were sent in just a few days during Hurricane Sandy. Data relevance is a challenging issue. Identifying important information potentially useful to disaster relief agencies is a challenge due to the huge volume of social media data generated during disasters. A major portion of the data contains unrelated information, such as commercials, and is also comprised of repeated, reposted messages (80% of the Chennai dataset consisted of retweeted messages). During disasters, people use different social media platforms (e.g. Twitter and Facebook) to share information. Integrating and unifying information from these different platforms is a challenge because of different data formats. The validity and quality of information also presents challenges. Rumors and fake news create panic among the people. Most importantly, given the nature of social media, these rumors also spread very fast (Mendoza et al. 2010; Oh et al. 2010; Starbird et al. 2014).

Geographical coordinates (Graham et al. 2014) facilitate filtering and retrieving the information from particular disaster-affected geographic regions. Metadata of a tweet provides geographic information in the form of latitudinal and longitudinal coordinates. Given the importance of geolocation information, systems such as emergency situation awareness (ESA) (Yin et al. 2012) tag the geoinformation for event detection, and systems like SensePlace2 (MacEachren et al. 2011) filter and extract geographic information from tweets to visualize the information in maps. However, according to previous research, and also based on our case study, the challenge is that only 1% of data contains geographic information. One explanation for this could be that the end users do not share location information due to privacy concerns. Applying different methods to infer location information is considered a privacy intrusion. Some of the social media platforms do not provide geolocation information at all. Facebook, for example, offers a "check in" option (Hannay and Baatard 2011), but does not provide geoinformation in the metadata.

Emergent digital groups are beginning to arise during disasters. People in affected areas play a big role in facilitating immediate rescue operations. The feeling of connectedness in distress situations is leading to the emergence of a new type of volunteer, who not only shares information and coordinates rescue activities through social media in affected areas, but also helps assess the damage or collaboratively collects the necessary information for affected individuals. The challenge is this new type of volunteers only emerges during disasters on an ad hoc basis, hence it is difficult to contact them and include them in the disaster relief activities.

Research communities have applied different methods, such as manual content analysis and computational methods, to analyze and categorize information. Manual content analysis is a very basic method, featuring human coders who the data and categorize the information. However, this method is only suitable when the dataset is rather "small," because analyzing millions of tweets is not humanly possible. So far, most of the DMA focus on manual content analysis (Wukich 2015). With the help of classifiers, data can be classified automatically by using a machine learning algorithm. Using well-defined training datasets and supervised algorithms, "huge" amounts of data can be classified or easily extracted. However, well-defined training sets are necessary for building the classifiers, and thus might require resources such as a human coder to prepare them. Given the nature of unstructured data, getting high accuracy from a classifier is difficult. Moreover, classifiers that are developed for one type of disaster (e.g. earthquakes) might not perform well with another type of disaster (e.g., floods).

Based on recent research (Reuter et al. 2016; Sheombar et al. 2015) emergency officials are still skeptical about fully integrating social media into their organizational routines. The reason could be the challenges that we mentioned previously regarding data and methods. In order to reach the next level, it is important for organizations to invest time, resources, and personnel (Wukich 2015) to harvest the benefits of social media. Most importantly, to gain value by using social media, organizations must align their goals with strategic initiatives. Various analytical techniques can be used to extract and understand the valuable information that is generated during disaster situations. Automated text-analytical methods are the best means of analyzing and understanding user-generated content in such situations.

5.6 Paper 6: The Development of a Temporal Information Dictionary for Social Media Analytics

Since the introduction of social media, unstructured, user-generated content has been created at an unprecedented rate, and has become an important source of information for social science researchers (Thelwall et al. 2008). The predominant research interest has focused on analyzing the textual content of social media messages, which contain people's opinions, expectations, feelings, and so forth. Sentiment analysis seems to be the research area that has gained the most prominence in this context (Read 2005). Sentiment analysis allows for the automated assessment of positive or negative feelings based on the tonality of a message. For example, by applying sentiment analysis, online retailers can get aggregated results while summarizing the collected reviews with an average sentiment score. In the financial domain, stock market experts can predict stock price fluctuations, based on average sentiments (Read and Carroll 2009). During disasters, sentiment analysis can help define feelings, concerns, or even the panic of affected people (Beigi et al. 2016). However, while sentiment analysis has illustrated its usefulness, an important dimension thus far not been addressed; namely, the time dimension the message refers to. Time also plays a role for another reason: the need to analyze social media data in close to real time, so that, for example, emergency response forces can prioritize responses along the urgencies of the different needs people suffer from in a disaster. For example, tweets containing words like "just," "urgent," or "immediately" might help disaster relief forces prioritize where to move in and help first if they can analyze the data quickly enough, meaning in an automated way, as the following two tweets illustrate:

"A massive earthquake just hit Everest. Basecamp has been severely damaged. Our team is caught in camp 1. Please pray for everyone."

"Urgent. Need 1000 water packets. Please contact us immediately."

Hence it is also important to differentiate what is more urgent in case of an emergency, which also requires the automated detection of time-indicating statements in such situations. Thus far, there has been only limited research focusing on extracting time-indicating statements from social media; furthermore, researchers have not devoted efforts to develop a dictionary or wordlist of time-relevant words and phrases like those available for sentiment analysis. Our research focuses on filling this gap by developing a time-indicating dictionary. Thus, the research question **Paper 6** focuses on is: *How can time-indicating expressions be captured in a dictionary to automatically assess social media data in close to real time?*

Paper 6 followed a methodology consisting of three phases. The first phase builds the T-wordlist by collecting words representing time. The second phase consists of data collection and preprocessing of social media data and applying the T-wordlist to it. The third phase primarily focuses on the validation of the data (extracted by the application of the T-wordlist) with the help of manual content analysis. We collected 12 publicly available different disaster datasets (Olteanu. 2017), each consisting of approximately 1,000 to 1,200 tweets from different disasters, such as wildfires, floods, shootings, etc. Altogether, the datasets consist of 12,831 tweets from disparate datasets. Moreover, these disasters occurred in different countries/geographical locations across the world, including Colorado, Philippines, Australia, Singapore, Los Angeles, etc.

Later, we applied the T-wordlist to the individual datasets. We segregated the tweets into two categories: the tweets that matched with words from our T-wordlist and the tweets that did not contain and were in no way related to any of the words from the T-wordlist. Around 45% of T-words were represented in the data, and, in total, we extracted 4,791 tweets (43% of the total tweets). At this stage, to ensure that the data represented temporal information, we validated the data. Data validation is an important part of any study, because, the same word can have different meanings in different contexts. In order to ascertain whether the tweet represented the time-relevant information as accurately as we intended, we conducted a manual content analysis. We randomly selected a small sample of tweets from four datasets. In the pilot study, one of the researchers analyzed a sample dataset individually to assess whether or not the tweets included temporal information. Later, both researchers looked into the results and discussed them. Again, to ensure validity, in the subsequent stage, both researchers analyzed a sample dataset of around 1,000 tweets from the matched tweets dataset representing different types of disasters. Based on our preliminary results, our future research will focus mainly on compiling a more complete T-wordlist to achieve more accuracy. To make our wordlist more rigorous and robust, we must tackle the problem of how to deal with words that change meanings depending on context. For example, prepositions like, on (Monday), in (the morning), at (night), by (11 o'clock) can indicate temporal information. However, the same prepositions might also indicate position or direction in different contexts. For example, in (the picture), on (the left), at (a concert), (standing) by. In this regard, by taking advantage of a specific crowdsourcing platform, we plan to make our wordlist more useful and generalizable for use with different data analyses. Furthermore, we will use crowdsourcing to categorize the words (based on meaning) into past/present/future, while words that are vaguely temporal being placed in a different category.

6. Contributions to Research and Practical Implications

This research has many theoretical, methodological, and practical implications. Overall, this dissertation contributes to the domain of disasters, disaster management and social media analytics.

6.1 Contributions to Research

First and foremost, this research contributes to the extant research on the use of social media for disaster management. In general, **Paper 1** contributes to the social media disaster research regarding how social media resonates with disaster management phases. In line with existing research, this dissertation focuses on natural disasters to analyze and extract the information types from social media data. Thus far, researchers have extracted different message types from different disasters. Most importantly, researchers have identified and analyzed online communications of police and fire services during Hurricane Sandy (Hughes et al. 2014). However, online communications of ordinary people have not yet been examined. **Paper 2**, in particular, not only focuses on information types but also focuses on who shared (information source) and what type of information was shared during Hurricane Sandy. One important finding of this study is that along with different disaster phases (pre-/during-/postdisaster), the source and type of information also changes.

While prior empirical investigations on man-made and natural disasters have focused on a number of countries, this dissertation is among the first to explore the effective use of social media in an Indian context particularly from the perspective of natural disasters. **Papers 3** and **4** are the first to focus on Indian context to analyze the social media disaster data and the role social media plays during disasters there, thus supporting the argument that social media use is increasing worldwide during disasters. Compared to the first empirical study findings on Hurricane Sandy (**Paper 2**), the study on Chennai floods found richer information available on the Twitter platform. In developing countries like India, when a disaster strikes the threat to life is high in comparison to disaster situations in developed countries; hence, at the community level, people use social media more effectively and efficiently for relief and rescue operations (Kaewkitipong et al. 2016). This community-level, on-site information is important for decision-makers. While previous studies have investigated different information types, in addition to focusing on moral and general awareness information, **Paper 3** specifically focuses on the needs and urgencies of affected individuals, and thus also focuses on immediate (immediacy related tweets) relief activities.

This dissertation also contributes to the research area of DMA adoption of social media. Thus far, this research area has focused on conducting surveys and interviews with employees of emergency response agencies. By analyzing case studies (**Paper 2, 3**) this dissertation provides insights into the current research (DMAs) by illuminating the opportunities as well as the relevant challenges as described in **Paper 5**. **Paper 6** contributes to the existing research on temporal information extraction and social media analytics. The research extends the current research by analyzing the disaster-related social media data to extract temporal information, which has not yet been explored. Most importantly, **Paper 6** fills a research gap by developing a dictionary (methodological contribution) developed to classify social media data based on temporal information; this will be helpful for researchers seeking to extract meaningful information from unstructured data.

Theoretical contributions: As part of the dissertation, **Papers 2** and **3** contribute theoretically to situation awareness and social presence theories, and **Paper 4** explores an emerging phenomenon by inductively deriving the latent concepts from the data (topics). There are only few studies focused on the theory of situation awareness in disasters thus far (Vieweg et al. 2010a; Vieweg 2012); **Paper 2** contributes to the theory of situation awareness, in general, by exploring different types of information relevant to disaster-management and especially by identifying the nature and characteristics of shared information through and analysis of social media data related to Hurricane Sandy.

To the best of our knowledge, previous research has not focused on applying the theory of social presence to the social media disaster data or the context of disasters. The previous research on online learning (Lowenthal 2009) and social media platforms (Han et al. 2015; Lim et al. 2015) have explored the theory of social presence using traditional survey and interview methods; in contrast, **Paper 3** of this dissertation explores the theory of social presence using the textual content (tweet content) of social media data in the context of disasters. **Paper 3** operationalizes the theoretical concepts of social presence to social media use during disasters, and also analyzes the message content through the theoretical lens of social presence in order to further classify the whole dataset by applying the machine learning method. Therefore, our research is the first to apply the social presence theory to understand the phenomenon of why people on social media go to extra lengths to help others during disasters despite the lack of face-to-face interaction. One important implication of our study is that during disasters people perceive the presence of others on social media; to fulfil their social connection needs (Han et al. 2015), users actively take part in social media by sharing and receiving information. Subsequently, these activities lead to real-time rescue and relief efforts at the community level. There is extant research about collective intelligence or collective behavior during disasters; however, no research thus far has attempted to explore or explain the mechanism of collective behavior as an emerging phenomenon. **Paper 4** explores the concepts of collective awareness, collective concern, collective empathy and collective support in order to understand this emerging phenomenon. We argue that these four concepts are predecessors for collective behavior at the community level. This study extends the understanding of how social media sites contribute to knowledge about what is happening at the societal level during disasters, and by identifying latent concepts, this dissertation contributes to the theory of collective behavior.

6.2 Practical Implications

As a whole, this dissertation provides valuable insights to practitioners about the role social media plays during disasters; this dissertation clarifies the importance of social media, along with the uses, challenges, and opportunities it provides. Social media has been identified as a real-time information provider from early social media disaster studies; hence, it is not surprising that our study confirms the increasing importance of this aspect. Social media is becoming a major source of situation awareness during disasters (**Paper 2**) and it is evident in the current dissertation that its use is increasing during disasters, even in developing countries (**Paper 3, 4**). The traditional disaster response systems are in place for inter- or intraorganizational purposes, whether it be for internal communication purposes or for information flow. But these systems are restricted in accessibility and the relational aspect. Most importantly, they are not able to extract real-time information on the ground. Social media acts as boundary object in its use during disasters (Tim et al. 2017). Therefore, though social media will not replace traditional practices of DMAs, its value could enhance organizational alertness in the identification of incidents that require an immediate response. In this context, **Paper 1** explains social media uses during different disaster management phases. Moreover, DMAs are, to some extent, successful in creating awareness by maintaining communication with the public. However, limited resources in DMAs impede the usage of social media during disaster management phases, and lack of time, especially, significantly affects the possibility of monitoring social media during the emergency response phase (van Gorp et al. 2015). Possible reasons for this include data volume, velocity, relevance and integration into the traditional informational systems. Most significant, perhaps are the challenging issues of rumors and fake news during disasters. The reflection of analysis on case studies in **Paper 5** provides relevant information in this regard to practitioners that will help them deal more effectively with social media. **Paper 2** illustrates the challenge in manual content analysis of extracting the relevant information, which is defined by the limits of human information processing capacity. Nevertheless, DMAs continue to use manual content analysis (Wukich 2015). Thus, to assist DMAs in overcoming this limitation, and making valuable information available, **Paper 3**, **Paper 4**, and **Paper 6** demonstrate that there are different computational techniques to classify large amounts of information. These papers not only discuss these methods, but the case studies provide evidence that more and more valuable information is shared by affected individuals themselves.

Though social media use in emergency response operations is still considered to be in its infancy, its use will perhaps be integrated at a later point in time (van Gorp et al. 2015). The general public will continue to increase its use of social media, especially during disasters. An in-depth understanding on how to embrace the power of social media is essential to derive effective disaster response actions (Tim et al. 2017). In this regard, DMA officials must understand social media data characteristics, such as volume and velocity, its relevance and integration, the challenges of fake news, and how rumors cascade, as discussed in **Paper 5.** Moreover, officials need easy-to-use, automatic methods to help extract the relevant messages from large datasets. As stated, having a wordlist, for example, VerbNet, is important for DMAs to process the data in real-time. As researchers argue, “while not perfect, systems incorporating these VerbNet codes is one step towards correctly validating data without human intervention” (Landwehr and Carley 2014). **Paper 6** demonstrates how to improve the intelligence obtained from social media data to access time-relevant information and situation updates. Applying sophisticated and easy-to-use methods for information extraction will help improve the decision-making capacities of DMAs, which will lead to saved lives and support for affected individuals. This dissertation offers the methodological applicability of different text analytics methods on social media disaster data, which will be useful to practitioners. In practice, the usefulness of our T-wordlist is of two-fold. First, social media data can be retrieved and analyzed by using the T-wordlist to understand the urgent needs of victims in disaster relief activities. Second, the same T-wordlist can be integrated into existing crowdsourced applications (such as MicroMappers, Sahana, Ushahidi) for disaster management for coordination of relief activities. DMAs should select the methods that are appropriate to their purposes and adjust the applied techniques to their context in order to obtain meaningful results from the analyses.

Another major practical contribution of this dissertation comprises providing DMA officials with reflections based on the analyses of case studies and offering actionable recommendations for the appropriate means to integrate social media into their routine processes. During disasters, geolocation information plays an important role in extracting valuable information from the affected zone, and also helps reduce noise. However, only 1% of data contains geocoordinates (**Paper 2**). Hence, DMAs should educate the public to enable geolocation information during disasters, at least. However, it should align with United Nations Office for the coordination of Humanitarian Affairs (OCHA) (Raymond et al. 2016) policy standards as a collective responsibility of the entire humanitarian data ecosystem regarding adequate data security, ethical standards, and privacy.

7. Limitations and Future Research

The meaningful insights drawn from our studies should be viewed in the context of their limitations and their relevance for future research.

First, we focus our analyses on a single microblogging platform, Twitter. Even though Twitter data presents rich information, it differs functionally from other social networking sites, because each platform is unique in terms of characteristics and functionalities. Moreover, we did not examine the use of other social media applications that could have complemented the use of Twitter in the investigated phenomena. Thus our methods and analysis should be replicated and refined in different social media contexts in order to be generalized in a way that could open new avenues for future research.

Second, we recognize some empirical limitations associated with our analysis of Twitter data use during Hurricane Sandy. For the first empirical study, we extracted the tweets based on a publicly available dataset. We do not know whether the researchers used hashtags or keywords to collect the data. For **Paper 2** we focused on geolocation information; hence, we had to sacrifice other data that might have contained valuable information. Only 1% of data contains geolocation information. Thus, we may have lost the opportunity to investigate other important information.

Third, we considered all the disasters as if they were in one category. There are many different types of natural disasters, such as earthquakes, tornadoes, hurricanes, and man-made disasters, such as terrorist attacks or industrial accidents. Each disaster is different in its own right depending on their characteristics: man-made verses natural, long response durations versus immediate responses, etc. Our study investigated only natural disasters (**Papers 2**, **3**, **4**)—hurricanes and floods, specifically. Hence the generalizability of our results to other types of disasters is unknown. Moreover, there are several secondary features of the disasters; for example, duration, scope, and magnitude of disaster impact that might not be directly dependent on the disaster phase, but are important variables to study (Kreps 1984). We have not examined that angle in our study. It is important to have a focus on secondary features, which might provide patterns to recognize and future research should focus on it. Even though we analyzed two different contexts, we did not compare results; for example, it might be interesting to explore the cultural differences in using social media during disasters. Furthermore, unlike sudden disasters like earthquakes, some disasters, like floods, are more predictable and have a slower speed of onset, so communities have some buffer time to react and respond online. Online behaviors and activities might also differ based on the differentiated aspects of disasters.

Fourth, we analyzed social media data and further developed a wordlist with an aim to extract the needs and urgencies of affected individuals which would be useful to DMAs. On the one hand, our focus was on analyzing user-generated social media content during disasters, and on the other hand, it concerned understanding the social media uses, methods, challenges, and opportunities associated with disasters, in light of their potential usefulness to DMAs. However, we did not conduct a case study where DMA uses social media in real time; this could have provided more insights and added value to our current dissertation. We evaluated our wordlist as applied to disaster-related social media data. Applying the wordlist to an actual scenario would provide more insights that could help improve the wordlist. From the beginning we have been in touch with the Danish Emergency Management Agency (DEMA). Recently DEMA has started using Twitter to communicate with the public. DEMA also plans to implement a social media listening platform. When DEMA rolls out these plans, we hope to present the research findings to them so that they can be successfully integrated in their processes. An action research study could offer another opportunity to further this research.

Finally, we neither directly experienced the online emerging phenomenon process at the research site, nor did we directly investigate the intimacy or immediacy. These aspects were all based on Twitter/archival data that was collected as the flooding was unfolding. In order to make the research more robust and rigorous, future research could complement the analytics techniques with in-depth qualitative interviews and surveys.

8. References

Abbasi, A., Chen, H., and Salem, A. 2008. "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Transactions on Information Systems (TOIS)* (26:3), p. 12.

Abbasi, M.-A., Kumar, S., Andrade Filho, J. A., and Liu, H. 2012. "Lessons Learned in Using Social Media for Disaster Relief-Asu Crisis Response Game," *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*: Springer, pp. 282-289.

Abdulla, N. A., Ahmed, N. A., Shehab, M. A., Al-Ayyoub, M., Al-Kabi, M. N., and Al-rifai, S. 2016. "Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis," in *Big Data: Concepts, Methodologies, Tools, and Applications*. IGI Global, pp. 1970-1986.

Acar, A., and Muraki, Y. 2011. "Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster," *International Journal of Web Based Communities* (7:3), pp. 392-402.

Ågerfalk, P. J. 2013. "Embracing Diversity through Mixed Methods Research." Springer.

Ahmed, A., and Sargent, J. 2014. "Analysis of Post-Crisis Twitter Communication: A Study of the Iquique, Chile Earthquake," *In: Proceedings of the Australasian Conference on Information Systems* ACIS.

Al-Ghaith, W. 2015. "Understanding Social Network Usage: Impact of Co-Presence, Intimacy, and Immediacy," *International Journal of Advanced Computer Science and Applications (IJACSA)* (6:8), p. 12.

Al-Saggaf, Y., and Simmons, P. 2015. "Social Media in Saudi Arabia: Exploring Its Use During Two Natural Disasters," *Technological Forecasting and Social Change* (95), pp. 3-15.

Alonso, O., Gertz, M., and Baeza-Yates, R. 2007. "On the Value of Temporal Information in Information Retrieval," *ACM SIGIR Forum*: ACM, pp. 35-41.

Altay, N., and Green, W. G. 2006. "Or/Ms Research in Disaster Operations Management," *European journal of operational research*, pp. 475-493.

Aragon, S. R. 2003. "Creating Social Presence in Online Environments," *New directions for adult and continuing education* (2003:100), pp. 57-68.

Argyle, M., and Dean, J. 1965. "Eye-Contact, Distance and Affiliation," *Sociometry* (28:3), pp. 289-304.

Avvenuti, M., Cimino, M. G., Cresci, S., Marchetti, A., and Tesconi, M. 2016. "A Framework for Detecting Unfolding Emergencies Using Humans as Sensors," *SpringerPlus* (5:1), p. 43.

Backfried, G., and Shalunts, G. 2016. "Sentiment Analysis of Media in German on the Refugee Crisis in Europe," *International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries*: Springer, pp. 234-241.

Beigi, G., Hu, X., Maciejewski, R., and Liu, H. 2016. "An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief," in *Sentiment Analysis and Ontology Engineering*. Springer, pp. 313-340.

Biocca, F., Harms, C., and Burgoon, J. K. 2003. "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence* (12:5), pp. 456-480.

Blei, D. M. 2012. "Probabilistic Topic Models," *Communications of the ACM* (55:4), pp. 77-84.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of machine Learning research* (3:Jan), pp. 993-1022.

Blumberg, B. C., Donald, R., and Schindler, P. S. 2005. "Business Research Methods 2 Nd Edition: Boston," *McGrawHill Higher Education*).

Bradley, M. M., and Lang, P. J. 1999. "Affective Norms for English Words (Anew): Instruction Manual and Affective Ratings," Technical report C-1, the center for research in psychophysiology, University of Florida.

Bunker, D., Mirbabaie, M., and Stieglitz, S. 2017. "Convergence Behaviour of Bystanders: An Analysis of 2016 Munich Shooting Twitter Crisis Communication," in: *Australasian Conference on Information Systems*. Hobart, Australia.

Burnap, P., Rana, O., Williams, M., Housley, W., Edwards, A., Morgan, J., Sloan, L., and Conejero, J. 2015. "Cosmos: Towards an Integrated and Scalable Service for

Analysing Social Media on Demand," *International Journal of Parallel, Emergent and Distributed Systems* (30:2), pp. 80-100.
CartoDB. 2015. from https://cartodb.com/
Chatfield, A. T., and Brajawidagda, U. 2013. "Twitter Early Tsunami Warning System: A Case Study in Indonesia's Natural Disaster Management," *System sciences (HICSS), 2013 46th Hawaii international conference on*: IEEE, pp. 2050-2060.
Chatfield, A. T., Scholl, H. J., and Brajawidagda, U. 2014. "# Sandy Tweets: Citizens' Co-Production of Time-Critical Information During an Unfolding Catastrophe," *2014 47th Hawaii International Conference on System Sciences*: IEEE, pp. 1947-1957.
Chong, T., Leong, C., Pan, S. L., Bahri, S., and Fauzi, A. 2014. "Use of Social Media in Disaster Relief During the Kuantan (Malaysia) Flood,").
Cobb, C., McCarthy, T., Perkins, A., Bharadwaj, A., Comis, J., Do, B., and Starbird, K. 2014. "Designing for the Deluge: Understanding & Supporting the Distributed, Collaborative Work of Crisis Volunteers," *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*: ACM, pp. 888-899.
Cobb, S. C. 2009. "Social Presence and Online Learning: A Current View from a Research Perspective," *Journal of Interactive Online Learning* (8:3), pp. 241-254.
Comfort, L. K., Ko, K., and Zagorecki, A. 2004. "Coordination in Rapidly Evolving Disaster Response Systems: The Role of Information," *American Behavioral Scientist* (48:3), pp. 295-313.
Cozzolino, A. 2012. *Humanitarian Logistics: Cross-Sector Cooperation in Disaster Relief Management*. Springer Science & Business Media.
Crowston, K. 2000. "Process as Theory in Information Systems Research," in *Organizational and Social Perspectives on Information Technology*. Springer, pp. 149-164.
Currion, P., Silva, C. d., and Van de Walle, B. 2007. "Open Source Software for Disaster Management," *Communications of the ACM* (50:3), pp. 61-65.
Day, J. M., Junglas, I., and Silva, L. 2009. "Information Flow Impediments in Disaster Relief Supply Chains," *Journal of the Association for Information Systems* (10:8), p. 637.
Debarati Guha-Sapir , P. H., Regina Below. 2016. "Statistical Review 2015 the Numbers and Trends," Centre for Research on the Epidemiology of Disasters (CRED), Brussels, Belgium.
Debortoli, S., Müller, O., Junglas, I., and vom Brocke, J. 2016. "Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial," *Communications of the Association for Information Systems (CAIS)*).
Deng, Q., Hine, M., Ji, S., and Sur, S. 2017. "Building an Environmental Sustainability Dictionary for the It Industry," *Proceedings of the 50th Hawaii International Conference on System Sciences*.
Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. 2011. "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter," *PloS one* (6:12), p. e26752.
Drabek, T. E., and McEntire, D. A. 2003. "Emergent Phenomena and the Sociology of Disaster: Lessons, Trends and Opportunities from the Research Literature," *Disaster Prevention and Management: An International Journal* (12:2), pp. 97-112.
Economist, T. 2015. "To the Rescue," *The Economist*).
Ehnis, C., and Bunker, D. 2012. "Social Media in Disaster Response: Queensland Police Service-Public Engagement During the 2011 Floods," *ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian Conference on Information Systems 2012*: ACIS, pp. 1-10.
Eismann, K., Posegga, O., and Fischbach, K. 2016. "Collective Behaviour, Social Media, and Disasters: A Systematic Literature Review," in: *European Conference on Information Systems*.
Elo, S., and Kyngäs, H. 2008. "The Qualitative Content Analysis Process," *Journal of advanced nursing* (62:1), pp. 107-115.

Endsley, M. R. 1988. "Design and Evaluation for Situation Awareness Enhancement," *Proceedings of the Human Factors Society annual meeting*: SAGE Publications Sage CA: Los Angeles, CA, pp. 97-101.

Endsley, M. R. 1995. "Toward a Theory of Situation Awareness in Dynamic Systems," *Human factors* (37:1), pp. 32-64.

Endsley, M. R., and Garland, D. 2000. "Theoretical Underpinnings of Situation Awareness: A Critical Review," *Situation awareness analysis and measurement* (1), p. 24.

Esuli, A., and Sebastiani, F. 2006. "Determining Term Subjectivity and Term Orientation for Opinion Mining," *EACL*, p. 2006.

Esuli, A., and Sebastiani, F. 2007. "Sentiwordnet: A High-Coverage Lexical Resource for Opinion Mining," *Evaluation*), pp. 1-26.

Ferro, L., Mani, I., Sundheim, B., and Wilson, G. 2001. "Tides Temporal Annotation Guidelines-Version 1.0. 2," *The MITRE Corporation, McLean-VG-USA*).

Firth, J. R. 1957. "{a Synopsis of Linguistic Theory, 1930-1955},").

Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S., and Stange, H. 2013. "Tracing the German Centennial Flood in the Stream of Tweets: First Lessons Learned," *Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information*: ACM, pp. 31-38.

Ga, #235, Dias, l. H., Hasanuzzaman, M., St, #233, Ferrari, p., and Mathet, Y. 2014. "Tempowordnet for Sentence Time Tagging," in: *Proceedings of the 23rd International Conference on World Wide Web*. Seoul, Korea: ACM, pp. 833-838.

Gao, H., Barbier, G., and Goolsby, R. 2011. "Harnessing the Crowdsourcing Power of Social Media for Disaster Relief," *IEEE Intelligent Systems* (26), pp. 10-14.

Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. 2013. "Comparing and Combining Sentiment Analysis Methods," *Proceedings of the first ACM conference on Online social networks*: ACM, pp. 27-38.

Graham, M., Hale, S. A., and Gaffney, D. 2014. "Where in the World Are You? Geolocation and Language Identification in Twitter," *The Professional Geographer* (66:4), pp. 568-578.

Grimmer, J., and Stewart, B. M. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political analysis* (21:3), pp. 267-297.

Gunawardena, C. N. 1995. "Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences," *International journal of educational telecommunications* (1:2), pp. 147-166.

Gunawardena, C. N., and Zittle, F. J. 1997. "Social Presence as a Predictor of Satisfaction within a Computer-Mediated Conferencing Environment," *American journal of distance education* (11:3), pp. 8-26.

Gupta, A., Lamba, H., Kumaraguru, P., and Joshi, A. 2013. "Faking Sandy: Characterizing and Identifying Fake Images on Twitter During Hurricane Sandy," *Proceedings of the 22nd international conference on World Wide Web*: ACM, pp. 729-736.

Halevy, A., Norvig, P., and Pereira, F. 2009. "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems* (24:2), pp. 8-12.

Hamad, E. O., Savundranayagam, M. Y., Holmes, J. D., Kinsella, E. A., and Johnson, A. M. 2016. "Toward a Mixed-Methods Research Approach to Content Analysis in the Digital Age: The Combined Content-Analysis Model and Its Applications to Health Care Twitter Feeds," *Journal of medical Internet research* (18:3).

Han, S., Min, J., and Lee, H. 2015. "Antecedents of Social Presence and Gratification of Social Connection Needs in Sns: A Study of Twitter Users and Their Mobile and Non-Mobile Usage," *International Journal of Information Management* (35:4), pp. 459-471.

Hannay, P., and Baatard, G. 2011. "Geointelligence: Data Mining Locational Social Media Content for Profiling and Information Gathering,").

Harris, Z. S. 1954. "Distributional Structure," *Word* (10:2-3), pp. 146-162.

Hasanuzzaman, M., Dias, G., Ferrari, S., and Mathet, Y. 2016. "Iterative Tempowordnet," in *The Semantic Web: Eswc 2016 Satellite Events, Heraklion, Crete, Greece, May 29 –*

June 2, 2016, Revised Selected Papers, H. Sack, G. Rizzo, N. Steinmetz, D. Mladenić, S. Auer and C. Lange (eds.). Cham: Springer International Publishing, pp. 40-45.
Heverin, T., and Zach, L. 2010. "Microblogging for Crisis Communication : Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma , Washington Area," in: *In: Proceedings of the 7th International ISCRAM Conference*. p. 5.
Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., Turner McGowen, S. E., Davis, R., Vaid, S., McElderry, J. A., and Griffith, S. A. 2015. "Social Media and Disasters: A Functional Framework for Social Media Use in Disaster Planning, Response, and Research," *Disasters* (39), pp. 1-22.
Hsieh, H.-F., and Shannon, S. E. 2005. "Three Approaches to Qualitative Content Analysis," *Qualitative health research* (15:9), pp. 1277-1288.
Huang, Q., and Cervone, G. 2016. "Usage of Social Media and Cloud Computing During Natural Hazards," *Cloud Computing in Ocean and Atmospheric Sciences*), pp. 297-324.
Hughes, A. L., St Denis, L. A., Palen, L., and Anderson, K. M. 2014. "Online Public Communications by Police & Fire Services During the 2012 Hurricane Sandy," *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*: ACM, pp. 1505-1514.
Imran, M., Castillo, C., Diaz, F., and Vieweg, S. 2015. "Processing Social Media Messages in Mass Emergency: A Survey," *ACM Computing Surveys (CSUR)* (47:4), p. 67.
Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S. 2014. "Aidr: Artificial Intelligence for Disaster Response," *Proceedings of the 23rd International Conference on World Wide Web*: ACM, pp. 159-162.
Java, A., Song, X., Finin, T., and Tseng, B. 2007. "Why We Twitter: Understanding Microblogging Usage and Communities," *In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*: ACM, pp. 56-65.
Jayaraman, V., Chandrasekhar, M., and Rao, U. 1997. "Managing the Natural Disasters from Space Technology Inputs," *Acta Astronautica* (40:2-8), pp. 291-325.
Kaewkitipong, L., Chen, C. C., and Ractham, P. 2016. "A Community-Based Approach to Sharing Knowledge before, During, and after Crisis Events: A Case Study from Thailand," *Computers in Human Behavior* (54), pp. 653-666.
Kaplan, A. M., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business horizons* (53:1), pp. 59-68.
Karimi, S., Yin, J., and Paris, C. 2013. "Classifying Microblogs for Disasters," *Proceedings of the 18th Australasian Document Computing Symposium*: ACM, pp. 26-33.
Kaufhold, M.-A., and Reuter, C. 2016. "The Self-Organization of Digital Volunteers across Social Media: The Case of the 2013 European Floods in Germany," *Journal of Homeland Security and Emergency Management* (13:1), pp. 137-166.
Kehrwald, B. 2008. "Understanding Social Presence in Text-Based Online Learning Environments," *Distance Education* (29:1), pp. 89-106.
Kondracki, N. L., Wellman, N. S., and Amundson, D. R. 2002. "Content Analysis: Review of Methods and Their Applications in Nutrition Education," *Journal of nutrition education and behavior* (34:4), pp. 224-230.
Kreps, G. A. 1984. "Sociological Inquiry and Disaster Research," *Annual review of sociology*), pp. 309-330.
Krippendorff, K. 1989. "Content Analysis," in *International Encyclopedia of Communication,* G.G. E. Barnouw, W. Schramm, T. L. Worth, & L. Gross (ed.). pp. 403-407.
Krstajic, M., Rohrdantz, C., Hund, M., and Weiler, A. 2012. "Getting There First: Real-Time Detection of Real-World Incidents on Twitter," *2nd Workshop on Interactive Visual Text Analytics : Task-Driven Analysis of Social Media Content with Visweek'12.*
Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. 2016. "Rapid Assessment of Disaster Damage Using Social Media Activity," *Science Advances* (2:3), p. e1500779.

Kumar, S., Barbier, G., Abbasi, M., and Liu, H. 2011. "Tweettracker: An Analysis Tool for Humanitarian and Disaster Relief," in: *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*. pp. 661-662.

Landwehr, P. M., and Carley, K. M. 2014. "Social Media in Disaster Relief," in *Data Mining and Knowledge Discovery for Big Data*. Springer, pp. 225-257.

Latonero, M., and Shklovski, I. 2011. "Emergency Management, Twitter, and Social Media Evangelism,").

Lee, J. B., Ybañez, M., De Leon, M. M., and Estuar, M. R. E. 2013. "Understanding the Behavior of Filipino Twitter Users During Disaster," *GSTF Journal on Computing (JoC)* (3:2), p. 37.

Lee, N.-S., Hirschmeier, S., Müller, S., and Luz, L. J. 2017. "Enablers in Crisis Information Management: A Literature Review,").

Leong, C. M. L., Pan, S. L., Ractham, P., and Kaewkitipong, L. 2015. "Ict-Enabled Community Empowerment in Crisis Response: Social Media in Thailand Flooding 2011," *Journal of the Association for Information Systems* (16:3), p. 174.

Lim, J. S., Hwang, Y., Kim, S., and Biocca, F. A. 2015. "How Social Media Engagement Leads to Sports Channel Loyalty: Mediating Roles of Social Presence and Channel Commitment," *Computers in Human Behavior* (46), pp. 158-167.

Lindsay, B. R. 2011. "Social Media and Disasters: Current Uses, Future Options, and Policy Considerations." Congressional Research Service Washington, DC.

Lingad, J., Karimi, S., and Yin, J. 2013. "Location Extraction from Disaster-Related Microblogs," in: *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*. pp. 1017-1020.

Liu, S. B., Palen, L., Sutton, J., Hughes, A., and Vieweg, S. 2008. "In Search of the Bigger Picture: The Emergent Role of on-Line Photo Sharing in Times of Disaster," *Proceedings of the information systems for crisis response and management conference (ISCRAM)*, pp. 969-980.

Lombard, M., Snyder-Duch, J., and Bracken, C. C. 2002. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability," *Human communication research* (28:4), pp. 587-604.

Lowenthal, P. R. 2009. "The Evolution Andinfluence of Social Presence Theory on Online Learning," *Social Computing: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*), p. 113.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., and Blanford, J. 2011. "Senseplace2: Geotwitter Analytics Support for Situational Awareness," *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*: IEEE, pp. 181-190.

Madhoushi, Z., Hamdan, A. R., and Zainudin, S. 2015. "Sentiment Analysis Techniques in Recent Works," *Science and Information Conference (SAI), 2015*: IEEE, pp. 288-291.

Markus, M. L., and Robey, D. 1988. "Information Technology and Organizational Change: Causal Structure in Theory and Research," *Management science* (34:5), pp. 583-598.

Meier, P. 2016a. "Micromappers Digital Disaster Response. With a Single Click!" Retrieved 2016-08-09, 2016, from http://www.micromappers.com/

Meier, P. 2016b. "The Value of Timely Information During Disasters (Measured In hours) ." from https://irevolutions.org/2016/04/06/the-value-of-timely-information-during-disasters-measured-in-hours/

Mendoza, M., Poblete, B., and Castillo, C. 2010. "Twitter under Crisis: Can We Trust What We Rt?," *Proceedings of the first workshop on social media analytics*: ACM, pp. 71-79.

Miller, G. A. 1995. "Wordnet: A Lexical Database for English," *Communications of the ACM* (38:11), pp. 39-41.

Mirbabaie, M., Ehnis, C., Stieglitz, S., and Bunker, D. 2014. "Communication Roles in Public Events," *Working Conference on Information Systems and Organizations*: Springer, pp. 207-218.

Mirbabaie, M., and Zapatka, E. 2017. "Sensemaking in Social Media Crisis Communication–a Case Study on the Brussels Bombings in 2016,").

Misra, T. 2015. "Chennai Floods: A Perfect Storm of Climate Change and Failed Planning." 2016, from http://www.citylab.com/weather/2015/12/chennai-flood-climate-change-urban-planning-cop21-paris/419349/

Morris, R. 1994. "Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations," *Journal of Management* (20:4), pp. 903-931.

Muralidharan, S., Rasmussen, L., Patterson, D., and Shin, J.-H. 2011. ""Hope for Haiti: An Analysis of Facebook and Twitter Usage During the Earthquake Relief Efforts"," *Public Relations Review* (37:2), pp. 175-177.

Nan, N., and Lu, Y. 2014. "Harnessing the Power of Self-Organization in an Online Community During Organizational Crisis," *Mis Quarterly* (38:4), pp. 1135-1157.

Narayanan, V., Arora, I., and Bhatia, A. 2013. "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model," in *Intelligent Data Engineering and Automated Learning–Ideal 2013*. Springer, pp. 194-201.

Nielsen, F. Å. 2011. "A New Anew: Evaluation of a Word List for Sentiment Analysis in Microblogs," *arXiv preprint arXiv:1103.2903*).

Oh, O., Agrawal, M., and Rao, H. R. 2011. "Information Control and Terrorism: Tracking the Mumbai Terrorist Attack through Twitter," *Information Systems Frontiers* (13:1), pp. 33-43.

Oh, O., Agrawal, M., and Rao, H. R. 2013. "Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises," *Mis Quarterly* (37:2).

Oh, O., Kwon, K. H., and Rao, H. R. 2010. "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter During the Haiti Earthquake 2010," *ICIS*, p. 231.

Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S. 2014. "Crisislex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises," *ICWSM*.

Olteanu, A., Vieweg, S., and Castillo, C. 2015. "What to Expect When the Unexpected Happens: Social Media Communications across Crises," *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*: ACM, pp. 994-1009.

Olteanu., A. 2017. "Crisislex.Org." from http://crisislex.org/

Palen, L., and Liu, S. B. 2007. ""Citizen Communications in Crisis: Anticipating a Future of Ict-Supported Public Participation"," *In: Proceedings of the SIGCHI conference on Human factors in computing systems*: ACM, pp. 727-736.

Panagiotopoulos, P., Ziaee Bigdeli, A., and Sams, S. 2012. "" 5 Days in August"–How London Local Authorities Used Twitter During the 2011 Riots," *Electronic Government*), pp. 102-113.

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and trends in information retrieval* (2:1-2), pp. 1-135.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2001. "Linguistic Inquiry and Word Count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates* (71:2001), p. 2001.

Powers, D. M. 2011. "Evaluation: From Precision, Recall and F-Measure to Roc, Informedness, Markedness and Correlation," 2229-3981, School of Informatics and Engineering, Flinders University , Adelaide, Australia.

Pustejovsky, J., Castano, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Katz, G., and Radev, D. R. 2003a. "Timeml: Robust Specification of Event and Temporal Expressions in Text," *New directions in question answering* (3), pp. 28-34.

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., and Ferro, L. 2003b. "The Timebank Corpus," *Corpus linguistics*, p. 40.

Qu, Y., Huang, C., Zhang, P., and Zhang, J. 2011. "Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake," *In: Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 25-34.
Qu, Y., Wu, P. F., and Wang, X. 2009. "Online Community Response to Major Disaster: A Study of Tianya Forum in the 2008 Sichuan Earthquake," in: *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences, HICSS*. pp. 1-11.
Raymond, N., Al Achkar, Z., Verhulst, S., and Berens, J. 2016. "Building Data Responsibility into Humanitarian Action," United Nations Office for the Coordination of Humanitarian Affairs, pp. 1-16.
Read, J. 2005. "Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification," *Proceedings of the ACL student research workshop*: Association for Computational Linguistics, pp. 43-48.
Read, J., and Carroll, J. 2009. "Weakly Supervised Techniques for Domain-Independent Sentiment Classification," *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*: ACM, pp. 45-52.
Reilly, W., Guarino, S. L., and Kellihan, B. 2007. "Model-Based Measurement of Situation Awareness," *Simulation Conference, 2007 Winter*: IEEE, pp. 1353-1360.
Reuter, C., Heger, O., and Pipek, V. 2013. "Combining Real and Virtual Volunteers through Social Media," *ISCRAM*.
Reuter, C., Ludwig, T., Kaufhold, M.-A., and Spielhofer, T. 2016. "Emergency Services′ Attitudes Towards Social Media: A Quantitative and Qualitative Survey across Europe," *International Journal of Human-Computer Studies* (95), pp. 96-111.
Rogstadius, J., Vukovic, M., Teixeira, C., Kostakos, V., Karapanos, E., and Laredo, J. A. 2013. "Crisistracker: Crowdsourced Social Media Curation for Disaster Awareness," *IBM Journal of Research and Development* (57:5), pp. 4: 1-4: 13.
Rourke, L., Anderson, T., Garrison, D. R., and Archer, W. 2001. "Methodological Issues in the Content Analysis of Computer Conference Transcripts," *International Journal of Artificial Intelligence in Education (IJAIED)* (12), pp. 8-22.
Sakaki, T., Okazaki, M., and Matsuo, Y. 2010. "Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors," in: *WWW '10: Proceedings of the 19th international conference on World wide web*. pp. 851-860.
Salmon, P., Stanton, N., Walker, G., and Green, D. 2006. "Situation Awareness Measurement: A Review of Applicability for C4i Environments," *Applied ergonomics* (37:2), pp. 225-238.
Schilder, F., and Habel, C. 2001. "From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages," *Proceedings of the workshop on Temporal and spatial information processing-Volume 13*: Association for Computational Linguistics, p. 9.
Schwartz, H. A., Park, G., Sap, M., Weingarten, E., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., Berger, J., and Seligman, M. 2015. "Extracting Human Temporal Orientation from Facebook Language," *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 409-419.
Seebach, C., Beck, R., and Pahlke, I. 2011. "Situation Awareness through Social Collaboration Platforms in Distributed Work Environments," *In: Proceedings of the International Conference on Information Systems*.
Shelton, T., Poorthuis, A., Graham, M., and Zook, M. 2014. "Mapping the Data Shadows of Hurricane Sandy: Uncovering the Sociospatial Dimensions of 'Big Data'," *Geoforum* (52), pp. 167-179.
Sheombar, A., Urquhart, C., Ndhlovu, T., and Ravesteijn, P. 2015. "Social Media in the Context of Development: A Case Study of Dutch Ngos," *ECIS*.
Shklovski, I., Palen, L., and Sutton, J. 2008. "Finding Community through Information and Communication Technology During Disaster Events," in: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, pp. 1-10.
Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*.

Simon, T., Aharonson-Daniel, L., El-Hadid, M., and Adini, B. 2015. "Cross-Border Emergency Coordination and Communications Using Social Media: Developing a Joint Israeli-Jordanian Standard Operating Procedure for Leveraging Social Media in Emergencies," *International Journal of Emergency Management* (11:2), pp. 169-190.

Sinnappan, S., Farrell, C., and Stewart, E. 2010. "Priceless Tweets! A Study on Twitter Messages Posted During Crisis: Black Saturday," *In: Proceedings of the 21st Australasian Conference on Information Systems*.

Sivarajah, U., Kamal, M. M., Irani, Z., and Weerakkody, V. 2017. "Critical Analysis of Big Data Challenges and Analytical Methods," *Journal of Business Research* (70), pp. 263-286.

Starbird, K., Dailey, D., Walker, A. H., Leschine, T. M., Pavia, R., and Bostrom, A. 2015. "Social Media, Public Participation, and the 2010 Bp Deepwater Horizon Oil Spill," *Human and Ecological Risk Assessment: An International Journal* (21:3), pp. 605-630.

Starbird, K., Maddock, J., Orand, M., Achterman, P., and Mason, R. M. 2014. "Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing," *iConference 2014 Proceedings*).

Starbird, K., and Palen, L. 2011. "Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM, pp. 1071-1080.

Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. 2010. "Chatter on the Red: What Hazards Thret Reveals About the Social Life of Microblogges Information," *In: Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*, pp. 241-250.

Stemler, S. 2001. "An Overview of Content Analysis," *Practical assessment, research & evaluation* (7:17), pp. 137-146.

Stiegler, R., Tilley, S., and Parveen, T. 2011. "Finding Family and Friends in the Aftermath of a Disaster Using Federated Queries on Social Networks and Websites," *Web systems evolution (WSE), 2011 13th IEEE international symposium on*: IEEE, pp. 21-26.

Stieglitz, S., Mirbabaie, M., Schwenner, L., Marx, J., Lehr, J., and Brünker, F. 2017. "Sensemaking and Communication Roles in Social Media Crisis Communication,").

Strötgen, J., and Gertz, M. 2013. "Multilingual and Cross-Domain Temporal Tagging," *Language Resources and Evaluation* (47:2), pp. 269-298.

Subba, R., and Bui, T. 2010. "An Exploration of Physical-Virtual Convergence Behaviors in Crisis Situations," *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*: IEEE, pp. 1-10.

Subba, R., and Bui, T. 2017. "Online Convergence Behavior, Social Media Communications and Crisis Response: An Empirical Study of the 2015 Nepal Earthquake Police Twitter Project," *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Sutton, J., Palen, L., and Shklovski, I. 2008. "Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires," *In: Proceedings of the 5th International ISCRAM Conference*: Washington, DC, pp. 624-632.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. 2011. "Lexicon-Based Methods for Sentiment Analysis," *Computational linguistics* (37:2), pp. 267-307.

Takahashi, B., Tandoc, E. C., and Carmichael, C. 2015. "Communicating on Twitter During a Disaster: An Analysis of Tweets During Typhoon Haiyan in the Philippines," *Computers in Human Behavior* (50), pp. 392-398.

Temnikova, I., Castillo, C., and Vieweg, S. 2015. "Emterms 1.0: A Terminological Resource for Crisis Tweets," *ISCRAM 2015 proceedings of the 12th international conference on information systems for crisis response and management*.

Thelwall, M., Buckley, K., and Paltoglou, G. 2011. "Sentiment in Twitter Events," *Journal of the American Society for Information Science and Technology* (62:2), pp. 406-418.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. 2010. "Sentiment Strength Detection in Short Informal Text," *Journal of the American Society for Information Science and Technology* (61:12), pp. 2544-2558.

Thelwall, M., Wouters, P., and Fry, J. 2008. "Information-Centered Research for Large-Scale Analyses of New Information Sources," *Journal of the American Society for Information Science and Technology* (59:9), pp. 1523-1527.

Tierney, K. J. 1993. "Disaster Preparedness and Response: Research Findings and Guidance from the Social Science Literature,").

Tim, Y., Pan, S. L., Ractham, P., and Kaewkitipong, L. 2017. "Digitally Enabled Disaster Response: The Emergence of Social Media as Boundary Objects in a Flooding Disaster," *Information Systems Journal* (27:2), pp. 197-232.

Torrey, C., Burke, M., Lee, M., Dey, A., Fussell, S., and Kiesler, S. 2007. "Connected Giving: Ordinary People Coordinating Disaster Relief on the Internet," *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*: IEEE, pp. 179a-179a.

Tu, C.-H. 2002. "The Impacts of Text-Based Cmc on Online Social Presence," *The journal of interactive online learning* (1:2), pp. 1-24.

Tu, C.-H., and McIsaac, M. 2002. "The Relationship of Social Presence and Interaction in Online Classes," *The American journal of distance education* (16:3), pp. 131-150.

van Gorp, A. F., Pogrebnyakov, N., and Maldonado, E. A. 2015. "Just Keep Tweeting: Emergency Responder's Social Media Use before and During Emergencies," *ECIS*.

Velichety, S., and Ram, S. 2013. "A Cross-Sectional and Temporal Analysis of Information Consumption on Twitter," *In: Proceedings of the International Conference on Information Systems*.

Venkatesh, V., Brown, S. A., and Bala, H. 2013. "Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems," *MIS quarterly* (37:1).

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., Schram, A., and Anderson, K. M. 2011. "Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency," *ICWSM*: Citeseer.

Vidulich, M., Dominguez, C., Vogel, E., and McMillan, G. 1994. "Situation Awareness: Papers and Annotated Bibliography," DTIC Document.

Vieweg, S., Castillo, C., and Imran, M. 2014. "Integrating Social Media Communications into the Rapid Assessment of Sudden Onset Disasters," *International Conference on Social Informatics*: Springer, pp. 444-461.

Vieweg, S., Hughes, A., Starbird, K., and Palen, L. 2010a. "Supporting Situational Awareness During Emergencies Using Microblogged Information," *SIGCHI 2010*).

Vieweg, S., Hughes, A. L. A., Starbird, K., and Palen, L. 2010b. "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*: ACM, pp. 1079-1088.

Vieweg, S., Palen, L., Liu, S. B., Hughes, A. L., and Sutton, J. 2008. "Collective Intelligence in Disaster: An Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shootings," *Proceedings of the information systems for crisis response and management conference (ISCRAM)*.

Vieweg, S. E. 2012. "Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications." University of Colorado at Boulder.

Vivacqua, A. S., and Borges, M. R. 2010. "Collective Intelligence for the Design of Emergency Response," *Computer Supported Cooperative Work in Design (CSCWD), 2010 14th International Conference on*: IEEE, pp. 623-628.

Walther, J. B., Loh, T., and Granka, L. 2005. "Let Me Count the Ways the Interchange of Verbal and Nonverbal Cues in Computer-Mediated and Face-to-Face Affinity," *Journal of language and social psychology* (24:1), pp. 36-65.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *Management Information Systems Quarterly* (26:2), p. 3.

Wen, M., Zheng, Z., Jang, H., Xiang, G., and Rosé, C. P. 2013. "Extracting Events with Informal Temporal References in Personal Histories in Online Communities," *ACL (2)*, pp. 836-842.

Wiener, M., and Mehrabian, A. 1968. *Language within Language: Immediacy, a Channel in Verbal Communication*. Ardent Media.

Wilson, T., Wiebe, J., and Hoffmann, P. 2005. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proceedings of the conference on human language technology and empirical methods in natural language processing*: Association for Computational Linguistics, pp. 347-354.

Woods, R. H., and Baker, J. D. 2004. "Interaction and Immediacy in Online Learning," *The International Review of Research in Open and Distributed Learning* (5:2).

Wukich, C. 2015. "Social Media Use in Emergency Management," *Journal of Emergency Management* (13:4), pp. 281-295.

Xu, C., Ryan, S., Prybutok, V., and Wen, C. 2012. "It Is Not for Fun: An Examination of Social Network Site Usage," *Information & Management* (49:5), pp. 210-217.

Yang, C. C., Tang, X., Wong, Y., and Wei, C.-P. 2010. "Understanding Online Consumer Review Opinions with Sentiment Analysis Using Machine Learning," *Pacific Asia Journal of the Association for Information Systems* (2:3).

Yang, Y. 1999. "An Evaluation of Statistical Approaches to Text Categorization," *Information retrieval* (1:1-2), pp. 69-90.

Yang, Y., and Liu, X. 1999. "A Re-Examination of Text Categorization Methods," *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*: ACM, pp. 42-49.

Yin, J., Karimi, S., Robinson, B., and Cameron, M. 2012. "Esa: Emergency Situation Awareness Via Microbloggers," in: *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12*. ACM, pp. 2701-2703.

Zagorecki, A. T., Johnson, D. E., and Ristvej, J. 2013. "Data Mining and Machine Learning in the Context of Disaster and Crisis Management," *International Journal of Emergency Management* (9:4), pp. 351-365.

Zhang, H., and Li, D. 2007. "Naive Bayes Text Classifier," *Granular Computing, 2007. GRC 2007. IEEE International Conference on*: IEEE, pp. 708-708.

Zubiaga, A. 2015. "Dataset: Hurricane Sandy Tweets." 2015, from http://www.zubiaga.org/datasets/hurricane-sandy-tweets/

Paper 1

Mukkamala, AM & Beck, R (2016)

Disaster management and social media use for decision making by humanitarian organizations

In: Proceedings of the 49th Hawaii International Conference on System Sciences **(HICSS 2016)**; Kauai, Hawaii, USA

Disaster management and social media use for decision making by humanitarian organizations

Alivelu Mukkamala
IT University of Copenhagen
Denmark
ammu@itu.dk

Roman Beck
IT University of Copenhagen
Denmark
romb@itu.dk

Abstract

In times of a crisis, disasters or extreme events, affected people use social media solutions to share information about the situation. Hence, the use of this information for supporting humanitarian operations is becoming a valuable source to develop a real time understanding about the situation even before they arrive at the ground. From a scholarly perspective, the use of social media management during and after a disaster has hit has been researched, but no special focus has been given to the needs of humanitarian organizations and how their management phases can be supported by social media analytics in real time. In this research, we will identify the few papers that have been concerned with this topic and will apply a theoretical lens from disaster management to investigate viable areas where social media research can support humanitarian organization in the different phases of disaster response.

1. Introduction

With the emergence of social media as widely used information and communication technologies (ICT), it is possible to have now for the first time real - time information about the situation at the ground of a crisis or disaster zone from the affected people themselves. Rather than waiting for non-governmental organizations (NGO) and governmental organizations to bring in technology to do the surveying and intelligence of the situation for managing the disaster relief efforts, the people affected quite often have still access to social media and provide information electronically. For example, a majority of social media users in Oklahoma turned their social media access into a tool to provide and find information about the damage but also about their friends and families wellbeing after a tornado struck the area [18].

Before that, communication of disaster information was unidirectional: disaster information was communicated from the governments and relief organizations to the public. However, the general public now uses social media accounts in a disaster in order to gain and contribute information [16]. Moreover, people are reorienting [25] their personal social media to propagate information and support the relief efforts in disaster situations. In the aftermath of hurricane Katrina, the public stood by their community by forming online forums to help and support affected victims [26]. Essentially, they switched to their personal communication channels when traditional communication channels failed, facing difficulties to operate during a catastrophe [20].

Humanitarian relief organizations are now able to receive real time and detailed information that once was difficult or impossible to access. Such user-generated content can be key in disaster situations. It contains valuable information regarding geographic information, which would otherwise not be accessible, such as the situation of the event, victims' needs, urgency of supplies and so on. Hence, research on social media during disaster situations is gaining prominence. Overall, one can observe a paradigm shift in the use of social media during disasters, away from the idea of bringing in intelligence technologies by humanitarian agency to help public towards an increasing focus on existing, already available technology such as social media possessed by the public.

To our knowledge, the transition from focusing on dedicated technologies for disaster relief towards the existing social media has not been specifically investigated yet. Moreover, it is unclear how the user-generated content can help humanitarian organizations in real-time for decision management and organizing their disaster management and operations. Thus, this study addresses this research gap by conducting a systematic literature analysis to answer essentially two research questions, namely, *how far does existing research reflect the real-time use of social media based information in disasters and how can the existing research findings be used to help humanitarian organizations to respond?*

The remainder of the paper is organized as follows: In the methodology section, we explain the structured literature review and the method we followed to analyze and categorize the articles. In the analysis section, we present the mapping of disaster social media uses [8] to disaster management phases [3], the analysis and our findings of articles and mapping them to disaster management phases. Finally in the last section we discuss the findings and conclusion.

2. Methodology

In order to find answers to our research questions, as a starting point, we conducted a systematic literature review [28] to gather the bibliography to analyze the articles. Moreover, this approach also sheds light on which different directions the research on the use of social media in disasters is heading to. The combined keywords used for the literature search were social media and disaster, social media and crisis, Twitter and crisis, Twitter and disaster, Facebook and disasters, Facebook and crisis, ICTs and disasters, and ICTs and crisis. Furthermore, it should also be noted that we not only selected the articles focusing on social media and disasters or crisis but also the articles that mentioned ICTs, for example web forums or blogs.

The literature review has been done in two stages. Initially, by using the combined keywords we searched for the scholarly literature on that topic published in the senior scholar basket of eight journals and also in information systems (IS) conference proceedings such as ICIS, ECIS, AMCIS, HICSS, and PACIS. Because of the nascent nature of the research area, we used the same keywords and conducted a search for relevant articles in Google scholar and also performed backward and forward search on the collected articles. The final sample of articles comprised 45 articles, including articles from various conference proceedings such as ISCRAM, CHI, WWW.

We analyzed all the 45 articles and started to cluster them to understand how many research articles are focusing on the real-time social media information analysis of users affected by a disaster. In our research, we consider content from social media users who are not only direct victims of a disaster but also of witnesses present at the time a disaster took place, sharing their observations of a disasters as it happens or shortly afterwards. Overall, the 45 articles are clustered into four different categories based on their research focus. In the first category, the focus is on real-time data coming from end-users through social media, and activities during a disaster. In the second category, even though the research was conducted on disaster data of end-users, the data collection and analysis were performed subsequently after a disaster took place. Especially in this category, the articles focused on rumors, user behavior, re-tweets, and uses of social media at the time of disaster and so on. In the third category, the research focus of the articles is about crowd-sourced/open source software applications, tools and systems that either evolved during the times of disasters or developed afterwards to analyze the user-generated content of social media witness accounts of the disaster, for example, to generate insights in the catastrophe that might be useful in tackling future disasters. A few articles belonging to the last category are dealing with the use of social media by the humanitarian organizations themselves, for example for fundraising or other communication purposes. In order to answer our research questions, we focused on the first and third paper categories, resembling together solely 15 articles.

For the analysis of the papers, we further sub-categorized the papers as being part of one of four disaster management phases, and started coding the articles based disaster social media uses drawn from the functions of disaster social media [8] as further discussed in the analysis section.

3. Analysis

A disaster is a sudden event that seriously affects the normal routine conditions of a community or society. It has not only an economic and environmental impact but also an important humanitarian component [9]. Disasters could be natural calamities such as earthquakes, tornadoes or hurricanes but also man-made destructive activities such as terrorist attacks or industrial accidents. These kinds of disasters occur suddenly, that demands immediate and fast relief activities in devastated areas. Other kinds of man-made or natural crises such as epidemic diseases or economic crises also have an impact on society but occur not as disruptive over time which is the reason while the role of real-time intelligence is not as important. Clearly, each type of disaster has to be managed differently [3], while for all disasters being able to reach victims as fast as possible to provide first aid and supplies are of paramount importance for any humanitarian organization. Thus, in order to perform the humanitarian operations efficiently and effectively, commanding over good intelligence for planning and organizing the disaster management activities is of vital importance.

Since disaster management activities are rather complex, comprising several sub-tasks and specialist decision making and skills in the different disaster management phases, humanitarian operations research is studied by many disciplines applying different perspectives. For example, research on the recovery phase is conducted from a process management perspective [14], while natural disaster management has been researched from a design science perspective [22], as well as from a humanitarian logistics and supply chain management point of view [3].

While some literature on the humanitarian logistics and disaster management phases [3] mentions that there are some disagreements about the structure and nomenclature of the disaster management phases, there seems to be a general agreement that the disaster management can be modeled into four phases, namely mitigation, preparedness, response, and reconstruction. Having a good strategy for each of the phases is essential for an efficient disaster management. In order to accomplish this, managers need proper information about the different activities within each of the four disaster management phases [10]. With social media information is now accessible in real-time, so those activities can be planned more accurately. Thus, we adopted the disaster management model and its phases in our research to map where and how social media information can be used to improve the decision making [3].

While social media platforms have not been designed to be used in disaster management, the people affected by a disaster nevertheless turn to those medias to communicate, thereby also changing the role of the general public from receiving information about a catastrophe to producing and sending important information [2, 16]. From a disaster communication point of view, Houston et al. [8] developed a comprehensive framework containing 15 disaster social media uses and their applicability from a communication perspective. They reviewed online information and official websites in addition to scientific literature to explain the disaster social media uses and users. Given the two-way communication nature of social media, the authors considered the users as producers of information and mainly categorized them as: individuals, organizations, governments, communities and news media. Since the research angle they applied primarily considers social media from a communication point of view, the social media uses are more generic. Even though we considered the 15 social media uses in our study, we only considered and integrated those uses which are relevant for the real-time use within the disaster management phases.

In order to set a theoretical basis for this research, we adopted two guiding frameworks, dealing with the functions of disaster social media use [8] and disaster management phases [3]. As mentioned before, we categorized the research of the collected articles into four disaster management phases, as shown in Figure 1. Instead of categorizing the existing research into disaster management phases directly, we added the disaster social media uses because they represent a more fine-grained perspective of social media within a disaster. Moreover, in applying the disaster social media uses to traditional disaster management phases allows us to integrate literature about potential social media activities into the disaster management model. This will act as a theoretical lens to categorize the existing research into disaster management phases.

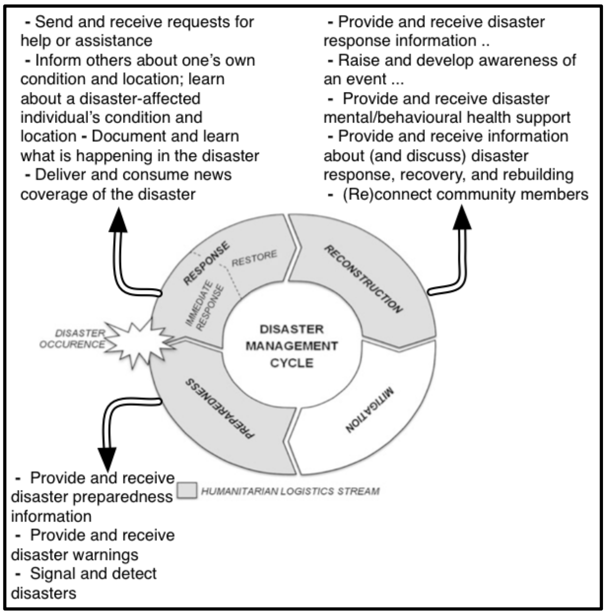

Figure 3 Use of social media in different disaster management phases (adapted from [8, 3])

Even though Houston et al. [8] proposed their framework from a communication point of view, it is useful to shed some light on the disaster situations where social media is facilitating communication during disasters from different actors or stakeholders. Moreover, most of the disaster social media uses where apparent in more than one disaster management phase in Houston et al. [8].

In order to categorize the discovered literature, firstly we mapped the disaster social media uses to disaster management phases as shown in Figure 1. We coded the literature along the two dimension (disaster management phase and disaster social media use) and whenever we believed an article could be categorized in one or more of the categories we did so. The resulting figure allowed us to understand how the social media in disaster classification actually resonates with the more commonly used disaster management phases. Since the disaster social media uses explain the fine grained activities performed during disaster, we coded 15 articles to these activities and thereby categorized them into different phases of disaster management. The result of the analysis is discussed below.

3.1 Categorization of articles into disaster management phases

In the following, the findings will be discussed now in more detail. Most of the disaster social media uses we found in literature have an overlap across different disaster management phases. Hence some of the articles are related and mapped to more than one phase.

Mitigation: The mitigation phase focuses on measures that either try to prevent the disaster or try to reduce the impact of the disaster [1]. We were not able to identify any articles that fit into that category, as we have confined ourselves to either research on real-time data during disaster or the applications that will be useful in future disasters. Thus, it could be due to our focus that we have not detected research on social media to support disaster management in the mitigation phase. An alternative explanation could be that there are only limited application areas imaginable how social media can be off use before a sudden, unexpected disaster takes place. In case of flooding or wildfire it is nevertheless thinkable that social media can be used to mitigate the most severe effects of such catastrophes if there is enough time to prepare the general public. However, here we have a clear overlap to the preparedness phase.

Preparedness: In the preparedness phase, the aim is to prepare people to respond to a disaster. The articles that are relevant to this phase falls into two categories in our framework as described below.

Twitter-based monitoring applications: The first category under preparedness mainly focuses on the recently developed applications and tools [11, 21] that monitor, track and analyze the tweets for event detection and location extraction. In other words, it gives situational awareness. One of those applications [11] provides information to first respondents and the other one acts as an earthquake responding system [21].

Twitter-based data analysis: In this category, research focus is mainly on micro blogging or Twitter data either to find out the process of information production and distribution by general public or to identify the information shared by local residents to enhance situational awareness. Twitter users and communities shared local disaster information such as flood levels, wind direction and fire paths, which lead to the preparatory activities [24, 27]. These situational updates are useful in humanitarian relief operations as well as to the affected community members to prepare themselves to respond actively for a disaster.

Response: The response phase is the most important stage where the concern is to preserve the community, environment and saving lives by deploying proper resources [1]. One can notice that most of the papers are categorized into the response phase. According to our framework, the articles falls into four categories.

Role of ICTs for disaster response: Communities and individuals seeking and providing disaster/crisis information for emergency response. The main key practices observed in the communication are sharing of real-time information, extending moral support to communities, proposing relief activities. Citizenry are also asking for help, suggesting to the officials what kind of actions need to be taken, and giving moral support to the community members through ICTs such as web forums, blogs, Facebook. The ICTs during disaster events [23], despite being back channels acted as main channels to fulfill the timely needs of the people [25]. Moreover whenever a crisis occurs in a networked world [17] online communities are responding positively to disasters [20].

Role of microblogging for disaster response: Twitter has been widely used for risk and crisis communication in crisis/disaster response events. End users by self-organizing themselves disseminate crisis/disaster related information as a way to reassure other victims and to help in relief activities. Some of the articles are dealing with the social life of microblogged information where authors discussed the information production on Twitter at the time of floods. For example, Twitter helped the end users to self-organize themselves by producing as well as distributing the flood relevant information [24]. Moreover, a microblogging site acts as a system to share different types of messages for different purposes such as situational updates, asking for help, expressing opinion and emotional support [19] and also for the crisis communication [7]. Others discuss the role of microblogging in response situations such as in the case of re-tweeted messages after the Fukushima nuclear radiation disaster [12]. One important finding of the Fukushima nuclear radiation disaster [12] is that, even though government organizations tried to create situational awareness and calmed the end-users via microblogging, the user engagement in re-tweeting the information shared by government has been very low due to lack of trust and increased fear among public. However, the information that is shared by the end users is useful to the humanitarian organizations to act in the response phase of any crisis or disaster.

Twitter-base applications for response: Monitoring applications based on tweets were developed to visualize the disaster-affected area and to provide geo location information. This situational awareness is also helpful in the disaster response phase. This helps humanitarian relief activities to act and reach the victims. The reviewed articles are concerned about location extraction from disaster-related microblogs [13], and emergency situation awareness [29].

Crowd-sourced applications for disaster response*:* The crowd-sourced open software applications collect data from different disparate sources and provide the visual information of affected areas as well as needs and urgencies of the victims. This information is especially helpful in humanitarian relief activities to coordinate and allocate the resources effectively [4, 6].

Reconstruction: In the aftermath of a disaster, the reconstruction phase involves both long-term and short-term activities to stabilize and bring the community to normal conditions [1]. In the reconstruction phase, the research is primarily focused on ICTs such as web forums and Twitter.

Role of ICTs and microblogging in the reconstruction phase*:* Mainly individuals used ICTs to inform others about their safety and also enquiring about others safety. People through web forums and Twitter trying to re-connect to their community [19, 20, 23, 25]. However, crowd-sourced applications are also useful in reconstruction phase to find their missing family members.

Given to the nature of disaster social media uses that are useful to disaster management phases, some of the articles are also categorized into more than one phase. However, primary motivation for clustering the discovered articles from preparedness phase to reconstruction phase, (for example role of ICTs in response, role of microblogging for response, so on) is to observe a paradigm shift in the gradual use of different ICTs during disasters, away from the idea of bringing technologies by NGOs or governmental organizations to help towards an increasing focus on social media possessed by the general public. Along with the technological developments general public also switching to their own accessible latest social media, for example, web forums to Twitter during disasters.

4. Discussion and Conclusion

In this research, our focus was to explore reflections of existing research on social media in disasters. The motive of our study is to affirm the transition where end-users started using social media for humanitarian disaster relief purposes. This transition is changing the way of doing things both at emergency management level as well as at end-users level/societal level. Moreover, applying the research findings of the collected articles to disaster management is an attempt to shed more light on the true potential of social media in the emergency management. Applicability of real-time information to disaster management is important because "social media is already making significant changes in emergency and crisis management" [5]. However, even though non-profit organizations used social media at the times of a disaster as a public relationship tool, it failed to exploit the two-way communication and true potential of the social media [15]. The reason could be either lack of proper analytical intelligence methodologies to extract necessary valuable information from the social media or lack of a structured development and use of social media with focus on the disaster management phases. However, our findings revealed that social media has the capability to assist as decisions support system through which NGOs and emergency organizations perform their decision making activities efficiently. Not only is this, for example, the fact in a wild fire case, if the information comes from people witnessing it, it also will help in evacuation instructions. However, based on our analysis, we state that there is a need for more research in the areas where social media is used but not exploited successfully in disaster management. Moreover there is also a need to encourage IS research in general to engage more with humanitarian organizations to help them to develop more aligned solutions.

Out of the 45 collected articles dealing with social media us in relation to disaster management, only 15 articles were found to be deal with the real-time use of social media to support humanitarian organizations for managing their disaster response activities. In the framework of functions disaster social media [8], the authors categorized disaster social media uses to disaster phases: pre event, event and post event. Most of the social media uses are categorized into event and post-event of disaster phases, but only few uses fall into the pre-event category. Similar to the work of Houston et al. [8], in our literature analysis, most of the articles are categorized into response and reconstruction phases primarily. None of the research articles are categorized into the mitigation phase, whereas only few articles are mapped to preparedness phase. The reason could be that the research in disaster social media is still in its infancy. Moreover, the end users are the information contributors in the disaster social media and their activity is mainly confined to the disaster response phase and partially to the preparedness and recovery phases.

Based on the systematic literature review we conclude that because social media is a relatively new technology and has not been designed in the first place to support humanitarian aid in disasters, research in that area is still scarce. Despite the high research interest in that area, very few articles are relevant or useful to emergency management. Moreover there is neither a holistic methodology/approach nor sufficient theoretical foundations on using social media to disaster management across the different phases. To the best of our knowledge, we did not come across any research work that tries to apply social media to all the phases of disaster management.

It can be concluded that currently, NGOs or other relief organizations are not able to exploit social media to its full potential since the areas social media can be used are rather fragmented and they are not aligned with the disaster management phases typically in use. Hence there is a need for further development of systematic methodologies for harnessing the social data. For example, there is a need for automation certain social media coding approaches so that they can provide real-time information rather than providing only analytical insights in the aftermath of a disaster. Furthermore, promising crowed-based approaches need to be analyzed regarding their real-time applicability: If a disaster is not able to garner enough attention then the crowd coding, for example, pictures taken from devastated areas might not be coded in time due to a lack of crowd-interest.

This study has certain limitations as we only tried to look for articles that analyzed the real-time data during disasters and application or tools. Hence we have limited ourselves to few articles. Another limitation of the study is that we considered all the disasters as in one category. Each disaster is different in its own right depending on their characteristics such as manmade verses natural, long response duration verses short and immediate response and so on.

5. References

[1] N. Altay and W. G. Green, "OR/MS research in disaster operations management", European journal of operational research, 175 (2006), pp. 475-493.

[2] N. Cook, Enterprise 2.0: how social software will change the future of work, Gower Publishing, Ltd., 2008.

[3] A. Cozzolino, Humanitarian logistics: cross-sector cooperation in disaster relief management, Springer Science & Business Media, 2012.

[4] P. Currion, C. d. Silva and B. Van de Walle, "Open source software for disaster management", Communications of the ACM, 50 (2007), pp. 61-65.

[5] G. F. For-mukwai, "The Transformative Power of Social Media on Emergency and Crisis Management", International journal of information systems for crisis response and management, 2 (2010), pp. 1.

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

[7] T. Heverin and L. Zach, "Microblogging for Crisis Communication : Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma , Washington Area", Proceedings of the 7th International ISCRAM Conference, 2010, pp. 5.

[8] J. B. Houston, J. Hawthorne, M. F. Perreault, E. H. Park, M. Goldstein Hode, M. R. Halliwell, S. E. Turner McGowen, R. Davis, S. Vaid, J. A. McElderry and S. A. Griffith, "Social media and disasters: a functional framework for social media use in disaster planning, response, and research", Disasters, 39 (2015), pp. 1-22.

[9] IFRC, "About disasters", The International Federation of Red Cross and Red Crescent Societies, 2015.

[10] V. Jayaraman, M. Chandrasekhar and U. Rao, "Managing the natural disasters from space technology inputs", Acta Astronautica, 40 (1997), pp. 291-325.

[11] S. Kumar, G. Barbier, M. Abbasi and H. Liu, "TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief", Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2011, pp. 661-662.

[12] J. Li, A. Vishwanath and H. R. Rao, "Retweeting the Fukushima nuclear radiation disaster", Communications of the ACM, 57 (2014), pp. 78-85.

[13] J. Lingad, S. Karimi and J. Yin, "Location extraction from disaster-related microblogs", WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web, 2013, pp. 1017-1020.

[14] P. Mousavi, O. Marjanovic and P. Hallikainen, "Disaster Recovery-The Process Management Perspective", PACIS, 2012, pp. 67.

[15] S. Muralidharan, L. Rasmussen, D. Patterson and J.-H. Shin, "Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts", Public Relations Review, 37 (2011), pp. 175-177.

[16] L. Palen and S. B. Liu, "Citizen communications in crisis: anticipating a future of ICT-supported public participation", Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 2007, pp. 727-736.

[17] L. Palen, S. Vieweg, S. B. Liu and A. L. Hughes, "Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event", Social Science Computer Review, 27 (2009), pp. 467-480.

[18] M. Price, "Oklahoma tornadoes: Social media used to warn, find information", NewsOK, 2013.

[19] Y. Qu, C. Huang, P. Zhang and J. Zhang, "Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake", Proceedings of the ACM 2011 conference on Computer supported cooperative work, 2011, pp. 25-34.

[20] Y. Qu, P. F. Wu and X. Wang, "Online community response to major disaster: A study of Tianya forum in the 2008 Sichuan earthquake", Proceedings of the 42nd Annual Hawaii International Conference on System Sciences, HICSS, 2009, pp. 1-11.

[21] T. Sakaki, M. Okazaki and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors", WWW '10: Proceedings of the 19th international conference on World wide web, 2010, pp. 851-860.

[22] G. Schryen and F. Wex, "IS Design Thinking in Disaster Management Research", System Science (HICSS), 2012 45th Hawaii International Conference on, IEEE, 2012, pp. 4102-4111.

[23] I. Shklovski, L. Palen and J. Sutton, "Finding Community Through Information and Communication Technology During Disaster Events", Proceedings of the 2008 ACM conference on Computer supported cooperative work, ACM, 2008, pp. 1-10.

[24] K. Starbird, L. Palen, A. L. Hughes and S. Vieweg, "Chatter on The Red: What Hazards Thret Reveals about the Social Life of Microblogges Information", Proceedings of the Conference on Computer Supported Cooperative Work (CSCW), 2010, pp. 241-250.

[25] J. Sutton, L. Palen and I. Shklovski, "Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires", Proceedings of the 5th International ISCRAM Conference, Washington, DC, 2008, pp. 624-632.

[26] C. Torrey, M. Burke, M. Lee, A. Dey, S. Fussell and S. Kiesler, "Connected giving: Ordinary people coordinating disaster relief on the Internet", System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on, IEEE, 2007, pp. 179a-179a.

[27] S. Vieweg, A. L. A. Hughes, K. Starbird and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness", Proceedings of the 28th international conference on Human factors in computing systems - CHI '10, ACM, 2010, pp. 1079-1088.

[28] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review", Management Information Systems Quarterly, 26 (2002), pp. 3.

[29] J. Yin, S. Karimi, B. Robinson and M. Cameron, "ESA: Emergency Situation Awareness via Microbloggers", Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12, ACM, 2012, pp. 2701-2703.

Paper 2
Mukkamala, AM & Beck, R (2016)
Enhancing Disaster Management Through Social Media Analytics to Develop Situation Awareness: What Can Be Learned from Twitter Messages About Hurricane Sandy?
In: Proceedings of the Pacific Asia Conference on Information Systems **(PACIS 2016)** Chiayi, Taiwan. (Completed research accepted as a short paper)

Enhancing Disaster Management Through Social Media Analytics To Develop Situation Awareness -What Can Be Learned From Twitter Messages About Hurricane Sandy?

Alivelu Mukkamala, Information Management Section, IT University of Copenhagen, Denmark, ammu@itu.dk

Roman Beck, Information Management Section, IT University of Copenhagen, Denmark, romb@itu.dk

Abstract

Twitter became an important channel to contribute and consume all kinds of information, especially in times of disasters, when people feel the need for fast, real-time flows of information. Given the wealth of information Twitter provides, that information can be used by practitioners and researchers alike to study what people affected by a disaster talk about, e.g., to develop a situation awareness and to coordinate disaster management accordingly. In our research, we analyze 11 million tweets that deal with hurricane Sandy, one of the strongest hurricanes that ever hit the US east coast in 2012. First, we extract the tweets by narrowing down the hurricane affected path along the US east coast, based on geo-spatial information. Further, drawing on the situation awareness literature and previous coding schemes, we analyze the nature and characteristics of the tweets. Our research reveals that there are significantly more tweets from original sources than from secondary sources and that individuals tend to share valuable personal experiences and observations at the time of disasters. In analyzing those individual level perceptions, we illustrate how one can generate situation awareness at the collective level. This situation awareness will enhance the decision-making of disaster management agencies at the time of uncertain and volatile situations.

Keywords: Social Media Analytics, Disaster Management, Situation Awareness

1. Introduction

@tweetuser (2012-11-01 03:28:59) "What's amazing is how Twitter and Facebook are more current and up to date with events on #Sandy then the actual news on tv...#media #fail"

Among the different social media platforms, Twitter is not only one of the largest but also one of the fastest when it comes to reliably disseminating news and information in the form of short messages known as tweets (Java et al. 2007;Velichety & Ram 2013). Twitter also allows us to enrich the messages exchanged by adding links, pictures, hashtags, and video clips thereby empowering users to exchange information they regard as important, which in turn makes Twitter an interesting source of information for researchers and practitioners alike (Acar & Muraki 2011;Ahmed & Sargent 2014;Ha & Ahn 2011;Oh et al. 2010;Sinnappan et al. 2010). Thus, Twitter is an effective medium to share information, opinions, and emotions in real-time (O'Connor et al. 2010;Pak & Paroubek 2010).

Initially, Twitter was designed to be used as a convenient way to share short messages with family and friends, but over time it became a platform to share impressions and news, as seen in the Arab spring with its democratic movements, e.g., in Tunisia and Egypt. Thus, Twitter users are able to learn about events taking place from exchanged messages, thereby decreasing their situational uncertainty while promoting certain collective goals (Lotan et al. 2011;Oh et al. 2012;Starbird & Palen 2012).

In our research, we are interested in the value of Twitter messages to develop a situation awareness in areas that are affected by a natural or man-made catastrophe. Moreover, it is in the human nature to communicate and get closer together in times of unsettling events, thereby creating a feeling of closeness with each other and thus share and read news and messages more often, e.g., when floods, wildfires (Starbird et al. 2010;Vieweg et al. 2010), or earthquakes (Qu et al. 2011) occur.

Despite an increasing interest in social media analytics in information systems research, there is only a limited number of publications on, e.g., the use of Twitter messages for disaster management. With our research, we are addressing this gap and analyse how different information categories shared in social media messages can be identified and used to provide disaster response agencies with better intelligence and situation awareness. In order to illustrate what a disaster management agency can learn from social media data, we analysed approximately 11 million tweets that have been sent between October 25th and November 5th 2012, dealing with hurricane Sandy that hit the US east coast and made landfall at New York City at the end of October 2012. We will illustrate how we processed the Twitter messages, including data cleansing and pre-processing before we conducted a manual coding to extract and classify information that allows for an improved situational awareness. In doing so, we strive to answer our research questions, namely *which different types of disaster management relevant information can be identified* and *what is the nature and characteristics of this shared information?*

The remainder of the paper is organised as follows. In section 2, we will explain the literature background, followed by methodology in section 3, where we describe the methodology where we systematically pre-processed and narrowed down the data using a series of methodologies, before we describe the qualitative content analysis that we applied for the coding of the data. The subsequent analysis of the data will be presented in section 3. In section 4, we will discuss the theoretical and practical implications, before we conclude our work with section 5 where we also discuss the limitations of our work as well as future research directions.

2. Literature Background

2.1 Social Media Analytics in Times of Disasters

In information systems, social media analytics and research on Twitter data in times of disasters has gained some prominence in recent years, focusing on both, man-made as well as natural disasters. While some early work was focused on information production and distribution, as illustrated by the Red river flood in central North America in 2009, where individuals not only shared information about their situation but also re-tweeted the available, useful information, later work already differentiated in original and secondary information, synthesized or derivative from prior sources (Starbird et al. 2010). In any case, it has been discovered that the information shared contains situational updates that are used to create situational awareness, such as illustrated in the Oklahoma wildfire case (Vieweg et al. 2010). However, Twitter is not only used by the people affected by a disaster, but also by local authorities to inform and manage the situation, as the Queensland flood case illustrates, where the interactions between Twitter users were analyzed (Cheong & Cheong 2011). The case clearly illustrates that Twitter is not only an important source to gain information and develop situational awareness but that it can also be used by disaster management agencies to actively manage a crisis.

When an earthquake struck Yushu in China, the people affected by the disaster not only provided information about the situation on the ground but they also used social media to ask for help or for showing empathy with those who suffered most losses (Qu et al. 2011). It furthermore seems to be irrelevant if the disaster is a natural catastrophe or a man-made disaster, the prime concern always seems to be to share information of the situation rather than sharing opinions and thoughts, as research on a police shooting case in Seattle revealed (Heverin & Zach 2010). The same finding has been made when Twitter messages dealing with the Boston marathon bombing case were analyzed (Venkatesan et al. 2014).

2.2 Situation Awareness

In order to support good decision making, accumulating information to develop a certain understanding about what is happening in a certain situation is important (Reilly et al. 2007), especially in times when disasters take place suddenly which require immediate action to respond and help affected people. As illustrated in 2.1, some empirical research uses Twitter data in times of disasters with the aim to shed light on the role of information creation and sharing on social media platforms to create situation awareness (SA). Situation awareness refers to the way human beings extract meaning from information about their surroundings to develop mental models of a situation by integrating the extracted information with their own knowledge to explore and anticipate further action (Seebach et al. 2011;Vidulich et al. 1994). Often, SA helps in emergency situations to implement the response strategies and to derive decisions to combat the crisis (Vieweg et al. 2010). Even though SA has often been analysed from an individual point of view, it can also be aggregated at the group level

(Seebach et al. 2011). It can be argued that Twitter is facilitating group level SA that can be subsequently used by disaster management agencies to develop an understanding about the situation on the ground. The situation updates that started at an individual level lead to engage at a collective level activity. This collective level sharing of knowledge enhances the SA in a broader prospect and helps emergency organizations act immediately in disaster response.In relation to situational awareness, currently, research is focusing on employing machine-learning approaches to extract the situational awareness information from Twitter data during disasters (Sen et al. 2015;Verma et al. 2011).

3. Methodology

3.1 Data Collection

For our research, we used a Twitter message dataset containing hurricane Sandy tweet IDs (Zubiaga & Ji 2014) that are publicly accessible (Zubiaga 2015). The original dataset contains nearly 15 million tweets posted on Twitter between October 25th and November 5th, 2012 when hurricane Sandy hit the east coast of the US. In order to comply with the terms of service of Twitter, the researchers only shared the tweet IDs and not the actual tweets from the dataset of hurricane Sandy. Using those tweet IDs, we were able to retrieve tweets between May and June 2015, using the Twitter Rest APIs (Twitter 2016). The originally shared dataset comprised nearly 15 million tweets, but when we downloaded the tweets we were only able to retrieve approximately 11 million tweets.

As mentioned, the total dataset of Hurricane Sandy contains approximately 11 million tweets that originated across the world. The descriptive statistics of the dataset are provided in Table 1. The Twitter data collection starts October 25th, 2012 and ends November 5th, 2012, which is the period when Hurricane Sandy hit the east coast of the US. Altogether, 3,983,288 unique Twitter accounts sent messages about Hurricane Sandy during this time, writing tweets in 61 different languages. As depicted in Table 1, only 0.93% of the tweets are written in English and contain geo-location information as well.

Twitter messages (tweets)	Absolute numbers	Percentage
Total tweets	1,658,279	100%
Original tweets	5,369,520	49.50%
Retweeted tweets	5,478,562	50.50%
Tweets with geo-location	115,800	1.07%
English tweets with geo-location	100,700	0.93%

Table 1 Descriptive Statistics of Hurricane Sandy dataset

Even though the majority of the tweets has been sent in English, a considerable percentage of tweets are also composed in Spanish and Portuguese.

3.2 Data Pre-processing

As a first step in data pre-processing, we used Tableau which is a business analytics and visualization software (Tableau 2015). Initially, with the help of Tableau we visualized the dataset of 11 million tweets to get an overview of how the tweets were distributed over the covered period.

In general, users' privacy settings play an important role in the availability of geo-location information in tweets as metadata in form of latitude and longitude coordinates. These longitude and latitude coordinates indicate the exact location of a Twitter user at the time of sharing the information (Graham et al. 2014). For our purposes, we used Cosmos software which provides the geo-located information (Burnap et al. 2015) from the tweets in our first phase of data pre-processing. Using Cosmos, in doing so, we identified 115,800 tweets, or 1.07% of our initial 11 million tweets, which actually contained such geo-location information

During the second phase, we used CartoDB to 1) filter the English language tweets, 2) visualize the data, and 3) narrow down the tweets geographically to the path hurricane sandy took along the coastline (CartoDB 2015). Since we were only interested in tweets sent in English language, the number of tweets further declined to 100,700. In subsequent steps, we narrowed down the focus on tweets from hurricane-affected areas of US east coast. As a result, we ended up with 68,800 tweets that originated from the hurricane-affected area between 25th October and 5th November, 2012. In addition to the message content, the obtained dataset also contains meta-data such as time stamps, geo-information, and user IDs.

In the third phase of data pre-processing, we manually screened the tweets to exclude the ones with commercial content such as advertisements, spam tweets, etc. We also excluded the tweets that had mentioned the word hurricane but were not related to the hurricane Sandy (e.g. talking about hurricane Katrina instead). And subsequently conducted the content analysis.

3.3 Qualitative content analysis

Subsequently, after the data pre-processing phase, we conducted a manual content analysis on the remaining 68,800 tweets. In this regard, we followed a directed content coding and analysis approach (Hsieh & Shannon 2005;Risius et al. 2015) based upon a coding classification framework we developed.

In the first step, we explored different coding schemes and adapted two different coding classifications from prior research (Qu et al. 2011;Starbird et al. 2010) to develop our coding framework. The primary reasons for choosing prior coding classifications are: firstly, both coding classifications were evolved from and developed for analysis of disaster-related Twitter data in an earthquake and floods respectively. Moreover, one of the coding schemes (Qu et al. 2011) was developed based on the previous coding scheme (Qu et al. 2009;Vieweg et al. 2010) that also originated from prior disaster management research on Twitter data. Secondly, these coding classifications were designed to understand the importance of social media content during disasters.

The first coding classification introduced by Starbird et al. (2010) where tweets were classified based on information source, which are: original source (reflecting the users' personal observations, experiences), secondarily synthesized (from other tweets and news sources), resourced (passing-on other online sources) and finally retweets (forwarding the tweets). In order to aggregate the tweets at a higher level, in our coding scheme we merged secondarily synthesized, resourced, and retweet categories into a single category as secondary source. Therefore, our first level of coding scheme consists of two information source categories: original and secondary.

The other coding scheme (Qu et al. 2011) evolved through a mixed process where classification was based on two previous coding schemes (Qu et al. 2009; Vieweg et al. 2010). Primarily it consists of four major categories based on the nature of messages: informational messages, action-related, opinion-related and emotion-related. Furthermore, we combined both the coding schemes to obtain an integrated classification of hurricane Sandy tweets. Due to this, we were able to aggregate the tweets on the highest level, thereby not following the more granular coding approach of Qu et al. (2011) since we were primarily interested in the real-time nature of the data. In the process of developing our coding scheme, both the authors were involved in the discussions from the beginning and reached consensus on the proposed integrated coding scheme.

Due to our new coding scheme, it is feasible to analyze a tweet along the two dimensions: information source and nature of message. Therefore, our classification of tweets provides a basic understanding of distribution of tweets among original and secondary sources. At the same time, it also presents the nature of messages shared by the individuals, in terms of informational, action-related, opinion-related and emotion-related. Our primary focus is to study the type and nature of information that will contribute to the situation awareness of a disaster to emergency management agencies. Primarily, original, and informational messages with personal experiences and observations about a certain situation will contribute to the situation awareness. Even though the secondary source information does not contain the personal status updates but it will still contribute to situation awareness indirectly by providing useful information about situation.

The qualitative content analysis of 68,800 geo-located tweets was conducted by the first author manually to screen for the tweets with relevant information, which resulted in 677 identified tweets. Subsequently, those 677 geo-located tweets with relevant information were further analyzed by both the authors based on the previously mentioned coding scheme. The results from the analysis were checked again by a linguistic expert from our research group.

4. Analysis

It is obvious that people share and discuss information before, during, and immediately after any major event such as disasters and the same can be witnessed while analyzing tweets from hurricane Sandy. As expected, the number of tweets gradually increased beginning in October 27th, just before the disaster hit New York City and the first peak with around 2.2 million tweets per day was noticed on October 29th, while the highest peak with 2.7 million messages per day was on October 30th 2012. Soon after that, the number of tweets decreased sharply for the following two days, but a trend of approximately half a million tweets per day continued until November 4th. As mentioned earlier, we extracted the tweets that originated from the affected area of hurricane Sandy along the US east coast, based on geo-location coordinates. The resulting tweet density map is shown in *Figure 4* where a visualization containing 68,800 tweets for a period of roughly two weeks in the months of October and November 2012. It is evident from the map that more and more tweets were generated along east coast, mainly from Florida, Connecticut, New Jersey, Massachusetts, and New York and so on.

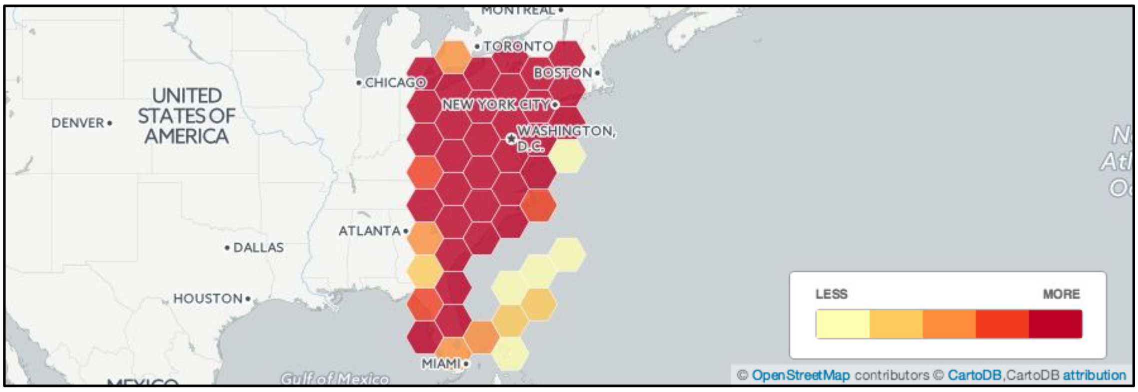

Figure 4 Sandy affected area (East coast of the US)

In the qualitative content analysis, we have manually coded the tweets according to our coding scheme and further narrowed down the tweets to 677 tweets, which is around 1% of the geo-located tweets. The graph in Figure 5 shows the representation of the 677 tweets that were categorized. As we have considered three separate phases of disaster; namely the pre-disaster, during-disaster and post-disaster phases, we considered the time line of 25th October to 5th November which encompasses the aforementioned phases. The tweets are classified based on whether the source of information in the tweets were original or secondary. Furthermore, based upon the content of the message and the links included in the tweets, we categorized them into original and secondary source categories. The tweets were put under heavy scrutiny to ensure effective categorization.

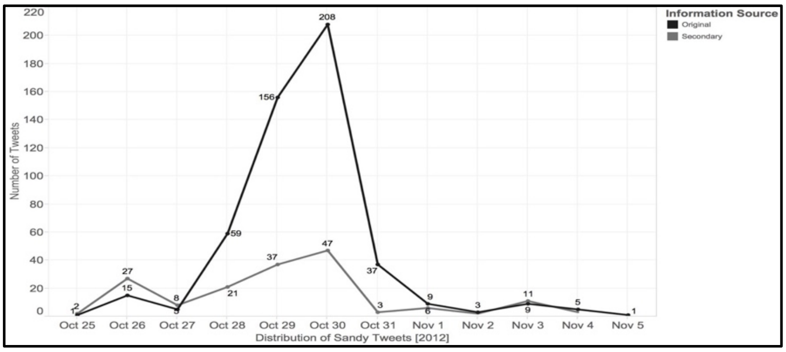

Figure 5 Information Source

Basically, if the content of a message contains first hand observations and experiences, then the tweet is considered as from "original source". In order to cross-verify the original tweets, if a tweet contains a URL or a link, then we followed the link in an effort to find out if the person who created the tweet was also the creator of the link itself. If this was the case, the tweet would be classified as an original tweet. On all other cases, the tweets were classified as a "secondary source" tweet.

Figure 5 shows distribution of coded tweets among original and secondary sources. Altogether 75% of the tweets represent original tweets, whereas 25% belong to the secondary tweet category. In the initial days of pre-disaster phase, the number of tweets from secondary source is more than the tweets from original source, as users were mostly resourcing and re-tweeting the information from other information sources such news articles, weather reports, or showing forecasted severity of the disaster. It is obvious that when people are awaiting an impending disaster that would potentially affect them, they tend to vent on their concerns, fears and also discuss the preventative measures that are being taken. In this process, they also share whatever information they gain from other sources. As the disaster unfolds slowly, people started sharing their own personal experiences, observations and also discussed the situation based on common knowledge or adapting from other sources (Starbird et al. 2010). Hence, from 27th onwards, the tweets in the original category started increasing till 31st October as on 29th hurricane Sandy hit New York City. During the disaster phase, people's curiosity to understand, know, share and estimate the impact of current situation increased considerably. This was more so the case for people who were affected by the hurricane.

As part of further classification, based on the nature of information in tweets, the tweets were classified into four categories such as informational messages, action-related, opinion-related and emotion-related. In this step, the tweets were classified based on the content, and distribution of four categories is shown in Figure 6. Approximately 73% of the tweets were informational messages, 15% tweets were emotion-related, whereas opinion and action-related tweets were 8% and 4%, respectively. As discussed previously, in our analysis, most of the informational messages were providing the situation updates, which are important for all phases of a disaster.

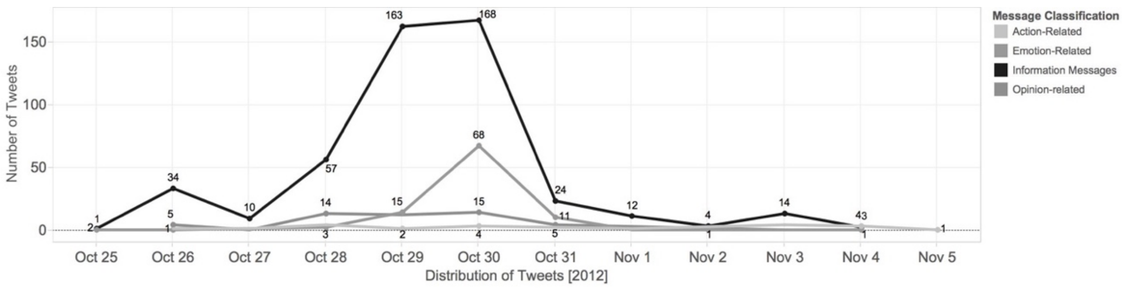

Figure 6 Message occurrence according to different classification

We further subdivided the tweets based on information source and nature of messages. In both, original and secondary source categories, information messages are predominant throughout the disaster period, even though the majority of the informational messages came from the original source category only. In respect to other message categories, emotion-related tweets started increasing from 28th October and peaked on 30th October. It is obvious that people expressed their sympathy as well as provided emotional and social support during the time the hurricane raged. Emotions are often personal hence we noticed them in only the original source. Not only emotion-related but also opinion and action-related discussions are often personal and therefore they appeared only in the original source category.

5. Discussion

In our current research, our focus was to explore the relevant type and nature of information in disaster twitter messages. For that we have collected 11 million tweets that were produced during the time of hurricane Sandy, which were further systematically narrowed down to a meaningful sample. The motive of our study is to affirm that self-organized, affected individuals' can develop individual level situation awareness from others, sharing their personal observations, experiences, and opinions that can also be used to develop situation awareness at the collective level among community members. This will be useful to the emergency management agencies.

5.1 Situation Awareness in Disaster Management

So far, the research focus on disaster twitter data has been predominantly on analyzing tweets from a communication point of view. Very scarce research has focused on disaster twitter data from the situation awareness point of view especially focusing on the real-time information to support disaster management agencies.

This is of crucial importance since it is evident that disaster management agencies depend on proper information in times of disasters, which can be gained from individuals which share information to decrease the uncertainty and anxiety among followers (Kaewkitipong et al. 2012). The realization of the potential use of social media during disasters made emergency agencies adopt and focus more on social media tools like Twitter. However, in initial attempts, disaster management agencies used social media, such as Twitter as broadcasting channel for their emergency information communication (Chavez et al. 2010) or to maintain public relationships, but failed to establish a dialogue with the community (Muralidharan et al. 2011;Waters et al. 2009). The reason could be that the emergency agencies do not have proper strategies and methodologies to extract and handle social media based information streams in close to real-time yet. However, this is going to change with more and more disaster management officials being aware of the importance of social media also in order to play an active role during crises (Cheong & Cheong 2011). Moreover, top management is playing a crucial role in the adoption of twitter for emergency services, as we have seen in a case from New South Wales (Fosso Wamba & Edwards 2014).

It is evident that Twitter is being used by ordinary people during times of disasters to provide up to date and real-time information of the disaster situation. This information is useful for disaster management agencies to keep track of real-time situations in the affected area. This type of community level situation awareness helps emergency management officials to make better decisions to handle the disaster response effectively.

Our research findings also revealed that most of the informational messages are situation updates, whether it is produced by the original source or secondary source. However, the percentage of original tweets is larger, indicating the affected individuals share the real-time information either to provide the situation updates or because they are in need of real-time help. This bottom-up generated information is needed by the emergency agencies.

5.2 Implications for Theory and Practice

Our study offers significant contributions to research and practitioners alike. We are not aware of any other research within IS that analyzed the disaster twitter data from a situation awareness point of view, specifically focusing on supporting disaster management agencies., We are among the first to advance the research to the field of SA from the real-time angle, while reflecting it in disaster management agencies. Furthermore, from an organizational point of view, our findings reveal that the individuals share situational updates hence disaster management agencies need to have social media integrated centralized information technology infrastructure to exploit the situation awareness.

5.3 Limitation and Future Work

This study has certain limitations as well. We were able to access most of the links or URLs posted in the tweet messages, but very few were not available, therefore based on the content of the message we divided them into original and secondary categories. We also were not able to analyze those messages that were sent without geo-location information. Thus, we had to sacrifice the majority of the Twitter message from peoples directly affected by the hurricane, but did not share their whereabouts.

We also included the tweets that were posted by the emergency agencies, even though they were very negligible, and classified them as messages in the secondary category. It is important and challenging to extract the valuable information from huge amount of data in real-time. Hence the future research should consider developing systematic methodologies and analytical capabilities to handle the huge amount of data. Our current research shed light on the real-time information that has practical and societal level implication of disaster management.

6. References

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

Ahmed, A. and Sargent, J. (2014). Analysis of Post-Crisis Twitter Communication: A Study of the Iquique, Chile Earthquake. *In: Proceedings of the Australasian Conference on Information Systems* ACIS.

Burnap, P., et al. (2015). COSMOS: Towards an integrated and scalable service for analysing social media on demand. *International Journal of Parallel, Emergent and Distributed Systems* 30(2) 80-100.

CartoDB. (2015). Available at: https://cartodb.com/.

Chavez, C., Repas, M. A. and Stefaniak, T. L. (2010). A New Way to Communicate with Residents: Local Government Use of Social Media to Prepare for Emergencies. the International City/County Manage- ment Association.

Cheong, F. and Cheong, C. (2011). Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods. *In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS)* 46-46.

Fosso Wamba, S. and Edwards, A. (2014). Factors Related to Social Media Adoption and Use for Emergency Services Operations: The Case of the NSW SES. *In proceedings of the 20th Americas Conference on Information Systems. Savannah.*

Graham, M., Hale, S. A. and Gaffney, D. (2014). Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer* 66(4) 568-578.

Ha, S. and Ahn, J. (2011). Why are you sharing others' tweets?: The impact of argument quality and source credibility on information sharing behavior. *In: Proceedings of the International Conference on information Systems*.

Heverin, T. and Zach, L. (2010). Microblogging for Crisis Communication : Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma , Washington Area. *In: Proceedings of the 7th International ISCRAM Conference*. 5.

Hsieh, H.-F. and Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research* 15(9) 1277-1288.

Java, A., et al. (2007). Why we twitter: understanding microblogging usage and communities. *In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, ACM.

Kaewkitipong, L., Chen, C. and Ractham, P. (2012). Lessons learned from the use of social media in combating a crisis: A case study of 2011 thailand flooding disaster. *In: Proceedings of the International Conference on information Systems*.

Lotan, G., et al. (2011). The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International journal of communication* 5 31.
Muralidharan, S., et al. (2011). Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts. *Public Relations Review* 37(2) 175-177.
O'Connor, B., et al. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*.
Oh, O., Eom, C. and Rao, H. R. (2012). Collective sense-making through the twitter service during the 2011 egypt revolution. *Research in Progress In the International conference on Information Systems*.
Oh, O., Kwon, K. H. and Rao, H. R. (2010). An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010. *ICIS*.
Pak, A. and Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREC*.
Qu, Y., et al. (2011). Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. *In: Proceedings of the ACM 2011 conference on Computer supported cooperative work*.
Qu, Y., Wu, P. F. and Wang, X. (2009). Online community response to major disaster: A study of Tianya forum in the 2008 Sichuan earthquake. *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences, HICSS*. 1-11.
Reilly, W., Guarino, S. L. and Kellihan, B. (2007). Model-based measurement of situation awareness. *Simulation Conference, 2007 Winter*, IEEE.
Risius, M., Akolk, F. and Beck, R. (2015). Differential Emotions and the Stock Market-The Case of Company-Specific Trading. *In: Proceedings of the European Conference on Information Systems*.
Seebach, C., Beck, R. and Pahlke, I. (2011). Situation awareness through social collaboration platforms in distributed work environments. *In: Proceedings of the International Conference on information Systems*.
Sen, A., Rudra, K. and Ghosh, S. (2015). Extracting situational awareness from microblogs during disaster events. *Communication Systems and Networks (COMSNETS), 2015 7th International Conference on*, IEEE.
Sinnappan, S., Farrell, C. and Stewart, E. (2010). Priceless tweets! A study on Twitter messages posted during crisis: Black Saturday. *In: Proceedings of the 21st Australasian Conference on Information Systems*.
Starbird, K. and Palen, L. (2012). (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*, ACM.
Starbird, K., et al. (2010). Chatter on The Red: What Hazards Thret Reveals about the Social Life of Microblogges Information. *In: Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*.
Tableau. (2015). Available at: http://www.tableau.com/.
Twitter. (2016). *Twitter REST APIs* [*online*]. Available at: https://dev.twitter.com/rest/public.
Velichety, S. and Ram, S. (2013). A Cross-Sectional and Temporal Analysis of Information Consumption on Twitter. *In: Proceedings of the International Conference on information Systems*.
Venkatesan, S., et al. (2014). Exploring Multiplexity in TWITTER–The 2013 Boston Marathon Bombing Case. *In: Proceedings of the International Conference on information Systems*.
Verma, S., et al. (2011). Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency. *ICWSM*, Citeseer.
Vidulich, M., et al. (1994). Situation awareness: Papers and annotated bibliography. DTIC Document.

Vieweg, S., et al. (2010). 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*, ACM.
Waters, R. D., et al. (2009). Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public Relations Review* 35(2) 102-106.
Zubiaga, A. (2015). *Dataset: Hurricane Sandy tweets* [*online*]. Available at: http://www.zubiaga.org/datasets/hurricane-sandy-tweets/.
Zubiaga, A. and Ji, H. (2014). Tweet, but verify: epistemic study of information verification on Twitter. *Social Network Analysis and Mining* 4(1) 1-12.

Paper 3

Mukkamala, AM & Beck, R (2017)

Presence of Social Presence during Disasters

In: Proceedings of the Pacific Asia Conference on Information Systems **(PACIS 2017)**
Langkawi, Malaysia

Presence of Social Presence during Disasters

Completed Research Paper

Alivelu Mukkamala
IT University of Copenhagen
Copenhagen S, Denmark
ammu@itu.dk

Roman Beck
IT University of Copenhagen
Copenhagen S, Denmark
romb@itu.dk

Abstract

During emergencies, affected people use social media platforms for interaction and collaboration. Social media is used to ask for help, provide moral support, and to help each other, without direct face-to-face interactions. From a social presence point of view, we analyzed Twitter messages to understand how people cooperate and collaborate with each other during heavy rains and subsequent floods in Chennai, India. We conducted a manual content analysis to build social presence classifiers comprising intimacy and immediacy concepts which we used to train a machine learning approach to subsequently analyze the whole dataset of 1.65 million tweets. The results showed that the majority of the immediacy tweets are conveying the needs and urgencies of affected people requesting for help. We argue that during disasters, the online social presence creates a sense of responsibility and common identity among the social media users to participate in relief activities.

Keywords: Social media, social presence, disasters, and disaster management.

1. Introduction

In recent years, social media has significantly changed the way we interact on the Internet. It has achieved popularity because it allows the creation and exchange of end-user-generated content, quick information dissemination, and most importantly, publicly available content (Kaplan and Haenlein 2010). Social media platforms such as Facebook or Twitter are even used in situations of emergency either to communicate or to share information with their followers. Through social media, people can collaboratively engage in disaster response situations during an emergency such as when hurricane Sandy hit the east coast of U.S (Kryvasheyeu et al. 2016), during earthquakes in Haiti (Sarcevic et al. 2012; Zook et al. 2010), China (Qu et al. 2011), and Japan (Toriumi et al. 2013), during wildfires in southern California (Sutton et al. 2008), grassfires in Oklahoma (Vieweg et al. 2010b), during floods in U.S (Starbird et al. 2010), Australia (Cheong and Cheong 2011), and Pakistan (Murthy and Longwell 2013), and more recently during floods in Chennai, India (Pandey 2015).

One social media platform gained specifically importance in emergencies for information dissemination, and that is Twitter. It enables people to share opinions, experiences, and information through short text messages. On Twitter, people involve themselves in conversations by posting messages, following others or retweeting others' messages. Most importantly, anyone can follow any other without having a mutual acquaintance. Overall, it facilitates interactions and conversations among its users (Boyd et al. 2010).

Although people use Twitter in their personal life for entertainment purposes, it is evident that at times of disasters people are using Twitter to share situation updates and to provide emotional support on a large scale (Verma et al. 2011; Vieweg et al. 2010a). Twitter has gained a lot of attention from research communities, especially its usage during times of disasters, to understand how information is generated (Starbird et al. 2010), how different types of messages are shared by people (Qu et al. 2011), and how credible the information is (Oh et al. 2010). Thus, Twitter is changing the traditional communication practices during emergencies because of real-time user generated content which is enhancing the collective collaboration (Vieweg et al. 2008) through building social capital. Considering the media-related component of social media, based on the degree of social presence, different social media applications such as blogs, Wikipedia, YouTube, second life, and social networking sites were classified on a continuum between low and high social presence. Based on it, social networking sites fall into the "medium" category because users can connect and communicate with others through text messages and can also upload videos, pictures, and other forms of information (URLs and website links) (Kaplan and Haenlein 2010). During disasters, despite the lack of face-to-face interactions, people ask for help, provide moral support, and even help each other on social media platforms. It is important to explore what makes people to perceive and feel for the others and lend their hand, by reading the short text messages. So far research focus was to understand the social presence on online learning (Gunawardena and Zittle 1997; Tu and McIsaac 2002) and on social media (Al-Ghaith 2015; Xu et al. 2012) by applying survey strategies or interviews. Especially during disasters, when people seek help and support on social media, other online users feel intimacy and immediacy for those affected people and provide support by sharing the relevant information online and actively participate in relief activities. Without the feeling of social presence neither people who are in need ask for help nor people come forward to lend their hand.

Emergency management agencies (EMA) such as FEMA or the Red Cross can exploit real-time information from self-organized communities on social media during disasters. It is also envisioned that due to the potential usefulness of social media, EMA will increase social media use (Kaewkitipong et al. 2016). It is quiet beneficial for EMA to collaborate with self-organized communities to implement the relief measures effectively. In doing so, it is important to understand what makes people build relationships, cooperate, and collaborate with each other to face the threatening conditions during times of emergencies. In this regard, using the theory of social presence as a methodological framework, this paper shows how seemingly ephemeral and hastily written text messages on Twitter can create a feeling of intimacy and immediacy which are an important aspect to share information at community level, which will be useful to disaster management agencies.

In order to grasp the severity of the situation and act collaboratively to organize and to participate in emergency response activities, people primarily need to perceive, sense and empathize with others on social media, which is similar to the related concepts of social presence (Short et al. 1976). While there are a few studies which have used a questionnaire-based approach in their social presence research on social media (Al-Ghaith 2015; Xu et al. 2012) there is no research so far that has explored social presence based on a content analysis of tweets during disasters. The perceived social presence by people on social media is important. Thus, our research question is:
How social presence can be detected through a content analysis of tweets and which role does it play for building relationships, cooperate, and collaborate during times of emergencies?

In order to understand how people express intimacy and immediacy as forms of social presence in times of disasters, we analyzed approximately 1.65 million tweets from a devastating flooding in Chennai, India, which took place in December 2015. We applied a supervised machine learning approach to analyze the content of the 1.65 million tweets. In doing so, we will also illustrate how machine learning can be applied to analyze large volumes of textual content for exploring theoretical concepts such as social presence.

The remainder of this paper is structured as follows: Section two provides the theory of social presence. The main focus in section three is about the methodology we applied which is of three-fold: 1) operationalizing social presence in social media, 2) conducting manual content analyses to develop training dataset for message classification, 3) training and using a Naïve Bayes machine-learning approach to classify our dataset. In section four, the empirical analysis and results will be presented and subsequently, we conclude the paper with discussion, limitations, and future research.

2. Literature Background

2.1 Social presence

Social presence can be traced back to telecommunications research in the 1970s (Lowenthal 2009; Short et al. 1976) where it was viewed as a media characteristic (Kehrwald 2008). Social presence can be defined as "the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships" (Short et al. 1976). In the communication context, the degree of salience indicates the perceived feeling or significance of the other person being present in the interaction (Kehrwald 2008). The quality of a communication medium plays an important role as it can determine the way people interact and communicate. Hence, from this perspective, the degree of social presence of a communication medium is assessed based on how well a medium can transmit information of non-verbal cues, facial expressions, posture and attire (Gunawardena 1995). It is believed that each communication medium is different in its degree of social presence, for example, audio vs. video (Lowenthal 2009). Social presence is typically associated mainly with two concepts; namely intimacy (Argyle and Dean 1965) and immediacy (Wiener and Mehrabian 1968). The former concept describes how people act and come close during social interactions while the latter concept focuses on interpersonal communication and communicative behavior (Short et al. 1976).

Equilibrium of intimacy develops between any pair of people, where the joint function of mutual exchange of smile, conversation, eye contact, or physical distance occurs. People alter their behavior to maintain the intimacy whenever one of the functions changes(Argyle and Dean 1965). However, immediacy is a measure of the psychological distance, which an individual places between himself and his target audience (Wiener and Mehrabian 1968). The selected communicative behavior of an individual leads to physical or psychological closeness in interpersonal communication (Wiener and Mehrabian 1968; Woods and Baker 2004). As an example, the use of television enhances intimacy to a greater degree than radio (Short et al. 1976). Along the same lines, it can be argued that a person can create an impression of formal or informal attitude while speaking with someone on the phone. In other words, a person can convey immediacy or non-immediacy through verbal and non-verbal communication while speaking with someone on the phone (Aragon 2003; Cobb 2009; Gunawardena and Zittle 1997; Tu 2002). Therefore, social presence can be viewed as an attribute of the media in question as well as that of the communicators and their presence in a sequence of interactions (Biocca et al. 2003; Gunawardena and Zittle 1997).

Over time, with the rise of computer-mediated communication (CMC), social presence theory altered and gained importance in online learning disciplines (Lowenthal 2009). CMC facilitates social interactions through text-based content. However, from the social presence theory point of view, a text-based CMC could be relatively low in social presence when compared to face-to-face interactions. Although the new technologies are more effective in information processing, transmission, and user experience, it is unclear yet "how the social meaning of interactions is affected in the absence of nonverbal cues when communicators substitute text-based electronic messaging for face-to-face encounters" (Walther et al. 2005, p. 1). However, when it comes to text-based media such as e-mail or chat, which are considered richer than face-to-face conversations (Walther 1992). Moreover, mediated communications play an important role to perceive the other person as real while communicating (Gunawardena 1995). Thus, the theory of social presence that originates from media studies has been often applied to examine interactions between students and teachers in the context of online learning (Tu and McIsaac 2002). Most importantly, in online learning, online social presence is conveyed through the messages sent by the online participants and the interpretation of those messages by others. The visible activities, such as posting messages, replying and responding to others, and participating in the activities of the group contain the cues of social presence of the individuals who send them and who receive them (Kehrwald 2008). This confirms the fact that despite the lack of existence of non-verbal cues in online environments, individuals grasp cues through language, style, and other cues to build relationships (Walther et al. 2005). Thus, the theory of social presence can be used to explain and understand how people interact within online learning environments while there is still a lack of understanding on how to properly detect and measure social presence in social media environments. In consequence, social presence has been defined in different ways in online learning research in the past (Kehrwald 2008; Lowenthal 2009). It is a known fact that social networking sites fall into the "medium" category because of the richness of different attributes such as textual content, videos, pictures, and other forms of information (URLs and website links) (Kaplan and Haenlein 2010). While social presence has been used in research on social networks before, it is unclear which role social presence plays in times of emergencies, e.g., to stimulate relief activities. However, previous studies mentioned that online users' interactions with other users and their engagement is directly related to social presence (Lim et al. 2015) and that it plays an important role in fulfilling social connection needs in online environments (Han et al. 2015).

3. Operationalization of Social Presence

On Twitter, a user can "follow" and "followed" by anyone without mutual acquaintance. A user can "retweet" (RT) and also "like" (favorite) some one's status updates. Users can form groups using "list" and can use "mention" to send public messages. The hashtags, with prefix "#" empowers users to connect with a group and enhances coordination among users for a common cause and also increases the search ability (Starbird et al. 2015). For example, at the times of emergency situations, a victim of a disaster or a person who is witnessing it can send a message for help or update his/her status by posting a message related to the disaster. Thus, a user can tweet a message for help or explain the situation. People who want to help can forward the same message as RT@ and also can reply using @xxx to provide help.

3.1 Social Presence

All the above features and the characteristics of Twitter are facilitating interactions and conversations with others. In general, if we perceive Twitter and its features as media-related component, then the features create a sense of intimacy and immediacy based on its content. However, a tweet could also create a feeling of intimacy and immediacy based on the textual content. The online social presence is conveyed through the messages sent by the online participants and the interpretation of those messages by others (Kehrwald 2008).

In order to operationalize the concept of "intimacy", we analyze it from an angle of affective intimacy that explains how people express their feeling of closeness through liking and emotional bonding by providing moral support (Hu et al. 2004; Tolstedt and Stokes 1983). In other words, at times of emergency situations, people post messages to express moral support to combat the situation, or in situations where people go through the same bad conditions, provide a feeling of closeness, e.g., through providing information about road blocks during earthquakes or floods, to create awareness. Although immediacy means perceived psychological distance people feel while communicating with others (Mehbabian 1967; Walther 1992), immediacy is expressed also through situations that give rise for a sense of urgency or excitement and involve people instantly in the action. In general, people perceive psychological closeness towards immediate family members or friends but whenever the whole community faces an emergency, people feel psychological closeness towards affected people such as neighbors or the community itself. Therefore, people come forward to provide help in times of emergency situations as they perceive the closeness, which creates a sense of urgency and importance to act immediately to the people who are in vulnerable conditions such as in need for shelter or food in life threatening situations. Some of the example tweets for social presence reflecting intimacy and immediacy are provided in Table 2.

Table 2 Overview of the Social Presence concepts and their operationalization

categories	Description	Tweet Content
Intimacy	Feeling closeness: sharing road closure info, asking for help on behalf of others	Is there any way to provide any form of support monetary and supplies? #ChennaiRains
	Moral support: stand by people, providing hope for best	Hats off to the fighting spirit of Chennai! A salute all those volunteers who have been helping relentlessly! #staystrongchennai #chenna
Immediacy	Urgent action is needed: different types of rescue requests	any doctors in mudichur area? One pregnant lady in labour...no access to boat. Here is the Contact: 9940203871#chennairains
	Sharing information: provide shelter, food and help	#food available at #tnagar gurdwara 9094790989#ChennaiRainsHelp #ChennaiMicro #ChennaiVolunteer https://t.co/JD1BAdWXSe

4. Disaster Description and Applied Methodology

Chennai, a city in southern India, received a devastatingly high amount of rainfall in December 2015. Especially in the first few days of December, the rainfall intensified and Chennai received 34 times more than the normal daily amount of rain (Misra 2015). According to reports, the flooding caused not only severe economic damages but also severely disrupted the whole city infrastructure such as homes, hospitals, roads, railway tracks and the international airport. During the flooding, people in Chennai used social media to coordinate and to cooperate in relief activities. For example, it was used to rescue people who were stranded in floods, for food distribution, to provide shelters, and also to reach out to people who needed help (Pandey 2015). For our analytics, we focus on user-generated content we harvested from Twitter.

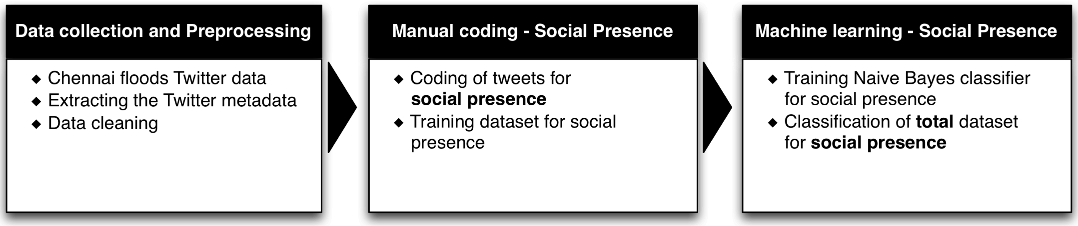

Figure 7 Overall Methodology of Empirical Study

In this section, we will illustrate the methodology to build the classifier and how we analyzed social presence through Twitter messages to have the sense of identity in emergency situations. Specifically, we conducted a manual coding approach as well as machine learning as shown in
Fig. 13. in order to classify tweet messages for social presence. In the following, we will discuss the depicted phases.

4.1 Data Collection and Preprocessing

We used the social media data collection tool Radian 6 to collect the Twitter messages. In case of disasters, hashtags are created during or soon after the incident unfolds to share and communicate information regarding disasters. Hence, we used the hashtags #TNflood, #chennaiRains, #chennafloods #chennaiRainsHelp, #IndiaWithChennai and #chennaiMicro to extract related tweets. The timeline for the collected data was from November 30th to December 16th, 2015 with a total dataset consisting of 1.65 million tweets posted by 209,644 unique users as shown in Table 3. The Radian 6 tool provided some Twitter attributes such as tweet ID, author, content, and followers count, but it does not provide metadata such as retweet status, retweet count, and original tweetId for retweets. Hence, we again downloaded the whole dataset via the open Twitter API using tweet-Ids with the help of a custom tool and reanalyzed the dataset to segregate the original tweets from the retweets based on the retweet status information. As part of the data pre-processing, we deleted the tweets containing some of the mentioned hashtags, but which fall outside the time period mentioned above. Moreover, we also extracted the original tweet-Id for retweets by processing the Twitter data.

Table 3 Descriptive Dataset Statistics

Twitter messages (tweets)	Absolute numbers	Percentage
Total tweets	1,658,220	100%
Retweets	1,226,098	73.94%
Original tweets of Retweets	141,941	08.56%
Tweets never got retweeted	290,181	17.50%
Mean Retweet Ratio	8.64	
Total Unique Twitter Users	209,644	

Most of our dataset consists of "retweets" as shown in Table 3 which constitutes around 74% of the total dataset. Moreover, only 8.5% tweets are original tweets that got retweeted (73.94%) many times, at an average retweet rate of 8.64.

4.2 Manual Coding and Content Analysis

We have conducted a manual content analysis in two phases as shown in Fig. 13 to analyze social presence. A content analysis helps researchers to have their own context of inquiry and constructs to make the texts more meaningful. Through this approach, one can make replicable, reliable, and valid inferences from data on an aggregate level that opens an avenue to understand trends, patterns, and differences (Krippendorff 1989; Lombard et al. 2002). For our approach, we followed a directed content analysis where we have drawn the coding scheme from existing theory (Hsieh and Shannon 2005; Risius et al. 2015; Risius and Beck 2014).

To establish principles along the entire process of our manual content analysis and to ensure obtained measures and results to be more valid and reliable, we followed Morris' 5-step process (Morris 1994) for our content analysis. This approach not only guided us through a step-wise iterative research process but also made the whole process more transparent. In general, the unit of analysis can be a word or a sentence. In the first step, we decided to take the whole tweet content as unit of analysis, as a tweet can be objectively identifiable by coders (Rourke et al. 2001). In the second step, drawing on existing theory, we developed the categories as coding scheme for the data classification. The categories for social presence are 1) intimacy (in), 2) immediacy (im), and 3) none (n). We introduced none label in categories to filter out the tweets that don't belong to the categories. To ensure validity, both researchers who upfront have intensively worked on the theoretical background discussed extensively what constitutes the categories and what does not. The third step basically enhances the coders' familiarity with the coding scheme and also acts as a training session for independent coders. Therefore, one of the coders coded a sample of 100 tweets in both phases, and later on both coders together analyzed the results to exclude the subjective bias and discrepancies.

In the fourth step, after reaching a consensus, both coders independently coded a randomly selected sample of approximately 500 (580) tweets for intimacy and immediacy. Afterwards, the results were compared and both coders discussed about the tweets to clear the discrepancies about the concepts. After an iteration of coding, the inter coder agreement matrix for the texts of social presence coded by the two different coders is presented in Table 4. The integer values in Table 4 represent the number of texts coded by each coder under different categories, whereas the decimal values shown in parentheses represent proportions of the categorized texts to the total coded texts. The inter coder agreement value can be simply measured by considering the agreement rates between different coders, which can be calculated by the proportion of the total number of agreed texts by the total number of texts coded. For example, in case of social presence, the inter-coder agreement value is (38 + 83 + 431) / 580 = 0.95. However, this proportion does not account for discrepancies that might have occurred by chance i.e. where the coder's agreement for the texts might have expected on the basis of chance.

In order to eliminate this limitation and to ensure the validity of inter-coder agreement, we have calculated Cohen's Kappa value (Cohen 1960; Stemler 2001). The Cohen's Kappa value varies from 0 to 1, where a value of 1 indicates perfectly reliable agreement and the value 0 indicates that there is no agreement between the coders other than what is expected on the basis of chance. The calculation of Cohen's Kappa involves two variables: proportion of texts where both coders agree (inter-coder agreement as calculated previously) and proportion of texts for which agreement happened by chance. The proportion of agreement by chance can be calculated by multiplying the proportions of marginal totals of each category and then summing them up. For example, in case of social presence, the proportion of agreement by chance = (0.08 * 0.08) + (0.16 * 0.16) + (0.76 * 0.76) = 0.62. Finally, Cohen's Kappa value for social presence can be calculated by using the standard formula as: (0.95 – 0.62) / (1 – 0.62) = 0.868.

<table>
<tr><th colspan="6">Table 4 Inter-Coder Agreement Matrix of Social Presence</th></tr>
<tr><td colspan="2" rowspan="2"></td><td colspan="3">Coder 1</td><td></td></tr>
<tr><td>Intimacy</td><td>Immediacy</td><td>None</td><td>Marginal Totals</td></tr>
<tr><td rowspan="3">Coder 2</td><td>Intimacy</td><td>38 (0.06)</td><td>7 (0.01)</td><td>2 (0.01)</td><td>47 (0.08)</td></tr>
<tr><td>Immediacy</td><td>7 (0.01)</td><td>83 (0.14)</td><td>6 (0.01)</td><td>96 (0.16)</td></tr>
<tr><td>None</td><td>1 (0.01)</td><td>5 (0.01)</td><td>431 (0.74)</td><td>437 (0.76)</td></tr>
<tr><td></td><td>Marginal Totals</td><td>46 (0.08)</td><td>95 (0.16)</td><td>439 (0.76)</td><td>580 (1.00)</td></tr>
</table>

In general, an inter-coder agreement of 0.40 to 0.80 is considered as a good indicator of valid agreement between coders (Stemler 2001). In the final step of content analysis, one of the researchers processed randomly selected 5000 tweets for social presence in the subsequent phase according to categorization of coding scheme to get the trained dataset. Thus, in this phase we conducted a manual coding of randomly selected tweets from the whole dataset to prepare a training dataset for social presence (intimacy and immediacy). Using the trained dataset of social presence, we applied machine learning to classify the whole dataset for social presence.

4.3 Tweet Classification using Naïve Bayes Classifier

We have adopted a supervised machine learning approach to classify the tweet content of the dataset. Text classification can be defined as a process that comprises of assigning a predefined category of labels to new texts or documents based on a probabilistic measure of likelihood using a training set of labeled texts (Yang 1999). We have used a Naïve Bayes classifier, which is a probabilistic classifier that will estimate the probability of a given text based on the joint probabilities of words and categories using a bag of words approach. The naïve part of the classifier is that it assumes that the conditional probability of a word given a category is independent from the conditional probabilities of other words given in that category. The Naïve Bayes assumption makes the classifier far more efficient and practical than the exponential complexity of other classifiers and also it works quite well for the text classification with a fair amount of accuracy and therefore it stands as one of the most used techniques for text classification (Yang and Liu 1999; Zhang and Li 2007). For classification of tweets using a Naïve Bayes classifier, we used the Natural Language Toolkit (NLTK) (Bird 2006) and Python as programming language to train the classifier using the training datasets containing manually coded tweets as explained in the previous section.

The classification of tweets was conducted for social presence containing labels as intimacy and immediacy. A training set of 5000 manually coded tweets was used, where 80% of the training set was used to train the classifier and the rest 20% of the tweets were used to test the accuracy of the classifier. The text classification is performed in several iterations and within each iteration adopting different strategies (Narayanan et al. 2013) to enhance the accuracy of the classifier. Even though, initially we started with a strategy of using all the words from the training set, but in subsequent iterations we employed strategies such as stop words removal, bigram association measures, feature selection and others to find out a suitable strategy that will suit to our text corpus and yields the best accuracy for Naïve Bayes classifier. Finally, we adopted bigram association measures with removal of stop words and proper nouns (such as names of the people, places etc.) which yielded the best accuracy out of all the strategies for the text classification.

In general, performance of a machine learning algorithm can be described by four measures: precision, recall, F-measure, and accuracy (Powers 2011; Yang 1999). These measures are built over the statistical variables True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). True and False Positives refer to the number of predicted values that were correctly identified and incorrectly identified, whereas True and False Negatives refer to the predicted values that are correctly rejected and incorrectly rejected respectively. Building on these variables, *precision* is defined as the ratio of predicted positive values that are correctly identified as real positive values i.e. TP/(TP+FP). Similarly, *recall* (also known as true positive rate) is defined as a ratio of correctly predicted positive values over all positive values which is TP/(TP+FN). Moreover, F-measure or F-score is a trade-off between precision and recall and it is defined as a single measure which is the harmonic mean of the precision and recall. All these measures precision, recall and F-measure provide performance information at the level of labels, whereas *accuracy* of the classifier provides information about overall performance of the classifier, which can be defined as (TP + TN) / (TP+FP+TN+FN).

5. Results and Analysis

We will present results and analysis of the text classification of tweet content of twitter data.

Table 5 Performance Measures of Text Classification using Naïve Bayes Classifier

Model	Labels	Precision	Recall	F-Measure	Accuracy
Social Presence	Intimacy	0.158	0.466	0.236	0.805
	Immediacy	0.626	0.520	0.568	
	None	0.943	0.858	0.898	

The performance measures about text classification of tweet content using Naïve Bayes classifier are presented in Table 5. The overall accuracy of the classifier is fairly high (i.e. around 80% of the prediction are accurate for correct predictions). In terms of precision and recall, *Immediacy r*eceived fairly good values than *intimacy*. In case of F-measure, a value around 0.6-0.7 indicates a fairly better performance and the F-measure values for all labels/categories indicate reasonably good performance except for the categories: *intimacy.*

Table 6 Most Informative Features for Social Presence

Model	Labels	Most informative features/words
Social Presence	Intimacy	elderly, child, glad, responsive, struck, requesting, praying, help, bravery, aged, catastrophic, wife, together, ...
	Immediacy	tonight, labour, pregnant, accommodate, rides, blankets, mall, ambulance, parcels, surgery, packets, inside, needed, evacuate, hot, boat, distribute, immediately
	None	media, india, victims, well, news, money, hope, government, chennaifloods, day, helpchennai, traffic

Furthermore, we also extracted the most informative features (i.e. the words with high probabilities) that the classifier used to classify the tweet messages into categories of social presence. A few of the top of such words from each category are gathered and presented in Table 6 for illustration of the nature of words that were used by the classifier. An obvious distinction between the words of intimacy and immediacy is that the words in the immediacy category are more related to the needs and urgencies of the effected people, such as *blankets*, *ambulance*, *pregnant*, and so on, whereas the words in the intimacy category are more related to expressing concerns and moral support. The words in both none categories are more related to opinions and criticism as the words indicate *media*, *news*, *government*, *citizens*, *India* and so on.

In the classification of tweets for social presence, we have applied the classifier on the whole dataset, with the results depicted in Table 7. Out of 1.65 million tweets, 37% of tweets are classified as social presence containing intimacy and immediacy categories, whereas the remaining 63% of tweets belong to the none category. This is consistent with previous research that states that during disasters people post and share different types of messages containing suggestions, comments, criticism, etc., (Qu et al. 2011; Vieweg 2012) and also discussions about either media or government, frustrations or anger which could be not categorized as social presence. However, when it comes to retweets which constitute 80% of the dataset, the proportion of social presence retweets is higher with around 42% of retweets belong to intimacy and immediacy, as elaborated in Table 7.

Table 7 Social Presence Categories in the Total Dataset

Social Presence	Total tweets	Percentage	Retweets	Percentage
Intimacy	285,292	17.20%	226,697	18.49%
Immediacy	335,877	20.26%	291,516	23.78%
None	1,037,051	62.54%	707,885	57.73%
Total	1,658,220	100%	1,226,098	100%

The reason for having more retweets in the dataset is that retweet “RT” allow users to mention the original tweet author’s name and also acts as re-broadcasting or forwarding original tweets. Moreover, originators of those retweets are in general tweeting from the affected zone (Starbird and Palen 2010) and through this informal communication, people want to pass on the important information of disastrous events such as shelter information, coordination activities, and personal experiences that are useful to other affected people (Starbird and Palen 2010; Vieweg et al. 2010a).

6. Discussion and Conclusion

Emergent support groups evolve immediately after a disaster or after a crisis unfolds and it has been identified that these groups collectively work together to cope with the situation (Drabek and McEntire 2003). Nowadays, during emergency situations, people are interacting and communicating rather intensively on social media. In this research, we shed light on how the online social presence creates a sense of responsibility and common identity during disasters among the social media users. According to the social presence theory, people perceive intimacy and immediacy on mediated communications (Short et al. 1976). One of the reason behind the active participation of people on online social media during emergencies could be that they perceive the online social presence through the messages sent by the online participants and the interpretation of those messages by others. In order to detect the social presence based on message content apply the theory of social presence on social media users, we analyzed the Twitter messages and segregated them into based on intimacy and immediacy categories. First, we operationalized the concepts and then conducted manual content analysis to prepare the training sets for the classifier. Since the dataset is huge (around 1.65 million), therefore, we adopted a machine learning technique to automatically categorize the data into intimacy, immediacy and none categories. The results indicate that during emergencies, people are drawn to Twitter to fulfill their social need for connections as they feel the presence of others. Moreover, people, who feel higher levels of social presence continue to use and interact more on Twitter (Han et al. 2015). Thus, social presence plays a significant role hence despite the lack of face-to-face communication, people feel a sense of bonding due to the intimacy and immediacy felt for each other and therefore offer assistance by simply reading and retweeting the tweets. As explained previously, intimacy tweets are more towards standing by the people while showing moral support and coming forward to provide help on online, while immediacy tweets especially make people to perceive the vulnerable situation so people start actively taking part in relief activities.

Research has also revealed that during emergency situations, local people often retweet to ensure that emergency relevant information (Starbird and Palen 2010) is forwarded to others through collective action and collective behavior. Our total dataset consists of around 80% of retweets and more importantly we noticed that our analyzed dataset of social presence also contains a large number of retweets. The results showed that, the majority of the immediacy tweets conveying the needs and urgencies of affected people requesting for help. Especially in Chennai, people estimated the severity of the situation and gathered information that lead to the active participation and orchestration of needs and urgencies for affected individuals. According to our findings, even though a number of tweets fall into the “n” category, it is evident that social presence is important factor on social media to reach out people. Our results support the theory that online participants perceive others, experience a feeling of closeness, and when situations give rise for a sense of urgency people collectively get involved in the grassroots activities as well. For example, people offered accommodation, supplied and provided needs and necessities like food, water, clothes and activity participated in volunteer activities. Large private places like cinema theaters and wedding halls were opened for shelters. Another interesting observation is that people even shared their personal mobile numbers online to allow those in need to get in contact with them. Most importantly the hashtag #chennaimicro was introduced by one of the active volunteers and requested others online to use that particular hashtag for food supply only (Pandey 2015). We argue that social presence which is conveyed through a message content especially immediacy is highly valuable in emergency situations for emergence support groups to combat the situation.

Our results are consistent with previous findings (Al-Ghaith 2015; Xu et al. 2012), where social presence has been found to have a positive impact on usage of social media, as we also observed that social media has been actively used for communication and coordination during times of disasters. Moreover, in line with previous research results, affected individuals distribute coordination and online collective action (Vieweg et al. 2008) by relying on relevant information while verifying it and later on structure the information in a meaningful way (Starbird 2013). It is also apparent from our research results that people coordinate actively support activities through social media by retweeting, as indicated by the fact that 74% of our dataset consists of retweets.

Our research contributions are of twofold. Firstly, our research contributes to the theory of social presence in particular, and to social media analytics in general. So far, the theory of social presence was predominantly used in online learning and the research focus on social presence on social media was minimal and by conducting surveys only. Most importantly, on online learning, it was mentioned that messages have cues which evokes certain feelings but none measured them from content point of view. Unlike the previous studies, which were mostly survey based methods, our work focused more from message content point of view. Thus, we are among the first to analyze tweets using social presence as theoretical lens. Furthermore, our analysis presents new insights how social presence is formed during disasters through the use of social media messages.

Our findings have also some practical implications. We argue that information in immediacy tweets which reflect the needs and urgencies of affected people is important and valuable to emergency management agencies to reach out and save people's lives. We were able to illustrate how immediacy tweets can be automatically extracted using a machine learning approach. For emergency management agencies, those tweets contain valuable information to coordinate their support activity in close to real time.

7. Limitations and Future Research

Our study has certain limitations based on our data collection method. We exclusively used hashtags (#) to collect the data . This could have resulted in missing other relevant tweets in our dataset. We applied a machine learning approach with a relatively good accuracy for one of our classifier concepts (immediacy). However, the accuracy for the intimacy is rather comparatively low. During disasters, often social media data (e.g. twitter data) is produced in huge volumes, therefore our focus is to analyze the whole dataset instead of considering a small sample of the huge dataset. So we applied a technique that helps to analyze huge volumes of textual content for exploring theoretical concept such as social presence. However, there is also a certain likelihood that tweets fall in both categories (immediacy and intimacy), but the machine learning approach forces the tweets in one category only. Furthermore, the dataset used comprises only tweets from a flooding disaster and thus our findings may not be generalizable to other types of disasters. As part of our future research, we will analyze the tweets which have fallen into the “n” category. In the research at hand, we limited ourselves to only carrying out a content analysis of tweets from a social presence point of view. However there are other Twitter features that have yet to be explored through our research lens. These features, which include mentions, replies, and follower and following counts, may also exhibit social presence and collective intelligence which have yet to be explored.

8. References

Al-Ghaith, W. 2015. "Understanding Social Network Usage: Impact of Co-Presence, Intimacy, and Immediacy," *International Journal of Advanced Computer Science and Applications (IJACSA)* (6:8), p. 12.

Aragon, S. R. 2003. "Creating Social Presence in Online Environments," *New directions for adult and continuing education* (2003:100), pp. 57-68.

Argyle, M., and Dean, J. 1965. "Eye-Contact, Distance and Affiliation," *Sociometry* (28:3), pp. 289-304.

Biocca, F., Harms, C., and Burgoon, J. K. 2003. "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence* (12:5), pp. 456-480.

Bird, S. 2006. "Nltk: The Natural Language Toolkit," *Proceedings of the COLING/ACL on Interactive presentation sessions*: Association for Computational Linguistics, pp. 69-72.

Boyd, D., Golder, S., and Lotan, G. 2010. "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*: IEEE, pp. 1-10.

Cheong, F., and Cheong, C. 2011. "Social Media Data Mining: A Social Network Analysis of Tweets During the 2010-2011 Australian Floods," *In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*), pp. 46-46.

Cobb, S. C. 2009. "Social Presence and Online Learning: A Current View from a Research Perspective," *Journal of Interactive Online Learning* (8:3), pp. 241-254.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* (20:1), pp. 37-46.

Drabek, T. E., and McEntire, D. A. 2003. "Emergent Phenomena and the Sociology of Disaster: Lessons, Trends and Opportunities from the Research Literature," *Disaster Prevention and Management: An International Journal* (12:2), pp. 97-112.

Gunawardena, C. N. 1995. "Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences," *International journal of educational telecommunications* (1:2), pp. 147-166.

Gunawardena, C. N., and Zittle, F. J. 1997. "Social Presence as a Predictor of Satisfaction within a Computer-Mediated Conferencing Environment," *American journal of distance education* (11:3), pp. 8-26.

Han, S., Min, J., and Lee, H. 2015. "Antecedents of Social Presence and Gratification of Social Connection Needs in Sns: A Study of Twitter Users and Their Mobile and Non-Mobile Usage," *International Journal of Information Management* (35:4), pp. 459-471.

Hsieh, H.-F., and Shannon, S. E. 2005. "Three Approaches to Qualitative Content Analysis," *Qualitative health research* (15:9), pp. 1277-1288.
Hu, Y., Wood, J. F., Smith, V., and Westbrook, N. 2004. "Friendships through Im: Examining the Relationship between Instant Messaging and Intimacy," *Journal of Computer-Mediated Communication* (10:1), pp. 00-00.
Kaewkitipong, L., Chen, C. C., and Ractham, P. 2016. "A Community-Based Approach to Sharing Knowledge before, During, and after Crisis Events: A Case Study from Thailand," *Computers in Human Behavior* (54), pp. 653-666.
Kaplan, A. M., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business horizons* (53:1), pp. 59-68.
Kehrwald, B. 2008. "Understanding Social Presence in Text-Based Online Learning Environments," *Distance Education* (29:1), pp. 89-106.
Krippendorff, K. 1989. "Content Analysis," in *International Encyclopedia of Communication,* G.G. E. Barnouw, W. Schramm, T. L. Worth, & L. Gross (ed.). pp. 403-407.
Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. 2016. "Rapid Assessment of Disaster Damage Using Social Media Activity," *Science Advances* (2:3), p. e1500779.
Lim, J. S., Hwang, Y., Kim, S., and Biocca, F. A. 2015. "How Social Media Engagement Leads to Sports Channel Loyalty: Mediating Roles of Social Presence and Channel Commitment," *Computers in Human Behavior* (46), pp. 158-167.
Lombard, M., Snyder-Duch, J., and Bracken, C. C. 2002. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability," *Human communication research* (28:4), pp. 587-604.
Lowenthal, P. R. 2009. "The Evolution Andinfluence of Social Presence Theory on Online Learning," *Social Computing: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*), p. 113.
Mehbabian, A. 1967. "Orientation Behaviors and Nonverbal Attitude Communication1," *Journal of communication* (17:4), pp. 324-332.
Misra, T. 2015. "Chennai Floods: A Perfect Storm of Climate Change and Failed Planning." 2016, from http://www.citylab.com/weather/2015/12/chennai-flood-climate-change-urban-planning-cop21-paris/419349/
Morris, R. 1994. "Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations," *Journal of Management* (20:4), pp. 903-931.
Murthy, D., and Longwell, S. A. 2013. "Twitter and Disasters: The Uses of Twitter During the 2010 Pakistan Floods," *Information, Communication & Society* (16:6), pp. 837-855.
Narayanan, V., Arora, I., and Bhatia, A. 2013. "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model," in *Intelligent Data Engineering and Automated Learning–Ideal 2013*. Springer, pp. 194-201.
Oh, O., Kwon, K. H., and Rao, H. R. 2010. "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter During the Haiti Earthquake 2010," *ICIS*, p. 231.
Pandey, V. 2015. "How Chennai Residents Are Helping Flood Affected Via Social Media." from http://www.bbc.com/news/world-asia-india-34981360
Powers, D. M. 2011. "Evaluation: From Precision, Recall and F-Measure to Roc, Informedness, Markedness and Correlation," 2229-3981, School of Informatics and Engineering, Flinders University , Adelaide, Australia.
Qu, Y., Huang, C., Zhang, P., and Zhang, J. 2011. "Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake," *In: Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 25-34.
Risius, M., Akolk, F., and Beck, R. 2015. "Differential Emotions and the Stock Market-the Case of Company-Specific Trading," *In: Proceedings of the European Conference on Information Systems*.
Risius, M., and Beck, R. 2014. "You Reap What You Sow? How Knowledge Exchange Effectiveness Is Affected by Different Types of Communication in Enterprise Social

Media," *System Sciences (HICSS), 2014 47th Hawaii International Conference on*: IEEE, pp. 540-549.
Rourke, L., Anderson, T., Garrison, D. R., and Archer, W. 2001. "Methodological Issues in the Content Analysis of Computer Conference Transcripts," *International Journal of Artificial Intelligence in Education (IJAIED)* (12), pp. 8-22.
Sarcevic, A., Palen, L., White, J., Starbird, K., Bagdouri, M., and Anderson, K. 2012. "Beacons of Hope in Decentralized Coordination: Learning from on-the-Ground Medical Twitterers During the 2010 Haiti Earthquake," *Proceedings of the ACM 2012 conference on computer supported cooperative work*: ACM, pp. 47-56.
Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*.
Starbird, K. 2013. "Delivering Patients to Sacré Coeur: Collective Intelligence in Digital Volunteer Communities," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM, pp. 801-810.
Starbird, K., Dailey, D., Walker, A. H., Leschine, T. M., Pavia, R., and Bostrom, A. 2015. "Social Media, Public Participation, and the 2010 Bp Deepwater Horizon Oil Spill," *Human and Ecological Risk Assessment: An International Journal* (21:3), pp. 605-630.
Starbird, K., and Palen, L. 2010. *Pass It On?: Retweeting in Mass Emergency*. International Community on Information Systems for Crisis Response and Management.
Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. 2010. "Chatter on the Red: What Hazards Thret Reveals About the Social Life of Microblogges Information," *In: Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*, pp. 241-250.
Stemler, S. 2001. "An Overview of Content Analysis," *Practical assessment, research & evaluation* (7:17), pp. 137-146.
Sutton, J., Palen, L., and Shklovski, I. 2008. "Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires," *In: Proceedings of the 5th International ISCRAM Conference*: Washington, DC, pp. 624-632.
Tolstedt, B. E., and Stokes, J. P. 1983. "Relation of Verbal, Affective, and Physical Intimacy to Marital Satisfaction," *Journal of Counseling Psychology* (30:4), p. 573.
Toriumi, F., Sakaki, T., Shinoda, K., Kazama, K., Kurihara, S., and Noda, I. 2013. "Information Sharing on Twitter During the 2011 Catastrophic Earthquake," *In: Proceedings of the 22nd international conference on World Wide Web companion*, pp. 1025-1028.
Tu, C.-H. 2002. "The Impacts of Text-Based Cmc on Online Social Presence," *The journal of interactive online learning* (1:2), pp. 1-24.
Tu, C.-H., and McIsaac, M. 2002. "The Relationship of Social Presence and Interaction in Online Classes," *The American journal of distance education* (16:3), pp. 131-150.
Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., Schram, A., and Anderson, K. M. 2011. "Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency," *ICWSM*: Citeseer.
Vieweg, S., Hughes, A., Starbird, K., and Palen, L. 2010a. "Supporting Situational Awareness During Emergencies Using Microblogged Information," *SIGCHI 2010*).
Vieweg, S., Hughes, A. L. A., Starbird, K., and Palen, L. 2010b. "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*: ACM, pp. 1079-1088.
Vieweg, S., Palen, L., Liu, S. B., Hughes, A. L., and Sutton, J. 2008. "Collective Intelligence in Disaster: An Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shootings," *Proceedings of the information systems for crisis response and management conference (ISCRAM)*.
Vieweg, S. E. 2012. "Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications." University of Colorado at Boulder.
Walther, J. B. 1992. "Interpersonal Effects in Computer-Mediated Interaction a Relational Perspective," *Communication research* (19:1), pp. 52-90.

Walther, J. B., Loh, T., and Granka, L. 2005. "Let Me Count the Ways the Interchange of Verbal and Nonverbal Cues in Computer-Mediated and Face-to-Face Affinity," *Journal of language and social psychology* (24:1), pp. 36-65.
Wiener, M., and Mehrabian, A. 1968. *Language within Language: Immediacy, a Channel in Verbal Communication*. Ardent Media.
Woods, R. H., and Baker, J. D. 2004. "Interaction and Immediacy in Online Learning," *The International Review of Research in Open and Distributed Learning* (5:2).
Xu, C., Ryan, S., Prybutok, V., and Wen, C. 2012. "It Is Not for Fun: An Examination of Social Network Site Usage," *Information & Management* (49:5), pp. 210-217.
Yang, Y. 1999. "An Evaluation of Statistical Approaches to Text Categorization," *Information retrieval* (1:1-2), pp. 69-90.
Yang, Y., and Liu, X. 1999. "A Re-Examination of Text Categorization Methods," *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*: ACM, pp. 42-49.
Zhang, H., and Li, D. 2007. "Naive Bayes Text Classifier," *Granular Computing, 2007. GRC 2007. IEEE International Conference on*: IEEE, pp. 708-708.
Zook, M., Graham, M., Shelton, T., and Gorman, S. 2010. "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake," *World Medical & Health Policy* (2:2), pp. 7-33.

Paper 4

Mukkamala, AM & Beck, R (2018)

THE ROLE OF SOCIAL MEDIA FOR COLLECTIVE BEHAVIOR DEVELOPMENT IN RESPONSE TO NATURAL DISASTERS

In: Proceedings of the European Conference on information systems **(ECIS 2018)**

(Under review)

The Role of Social Media for Collective Behavior Development in Response to Natural Disasters

Research paper

Alivelu Mukkamala, Information Management Section, IT University of Copenhagen, Denmark, ammu@itu.dk

Roman Beck, Information Management Section, IT University of Copenhagen, Denmark, romb@itu.dk

Abstract

With the emergence of social media, user-generated content from people affected by the disaster gained significant importance. So far, research has focused on identifying categories and taxonomies of the types of information being shared among users during times of disasters. However, there is a lack of theorizing with regards to the dynamics of and relationships between the identified concepts. In our current research, we applied probabilistic topic modelling approach to identify topics from Chennai disaster Twitter data. We manually interpreted and further clustered the topics into generic categories and themes, and traced their development over the days of the disaster. Finally, we build a process model to explore an emerging phenomenon on social media during a disaster. We argue that the conditions/activities such as collective awareness, collective concern, collective empathy and support are necessary conditions for people to feel, respond, and act as forms of collective behaviour.

Keywords: Social media, Topic modelling, Disasters, Collective behavior

1. Introduction

In times of a crisis, fast and effective sharing of relevant information is of utmost importance to rescue the affected people. With the emergence of social media, disaster-related, user-generated content gained significant importance, as it contains real-time information about the disaster situations (Palen et al. 2009; Qu et al. 2011; Sutton et al. 2008). Such real-time information is not only valuable for emergency management agencies to organize disaster response, but also helps directly affected people in self-organize neighbourhood support (Kaufhold and Reuter 2016; Starbird and Palen 2011). Several research results elaborated already on the merits of social media during disasters (Starbird et al. 2015; Stieglitz et al. 2017; Vieweg et al. 2010a; Vieweg et al. 2010b), which makes social media interesting for emergency management agencies as an additional channel to communicate disaster-related information (Bruns et al. 2012; Ehnis and Bunker 2012; Reeder et al. 2014). Furthermore, it is argued that social media also helps the inter-organizational collaboration efforts of emergency organizations (Simon et al. 2015a) to gain a better understanding of the on-going emergency situation during times of disaster (Carter et al. 2014). Thus, one can state that the information shared by affected people over social media not only helps to create broader awareness about the situation, but also provides the foundation for collectively responding and organizing relief activities (Eismann et al. 2016; Pandey 2015).

Although there are several social media platforms (such as, Flicker, Instagram, or Facebook) due to the swiftness of information exchange, Twitter has garnered the scholarly attention (Chatfield et al. 2014) as the most effective channel used during times of extreme events (Mirbabaie et al. 2014). In several research papers, disaster-related tweets have already been analysed, for example, around the Iquique earthquake in Chile (Ahmed and Sargent 2014), floods in Australia, Pakistan, India and the Philippines (Ehnis and Bunker 2012) (Lee et al. 2013; Mukkamala and Beck 2017; Murthy and Longwell 2013), hurricane Sandy in the US (Mukkamala and Beck 2016; Shelton et al. 2014), the Boston bombings in US (Starbird et al. 2014), drug wars in Mexico (Monroy-Hernández et al. 2015), or terrorist attacks in Mumbai, India (Oh et al. 2011) or the more recent bombing attacks in Brussels (Mirbabaie and Zapatka 2017) and shootings in Munich (Bunker et al. 2017).

Disaster-related social media information typically not only comprises people's own observations and experiences, referred to as original sources, but also comprise re-tweets, URLs and @mentions from secondary sources (Starbird and Palen 2010). On Twitter, while communicating with others, users perceive the presence of others (Mukkamala and Beck 2017), play different roles (Lee et al. 2013; Reuter et al. 2013) while converging their behaviours to make sense of the situation at hand in order to cope with it (Bunker et al. 2017; Mirbabaie and Zapatka 2017; Stieglitz et al. 2017). Therefore, we argue that Twitter and it's user-generated content plays an important role within the area affected by a catastrophe but that it is also useful for learning more about the discourse in the general public ((Eismann et al. 2016; Takahashi et al. 2015).

Prior research has focused on identifying categories and taxonomies of the types of information being shared among users during times of disasters (Olteanu et al. 2015; Qu et al. 2011; Vieweg 2012). However, there is a lack of theorizing with regards to the dynamics of and relationships between the identified categories. Against this background, the main objective of our research is to identify the topics that are shared via social media during disasters to stimulate collective behaviour and to investigate how users' information sharing behaviour is changing as the event unfolds. In other words, we aim at understanding a process, comprising temporal and logical relationships between the topics that are shared on social media and corresponding activities during disasters.

To identify how collective behaviour develops, we have chosen an unsupervised method, topic modelling, where manual content analysis is not necessary to train the system, as it can be applied directly to extract the topics (Debortoli et al. 2016). Topic modelling uses a pattern recognition approach to discover the hidden patterns in data, a process which requires that the results are further analysed and interpreted by the researchers. This is the reason why this research is also referred to as the computational grounded theory method, as interpretation and coding of the underlying data analytics is necessary (Berente and Seidel 2014; Yu et al. 2011). In this research, we apply topic modelling as the computational grounded theory approach and analyse the topics that have been extracted from the data. Since the topics are data-driven, we manually interpreted and further clustered these topics into generic categories, and then further into themes, and traced their development over the period of the disaster. We focus on analysing the Twitter data of the Chennai flooding using topic modelling and subsequently code the results in order to conceptualise a collective behavioural process model.

The remainder of the paper is organised as follows. Section two provides the background of the literature review on Twitter during disasters. The main focus in section three is about methodology, which is twofold: first, we explain our data collection process and the topic modelling method we applied on the data. Second, we explain interpretation and coding of topics. In section four we explain about conceptualization of collective behaviour during disasters. Finally, in section five we discuss the results and the conclusions that can be drawn from them.

2. Literature Background

Because of its short message characteristics and low bandwidth requirements, Twitter is an efficient medium to share information during disasters in a timely manner (Chatfield et al. 2014; Java et al. 2007). Hence, it's usage during disasters has increased and the nature of its use (Reuter and Spielhofer 2016). During disasters, social media platforms are enhancing the opportunities and back channel communication among community members (Shklovski et al. 2008; Sutton et al. 2008). Information produced and shared by the people from the affected areas during disasters not only contain original information, such as eyewitness reports or other personal observations, but also contains re-tweets or links (Fuchs et al. 2013; Simon et al. 2015b; Starbird and Palen 2010; Starbird et al. 2010). Users also change their communication mode on Twitter during disasters by sharing more factual information (Toriumi et al. 2013), while converging on online (Bunker et al. 2017;

Subba and Bui 2017) yet the problem remains as not all the messages shared are authentic. The credibility of messages is often drawn into question (Murakami and Nasukawa 2012; Starbird et al. 2014). However, it has been observed that users more often question rumors and try to verify the trustworthiness of information shared via Twitter than, more credible and traditional information sources (Mendoza et al. 2010).

So far, different methods have been applied to analyse disaster-related Twitter data. The widely used methods are qualitative content analysis techniques, quantitative computational methods and social network analysis and to name a few (Landwehr and Carley 2014). Initially qualitative content analysis techniques are applied to examine the content of tweets. Through bottom-up approaches different categorization schemes and taxonomies were developed. For example, during the Yushu earthquake in China people shared different types of messages, which can be classified as information-, opinion-, emotion-, and action-related categories (Qu et al. 2011). Likewise, in the case of the Red River Floods and Oklahoma Grassfires in the US, most of the emergency-related messages are situation updates (Vieweg et al. 2010b). In particular, depending on the type of disaster, different types of situation updates were identified, to name a few, fire line information, hazard location, visibility, road conditions, or flood level and so on (Olteanu et al. 2015). Despite the fact that the amount and type of information shared on Twitter varies based on the disaster type, it is evident that there are certain commonalities (Olteanu et al. 2015). Moreover, it was also noticed that geo-location tags and other location references are of importance to extract the real-time information from the affected area (De Albuquerque et al. 2015; Graham et al. 2014; Mirbabaie et al. 2016; Mukkamala and Beck 2016). More recently, social network analyses (SNA) were also conducted to understand the information diffusion in the Twitter users' network, and also to understand influential actors and their different contributions in the form of information seeking, offering and sharing, which is also known as sense-making. Online users play different roles as information starters, amplifiers and transmitters in sense-making processes to face uncertain situations (Mirbabaie and Zapatka 2017).

Due to the volume of social media data generated during disasters, supervised (Imran et al. 2013; Mukkamala and Beck 2017; Sen et al. 2015; Verma et al. 2011) and unsupervised research methods (Imran and Castillo 2015; Kireyev et al. 2009; Lee et al. 2013) are applied to extract information and patterns. These methods are useful when one requires an analysis of the whole dataset instead of a sample data. To extract situation-awareness relevant tweets from disaster events in the US, supervised classifiers were trained to automatically categorize tweets into different situational-awareness classes and found that situational awareness information is often objective, impersonal, and formal (Verma et al. 2011). In order to automatically sort messages into different classes, such as, caution and advice, information source, donation, and causalities & damage, classifiers were trained to extract the intended information (Imran et al. 2013). To understand what makes people to share information on Twitter and go the extra length to help each other during disasters, tweets were analysed through social presence concepts and in addition, classifiers were built to extract the data (Mukkamala and Beck 2017). In contrast to supervised learning methods, techniques like topic modelling do not require pre-defined classes in order to cluster messages into groups with similar content, but are able to detect categories in an indicative way. Hence, these techniques can be compared to manual open coding approaches, in which codes and themes are suggested by the data, and not derived from literature and theory. An example of this approach is the study about Twitter use during the 2012 flooding in the Philippines, in which eight prevalent topics from the data have been extracted, such as, traffic updates, weather agency updates, suspension of classes, prayers & rescue, and relief goods & rescue (Lee et al. 2013). Overall, it can be noted that recent studies have successfully applied supervised classifiers as well as unsupervised topic modelling to analyse the content of the large amounts of social media messages being shared online during disasters.

~~3.~~ Methodology

3.1 Data Collection

The Twitter dataset used for this research was collected using the social media monitoring tool Radian6. As a specific case, we used the Chennai flooding that took place between the last week of November and first week of December 2015. Soon after the disaster unfolded, one was able to not only notice that Twitter was used for sharing information but also for coordinating the disaster relief activities by the affected people themselves.

In general, the hashtags, indicated by the prefix "#" make it easier to extract tweets of a particular event and also increases the search ability of intended information. Most importantly, hashtags are established and recommended by the organizations that are affiliated to a particular event. For example, the American Red Cross encouraged individuals to use the #Haiti hashtag after the 2010 Haitian earthquake to ask questions and to share the information about their relief efforts (Lovejoy et al. 2012). Therefore, collecting Twitter data using hashtags especially during events like disasters will provide an opportunity to collect tweets related to the event. The hashtags we have used in our research to collect tweets are #TNflood, #chennaiRains, #chennafloods #chennaiRainsHelp, #IndiaWithChennai and #chennaiMicro. We collected Twitter data from November 30th to December 16th 2015, which covers the entire period of Chennai floods, resulting in an initial dataset consisting of 1.65 million tweets. However, the data collected using Radain 6 only provide a subset of all attributes of a tweet such as tweet Id, tweet text and so on. Therefore, in order to overcome this limitation, we have re-fetched the Twitter data using the Twitter search (Twitter 2016) by using the tweet Id from the data provided by Radian 6 to get complete data of a tweet provided by Twitter for the whole dataset of 1.65 M tweets.

After an exploratory data analysis, we noticed that around 74% of the total dataset are re-tweets. When it comes to the language of tweets, the majority of the tweets were written in the English language, while some of the tweets used the local Tamil language. Therefore, as part of data pre-processing, we have filtered out the tweets written in languages other than English and also filtered the retweets so as to not distort the analysis due to the duplication of some of the data caused by retweets. As our primary goal is to uncover the hidden information patterns or the topics that emerged during the disaster, considering the original tweets for the analysis is fairly reasonable. After the pre-processing step, we ended up with 171,314 original tweet messages written only in English.

3.2 Topic Modelling

As mentioned before, our primary goal is to uncover the hidden topics that were emerged during Chennai floods, therefore we have chosen unsupervised methods to perform text analysis on the tweets. Under text analytics, unsupervised methods can learn the underlying text features from a text corpus by using clustering methods without explicitly imposing the need for specifying the categories of interest before performing the textual analysis (Grimmer and Stewart 2013). Topic modelling is a popular unsupervised clustering method for text analysis that provides a quantitative technique for the analysis of qualitative data. Although the automated computational analysis of textual data is constrained by a computers' limited ability to process the meaning of human language, it has shown to be a valid and reliable tool when fed with sufficiently large data sets (Halevy et al. 2009). Hence, statistical techniques like topic modelling are emerging as a novel and complimentary strategy of inquiry for researchers interested in analysing large collections of qualitative data in a scalable and reproducible manner.

Over the last couple of years, the Latent Dirichlet Allocation (LDA) has become a popular algorithm for unsupervised topic modelling in IS research (Debortoli et al. 2016). LDA is able to inductively identify topics running through a large collection of documents and to assign individual documents to these topics (Blei 2012; Blei et al. 2003). The idea behind LDA is rooted in the distributional hypothesis of linguistics (Firth 1957; Harris 1954), which posits that words that repeatedly co-occur in similar contexts (e.g., documents, paragraphs, sentences) tend to share meaning and, hence, can be used as proxies for describing the content of a text. For example, the co-occurrence of words like "temperature", "wind", "rain", and "sunshine" in a set of tweets can be interpreted as a marker for a common topic of these tweets, namely "weather". In contrast to hard classification or clustering methods, which assign each document to exactly one category, probabilistic topic modelling algorithms like LDA allow that documents belong to multiple categories (topics) with a varying degree of member- ship. So statistically speaking, LDA represents documents through a probability distribution over a fixed set of topics, and each topic, in turn, through a probability distribution over a fixed vocabulary of words. For example, a tweet may be 60% about the topic "weather", which, in turn, is represented by words such as the ones mentioned above, and 40% about the topic "New York City", which might contain words like "nyc", "manhatten", and "big apple". Grouping and aggregating the topic distributions of a large number by metadata (e.g., author, time, geography) allows to quantitatively summarize their content, detect differences in content between subgroups of documents, or to track and trace the development of topics over time.

As described in the previous section, we applied the LDA topic-modelling algorithm[1] to the dataset of 171,314 tweets collected during the Chennai Floods in order to identify topics that were shared during the event via Twitter and to track and trace the development of these topics over the course of the disaster. As the LDA algorithm can be sensitive to variations in its input parameters and the input data, we performed the analysis in multiple iterative cycles. We followed the guidelines provided by (Debortoli et al. 2016), paid special attention to the pre-processing of the data (e.g., tokenization, stop- word removal, lemmatization) and to finding the appropriate number of topics to be extracted from the data. The number of topics one wants to extract is the most crucial LDA parameter (Blei et al. 2003; Boyd-Graber et al. 2014). If too many topics are chosen, the algorithm finds a multitude of only minimally distinct topics (e.g., topics differ in writing style but not in content), where as choosing too few topics unnecessarily constrain the exploratory potential of the approach. We estimated models ranging between 10 and 100 topics and found 50 to be an appropriate number, as more fine-granular topic models (i.e., models with more topics) included an increasing number of near duplicate topics.

After automatically extracting the 50 topics from the dataset, we manually inspected, interpreted, la- belled, and grouped them into overarching themes. For that, we not only tried to make sense of the word probabilities of each topic (Figure 8), but also inspected the 50 most strongly associated tweets for each topic. Finally, we grouped the tweets and their topic probability distributions by days (30th November 2015 and 7th December 2015) and calculated the mean prevalence of each topic at each day in order to trace the development of topics over time. The procedures and results for interpretation, labelling, and grouping of the topics will be discussed in more detail in the following sections.

3.3 Interpretation and coding of topics

During the process of interpreting and coding of the identified topics, two researchers independently inspected all 50 topics generated for each day along with the relevant tweet content and subsequently compared and synthesized their results in order to minimize errors due to subjective biases. For most of the topics (around 80%) the researchers generated very similar labels or, in some cases, even identical labels; the remaining disagreements in interpretation were discussed among the researchers until consensus was reached.

In order to illustrate the procedure and results of this step of the analysis, we have chosen Topic 5, which accounts for about 2.1% of all the words written in our dataset. The topic comprises highly likely words like “road” (10.64%), “route” (2.84%), “avoid” (1.68%), “safe” (2.34%), “travel” (2.25%), “clear” (2.04%), and “water” (1.9%) (Figure 1). Tweets associated with this topic contain phrases like “Shastry Ngr Road leading to Besant Ngr Bus Stop Besant Ngr beach Adyar Bridge are safe to travel. no water logging” or “Latest Traffic Update - Loyola bridge - fully drained”. By co-examining the word distribution and the tweets associated to this topic we decided to label this topic “updates on routes”. People shared information about road

[1] We used cloud-based topic modelling tool www.MineMyText.com for pre-processing, analysing, and visualizing the dataset.

conditions to help and to create awareness regarding the routes that are safe or passable, and the routes are flooded and thus blocked. Such cautionary information and situation updates are vital during heavy rainfall and has been shared predominantly by people who witnessed the road situation themselves (primary sources).

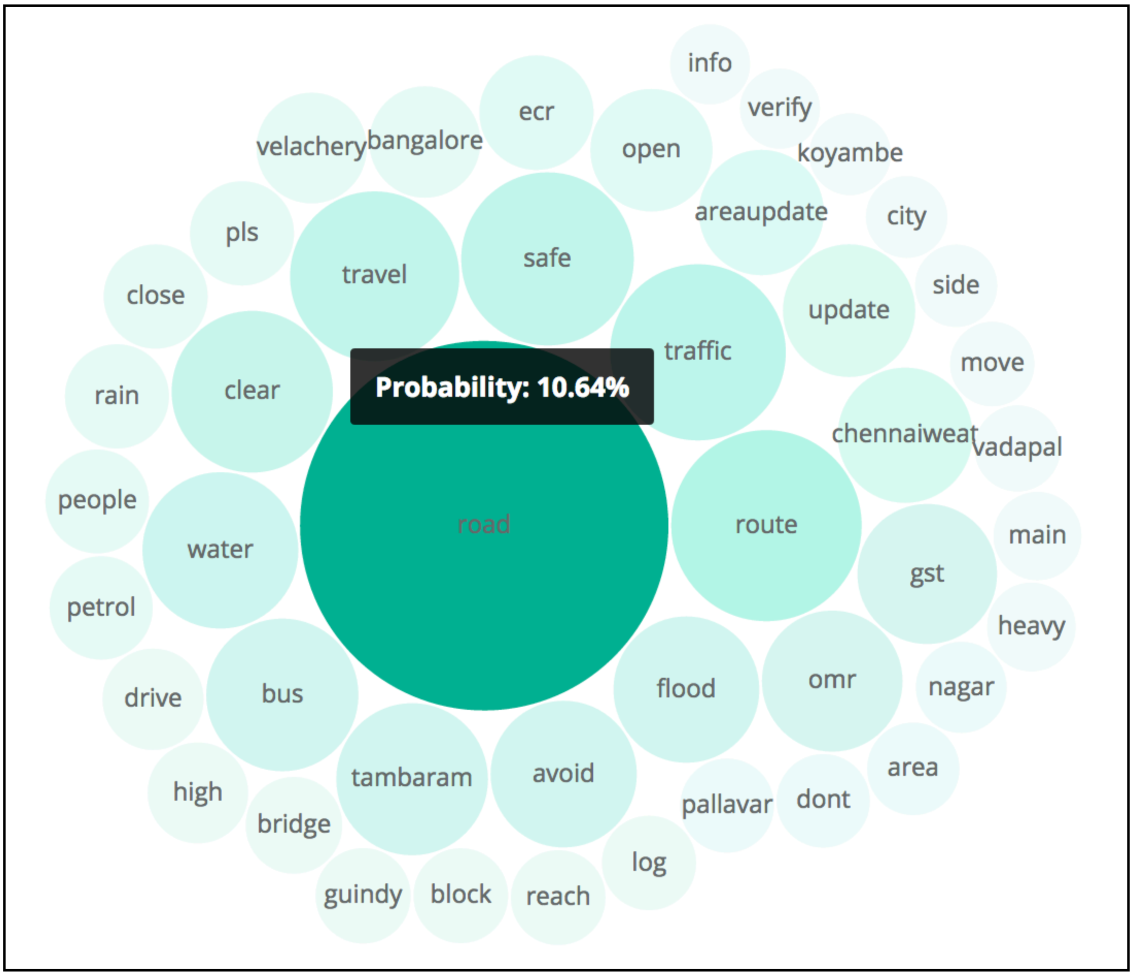

Figure 8 Bubble chart representing the most probable words for the topic labelled "Updates on routes" (volume and colour of bubbles represent the likelihood of words for this topic)

Table 8 provides some additional examples of topics, along with their most likely words, examples of associated tweets, and the researcher-generated labels. It can be seen that most of the topics are related to typical disaster situations, which reconfirms the findings reported in previous studies (Starbird and Palen 2010; Starbird et al. 2010; Vieweg et al. 2010a). Most of the topics represent vital concepts in situations of disaster, for example, "public places open for shelter", "emergency numbers", "sharing information of relief supplies", "stay safe", "request for help for/by stranded people", "rescue boat service", or "trying to reach friends and family".

We also found that certain topics were not directly related to the Chennai Floods (e.g., spam and advertising) or represented stylistic elements of tweets. For example, Topic 37 is almost exclusively made-up of purely hashtags, such as, #ChennaiFloods #ChennaiRescue #ChennaiVolunteer #ChennaiHelpRains #Helphennai #CuddaloreRains #CuddaloreRains. These topics do not provide any additional information as they contain only hashtags without any relevant content that is useful for the purpose of our study, therefore we excluded them in the subsequesnt analysis.

Topic Id	Most probable terms	Examples of strongly associated tweets	Topic label
12	Open, shelter, people, mall, cinema, door, mosque, hall, phoenix, food, accommodate, tonight, flood	"Sathyam Cinemas Royapettah AGS Cinemas Villivakkam Phoenix Mall Mayajaal OMR - open for people stuck in rain. #ChennaiRains"	Public places open for shelter
46	Update, friend, family, parent, reach, contact, pls, are, status, situation, unable	"People need urgent info of a friends Mom from #kotturpuram area. Stays in ground floor & no news since 24 hours #ChennaiRainsHelp"	Trying to reach friends and family
50	Helpline, number, emergency, army, navy, boat, rescue, contact, call	"#chennairains Indian Navy positioned wit rescue personnel wit boats at Gandhi Nagar Adyar. This is officially verified. Contact 04425394240"	Emergency numbers

Table 8 Selected examples of identified topics and related tweets

In order to identify how the information sharing behaviour on Twitter changes as the disaster situation unfolds, we plotted the mean prevalence of all topics over the days of the disaster, an example is shown in Figure 9.

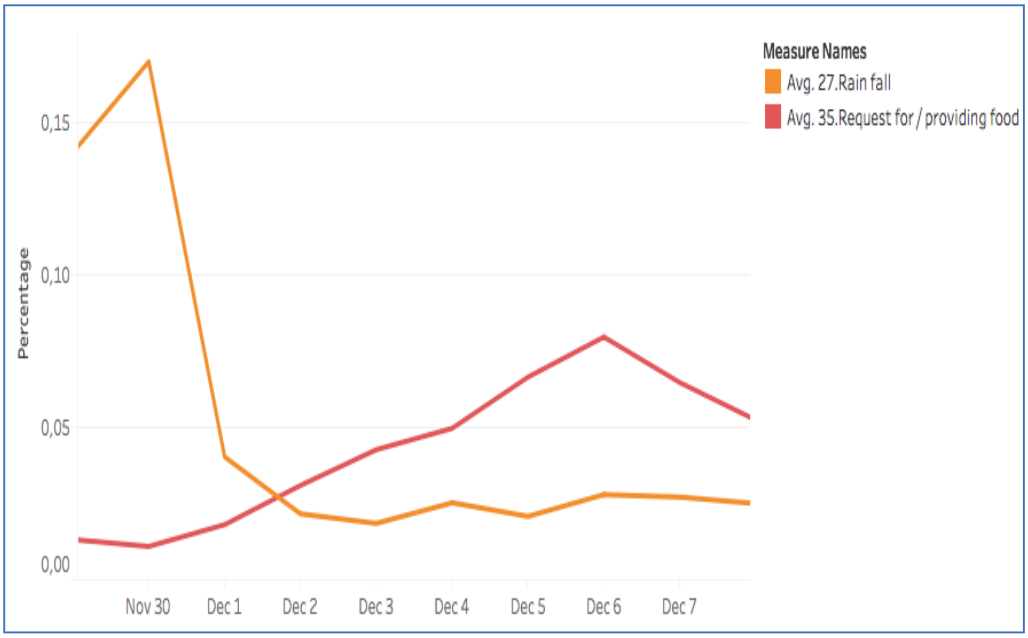

Figure 9 Mean prevalence of the example topics over the disaster period

In order to get a higher level of abstraction over the topics, we ordered all the 50 topics that were identified for a given day in the descending order of their probabilities of prevalence for that day. We then focused on the top ten most prevalent topics for each day and tried to manually cluster them into higher-order categories in order to get an overview of the most prevalent categories for each day. For example, we clustered Topic 5 (“updates on routes”), Topic 28 (“water level”), and Topic 42 ("public and private organizations opened / closed") into the category situation updates, as all of these topics are related to the sharing of information about the condition of in real-time. Similarly, we followed the same procedure to cluster all the topics into different categories.

Categories	Topics	Date
Initial information about disaster	Sharing news headlines; past and future rain fall	30th November
Situation updates	Updates on routes; updates about water levels; information of opened/closed public and private institutes	30th November
Criticism about insufficient attention	Self-criticism and criticizing the politicians; criticizing the media channels	30th November
Moral support	Praying; ask to stay safe	1st December
Preparations	Community information for shelter & food; leveraging social media for relief activities; emergency numbers	1st December
Criticism and control rumours	Criticising the media channels; trying to control rumours	1st December
Help request	Request for help for/by stranded people; request for food	2nd December
Offering help	Free mobile recharging; rescue boat service	2nd December
Self-organising support	Sharing information of relief supplies; trying to reach friends/family	3rd - 7th December
Active volunteerism	Coming forward to volunteer, Collecting information of Volunteers, Volunteering	3rd - 7th December

Table 9 Clustering of the identified topics into sub-categories

Finally, we aggregated all the most prevalent topics into ten categories as shown in Table 9. In contrast to prior research using topic modelling (Kireyev et al. 2009), we clustered the topics day wise to see the temporal evolution of topics during the disaster and also grouped the topics into categories to identify the emergence of the most prominent categories during the disaster timespan. Hence, for example, categories such as "criticism about insufficient attention" and "criticism and control rumours", have been found to be prominent over two consecutive days.

4. Research results

4.1 Conceptualisation of Collective Behaviour during Disasters

We plotted topics along the timeline of the disaster and further clustered the topics into categories for two reasons: First, to recognize which categories emerge at which point in time, thereby signalling relevance while the disaster unfolds, as shown in Table 9; second, to develop a process model that explains collective engagement of social media users in times of disasters. After aligning the categories day-wise, on a more abstract level, we derived the themes and defined their names. For example, the categories (a) initial information about disaster, (b) situation updates, and (c) criticism about insufficient attention explicitly provide information on rain updates, road updates etc., but implicitly this information dissemination creates awareness among social media users. Hence, we grouped the three categories under the theme "collective awareness". In the same manner, we created the rest of the themes, based on the prominent categories in each day. The main themes that emerged are: collective awareness, collective concern, collective empathy, and collective support. In the following sub-sections, we will describe them in more detail.

4.1.1 Collective awareness

Social media platforms allow for the creation of collective awareness among people. Especially information from the surrounding environment can lead to increased awareness among individuals and this aware- ness helps them to control their actions appropriately (Kellogg and Erickson 2002). Specifically, during disasters, online activity of social media users increases rapidly for information seeking and sharing. One of the uses of social media during a disaster is to "provide and receive disaster preparedness information" (Houston et al. 2015). On social media, people share news reports and various media sources to disseminate the information about anticipated events (Takahashi et al. 2015), and act as information brokers (Palen 2008) to create awareness. In the pre-event stage of a disaster people share secondary sources like media links or news report to create awareness of the event (Starbird et al. 2010), because these are credible sources of information about anticipated events. Most importantly, people share the caution and advice (Vieweg 2012) information regarding road closures and rising water levels to create awareness of the disaster. Since social media is a many-to-many medium, the awareness of disaster is created collectively among users. Information about closure of public and private organizations is also important for the people to gain the understanding about the severity of the situation, as the following examples illustrate:

- *"Watch out. Heavy spell on the way. Intense storms near #chennai coast. #chennaiweather #chennairains https://t.co/hgE3UskvNS"*
- *"All trains from Chennai Central and Chennai Egmore stands cancelled till 05 December 2015-12 PM announces Southern Railway #ChennaiFloods"*

In addition to initial information and situational updates, criticism also creates awareness of responsibility for the event and of "socio-political causes and implications of events" (Qu et al. 2011; Takahashi et al. 2015). Even though this information is often not directly related to the events of the disaster, users' opinion-related criticism nonetheless creates awareness. Moreover, frustration may arise among people because of a perceived lack of proper attention given to the disaster, either by media or by politicians:

- *"Its a shame that the so called national news channels arent even showing any news about #chennairains #shame @TimesNow @ndtv @abpnewstv"*

Collective awareness makes people more attentive towards the disaster. This attentiveness slowly leads to concern for people who are going to be affected by it. The concern may be shown in different ways on social media, which we will discuss in the following section.

4.1.2 Collective concern

Collective awareness of a disaster often leads to concern. During disasters, people start showing concern towards others, or among themselves, since everyone faces the same challenging conditions. The fears and concerns start when people notice worsening situations. On social media, the feeling of concern unfolds through expressing emotion-related (Qu et al. 2011) support for the community by showing moral support in the form of tweets (Takahashi et al. 2015). Being active online, people share up-to-date information, assist in preparation for the flood for those in need, share emergency contact details in advance and also share community information for shelter (Palen 2008). Examples for collective concern are:

- *“My prayers are with you Chennai. Be strong n tight until rains blows off. #chennairains ??”*
- *"Sathyam Cinemas Royapettah AGS Cinemas Villivakkam Phoenix Mall Mayajaal OMR -*

  open for people stuck in rain. #ChennaiRains"

Furthermore, concern makes people to look for other means to support the community. Previous re- search shows that, when people lose trust in government agencies, different social media platforms emerge and people use them as a source and medium to share information (Takahashi et al. 2015). However, there is also the issue of false and misleading information that spreads very quickly on social media and can trigger anxiety among people. Yet, “rumor is a form of collective behaviour surrounding information and psychology" (Oh et al. 2010). Users question the unauthenticated information more often than they question credible sources (Mendoza et al. 2010) to stop spreading the rumours. The anxiety could be controlled by reliable information with credible sources (Oh et al. 2010). Importantly, in order to control the panic, people try to control the flow of rumours and false information as conveyed by the tweet below:

- *"People forwarding pictures on Whatsapp of crocodiles in #Chennaifloods please stop! Noth- ing of the sort has happened. Stop spreading panic."*

4.1.3 **Collective empathy**

Disasters disrupt normalcy and make societies vulnerable. The actual impact caused during and immediately after a disaster is observable online and in real-time through social media messages. The messages reflect the situations and in these situations, firstly, people request evacuation and then food, which evoke emotional responses among the online community. The potent source of emotional response displayed by many to help their fellow Samaritans is the empathy they feel towards others, even strangers (Muller et al. 2014). Social media users are bonded by common identity and purposes. People feel “what is important to me is also important to others” (Kaewkitipong et al. 2016). The help requests (the examples are presented below) create a perceived need, which is an antecedent to feeling empathy and increases helping behaviour (Batson et al. 2007). Furthermore, people take it upon themselves to ask for help on behalf of those in need. In addition, or alternatively, they actively provide help in completing requests made by those in need and start participating in relief activities such as evacuation. Along with perceived need, valuing others’ welfare is also one of the antecedents of feeling empathy for people in need of help and increases the helping nature in individuals and produces their pro-social motivation (Batson et al. 2007).

- *"#ChennaiRainsHelp 5 families incl elderly kids stranded top floors no food water for 2 days. Urgent. 1/5 5th Cross st east CIT Nagar Ch35"*
- *"Need emergency boat cont number for Jafferkhanpet (Ashok Nagar). 6 month pregnant wom- en needs rescue #ChennaiRescue #ChennaiRainsHelp"*
- *"food packets needed: ashok nagar ALL HOMES raghavan colony 2nd cross st pls help #ashoknagar #ChennaiFloods #ChennaiRainsHelp"*

The motivation arising from empathy is taken to another level by actively participating in real-time relief coordinating activities, and are discussed in the following section.

4.1.4 Collective support

Positive cooperative social interactions on social media increase the real-time activities related to helping the affected individuals. The social practices, such as, gathering, reasoning, curating, stewarding and orchestrating, and acting are performed by different stakeholders during a disaster: the community as a whole, emergency responders, and affected people (Büscher et al. 2014). In order to support the victims, social structures emerge within online communities and people create their own ways to collaborate and cooperate using their own rules and resources (Kaewkitipong et al. 2016). Empathy and responsibility are the two main intrinsic motivators for volunteering (McDonald et al. 2015). During disasters, on social media a few active volunteers gather and collect the information of needs and urgencies pertaining to the affected individuals. Subsequently, they mobilize supplies and volunteers to complete the help requests of those in need. Volunteers self-organize themselves by sharing and categorizing the information regarding the location of the request and the collection and distribution of the necessities involved in completing the request. While coordinating the activities among themselves, volunteers take active, actionable, and real-time response and relief activities (Takahashi et al. 2015) (Takahashi et al. 2015), as illustrated by the following messages:

- *"#VERIFIED truck full of essential Supplies reaching North Chennai. Pls CT Hijaz - +91 9995000009 for distribution and place. #ChennaiMicro"*
- *"Have 25 water bottles 25 biscuit packets. Can provide them to Ambattur Mogappair Anna Nagar. Ping ASAP. #ChennaiMicro #ChennaiRainsHelp"*
- *"10 people with car ready to transport food and supplies for the people in need Contact 9884400543 #ChennaiRainsHelp @ActorMadhavan"*
- *"Perambur food is ready for 500 ppl. Contact 9884386734 any volunteers avilable in north Madras plz help #ChennaiMicro @iamVikramPrabhu"*

4.1.5 From collective awareness to collective support

So far, we have explained our method to extract topics, discussed how we grouped topics into categories, and explained the main themes subsuming these categories. In this section, we present a process model to explain how collective behaviour emerges through a sequence of activities. In general, a process model explains an emerging phenomenon and its outcomes with a set of sequential activities. Moreover, the observations enhance our understanding of the sequence of activities leading up to the outcomes (Crowston 2000). Although a process model typically describes necessary and not sufficient conditions, it still provides a form of explanation which combines and strings together conditions and activities (Markus and Robey 1988).

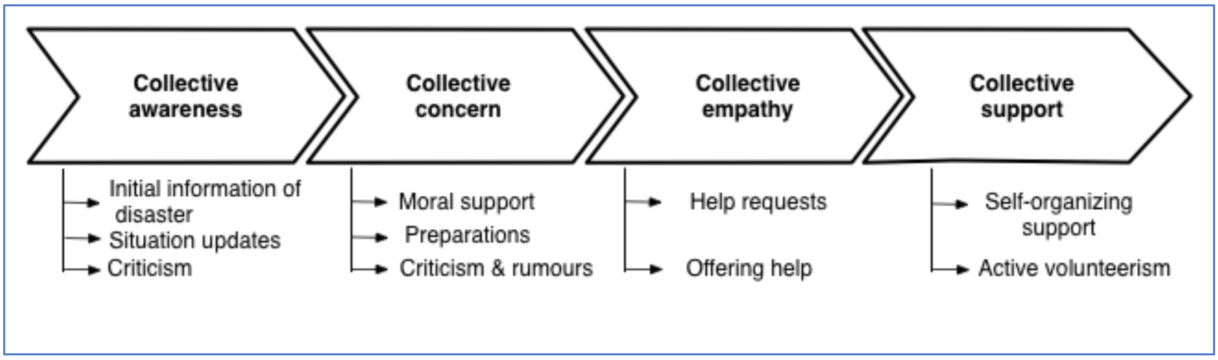

Figure 10 Change in collective behaviour during times of disaster

The process model depicted in Figure 3 illustrates the phases of collective behaviour as discovered in our research: 1) collective awareness, 2) collective concern, 3) collective empathy, and 4) collective support. The model illustrates how information sharing activities on social media turn into actions as the disaster unfolds. As mentioned above, we argue that the conditions/activities (collective awareness, collective concern, collective empathy and support) are necessary conditions for people to feel, respond, and act as forms of collective behaviour. The model provides an explanation for the people's actions and how their altruistic behaviour unfolds. Social media not only facilitates reaching and connecting with locally affected people, but also connects them with wider audiences who are not directly affected by the disaster ((Eismann et al. 2016; Takahashi et al. 2015). On social media, users' interactions, coordination, and coping mechanisms create dynamic and spontaneous social structures during disasters. Self-organized volunteer activities emerge because of a lack of top-down structures to guide them either in the information exchange process or in relief coordination activities. Hence, the by-product of social media's embedded features is bottom-up self-organized communities (Kaewkitipong et al. 2016).

5. Discussion

In recent years, a new form of digital volunteering emerged during disasters. Particularly, social media is facilitating the new forms of volunteerism during disasters (Starbird and Palen 2011) where groups of people come together to take active part in disaster response activities (Quarantelli and Dynes 1977). This collective behaviour of self-organized digital volunteerism during disasters (Kaufhold and Reuter 2016; Reuter et al. 2013; Starbird and Palen 2011) was explained through the theoretical lens of Krep's framework (Kreps 1984) by conducting a literature review (Eismann et al. 2016). However, the study did not explicitly focus on how collective behaviour on social media unfolds over time. We also noticed that the virtually formed groups not only take part on online response activities to coordinate but also actively take part to collaborate on the ground in relief activities. Our study aimed to identify what type of information is shared via social media during disasters and to investigate how users' information sharing behaviour is changing as the disaster event unfolds. We intended to build a process model to explore an emerging phenomenon on social media during a disaster, considering the flood in Chennai as a reference case. For this purpose, we applied topic modelling as a method to understand the emerging topics in a disaster. Later, we interpreted, coded, and finally grouped the topics into categories in a chronological order. Subsequently, we derived the four main themes from our analysis: collective awareness, collective concern, collective empathy and collective support, which are the forms of collective behaviour.

In our study, we have identified the sequence of activities in the process of unfolding collective behaviour. In contrast to previous research (Eismann et al. 2016), we applied the unsupervised machine leaning approach to inductively extract the topics from disaster related social media data, to understand the evolving phenomena. The topics found in our research are similar to the topics identified in the recent research (Lee et al. 2013) that applied the topic modelling method on Twitter data of 2012 flooding in the Philippines. However, in contrast to their work, in our study, we analysed the topics that emerged over time, aligned and clustered them to understand the process with which collective behaviour unfolds. However, to the best of our knowledge, there is no research that explored the emergence of collective behavior over the temporal dimension during a disaster.

The collective behaviour process has enhanced our understanding of the emerging phenomena on social media during a disaster situation. The process or sequence of categories explains firstly, what types of information people share on social media over the period of a disaster. Secondly, why and how peoples' information sharing behaviour changes and further leads to collaborative and cooperative activities as the event unfolds. In addition, the discovered process reveals situations and performed activities by affected as well as witnessing individuals. Moreover, there is a possibility to infer latent behavioural patterns through the activities performed or perceived by the individuals on social media during the disaster. To some extent the exploration of themes through categories enhanced our understanding of those latent behavioural patterns, which is the sequence starting with collective awareness, concern, and empathy, finally leading to collective support. Social media is enhancing and influencing communities resilience in a positive way (Kaufhold and Reuter 2016).

In our research, we shed light on the causal mechanisms through the process model we developed to show how collective behaviour progresses in virtual groups on social media (Reuter et al. 2013). The awareness about a disaster in an early phase makes social media users attentive in sharing and receiving information about impending disaster. The collective awareness triggers or causes concern for "to be affected individuals". Because of the concern, users start taking part in online preparation activities such as sharing emergency numbers in advance, list information of shelters and so on. This concern further leads to empathy towards the people who are in need of help. Feeling empathy motivates individuals to take active part in relief coordinating activities both on the online and offline platforms. We also noticed that as the disaster unfolds the needs of affected people change and so do the message types (information types) that are shared on social media.

It is evident from extant literature that during disasters groups of people come together to take active part in disaster response activities (Quarantelli and Dynes 1977) . We have been able to show how collective behaviour progresses in virtual groups on social media in order to help the affected people during disasters and also our results are consistent with previous research as social media facilitates the synergies between virtual and emergent volunteer groups (Reuter et al. 2013) while self-organizing (Kaufhold and Reuter 2016) . However, we also argue that virtually formed groups not only take part on online response activities to coordinate but also actively take part in the ground relief activities.

Our contributions are twofold. Firstly, to the theory of collective behaviour during disasters by explaining how it evolves through information sharing behaviour of users on social media. We especially noticed that from our case study, when people are challenged by the natural calamities, people come forward to self-organize themselves by sharing the disaster-related information such as collection and distribution of necessities and also to coordinate and volunteer themselves to take an active part in actionable, and real-time response relief activities. Secondly, our contribution is to the area of process theories, which basically explains the causal relationships and emerging phenomena. From our case study, we have illustrated that using unsupervised topic modelling approach on textual content, one could identify and uncover the prominent topics and categories that are hidden in the information to transform into a process that reveal behaviour patterns of the people involved in generating the textual content. Moreover, since the process visualizes a sequence of activities, the disaster management officials can make use of information that is relevant for their tactical decision making during disasters.

This study has a limitation as we developed a process model based exclusively on the data extracted from a flood situation. This affects generalizability of the findings since different types of disaster (flooding, earthquake, hurricane) unfold in different ways and each of them has specific characteristics both in terms of magnitude of the damage they can cause to the people and also in terms of duration of disasters.

In our future work, we would like to take this work further by supplementing with supervised machine-learning techniques on the textual content of Twitter data to do an in-depth analysis of the emerged collective behaviour. For example, we plan to develop domain specific models for text classification for each of the process steps: collective awareness, collective concern, collective empathy, collective support and use manual content analysis to code the tweets for these models and apply supervised machine-learning algorithms to analyse the textual content for a more in-depth analysis.

6. References

Ahmed, A., and J. Sargent. (2014). "Analysis of Post-Crisis Twitter Communication: A Study of the Iquique, Chile Earthquake," In: Proceedings of the Australasian Conference on Information Systems ACIS.

Batson, C. D., J. H. Eklund, V. L. Chermok, J. L. Hoyt, and B. G. Ortiz. (2007). "An Additional Antecedent of Empathic Concern: Valuing the Welfare of the Person in Need," Journal of personality and social psychology (93:1), p. 65.

Berente, N., and S. Seidel. (2014). "Big Data & Inductive Theory Development: Towards Computational Grounded Theory?,").

Blei, D. M. (2012). "Probabilistic Topic Models," Communications of the ACM (55:4), pp. 77-84.

Blei, D. M., A. Y. Ng, and M. I. Jordan. (2003). "Latent Dirichlet Allocation," Journal of machine Learning research (3:Jan), pp. 993-1022.

Boyd-Graber, J., D. Mimno, and D. Newman. (2014). "Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements," Handbook of Mixed Membership Models and Their Applications), pp. 3-34.

Bruns, A., J. E. Burgess, K. Crawford, and F. Shaw. (2012). "# Qldfloods and@ Qpsmedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods,").

Bunker, D., M. Mirbabaie, and S. Stieglitz. (2017). "Convergence Behaviour of Bystanders: An Analysis of 2016 Munich Shooting Twitter Crisis Communication," in: Australasian Conference on Information Systems. Hobart, Australia.

Büscher, M., M. Liegl, and V. Thomas. (2014). "Collective Intelligence in Crises," in Social Collective Intelligence. Springer, pp. 243-265.

Carter, L., J. B. Thatcher, and R. Wright. (2014). "Social Media and Emergency Management: Exploring State and Local Tweets," 2014 47th Hawaii International Conference on System Sciences: IEEE, pp. 1968-1977.

Chatfield, A. T., H. J. Scholl, and U. Brajawidagda. (2014). "# Sandy Tweets: Citizens' Co-Production of Time-Critical Information During an Unfolding Catastrophe," 2014 47th Hawaii International Conference on System Sciences: IEEE, pp. 1947-1957.

Crowston, K. (2000). "Process as Theory in Information Systems Research," in Organizational and Social Perspectives on Information Technology. Springer, pp. 149-164.

De Albuquerque, J. P., B. Herfort, A. Brenning, and A. Zipf. (2015). "A Geographic Approach for Combining Social Media and Authoritative Data Towards Identifying

Useful Information for Disaster Management," International Journal of Geographical Information Science (29:4), pp. 667-689.
Debortoli, S., O. Müller, I. Junglas, and J. vom Brocke. (2016). "Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial," Communications of the Association for Information Systems (CAIS)).
Ehnis, C., and D. Bunker. (2012). "Social Media in Disaster Response: Queensland Police Service-Public Engagement During the 2011 Floods," ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian Conference on Information Systems 2012: ACIS, pp. 1-10.
Eismann, K., O. Posegga, and K. Fischbach. (2016). "Collective Behaviour, Social Media, and Disasters: A Systematic Literature Review," in: European Conference on Information Systems.
Firth, J. R. (1957). "{a Synopsis of Linguistic Theory, 1930-1955},").
Fuchs, G., N. Andrienko, G. Andrienko, S. Bothe, and H. Stange. (2013). "Tracing the German Centennial Flood in the Stream of Tweets: First Lessons Learned," Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information: ACM, pp. 31-38.
Graham, M., S. A. Hale, and D. Gaffney. (2014). "Where in the World Are You? Geolocation and Language Identification in Twitter," The Professional Geographer (66:4), pp. 568-578.
Grimmer, J., and B. M. Stewart. (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," Political analysis (21:3), pp. 267-297.
Halevy, A., P. Norvig, and F. Pereira. (2009). "The Unreasonable Effectiveness of Data," IEEE Intelligent Systems (24:2), pp. 8-12.
Harris, Z. S. (1954). "Distributional Structure," Word (10:2-3), pp. 146-162.
Houston, J. B., J. Hawthorne, M. F. Perreault, E. H. Park, M. Goldstein Hode, M. R. Halliwell, S. E. Turner McGowen, R. Davis, S. Vaid, J. A. McElderry, and S. A. Griffith. (2015). "Social Media and Disasters: A Functional Framework for Social Media Use in Disaster Planning, Response, and Research," Disasters (39), pp. 1-22.
Imran, M., and C. Castillo. (2015). "Towards a Data-Driven Approach to Identify Crisis-Related Topics in Social Media Streams," Proceedings of the 24th International Conference on World Wide Web: ACM, pp. 1205-1210.
Imran, M., S. M. Elbassuoni, C. Castillo, F. Diaz, and P. Meier. (2013). "Extracting Information Nuggets from Disaster-Related Messages in Social Media," Proc. of ISCRAM, Baden-Baden, Germany).
Java, A., X. Song, T. Finin, and B. Tseng. (2007). "Why We Twitter: Understanding Microblogging Usage and Communities," In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis: ACM, pp. 56-65.
Kaewkitipong, L., C. C. Chen, and P. Ractham. (2016). "A Community-Based Approach to Sharing Knowledge before, During, and after Crisis Events: A Case Study from Thailand," Computers in Human Behavior (54), pp. 653-666.
Kaufhold, M.-A., and C. Reuter. (2016). "The Self-Organization of Digital Volunteers across Social Media: The Case of the 2013 European Floods in Germany," Journal of Homeland Security and Emergency Management (13:1), pp. 137-166.
Kellogg, W., and T. Erickson. (2002). "Social Translucence, Collective Awareness, and the Emergence of Place," Proc. of CSCW: Citeseer.
Kireyev, K., L. Palen, and K. Anderson. (2009). "Applications of Topics Models to Analysis of Disaster-Related Twitter Data," NIPS Workshop on Applications for Topic Models: Text and Beyond: Canada: Whistler.
Kreps, G. A. (1984). "Sociological Inquiry and Disaster Research," Annual review of sociology), pp. 309-330.
Landwehr, P. M., and K. M. Carley. (2014). "Social Media in Disaster Relief," in Data Mining and Knowledge Discovery for Big Data. Springer, pp. 225-257.

Lee, J. B., M. Ybañez, M. M. De Leon, and M. R. E. Estuar. (2013). "Understanding the Behavior of Filipino Twitter Users During Disaster," GSTF Journal on Computing (JoC) (3:2), p. 37.

Lovejoy, K., R. D. Waters, and G. D. Saxton. (2012). "Engaging Stakeholders through Twitter: How Nonprofit Organizations Are Getting More out of 140 Characters or Less," Public Relations Review (38:2), pp. 313-318.

Markus, M. L., and D. Robey. (1988). "Information Technology and Organizational Change: Causal Structure in Theory and Research," Management science (34:5), pp. 583-598.

McDonald, L. M., M. Creber, H. Sun, and L. Sonn. (2015). "Developing Public Disaster Communication for Volunteer Recruitment: Understanding Volunteer Motivations,").

Mendoza, M., B. Poblete, and C. Castillo. (2010). "Twitter under Crisis: Can We Trust What We Rt?," Proceedings of the first workshop on social media analytics: ACM, pp. 71-79.

Mirbabaie, M., C. Ehnis, S. Stieglitz, and D. Bunker. (2014). "Communication Roles in Public Events," Working Conference on Information Systems and Organizations: Springer, pp. 207-218.

Mirbabaie, M., S. Stieglitz, and S. Volkeri. (2016). "Volunteered Geographic Information and Its Implications for Disaster Management," System Sciences (HICSS), 2016 49th Hawaii International Conference on: IEEE, pp. 207-216.

Mirbabaie, M., and E. Zapatka. (2017). "Sensemaking in Social Media Crisis Communication–a Case Study on the Brussels Bombings in 2016,").

Monroy-Hernández, A., E. Kiciman, D. Boyd, and S. Counts. (2015). "Narcotweets: Social Media in Wartime," arXiv preprint arXiv:1507.01290).

Mukkamala, A., and R. Beck. (2016). "Enhancing Disaster Management through Social Media Analytics to Develop Situation Awareness What Can Be Learned from Twitter Messages About Hurricane Sandy?," in: The Pacific Asia Conference on Information Systems (PACIS) AIS Electronic Library (AISeL), p. 165.

Mukkamala, A., and R. Beck. (2017). "Presence of Social Presence During Disasters," Pacific Asia Conference on Information Syatems.

Muller, A. R., M. D. Pfarrer, and L. M. Little. (2014). "A Theory of Collective Empathy in Corporate Philanthropy Decisions," Academy of Management Review (39:1), pp. 1-21.

Murakami, A., and T. Nasukawa. (2012). "Tweeting About the Tsunami?: Mining Twitter for Information on the Tohoku Earthquake and Tsunami," In: Proceedings of the 21st international conference companion on World Wide Web: ACM, pp. 709-710.

Murthy, D., and S. A. Longwell. (2013). "Twitter and Disasters: The Uses of Twitter During the 2010 Pakistan Floods," Information, Communication & Society (16:6), pp. 837-855.

Oh, O., M. Agrawal, and H. R. Rao. (2011). "Information Control and Terrorism: Tracking the Mumbai Terrorist Attack through Twitter," Information Systems Frontiers (13:1), pp. 33-43.

Oh, O., K. H. Kwon, and H. R. Rao. (2010). "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter During the Haiti Earthquake 2010," ICIS, p. 231.

Olteanu, A., S. Vieweg, and C. Castillo. (2015). "What to Expect When the Unexpected Happens: Social Media Communications across Crises," Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing: ACM, pp. 994-1009.

Palen, L. (2008). "Online Social Media in Crisis Events," Educause Quarterly (31:3), pp. 76-78.

Palen, L., S. Vieweg, S. B. Liu, and A. L. Hughes. (2009). "Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event," Social Science Computer Review (27), pp. 467-480.

Pandey, V. (2015). "How Chennai Residents Are Helping Flood Affected Via Social Media." from http://www.bbc.com/news/world-asia-india-34981360

Qu, Y., C. Huang, P. Zhang, and J. Zhang. (2011). "Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake," In: Proceedings of the ACM 2011 conference on Computer supported cooperative work, pp. 25-34.

Quarantelli, E. L., and R. R. Dynes. (1977). "Response to Social Crisis and Disaster," Annual review of sociology), pp. 23-49.
Reeder, H. T., T. H. McCormick, and E. Spiro. (2014). "Online Information Behaviors During Disaster Events: Roles, Routines, and Reactions," Center for Statistics and the Social Sciences, University of Washington, Working paper:144).
Reuter, C., O. Heger, and V. Pipek. (2013). "Combining Real and Virtual Volunteers through Social Media," ISCRAM.
Reuter, C., and T. Spielhofer. (2016). "Towards Social Resilience: A Quantitative and Qualitative Survey on Citizens' Perception of Social Media in Emergencies in Europe," Technological Forecasting and Social Change).
Sen, A., K. Rudra, and S. Ghosh. (2015). "Extracting Situational Awareness from Microblogs During Disaster Events," 2015 7th International Conference on Communication Systems and Networks (COMSNETS): IEEE, pp. 1-6.
Shelton, T., A. Poorthuis, M. Graham, and M. Zook. (2014). "Mapping the Data Shadows of Hurricane Sandy: Uncovering the Sociospatial Dimensions of 'Big Data'," Geoforum (52), pp. 167-179.
Shklovski, I., L. Palen, and J. Sutton. (2008). "Finding Community through Information and Communication Technology During Disaster Events," in: Proceedings of the 2008 ACM conference on Computer supported cooperative work. ACM, pp. 1-10.
Simon, T., L. Aharonson-Daniel, M. El-Hadid, and B. Adini. (2015a). "Cross-Border Emergency Coordination and Communications Using Social Media: Developing a Joint Israeli-Jordanian Standard Operating Procedure for Leveraging Social Media in Emergencies," International Journal of Emergency Management (11:2), pp. 169-190.
Simon, T., A. Goldberg, and B. Adini. (2015b). "Socializing in Emergencies—a Review of the Use of Social Media in Emergency Situations," International Journal of Information Management (35:5), pp. 609-619.
Starbird, K., D. Dailey, A. H. Walker, T. M. Leschine, R. Pavia, and A. Bostrom. (2015). "Social Media, Public Participation, and the 2010 Bp Deepwater Horizon Oil Spill," Human and Ecological Risk Assessment: An International Journal (21:3), pp. 605-630.
Starbird, K., J. Maddock, M. Orand, P. Achterman, and R. M. Mason. (2014). "Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing," iConference 2014 Proceedings).
Starbird, K., and L. Palen. (2010). Pass It On?: Retweeting in Mass Emergency. International Community on Information Systems for Crisis Response and Management.
Starbird, K., and L. Palen. (2011). "Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: ACM, pp. 1071-1080.
Starbird, K., L. Palen, A. L. Hughes, and S. Vieweg. (2010). "Chatter on the Red: What Hazards Thret Reveals About the Social Life of Microblogges Information," In: Proceedings of the Conference on Computer Supported Cooperative Work (CSCW), pp. 241-250.
Stieglitz, S., M. Mirbabaie, L. Schwenner, J. Marx, J. Lehr, and F. Brünker. (2017). "Sensemaking and Communication Roles in Social Media Crisis Communication,").
Subba, R., and T. Bui. (2017). "Online Convergence Behavior, Social Media Communications and Crisis Response: An Empirical Study of the 2015 Nepal Earthquake Police Twitter Project," Proceedings of the 50th Hawaii International Conference on System Sciences.
Sutton, J., L. Palen, and I. Shklovski. (2008). "Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires," In: Proceedings of the 5th International ISCRAM Conference: Washington, DC, pp. 624-632.
Takahashi, B., E. C. Tandoc, and C. Carmichael. (2015). "Communicating on Twitter During a Disaster: An Analysis of Tweets During Typhoon Haiyan in the Philippines," Computers in Human Behavior (50), pp. 392-398.
Toriumi, F., T. Sakaki, K. Shinoda, K. Kazama, S. Kurihara, and I. Noda. (2013). "Information Sharing on Twitter During the 2011 Catastrophic Earthquake," In:

Proceedings of the 22nd international conference on World Wide Web companion, pp. 1025-1028.
Twitter. (2016). "Twitter Rest Apis." from https://dev.twitter.com/rest/public
Verma, S., S. Vieweg, W. J. Corvey, L. Palen, J. H. Martin, M. Palmer, A. Schram, and K. M. Anderson. (2011). "Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency," ICWSM: Citeseer.
Vieweg, S., A. Hughes, K. Starbird, and L. Palen. (2010a). "Supporting Situational Awareness During Emergencies Using Microblogged Information," SIGCHI 2010).
Vieweg, S., A. L. A. Hughes, K. Starbird, and L. Palen. (2010b). "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: ACM, pp. 1079-1088.
Vieweg, S. E. (2012). "Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications." University of Colorado at Boulder.
Yu, C. H., A. Jannasch-Pennell, and S. DiGangi. (2011). "Compatibility between Text Mining and Qualitative Research in the Perspectives of Grounded Theory, Content Analysis, and Reliability," The Qualitative Report (16:3), p. 730.

Paper 5

Mukkamala, AM & Beck, R (2017)

Social media for disaster situations: Methods, opportunities and challenges

In: Proceedings of IEEE Global Humanitarian Technology Conference **(GHTC 2017)** San Jose, California, USA

Social Media for Disaster Situations

Methods, Opportunities and Challenges

Alivelu Mukkamala, Information Management Section, IT University of Copenhagen, Denmark, ammu@itu.dk

Roman Beck, Information Management Section, IT University of Copenhagen, Denmark, romb@itu.dk

Abstract— *During disasters people start using social media as a platform to share and disseminate real-time disaster information to wider audiences. In order to understand the type of information that is being shared during disasters and how communities are using the technologies to respond to the disasters, we analyze two-different case studies on natural disasters using Twitter as a platform for gathering and sharing information. In both the case studies, we applied different content analysis methods, both manual and automated, to analyze the valuable information from the user-generated content produced during disaster situations. Based on our findings, we argue that social media platforms are facilitating collective level situation awareness among people and valuable information for disaster management agencies. However, in order to integrate social media in organizational work routines and processes, understanding the opportunities along with challenges is a key.*

Keywords—Disasters; Social Media; Twitter

1. Introduction

Humankind is prone to different types of disasters such as earthquakes, floods, epidemics, etc., and had to deal with devastating consequences of disasters in our history. Disasters are events that disrupt a communities' normal functioning and have an impact on people's lives, economy and environment. When compared to epidemic diseases or economic crises, both natural and man-made disasters occur suddenly and require fast relief activities [1]. To mitigate the damage, disasters typically are managed through four disaster management phases such as mitigation, preparedness, response and recover/reconstruction [2, 3].

In recent times, we have seen a noteworthy rise in the number of natural disasters throughout the world [4, 5]. Along with natural disasters, man-made disasters during the recent decades are also disrupting and making communities vulnerable. During disasters, the availability of real-time information is essential for both emergency management agencies as well as humanitarian organizations for effective decision making and coordinating their immediate response activities. The quicker the organizations react and respond; the more effective are the relief efforts. The rescue teams can save the affected population and assess the infrastructure damage, to estimate the economic loss and estimate the fatalities. Gauging who needs help, where to reach them, and what kind of help is needed, is limited, in reality, by the constraints of the difficulty in tracing and tracking the information needs of people [6]. However, compared to the 1990's, nowadays the time to get the information from the disaster zone has decreased tremendously because of technological developments such as Internet, mobile communication and most importantly social media [7].

Social media is playing an important role both in our personal and professional lives. The new digital technologies are facilitating the two-way communication as people become both consumers as well as producers of the information [8]. As discussed previously, disasters often disrupt traditional communication systems as infrastructure might get damaged. In these critical situations people often look for other means to share and receive the information. Hence social media become a privileged platform to share and disseminate real-time disaster information to an even wider audience [9] . In addition to individuals, government organizations, media and NGOs are also sharing information during different crisis events. For example, during Alberta floods, Boston bombing, Oklahoma tornado, West Texas explosion, affected people and others switched to social media like Twitter to gather information about a crisis situation [10-12]. The user-generated and unstructured data contains a large amount of valuable information. This information comprises, for example, rescue requests, situational awareness information, or coordinating relief activities. Moreover, affected communities are using social media to share information, to mobilize and self-organize to coordinate activities among themselves [13, 14]. Hence, it is very important for disaster management agencies to listen to the voices of affected communities to organize their relief activities because of the benefit of real time information [15]. However, the volume and the velocity of social media data makes it challenging to monitor and extract valuable information. It is important to develop methodologies and techniques that will automate the process to figure out the rescue requests and the urgencies of the needs of the affected people, which include shortages of food, clothing and shelter. In this regard, by taking example of two case studies that we have explored, we discuss the applied methods and their opportunities and challenges regarding analyzing disaster social media data to extract important information that will be an asset to the disaster management agencies. The two case studies are hurricane Sandy that happened in 2012 in US and Chennai floods that happened in India during 2015. Both these case studies belong to two different geographical locations with different cultural norms, but however, it is interesting to observe the commonalities between two different contexts. Moreover, the first disaster took place in a developed country where infrastructures and support/relief systems are at an advance level and the other case study is from the developing country where infrastructures as well as relief systems are at a minimal level. Our motivation for using these case studies is also to explore how social media is being used by the people in crisis situations, in different contexts, to share and receive the information.

2. Related work

The new trend of using technology to share disaster relevant information by the people during disasters, has gained the attention of research community way back in 2008 [16-18]. People used available technologies of their reach such as web-forums, blogs, Internet to locate their loved ones and to gain or disseminate information about disasters [16]. Social networking sites such as Twitter and Facebook penetrated into the personal lives of people and then also during times of disasters. People started using these networking sites tremendously to share the real-time information, which has changed the traditional disaster communication practices [16, 19]. Among the other social networking sites, research focused more on disaster related publicly available Twitter data because of its' message characteristics, public nature, and the speed with which the information reaches the wider audiences [20-22]. Since people are sharing and retweeting real-time information from the affected areas, significant research has focused on identifying the type and nature of information [23-28]. In addition to the opinions and emotion related messages, the social media messages also contain important situation updates, which can help in assessing the situation and creates situational awareness [13, 28].

Based on the prior research, for example, hurricane sandy [29], red river valley floods and wild fires [30], and floods in Pakistan [31] it is evident that disaster related social media data of all types of disasters contain valuable information. However, with a view to extract situational awareness information from the huge amounts of social media data, classifiers were built using different machine learning algorithms with the help of labelled data as training sets [32]. In order to understand the practical relevance and the importance of usage of social media by the emergency management agencies, interviews were conducted to understand officials' opinions and concluded that social media does provide invaluable information, however the organizations are still struggling to integrate it into their day to day practices [33]. However, there are certain success stories where disaster management agencies started using social media successfully. For example, during the Queensland floods, emergency officials used social media to communicate flood updates and information, yet they have ended up mostly with one-way communication [34]. Moreover, applications and tools have been developed to identify the disaster or to analyze the disaster related social media data, to name a few, SensePlace2 [35] ,Twitcident [36], Tweedr [37] CrisisTracker [38], or Artificial Intelligence for Disaster Response (AIDR) [39]. These are the few examples which shed light on the way social media has been exploited during disasters and for disaster management.

3. Case Study-Hurricane Sandy

In our first case study [29], we wanted to understand what people discussed about during the storm and what type of information is being shared by the people. We especially wanted to understand if people share their personal experiences and observations (original source) or retweet others' tweets or media information and weather reports (secondary source) [30]. For this purpose, we analyzed 11 million tweets that were posted on Twitter when hurricane Sandy hit the east coast of US in 2012.

3.1 Hurricane Sandy

Hurricane Sandy initially developed in the Caribbean waters on October 22, 2012. While gaining forward momentum and slowly intensified and developed into a superstorm affecting 60 million people across 24 states in the U.S. It made landfall on 29th of October and devastated New York city with power outages, flooded subway systems, disrupted communication systems, and also led to the shortages of gasoline, commodities and food [40]. However, people used social media extensively to share information about hurricane, for example, @tweetuser (2012-11-01 03:28:59)

> *"What's amazing is how Twitter and Facebook are more current and up to date with events on #Sandy then the actual news on tv...#media #fail"*

3.2 Data collection and pre-processing

In this case study, we analyzed tweets that originated when the U.S. east coast was hit by hurricane Sandy in October 2012. Hurricane Sandy Twitter dataset contains 15 million tweet IDs that are made available publicly [41] and out of the total dataset, we were only able to retrieve 11 million tweets between May and June 2015, using the Twitter Rest APIs. The number of tweets generated during the disaster situation indicates the interest among people including the non-effected, on hurricane Sandy, however, the tweets posted by the people in and around the hurricane impacted area are most useful and contains valuable situational awareness information. Hence our aim was to extract tweets that are produced and generated by the people in the affected areas. So, in our work, to extract and visualize the tweets from the affected area of hurricane Sandy, we used different tools that can process and filter the tweets that contain geo-location information.

Firstly, with the help of Tableau [42], a business analytics and visualization tool, first we visualized and identified the geographical spread of 11 million tweets. In general, if the twitter data contains longitude and latitude coordinates, it indicate the location of a Twitter user at the time of sharing the information [43]. With the help of Cosmos software [44], which can process geo-located information, we filtered and excluded tweets that do not contain geo-location information. Hence, we identified that out of 11 million tweets only 115,800 tweets, or 1.07% of tweets contain geo-location information. The next step was to extract tweets that originated from the hurricane Sandy path, that affected the coast-line, in addition to this we had to filter and extract tweets that were written in the English language. For this purpose, we used CartoDB

[45] to identify the English language tweets and to narrow down the tweets that are originated from the hurricane Sandy path based on the geo-location information of the tweet as shown in Fig. 11. Finally, we extracted 68,800 tweets from the hurricane Sandy affected area that were produced between 25th October and 5th November, 2012, in the east coast of the U.S. The overall descriptive statistics of the twitter dataset that was extracted during hurricane Sandy is presented in Table I.

TABLE I. DESCRIPTIVE STATISTICS OF HURRICANE SANDY DATASET

Twitter messages (tweets)	Absolute numbers	Percentage
Total tweets	11,658,279	100%
Original tweets	5,369,520	49.50%
Retweeted tweets	5,478,562	50.50%
Tweets with geo-location	115,800	1.07%
English tweets with geo-location	100,700	0.93%

3.3 Methodology

To explore whether valuable and relevant information exists in the final extracted dataset containing 68,800 tweets, one of the authors manually read through all the tweets for the important information and identified 677 important tweets. Further to analyze the resulted tweets, we developed our own coding scheme based on two different prior coding schemes: original/secondary sources [26] and nature of messages [30]. The coding scheme developed by us helped us to analyze a tweet along the two dimensions: information source and nature of message. With the help of information source a tweet can be classified as 1) original source or 2) secondary source. Original source refers to an individuals' personal observations and experiences, and secondary source refers to media information and its' links, retweets, and other online sources. The coding scheme under nature of messages consists of 1) informational messages, 2) action-related, 3) opinion-related and 4) emotion-related. Hence along with information source dimension, a tweet produced by a user from the hurricane sandy affected area can be classified with a label from one of the 4 types i.e. informational, or action-related, opinion-related or emotion-related. Based on our new coding scheme, further both the authors conducted the content analysis of 677 tweets. In between the coded tweets were discussed and cross validated and discrepancies were sorted out. The results from the analysis were later checked again by another member from our research group.

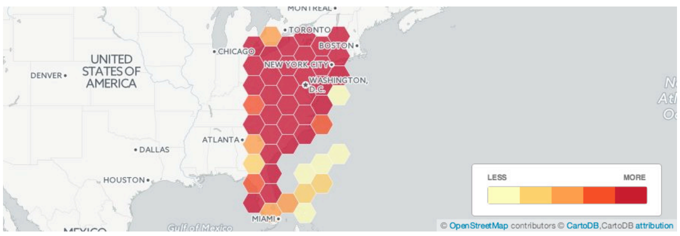

Fig. 11 Sandy affected area (East coast of the U.S.)

3.4 Data analysis and results

In our analysis, we identified a gradual increase in the number of tweets that began from 27th of October (which is before the landfall of hurricane Sandy), with a peak of 2.2 million tweets on the day of October 29th, when it actually made landfall. On October 30th 2012, this number rose to 2.7 million messages per day shared resulting in the highest peak. Fig. 11 shows a visualization of density map containing the 68,800 tweets, that were generated along the east coast, mainly from Florida, Connecticut, New Jersey, Massachusetts, New York. However, among a total of 68,800 tweets only 677 tweets contained valuable and relevant information pertaining to hurricane Sandy. The results revealed that during pre-disaster phase (27th of October) people shared hurricane relevant information from media sources, retweeted other's updates and shared weather reports, but just before the landfall of the hurricane, people started sharing real-time information.

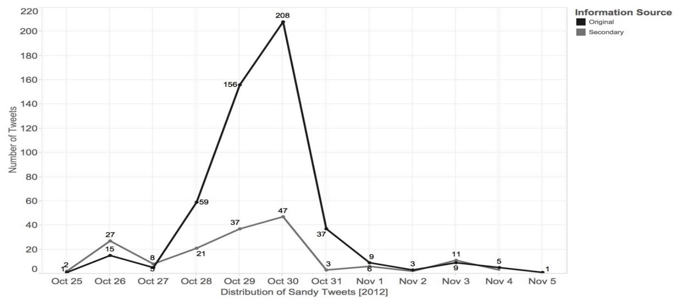

Fig. 12 Information Source

A significant amount of the information came from original sources, which means that people started sharing their valuable personal experiences and observations. The graph in Fig. 12 shows the representation of the 677 tweets that were coded among original and secondary sources. Along with information messages, people also shared their opinions, emotion-related and action related messages. However, information-related messages with situation updates are dominant among the dataset. Based on this case study, we argue that Twitter as a platform facilitates collective level situation awareness among people [29]. Situation awareness is important aspect during disasters either for government organizations, NGOs or relief organizations, to act quickly in disaster response and recovery phases.

4. Case Study- Chennai Rains

In our second case study, from the theoretical perspective of social presence, we wanted to understand what drives people to feel for others and lend their helping hand during disasters by simply reading hastily written messages on social media. For this purpose, we analyzed 1.6 million tweets that were posted on Twitter when Chennai, a southern Indian city, was significantly affected by heavy rains and subsequent flooding in 2015.

4.1 Chennai rains and floods-2015

TABLE II. DESCRIPTIVE DATASET STATISTICS

Twitter messages (tweets)	Absolute numbers	Percentage
Total tweets	1,658,220	100%
Retweets	1,226,098	73.94%
Original tweets of Retweets	141,941	08.56%
Tweets never got retweeted	290,181	17.50%
Mean Retweet Ratio	8.64	
Total Unique Twitter Users	209,644	

Chennai received 34 times the normal daily amount of rain in the first week of December 2015. The downpour intensified on December 2nd, leading to massive flooding. Subsequently it affected homes, hospitals, roads, railway tracks and the city's airport. Three million people suffered due to the lack of access to food, and drinking water [46]. Social media played an important role in times of distress as people used it extensively to reach out to the affected people, coordinated their search and rescue activities, and also for food distribution [47]. Hence, we analyzed Twitter data of the Chennai floods to understand how social presence on social media helped affected people to participate in relief activities [48].

4.2 Data collection

In this empirical work, we analyze the tweets that were shared when Chennai was affected by heavy rains and subsequent flooding. We used Radian 6 tool to collect the Twitter messages by using hashtags #TNflood, #chennaiRains, #chennafloods #chennaiRainsHelp, #IndiaWithChennai and #chennaiMicro. We collected data from November 30th to December 16th, 2015. The total dataset consists of 1.65 million tweets with 209,644 unique users. Since the radian 6 does not provide some of the Twitter attributes like retweet status, retweet count, and original tweet Id for retweets, with the help of tweet Ids using open Twitter API we again downloaded the whole dataset and separated the original tweets from the retweets based on the retweet status information. Interestingly, 74% of total dataset consists of retweets. The descriptive statistics are presented in Table II.

4.3 Methodology

Fig. 13 represents the overall methodology of our research work in this case study. From the media related characteristics, social media falls into the medium category from social presence point of view.

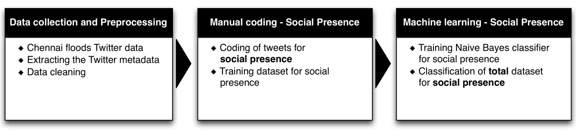

Fig. 13 Overall Methodology of Empirical Study

People perceive the presence of others on communicated medium [49]. In general, the concepts associated with social presence [50] are intimacy [51] and immediacy [52]. We operationalized the concepts *intimacy* and *immediacy* according to the disaster situations to conduct the content analysis. Our interest is to understand how these concepts were expressed in a tweet content.

TABLE III. OVERVIEW OF THE SOCIAL PRESENCE CONCEPTS AND THEIR OPERATIONALIZATION

categories	**Description**	**Tweet Content**
Intimacy	Feeling closeness: sharing road closure info, asking for help on behalf of others	Is there any way to provide any form of support monetary and supplies? #ChennaiRains
	Moral support: stand by people, providing hope for best	Hats off to the fighting spirit of Chennai! A salute all those volunteers who have been helping relentlessly! #staystrongchennai #chenna
Immediacy	Urgent action is needed: different types of rescue requests	any doctors in mudichur area? One pregnant lady in labour...no access to boat. Here is the Contact: 9940203871#chennairains
	Sharing information: to provide shelter, food and help	#food available at #tnagar gurdwara 9094790989#ChennaiRainsHelp #ChennaiMicro #ChennaiVolunteer https://t.co/JD1BAdWXSe

The operationalization of concepts for social presence is shown in Table III. Tweets displaying moral support, creating a feeling of closeness are placed into intimacy category. The tweet content that evokes a feeling of need for urgency and demands an immediate action placed into immediacy category. We used these concepts as our coding scheme in the subsequent step.

To automate the classification of the entire dataset, constituting 1.65 million tweets, along the intimacy and immediacy concepts, we have adopted the automated text classification approach using a supervised machine learning algorithm, the Naïve Bayes classifier. As mentioned above we used 1) Intimacy, 2) Immediacy and 3) None categories as coding scheme for the automated text classification and to prepare training dataset for social presence. The tweets that do not belong to either Intimacy or Immediacy are placed into the "None" category. Initially to ensure validity, authors discussed the concepts behind Intimacy and Immediacy extensively by taking examples, what constitutes the categories and what does not. Further authors independently coded randomly selected sample of approximately 500 tweets. Afterwards, the results were compared and both coders discussed to solve the discrepancies about the concepts. The inter coder agreement matrix for the text coded by the authors is presented in Table IV.

TABLE IV. INTER-CODER AGREEMENT MATRIX OF SOCIAL PRESENCE

		Coder 1			
		Intimacy	Immediacy	None	Marginal Totals
Coder 2	Intimacy	38 (0.06)	7 (0.01)	2 (0.01)	47 (0.08)
	Immediacy	7 (0.01)	83 (0.14)	6 (0.01)	96 (0.16)
	None	1 (0.01)	5 (0.01)	431 (0.74)	437 (0.76)
	Marginal Totals	46 (0.08)	95 (0.16)	439 (0.76)	580 (1.00)

In general, an inter-coder agreement of 0.40 to 0.80 is considered as a good indicator of valid agreement between coders [53]. In case of social presence, the proportion of agreement by chance = (0.08 * 0.08) + (0.16 * 0.16) + (0.76 * 0.76) = 0.62. and Cohen's Kappa value for social presence can be calculated by using the standard formula as: (0.95 – 0.62) / (1 – 0.62) = 0.868. In the final step, to get the trained dataset one of the authors conducted content analysis of randomly selected 5000 tweets for social presence concepts in the subsequent phase according to our coding scheme. Among the coded tweets, 80% of training set was used to train the classifier and the remaining 20% of the tweets were used to test the accuracy of the classifier.

TABLE V. PERFORMANCE MEASURES OF TEXT CLASSIFICATION USING NAÏVE BAYES CLASSIFIER

Model	**Labels**	**Precision**	**Recall**	**F-Measure**	**Accuracy**
Social Presence	Intimacy	0.158	0.466	0.236	0.805
	Immediacy	0.626	0.520	0.568	
	None	0.943	0.858	0.898	

4.4 Data analysis and results

In general, performance of a machine learning algorithm can be described by four measures: precision, recall, F-measure, and accuracy [54, 55]. The performance measures about text classification of tweet content using Naïve Bayes classifier are presented in Table V. The overall accuracy of the classifier is fairly high (i.e. around 80% of the prediction are accurate for correct predictions). In terms of precision and recall, immediacy received fairly good values in contrast to intimacy. In case of the F-measure, a value around 0.6-0.7 indicates a fairly better performance and the F-measure values for all labels/categories indicate reasonably good performance except for the categories: intimacy. The social presence categories in the total dataset are presented below in Table VI.

TABLE VI. SOCIAL PRESENCE CATEGORIES IN THE TOTAL DATASET

Social Presence	Total tweets	Percentage	Retweets	Percentage
Intimacy	285,292	17.20%	226,697	18.49%
Immediacy	335,877	20.26%	291,516	23.78%
None	1,037,051	62.54%	707,885	57.73%
Total	1,658,220	100%	1,226,098	100%

From the total dataset, 37% of tweets are classified as social presence categories. The remaining 63% of tweets belongs to the none category. The reason could be as mentioned in the previous studies [26, 28] people tend to post suggestions, comments, criticism and also discuss or vent their frustration about media or government. However, among the 37% classified tweets, 20% tweets belong to immediacy category and among the retweet category around 24% of retweets belongs to immediacy. Most importantly, 74% of the total dataset constitute retweets and the proportion of social presence retweets among the total retweets is higher with around 42% of retweets belonging to intimacy and immediacy, which indicates that social presence tweets are retweeted more than the tweets that belong to none category. Our results reveal that most of the immediacy tweets are conveying needs and urgencies of people. People perceive the online social presence through the messages sent by the online users and hence are drawn to Twitter to fulfill their social need for connections. Moreover, because of the perceived higher levels of social presence, people continue to use and interact more on Twitter [56].

5. Discussion

In this section, we will discuss about the opportunities and challenges from both social media data point of view and the methods point of view. The following observations are beneficial to practitioners and academics who are interested in social media analytics in general. In particular, these insights might be useful to the emergency management agencies who wanted to integrate social media in their routines and processes.

5.1 Social media-Provider of information

Social media has become a potential platform to gain and share valuable information during the different phases of a disaster. People are using different social media platforms, such as Twitter [57], Facebook [57, 58], Flickr [59] during disasters. Based on our first case study [29], it is evident that during disasters people share first-hand observations and experiences on social media. Moreover, people who are in the vicinity of an event or witnessing an unfolding event immediately share information on social media [15]. Hence social media has become a good source of information to recognize and tap into an unfolding or ongoing crisis/event. In one way, it complements the existing early warning systems because of rapid information dissemination [21] and in the other way, early warnings can be detected on social media by monitoring information [60].
For example, emergency situation awareness (ESA) system was developed to detect the earthquakes based on tweet burst detection method on Twitter [61]. Keywords play an important role in the burst detection in the twitter stream. Often, researchers with the help of keywords and hashtags extract the data using Twitter application

programming interface (API) or tools like Radian 6 [29, 62]. Along with textual data, rich multimedia data is also available on social media platforms in the form of pictures and videos [59, 62]. Given the importance of information embedded in the images, micro mappers [63], with the help of crowds assessing the disaster damages by viewing the pictures. We believe that the disaster related social media data is useful in creating situation awareness and also in disseminating information during disasters. Situation awareness is important for disaster management agencies to make appropriate decisions in disaster response activities. Similarly, creating awareness among the people by disseminating disaster preparedness and warnings using social media is also quite important. However, we also noticed certain challenges in using disaster social media data.

- Data volume and velocity: These two aspects are major issues in handling and processing social media data in real-time as ongoing social media activity explodes during disasters. For example, 11 million tweets were produced just in a few days during hurricane Sandy.
- Data Relevance: From the huge volumes of social media data generated during disasters, identifying important useful information for disaster relief agencies is a challenge, as a major portion of the data contains unrelated information such as commercials and also repeated, re-posted messages (74% of the Chennai dataset consists of retweeted messages).
- Data Integration: During disasters, people use different social media platforms (e.g. Twitter and Facebook) to share information. Integrating and unifying information from these different platforms is a challenge due to different data formats. Integrating data generated from different sources (eyewitnesses, traditional media, outsiders) and different languages [12] is also a challenge.
- Rumors and Fake News: Validity and quality of information is also a challenge. Rumors and fake news create panic among the people. Most importantly, given the nature of social media, these rumors also spread very fast [64-66].

5.2 Geo-information

Disaster related social media data originates from different parts of the world, geographical coordinates [43] allows to filter and retrieve the information from particular disaster affected geographic region. Meta data of a tweet provides geographic information in the form of latitude and longitudinal coordinates. With the help of geo-location information one can select the tweets produced from one specific geographic location and visualize the information of an event on a map as presented in Fig. 11 as demonstrated in our first case study [29]. Given the importance of geo-location information, systems such as emergency situation awareness (ESA) [61] tag the geo information for event detection and systems like SensePlace2 [35] filter and extract geographic information from tweets to visualize the information in maps. Real time information from the disaster affected location is very important for emergency management officials to make tactical decisions. However, some of the challenges to consider:

- According to prior research and also based on our case study, only 1% of data consists of geographic information. An explanation could be that the end users are not revealing the location information due to privacy concerns.

- Applying different methods to infer the location information is considered as privacy intrusion.
- Some of the social media platforms do not provide geo-location information at all. For example Facebook offers a “check in” option [67], but does not provide geo-information in the metadata.

5.3 Digital emergent groups

During disasters, people in the affected area play a big role in taking part in immediate rescue operations. However, it is evident from our second case study, despite the lack of face to face interactions, people are perceiving the presence of others on the social media platforms. The feeling of connectedness in distress situations facilitating the emergence of a new type of volunteers who are not only sharing information and coordinating the rescue activities through social media in affected areas but are also helping in assessing the damage or collaboratively collecting the necessary information that is useful to the affected individuals. So far based on prior literature [33, 34, 68] we know that organizations are interested in social media to communicate disaster relevant information to people. However, new structures are evolving at the community level [58], hence it provides an opportunity for disaster management agencies while monitoring the information on social media can also engage with digital volunteers directly in conversations and in collaboration to perform relief operations [68]. However, the challenge is:

- These new types of volunteers only emerge during the times of disasters on an ad-hoc basis, hence it is a challenge to get in contact with them and include them in the disaster relief activities.

5.4 Methods to analyze the textual data

In general, in order to analyze the textual data from disaster related social media content, different methods have been applied by the research communities: manual content analysis and computational methods. As part of computational methods, two techniques: supervised and unsupervised machine learning techniques are in use. Classifiers fall into the supervised machine learning technique and topic modeling is considered as belonging to the unsupervised approach. However, in our case studies, we have applied both manual content analysis and classifiers. In our first case study, we analyzed the data manually by applying manuals content analysis, but in our second case study we applied supervised machine learning technique to classify the data for identifying social presence concepts in tweet content.
Manual content analysis is a very basic method, where human coders analyze the data and categorize the information. With the help of manual content analysis either one could derive categories from the data by analyzing it or one could analyze and extract the information based on pre-defined categories. One could also use crowd sourcing platforms, for example, Amazon Mechanic Turk or Crowd flower, to analyze the data with the help of crowds by providing instructions and coding schemes. However, this method is perfect and feasible when the dataset is rather “small”, because analyzing millions of tweets is humanly not possible. However, there are systems like CrisisTracker [38] and Artificial Intelligence for Disaster Response (AIDR) [39] take the help of crowds to annotate the data. However, with the help of annotated data AIDR classifies the rest of the data automatically.

With the help of classifiers, data can be classified automatically by using a machine learning algorithm. Using well defined training dataset and supervised algorithm, one can classify “huge amounts” of data or to extract the information easily. In our second case study, with the help of classifiers, we tested the social presence theory and extracted intimacy and immediacy tweets from the huge amounts of data. Previous research used classification approaches to extract situation awareness and informative messages from social media data [32, 69] and there are also a few systems developed based on text classification approach: Tweedr [37], ESA [61]. Some of the challenges associated with classifiers are:

- Well-defined training sets are necessary for building the classifiers, which might need resources such as a human coder to prepare them.
- Given the nature of unstructured data, getting the high accuracy of classifier is difficult.
- The classifier developed in one disaster scenario might not work well in another situation. For example, classifiers that are developed for earthquakes might not be suitable for floods.

6. Conclusion

In this paper, based on the two case studies and the discussion above, we argue that emergency managers can use social media to disseminate disaster relevant information, gain situation awareness from the affected individuals, and can start a dialogue with people to reduce the risk [68]. Based on recent research [33, 70] emergency officials are still skeptical on fully integrating social media in their organizational routines. The reason could be the challenges that we mentioned previously regarding data and methods. However, in few cases, emergency organizations started using social media for disseminating information and can be coined as aspirational or early adopters of social media [71]. In order to reach to the next level, it is important for organizations to invest time, resources and personnel [68] to harvest the benefits of social media. Most importantly to gain value by using social media, organizations need to align their goals with the strategic initiatives.

However, based on our research we addressed certain challenges along with opportunities. Firstly, we argue that the various analysis techniques can be used to extract and understand the valuable information that is generated during the disaster situations. At the same time focusing on different case studies, we tried to understand how people used social media in different contexts during natural disasters such as floods and hurricanes. Based on case studies we also strongly argue that automated text analytical methods can be used to analyze and understand the user generated content. Most importantly, our current research helped us towards developing or applying other automated methods. For example, in future we would like to apply unsupervised topic modeling method to extract the hidden information from the user-generated texts without any human intervention. At the same time, we argue that the research community needs to focus on supervised and dictionary based approaches to classify the social media data, specifically focusing on extracting time relevant information with needs and urgencies that is embedded in social media data.

7. References

[1] A. Mukkamala and R. Beck, "Disaster management and social media use for decision making by humanitarian organizations," in System Sciences (HICSS), 2016 49th Hawaii International Conference on, 2016, pp. 1379-1385.

[2] N.-S. Lee, S. Hirschmeier, S. Müller, and L. J. Luz, "Enablers in Crisis Information Management: A Literature Review," 2017.

[3] A. Cozzolino, Humanitarian logistics: cross-sector cooperation in disaster relief management: Springer Science & Business Media, 2012.

[4] R. Stiegler, S. Tilley, and T. Parveen, "Finding family and friends in the aftermath of a disaster using federated queries on social networks and websites," in Web systems evolution (WSE), 2011 13th IEEE international symposium on, 2011, pp. 21-26.

[5] P. H. Debarati Guha-Sapir , Regina Below, "Statistical Review 2015 The numbers and trends," Centre for Research on the Epidemiology of Disasters (CRED), Brussels, Belgium2016.

[6] L. K. Comfort, K. Ko, and A. Zagorecki, "Coordination in rapidly evolving disaster response systems: The role of information," American Behavioral Scientist, vol. 48, pp. 295-313, 2004.

[7] P. Meier. (2016). The Value of Timely Information During Disasters (Measured in Hours). Available: https://irevolutions.org/2016/04/06/the-value-of-timely-information-during-disasters-measured-in-hours/

[8] N. Cook, Enterprise 2.0: how social software will change the future of work: Gower Publishing, Ltd., 2008.

[9] T. Simon, A. Goldberg, and B. Adini, "Socializing in emergencies—A review of the use of social media in emergency situations," International Journal of Information Management, vol. 35, pp. 609-619, 2015.

[10] A. Olteanu, C. Castillo, F. Diaz, and S. Vieweg, "CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises," in ICWSM, 2014.

[11] A. Olteanu. (2017). CrisisLex.org. Available: http://crisislex.org/

[12] A. Olteanu, S. Vieweg, and C. Castillo, "What to expect when the unexpected happens: Social media communications across crises," in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015, pp. 994-1009.

[13] S. Vieweg, A. Hughes, K. Starbird, and L. Palen, "Supporting situational awareness during emergencies using microblogged information," SIGCHI 2010, 2010.

[14] K. Starbird and L. Palen, "Voluntweeters: Self-organizing by digital volunteers in times of crisis," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 1071-1080.

[15] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting there first: Real-time detection of real-world incidents on twitter," in 2nd Workshop on Interactive Visual Text Analytics : Task-Driven Analysis of Social Media Content with Visweek'12., 2012.

[16] J. Sutton, L. Palen, and I. Shklovski, "Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires," in In: Proceedings of the 5th International ISCRAM Conference, 2008, pp. 624-632.

[17] I. Shklovski, L. Palen, and J. Sutton, "Finding Community Through Information and Communication Technology During Disaster Events," in Proceedings of the 2008 ACM conference on Computer supported cooperative work, ed: ACM, 2008, pp. 1-10.

[18] L. Palen, "Online social media in crisis events," Educause Quarterly, vol. 31, pp. 76-78, 2008.

[19] L. Palen, S. Vieweg, S. B. Liu, and A. L. Hughes, "Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event," Social Science Computer Review, vol. 27, pp. 467-480, 2009.

[20] A. T. Chatfield, H. J. Scholl, and U. Brajawidagda, "# Sandy Tweets: Citizens' Co-Production of Time-Critical Information during an Unfolding Catastrophe," in 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1947-1957.

[21] A. T. Chatfield and U. Brajawidagda, "Twitter early tsunami warning system: A case study in indonesia's natural disaster management," in System sciences (HICSS), 2013 46th Hawaii international conference on, 2013, pp. 2050-2060.

[22] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, 2007, pp. 56-65.

[23] A. Ahmed and J. Sargent, "Analysis of Post-Crisis Twitter Communication: A Study of the Iquique, Chile Earthquake," in In: Proceedings of the Australasian Conference on Information Systems 2014.

[24] T. Heverin and L. Zach, "Microblogging for Crisis Communication : Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma , Washington Area," presented at the In: Proceedings of the 7th International ISCRAM Conference, 2010.

[25] S. Muralidharan, L. Rasmussen, D. Patterson, and J.-H. Shin, ""Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts"," Public Relations Review, vol. 37, pp. 175-177, 2011.

[26] Y. Qu, C. Huang, P. Zhang, and J. Zhang, "Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake," in In: Proceedings of the ACM 2011 conference on Computer supported cooperative work, 2011, pp. 25-34.

[27] S. Vieweg, A. L. A. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in In: Proceedings of the 28th international conference on Human factors in computing systems - CHI '10, 2010, pp. 1079-1088.

[28] S. E. Vieweg, "Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications," PhD thesis, University of Colorado at Boulder, 2012.

[29] A. Mukkamala and R. Beck, "Enhancing Disaster Management Through Social Media Analytics To Develop Situation Awareness What Can Be Learned From Twitter Messages About Hurricane Sandy?," presented at the The Pacific Asia Conference on Information Systems (PACIS) 2016.

[30] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg, "Chatter on The Red: What Hazards Thret Reveals about the Social Life of Microblogges Information," in In: Proceedings of the Conference on Computer Supported Cooperative Work (CSCW), 2010, pp. 241-250.

[31] D. Murthy and S. A. Longwell, "Twitter and disasters: The uses of Twitter during the 2010 Pakistan floods," Information, Communication & Society, vol. 16, pp. 837-855, 2013.

[32] S. Verma, S. Vieweg, W. J. Corvey, L. Palen, J. H. Martin, M. Palmer, et al., "Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency," in ICWSM, 2011.

[33] C. Reuter, T. Ludwig, M.-A. Kaufhold, and T. Spielhofer, "Emergency services′ attitudes towards social media: A quantitative and qualitative survey across Europe," International Journal of Human-Computer Studies, vol. 95, pp. 96-111, 2016.

[34] C. Ehnis and D. Bunker, "Social media in disaster response: Queensland Police Service-public engagement during the 2011 floods," in ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian Conference on Information Systems 2012, 2012, pp. 1-10.

[35] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, et al., "Senseplace2: Geotwitter analytics support for situational awareness," in Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on, 2011, pp. 181-190.

[36] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, "Twitcident: fighting fire with information from social web streams," in Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 305-308.

[37] Z. Ashktorab, C. Brown, M. Nandi, and A. Culotta, "Tweedr: Mining twitter to inform disaster response," in ISCRAM, 2014.

[38] J. Rogstadius, M. Vukovic, C. Teixeira, V. Kostakos, E. Karapanos, and J. A. Laredo, "CrisisTracker: Crowdsourced social media curation for disaster awareness," IBM Journal of Research and Development, vol. 57, pp. 4: 1-4: 13, 2013.

[39] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, "AIDR: Artificial intelligence for disaster response," in Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 159-162.

[40] Y. Neria and J. M. Shultz, "Mental health effects of Hurricane Sandy: Characteristics, potential aftermath, and response," Jama, vol. 308, pp. 2571-2572, 2012.

[41] A. Zubiaga. (2015). Dataset: Hurricane Sandy tweets. Available: http://www.zubiaga.org/datasets/hurricane-sandy-tweets/

[42] Tableau. (2015). Available: http://www.tableau.com/

[43] M. Graham, S. A. Hale, and D. Gaffney, "Where in the world are you? Geolocation and language identification in Twitter," The Professional Geographer, vol. 66, pp. 568-578, 2014.

[44] P. Burnap, O. Rana, M. Williams, W. Housley, A. Edwards, J. Morgan, et al., "COSMOS: Towards an integrated and scalable service for analysing social media on demand," International Journal of Parallel, Emergent and Distributed Systems, vol. 30, pp. 80-100, 2015.

[45] CartoDB. (2015). Available: https://cartodb.com/

[46] T. Misra. (2015). Chennai Floods: A Perfect Storm of Climate Change and Failed Planning. Available: http://www.citylab.com/weather/2015/12/chennai-flood-climate-change-urban-planning-cop21-paris/419349/

[47] V. Pandey. (2015). How Chennai residents are helping flood affected via social media. Available: http://www.bbc.com/news/world-asia-india-34981360

[48] A. Mukkamala and R. Beck, "Presence of Social Presence during Disasters," in Pacific Asia Conference on Information Syatems, 2017.

[49] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Business horizons, vol. 53, pp. 59-68, 2010.

[50] J. Short, E. Williams, and B. Christie, The social psychology of telecommunications, 1976.
[51] M. Argyle and J. Dean, "Eye-contact, distance and affiliation," Sociometry, vol. 28, pp. 289-304, 1965.
[52] M. Wiener and A. Mehrabian, Language within language: Immediacy, a channel in verbal communication: Ardent Media, 1968.
[53] S. Stemler, "An overview of content analysis," Practical assessment, research & evaluation, vol. 7, pp. 137-146, 2001.
[54] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," School of Informatics and Engineering, Flinders University , Adelaide, Australia. 2229-3981, 2011.
[55] Y. Yang, "An evaluation of statistical approaches to text categorization," Information retrieval, vol. 1, pp. 69-90, 1999.
[56] [56] S. Han, J. Min, and H. Lee, "Antecedents of social presence and gratification of social connection needs in SNS: a study of Twitter users and their mobile and non-mobile usage," International Journal of Information Management, vol. 35, pp. 459-471, 2015.
[57] K. Eismann, O. Posegga, and K. Fischbach, "COLLECTIVE BEHAVIOUR, SOCIAL MEDIA, AND DISASTERS: A SYSTEMATIC LITERATURE REVIEW," presented at the European Conference on Information Systems, 2016.
[58] L. Kaewkitipong, C. C. Chen, and P. Ractham, "A community-based approach to sharing knowledge before, during, and after crisis events: A case study from Thailand," Computers in Human Behavior, vol. 54, pp. 653-666, 2016.
[59] S. B. Liu, L. Palen, J. Sutton, A. Hughes, and S. Vieweg, "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, pp. 969-980.
[60] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in WWW '10: Proceedings of the 19th international conference on World wide web, ed, 2010, pp. 851-860.
[61] J. Yin, S. Karimi, B. Robinson, and M. Cameron, "ESA: Emergency Situation Awareness via Microbloggers," in Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12, ed: ACM, 2012, pp. 2701-2703.
[62] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," ACM Computing Surveys (CSUR), vol. 47, p. 67, 2015.
[63] P. Meier. (2016-08-09). MicroMappers Digital Disaster Response. With a Single Click! Available: http://www.micromappers.com/
[64] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: can we trust what we RT?," in Proceedings of the first workshop on social media analytics, 2010, pp. 71-79.
[65] O. Oh, K. H. Kwon, and H. R. Rao, "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010," in ICIS, 2010, p. 231.
[66] K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. M. Mason, "Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing," iConference 2014 Proceedings, 2014.
[67] P. Hannay and G. Baatard, "GeoIntelligence: Data mining locational social media content for profiling and information gathering," 2011.
[68] C. Wukich, "Social media use in emergency management," Journal of Emergency Management, vol. 13, pp. 281-295, 2015.
[69] M. Imran, S. M. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Extracting information nuggets from disaster-related messages in social media," Proc. of ISCRAM, Baden-Baden, Germany, 2013.
[70] A. Sheombar, C. Urquhart, T. Ndhlovu, and P. Ravesteijn, "Social Media in the Context of Development: A Case Study of Dutch NGOs," in ECIS, 2015.
[71] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value," MIT sloan management review, vol. 52, p. 21, 2011.

Paper 6

Mukkamala, AM & Beck, R (2017)

The Development of a Temporal Information Dictionary for Social Media Analytics

In: Proceedings of International Conference on Information Systems **(ICIS 2017)** Seoul, South Korea. (Research-in-Progress)

The Development of a Temporal Information Dictionary for Social Media Analytics

Short Paper

Alivelu Mukkamala, Information Management Section, IT University of Copenhagen, Denmark, ammu@itu.dk

Roman Beck, Information Management Section, IT University of Copenhagen, Denmark, romb@itu.dk

Abstract

Dictionaries have been used to analyse text even before the emergence of social media and the use of dictionaries for sentiment analysis there. While dictionaries have been used to understand the tonality of text, so far it has not been possible to automatically detect if the tonality refers to the present, past, or future. In this research, we develop a dictionary containing time-indicating words in a wordlist (T-wordlist). To test how the dictionary performs, we apply our T-wordlist on different disaster related social media datasets. Subsequently we will validate the wordlist and results by a manual content analysis. So far, in this research-in-progress, we were able to develop a first dictionary and will also provide some initial insight into the performance of our wordlist.

Keywords: *Method, Social media, Dictionary, Temporal information*

1. Introduction

Since the introduction of social media, unstructured, user-generated content has been created at an unprecedented scale, which became an important source of information for researcher in social science (Thelwall et al. 2008). The predominant research interest has been on analyzing the textual content of social media messages, which contain people's opinions, expectations, feelings, and so on. Different scientific disciplines have used different approaches to analyze text such as manual and automated content analysis, information retrieval, or natural language processing, to mention a few (Loughran and McDonald 2011). The research area that gained most prominence is possibly the sentiment analysis (Read 2005), which allows the automated assessment of positive or negative feelings based on the tonality of a message.

The term sentiment analysis, sometimes referred to as sentiment classification (Read 2005; Read and Carroll 2009), opinion mining (Pang and Lee 2008; Thelwall et al. 2011; Thelwall et al. 2010), subjectivity analysis (Riloff and Wiebe 2003), or polarity classification (Read and Carroll 2009) describes more or less the same concept of carrying out a tonality analysis of a more positive or more negative tone used in a message (Liu 2012). The general assumption is that in general people have certain opinions on topics, and these opinions can be capture by a sentiment analysis (Kim and Hovy 2004). Hence, sentiment analysis is mainly focused at automatically extracting and summarizing the people's opinions about various topics by classifying them (Beigi et al. 2016). Being able to detect the sentiment of a message has been found to be important: For example, by applying sentiment analysis, online retailers can get aggregated results while summarizing the collected reviews with an average sentiment score. In the financial domain, stock market experts can predict stock price fluctuations, based on average sentiments (Read and Carroll 2009). During disasters, a sentiment analysis can help to understand feelings, concerns, or even panic of affected people (Beigi et al. 2016).

While sentiment analysis has illustrated its usefulness, an important dimension so far has been not taken into consideration, and that is to which time dimention the message is referring to. A tweet talking about how much someone loved his first BMW would result in a positive sentiment towards BMW, although the writer of this tweet might not be inclined to buy one nowadays. The same is true for statements such as "IBM was a great company": most sentiment analyses approaches would interpret this as a positive sentiment towards IBM, ignoring the fact that the author might have changed his sentiment in the meanwhile, making such a statement less valuable to build upon a buying decision for IBM stocks.

Time plays also a role in another dimension, namely the need to analyze social media data in close to real time, for example to be able to prioritize emergency response forces along the urgencies of the different needs people in a disaster are suffering from. For example, tweets containing words like "just", "urgent" or "immediately" might help disaster relief forces to prioritize where to move in and help first, if they can analyze the data quick enough, meaning in an automated way, as the following two tweets illustrate.

“A massive earthquake just hit Everest. Basecamp has been severely damaged. Our team is caught in camp 1. Please pray for everyone”.

“Urgent. Need 1000 water packets. Please contact us immediately”.

Thus, we claim that detecting time-indicating statements in social media is important to refine, for example if a positive sentiment is related to the past, the presence, or the future. It is also important to differentiate what is more urgent in case of an emergency, which also requires an automated detection of time-indicating statements. So far, there has been only limited research that focused on extracting time-indicating statements from social media nor has there been a similar effort as in the case of sentiment analysis to develop a dictionary or wordlist, comprising time-relevant words and phrases. Thus, our research focusses on filling this gap by developing a time-indicating dictionary. For that to do so, in a first step we had to develop a “dictionary development methods”, since there are many dictionaries in use in sentiment analysis, but it seems there is no structured approach how to develop one. Thus, in this paper our research question is:
How can time-indicating expressions be captured in a dictionary to automatically assess social media data in close to real time?

The remaining research in progress paper is organized as follows. In section two, we provide a literature review on prior research on detecting time in textual statement, as well as the dictionaries have been create so far using supervised approaches under data analytics. A detailed description of the methodology we developed to derive a dictionary will be presented in the third section. In section four, we present our preliminary results regarding the development of our time-indication dictionary and the data sources we used to extract our time-indicating expressions. Based on preliminary results, in the last section we discuss our future research where we explain what we intend to do in order to make our wordlist more robust and rigorous to extract more appropriate data from the given content so that it can be generalizable to other domains as well.

2. Literature review

2.1 Temporal Information

The temporal (time) information about events embedded in different types of text is of utmost importance to understand exact time or time period the text is referring to (Alonso et al. 2007) or to answer the questions in a given news article regarding events (Pustejovsky et al. 2003a). Since time is playing an important dimension, extraction and normalization of temporal information has been considered as an essencial task. For this purpose, temporal information extraction and retrieval gained attention back (Allen 1983) where James F. Allen proposed an interval-based temporal logic to reason about temporal knowledge and temporal intervals using the computational approach. In general, one can understand and group the temporal information into four categories: date, time, duration, and set (Pustejovsky et al. 2005) as discussed further. In contrast to time and date expressions which provide specific information about a point in time, duration expressions mainly give information about the length of an interval (Strötgen and Gertz 2013). Furthermore, according to the study of (Schilder and Habel 2001), time denoting expressions in a document come in three different types: explicit reference, index reference, and vague reference. Date expressions such as “18.08.71” provide an explicit reference and point to a precise moment and thus can be normalized easily. As part of indexical reference, temporal expressions (such as “today”, “by last week” etc.) can only be evaluated via the presence of a time stamp in the document. Other types of temporal expressions (such as “in several weeks”, “in the evening”, etc.) express vague temporal information that is difficult to place on a timeline. However, other studies refer to indexical references as relative expressions, where it is argued that context information such as document creation time or another temporal information is necessary to normalize the temporal expression in the documents (Alonso et al. 2011; Strötgen and Gertz 2013). Moreover, implicit expressions such as names of holidays (Christmas 2016) and events can be normalized by their temporal semantics. “The normalization task of a temporal tagger is to assign the same value to all expressions carrying the same semantics or referring to the same point in time” (Strötgen and Gertz 2013).

In continuation of the above discussion, there has been a great amount of interest from the Natural Language Processing (NLP) community in extracting temporal relationships and events from textual corpura such as news media and other formally written texts. The main focus of research in this direction is to identify temporal events and expressions from documents and to establish temporal relationships between such time events and time-dependent facts. One of the first research initiatives, TIMEX2 (Ferro et al. 2001) a standard for annotation of temporal expressions was initially developed as part of TIDES (Translingual Information Detection, Extraction, and Summarization) program. Based on the TIDES TIMEX2 annotation effort, TimeML (Pustejovsky et al. 2003a), a temporal markup language and Timebank corpus (Pustejovsky et al. 2003b) containing annotated events, times and temporal relations was developed to identify events and temporal expressions in natural language texts. After several iterations, TimeML language has become a gold standard for annotating temporal information (Verhagen et al. 2007). Several research groups have also developed tools and toolkits for performing temporal analysis on texts. Based on Timex3 (Group 2009) annotation standards for temporal information, temporal taggers like HeidelTime (Strötgen and Gertz 2010), Stanford temporal tagger (Chang and Manning 2012) were developed to recognize and normalize the time expressions in the textual documents. Stanford temporal tagger is a rule based tagger that is built on top of Stanford POS-tagger and named-entity taggers and offers a good accuracy in identifying the temporal expressions from text.

Moreover, as part of semantic evaluation initiative from NLP community (SemEval 2007-2017), the tasks of time annotation has received greater interest among the NLP community (Group 2009; UzZaman et al. 2012; Verhagen et al. 2007; Verhagen et al. 2010) for evaluating time expressions, events, and temporal relations among the multiple languages. To detect features of time series of facts from a large number of documents, techniques such as joint inference for temporal scoping was used (Ling and Weld 2010; Talukdar et al. 2012). To enhance summarization of multi documents, the study (Ng et al. 2014) focused on the temporal information in the form of timelines. In order to generate a timeline, an automated processing system was employed, through which three features were derived to measure and recognize the importance of sentences. To overcome any potential errors from underlying temporal processing system, a reliability filtering metric was used to decide when the important temporal information should be used. A multi-document summarization is useful where the event occurrences happen in a chronological order (Mani and Wilson

2000). To extract temporal information, in addition to news articles (Ferro et al. 2005; Pustejovsky et al. 2003a), Wikipedia documents (Strötgen and Gertz 2013), and scientific documents (Strötgen and Gertz 2012) were analyzed. To some extent, informal discussions of online communities were also used, to tag, retrieve and normalize temporal information (Wen et al. 2013). Most importantly the challenges in extracting temporal orientation from social media messages such as Facebook messages (Schwartz et al. 2015) were explored and discussed.

Distinct from the previous research, a temporal ontology, TempoWordNet (Ga et al. 2014; Hasanuzzaman et al. 2016) was constructed automatically by adding temporal information to the words from WordNet (Miller 1995) using a two-step classification approach. Using the similar classification approach, another research work (Kolomiyets et al. 2011) explored the task of recognizing time expressions using a number of bootstrapping strategies to generate additional training set documents that are supplemented with temporal words taken from WordNet (Miller 1995) and Latent Word Language Model (Deschacht and Moens 2009). However, we argue that if the time related words are extracted automatically from the the WordNet, these words are not representative of the words used by ordinary people in their daily communications. Hence, in our initial step, we prefer to construct the dictionary by manually collecting temporal words, and then later on we compare and add words from other sources.

Our approach differs in two aspects when compare with all the above mentioned NLP methods. First, most of the above mentioned methods use advanced NLP techniques such as parsing, classification etc, to identify and extract temporal events and thereby to find temporal relationships between the events. On the contrary, our approach uses a simple, easy lexicon-based approach using manually collected time words to identify and filter the texts containing time-indicating information. Second, most of the above mentioned temporal work primarily targeted for news media and articles (such as Wikipedia) where the language styles are generally formal, hence as also indicated in (Wen et al. 2013), applying these techniques to more informally written texts such as social media posts is challenging. As our primary focus is to identify time-indicating expressions in social media texts, we resorted to a lexicon-based approach that is suitable for processing of social media texts in close to real-time.

2.2 Sentiment Analysis and Dictionaries

In general, in order to perform sentiment analysis, research focused on two types of approaches which are unsupervised, dictionary-based approaches (Abdulla et al. 2016; Backfried and Shalunts 2016; Taboada et al. 2011; Thelwall et al. 2011; Thelwall et al. 2010), and supervised, machine learning approaches (Abbasi et al. 2008; Gonçalves et al. 2013; Pang and Lee 2008; Read 2005; Yang et al. 2010). The latter approaches are used to build the classifiers, where manually labelled data is used as training set for a supervised machine learning approach. It is argued that classifiers give high accurate results in detecting the sentiment and polarity of a given text (Boiy et al. 2007; Chaovalit and Zhou 2005). However, classifiers are domain specific, hence when applied in another context its' performance drops considerably (Aue and Gamon 2005; Taboada et al. 2011). A recent study classified the sentiment analysis approaches on a much broader spectrum by adding unsupervised, semi-supervised and hybrid approaches, in addition to the approaches mentioned above (Madhoushi et al. 2015).

In a dictionary-based approach, sentiment scores are assigned to a list of words to measure either semantic orientation or polarity (positive or negative), or strength (valence) of a given text (Nielsen 2011; Taboada et al. 2011)). Constructing a sentiment dictionary manually is labor intense and time consuming hence most of the sentiment analysis research depends on preexisting, manually constructed dictionaries. For example, the Linguistic Inquiry and Word Count (LIWC) application consists of an internal dictionary which was compiled manually (Pennebaker et al. 2001). It consists of 4,500 words and word stems where words initially were collected from different sources while more words were added to it over the time. The words were grouped into, emotion or affective subcategories and arranged hierarchically. Altogether, there are around 66 categories with varying numbers of words where the words are categorized either into positive or negative. In recent versions, the dictionary was updated with functional words, to name a few, conjunctions, adverbs, quantifiers, or commonly used verbs. Moreover, some original categories were removed as they are not used so often (Pennebaker et al. 2001). Another lexicon, Affective Norms for English Words (ANEW) has 2,477 unique words where the words are scored for valency that range from −5 (very negative) to +5 (very positive) (Bradley and Lang 1999). A few other widely used dictionaries in sentiment analysis are Opinion lexicon (Wilson et al. 2005), SentiWordNet (Esuli and Sebastiani 2007), WordNet (Miller et al. 1990), WordNet-Affect (Esuli and Sebastiani 2007), and AFINN-96 (Nielsen 2011). There are also certain domain specific dictionaries (Olteanu et al. 2014; Temnikova et al. 2015) and language specific dictionaries (Madhoushi et al. 2015) as well. Even though there are many dictionaries, it is still unclear how to build a dictionary in the best possible way (Nielsen 2011) and there are no standardized procedures or commonly accepted methods in place how to build them (Deng et al. 2017). Unlike the sentiment analysis dictionaries as mentioned previously, there are no existing dictionaries available for time-related information.

3. Methodology-how we developed our dictionary

To achieve our research objectives, we followed a methodology that consists of three phases as shown in Figure 14. The first phase comprises the building of a T-wordlist by collecting the words representing time. The second phase consists of data collection and pre-processing of social media data and applying the T-wordlist on it. Through the use of a manual content analysis, the third phase primarily focusses on the validation of the data extracted by the application of T-wordlist.
Regarding alternative approaches to our methodology, we could have chosen different approaches such as classification based methods using supervised machine learning, temporal taggers using advanced NLP methods such as Stanford Temporal Tagger (Chang and Manning 2012) etc. But since our focus is on developing a simple and transparent methodology that is easy to understand and adopt by disaster and other relief organizations, we rather focused on developing our T-wordlist using the lexicon-based approach using the manually collected words indicating time information. Moreover, our chosen lexicon-based approach is also inline with the methodology of CrisisLex lexicon (Olteanu et al. 2014) and in fact our methodology works very well in conjunction with CrisisLex lexicon by complementing it to identify time-indicating expressions from crisis-related messages during disasters.

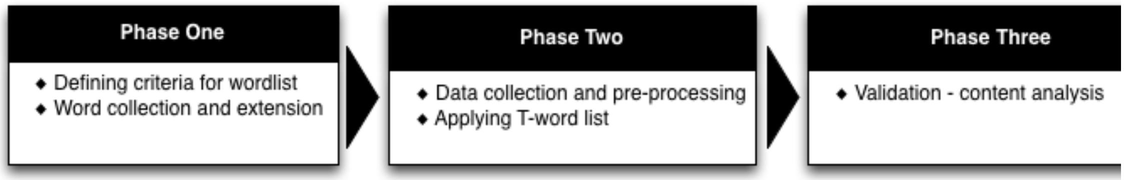

Figure 14. T-wordlist building process

3.1 The criteria for a time-indicating wordlist

In order to answer our research question, as a first step we developed a preliminary version of a wordlist containing time indicating and representing words (time indicating words) such as after, soon, tomorrow, later, which can extract temporal information from the data. Since there are no existing dictionaries available we had to start from the scratch to develop our T-wordlist. The selection of words defining our T-wordlist involved several iterations. The initial idea was to collect a group of words that explains the temporal information from the disaster social media data. However, over time we expanded the T-wordlist considerably by defining the criteria and adding more words as explained below.

Step 1. Defining criteria: In the development of the T- wordlist, most important requirement is to consider the right words and phrases that contain time information. So firstly, we focused on inclusion and exclusion of words. After several discussions, both the authors decided and defined the criteria for time relevant words as follows. The first criteria: we included words which indicate direct temporal expressions such as years ago, last month, tomorrow and so on. Second criteria: words which signal and help in interpretation of temporal expressions such as temporal prepositions (such as during on, at, for) and connectives (such as before, after, while) (Pustejovsky et al. 2003b). Third criteria: key time words that indirectly signal the time information such as immediately, urgent, now, and so on. Our focus was to analyse data from social media during times of disaster data and not news articles or text summarization (Chambers et al. 2007). Therefore, we decided to exclude the tens verbs such as has left, was affected, etc., which indicate the event expressions (Pustejovsky et al. 2003b). Once we defined our criteria, we proceeded to the next step of collecting words from an online dictionary.

Step 2. Word collection and extension: As mentioned previously, unlike dictionaries for sentiment analysis (Esuli and Sebastiani 2007; Nielsen 2011; Strapparava and Valitutti 2004) and domain specific dictionaries (Abdulla et al. 2016; Loughran and McDonald 2011) which are mostly built on already existing dictionaries (Miller 1995; Miller et al. 1990; Stone and Hunt 1963), there are no dictionaries available specifically pertaining to time words only. Therefore, we had to rely on online dictionaries to collect the words and used dictionary.com and Oxford dictionary as our main sources to collect the time relevant words. Initially, we collected only a few common words that provide time information. Later on we extended our wordlist by looking at synonyms and antonyms. In this process, we also used thesauruses to collect more words and phrases. Altogether our initial set of T-wordlist contains around 476 words. The T-wordlist does not only consist unigrams but also phrases. Once the T-wordlist is now ready, so in step 3 we apply it on disaster-related social media data to extract the messages with time relevant information.

3.2 Data collection, preprocessing and application of T-wordlist

In this phase, we applied the T-wordlist on disaster social media data. The social media data that we focused on and collected was the Twitter data from both manmade and natural disasters, such as wild fires, bush fire, floods, shootings, earthquake and so on. Moreover, these disasters occurred in different countries across the world, e.g., Colorado, Philippines, Australia, Singapore, Los Angeles and so on. We will explain in detail the data collection and pre-processing in the following section. We collected publicly available 12 different disaster datasets (Olteanu. 2017) each consisting of approximately 1,000 to 1,200 tweets from different disasters such as wild fires, floods, shootings and so on. The detail information of datasets is provided below Table 10. All together the datasets consist of 12,831 tweets from disparate datasets.
After collecting the data, we randomly took a sample data from each dataset to examine whether or not the dataset consists tweets in English language only. While performing this task, we noticed there are few tweets from languages other than English. For example, in the datasets of Philipinnes floods and typhoon Pablo. Hence we realised pre-processing is an important and required step to proceed to the further analysis. In order to clean and reduce the noise in the data, we pre-processed each dataset to eliminate the tweets other than English language. Of the 12 datasets, we identified all the datasets have language other than English. The results are presented in Table 10.

Table 10. Data Description

Name of dataset/Year	Total number of tweets	English Tweets	Non-English Tweets
Colorado wildfires – 2012	1,200	1,163	37
Philipinnes floods - 2012	1,000	851	149
Typhoon Pablo - 2012	1,000	821	179
Australia bushfire - 2013	1,200	1,174	26
Bohol earthquake - 2013	1,000	837	163
Brazil nightclub fire - 2013	1,000	407	593
Glasgow helicopter crash - 2013	1,100	1,079	21
Los Angeles airport shootings - 2013	1,032	938	94
New York train crash - 2013	1,000	968	32
Savar building collapse -2013	1,250	1,107	143
Singapore haze - 2013	1,000	947	53
Typhoon Yolanda - 2013	1,049	946	103

Table 1. Data Description

Later, we applied the T-wordlist on the individual datasets. We segregated the tweets into two categories. The tweets that are matching with words of our T-wordlist and the tweets that are no way related or containing any of the words from our T-wordlist. Around 45% of T- words represented in the data, will be discussed in the results section. All together we extracted 4,791 number of tweets (43% of total tweets) as presented in Table 11Table 10. At this stage, to ensure whether or not the data is representing temporal information, validation is important.

3.3 Validation of the T-wordlist

Data validation is important part of any study because previous studies mentioned that the same word provides different meaning in different context. In order to ascertain whether the tweet is accurate enough in representing the time relevant information as we intended, we conducted a manual content analysis. In general, on an aggregated level to identify meaningful insights out of data, and to obtain replicable, reliable and valid inferences, researchers often apply a content analysis (Krippendorff 1989). This approach is often used in IS research either to figure out the categories inductively or to classify the information based on pre-defined categories. However, in our current study we neither wanted to categorise the data nor did we classify new information: as mentioned earlier, we wanted to check whether or not a tweet is containing time-relevant information that one can extract using the T-wordlist. Both researchers intensively worked on time-relevant words and discussed rigorously what constitutes temporal information and what does not. However, we felt a pilot study is required before we analyse a sample from all the twelve datasets. For this, we randomly selected a small sample of tweets from four datasets. In the pilot study, one of the researchers analysed a sample dataset individually to assess whether or not tweets are representing temporal information. Later, both researchers looked into the results and discussed them. Again, to ensure validity, in the subsequent stage, both researchers analysed a sample data of around 1,000 tweets from the matched tweets dataset representing different types of disasters. The preliminary results are discussed in the following section.

In the future, as part of a further evaluation of our T-wordlist and also to bench mark it against established toolkits such as TempoWordNet (Ga et al. 2014), we will apply both the T-wordlist and other toolkits on the disaster social media datasets from CrisisLex.org (Olteanu. 2017) and also on the data sets collected by us (such as Chennai floods Twitter data). In order to evaluate the results, we will also employ services from crowdsourcing platforms such as MicroWorker, Mechanical Turk etc. and in this way, we will compare the results and bench mark our T-wordlist against the existing methods and toolkits. After the benchmarking process, we will also enrich the T-wordlist by supplementing with suitable temporal expression words taken from TempoWordNet (Ga et al. 2014) and other relevant sources to make it more robust and useful.

4. Results and future work

As base of our research in progress study, we used our T-wordlist which consists of 470 time-indicating phrases and a Twitter dataset consisting of 11,238 tweets in English language, as has been shown in Table 10. After applying the T-wordlist we extracted the tweets consisting time indicating phrases (Table 11). Only 210 words from the T-wordlist matched with words and phrases in our Twitter dataset. The frequency of words and phrases ranged from a single match to 628 matches in the 4791 tweets that were matched by the T-wordlist. Furthermore, 87 words of our T-wordlist appeared more than 10 times in the matched Twitter dataset. As a next step, we have to cross-check the data to make sure whether the tweet really contains the time-related information and also have to have a look into the tweets which had no time-related information, to check if we have missed important phrases or words.

Table 11. Matching of T-words

Name of dataset/Year	English Tweets	Match (T-word)	Non-Match
Colorado wildfires - 2012	1,163	490	673
Philipinnes floods - 2012	851	307	544
Typhoon Pablo - 2012	821	340	481
Australia bushfire - 2013	1,174	522	652
Bohol earthquake - 2013	837	331	506
Brazil nightclub fire - 2013	407	198	209
Glasgow helicopter crash - 2013	1,079	441	638
Los Angeles airport shootings - 2013	938	450	488
New York train crash - 2013	968	414	554
Savar building collapse -2013	1,107	407	700
Singapore haze - 2013	947	505	442
Typhoon Yolanda - 2013	946	386	560

Table 2. Matching of T-words

Based on our preliminary results, in our future research we mainly will focus on compiling a more complete T-wordlist to achieve more accuracy. To make our wordlist more rigorous and robust, we have to tackle the problem how to deal with changing meanings of words, depending on the context in which they are used. For example, prepositions like, *on* (Monday), *in* (the morning), *at* (night), *by* (11 o'çlock) indicate temporal information. However, the same prepositions might also indicate the position and direction in different context. For example, *in* (the picture), *on* (the left), *at* (a concert), (standing) *by*. In this regard, by taking advantage of the support of one of the crowd sourcing platforms, we plan to make our wordlist more useful and generalizable to use it for different data analyses. Furthermore, we will use the services from crowd sourcing platforms to categorise the words (based on meaning) into past/present/future categories and also to benchmark out T-wordlist against existing temporal tools and toolkit as explained in the validation section.

5. References

Abbasi, A., Chen, H., and Salem, A. 2008. "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," ACM Transactions on Information Systems (TOIS) (26:3), p. 12.

Abdulla, N. A., Ahmed, N. A., Shehab, M. A., Al-Ayyoub, M., Al-Kabi, M. N., and Al-rifai, S. 2016. "Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis," in Big Data: Concepts, Methodologies, Tools, and Applications. IGI Global, pp. 1970-1986.

Allen, J. F. 1983. "Maintaining Knowledge About Temporal Intervals," Communications of the ACM (26:11), pp. 832-843.

Alonso, O., Gertz, M., and Baeza-Yates, R. 2007. "On the Value of Temporal Information in Information Retrieval," ACM SIGIR Forum: ACM, pp. 35-41.

Alonso, O., Strötgen, J., Baeza-Yates, R. A., and Gertz, M. 2011. "Temporal Information Retrieval: Challenges and Opportunities," Twaw (11), pp. 1-8.

Aue, A., and Gamon, M. 2005. "Customizing Sentiment Classifiers to New Domains: A Case Study," Proceedings of recent advances in natural language processing (RANLP): Citeseer, pp. 2-1.

Backfried, G., and Shalunts, G. 2016. "Sentiment Analysis of Media in German on the Refugee Crisis in Europe," International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries: Springer, pp. 234-241.

Beigi, G., Hu, X., Maciejewski, R., and Liu, H. 2016. "An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief," in Sentiment Analysis and Ontology Engineering. Springer, pp. 313-340.

Boiy, E., Hens, P., Deschacht, K., and Moens, M.-F. 2007. "Automatic Sentiment Analysis in on-Line Text," ELPUB, pp. 349-360.

Bradley, M. M., and Lang, P. J. 1999. "Affective Norms for English Words (Anew): Instruction Manual and Affective Ratings," Technical report C-1, the center for research in psychophysiology, University of Florida.

Chambers, N., Wang, S., and Jurafsky, D. 2007. "Classifying Temporal Relations between Events," Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions: Association for Computational Linguistics, pp. 173-176.

Chang, A. X., and Manning, C. D. 2012. "Sutime: A Library for Recognizing and Normalizing Time Expressions," LREC, pp. 3735-3740.

Chaovalit, P., and Zhou, L. 2005. "Movie Review Mining: A Comparison between Supervised and Unsupervised Classification Approaches," System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on: IEEE, pp. 112c-112c.

Deng, Q., Hine, M., Ji, S., and Sur, S. 2017. "Building an Environmental Sustainability Dictionary for the It Industry," Proceedings of the 50th Hawaii International Conference on System Sciences.

Deschacht, K., and Moens, M.-F. 2009. "Semi-Supervised Semantic Role Labeling Using the Latent Words Language Model," in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1. Singapore: Association for Computational Linguistics, pp. 21-29.

Esuli, A., and Sebastiani, F. 2007. "Sentiwordnet: A High-Coverage Lexical Resource for Opinion Mining," Evaluation), pp. 1-26.

Ferro, L., Gerber, L., Mani, I., Sundheim, B., and Wilson, G. 2005. "Tides 2005 Standard for the Annotation of Temporal Expressions,").

Ferro, L., Mani, I., Sundheim, B., and Wilson, G. 2001. "Tides Temporal Annotation Guidelines-Version 1.0. 2," The MITRE Corporation, McLean-VG-USA).

Ga, #235, Dias, l. H., Hasanuzzaman, M., St, #233, Ferrari, p., and Mathet, Y. 2014. "Tempowordnet for Sentence Time Tagging," in: Proceedings of the 23rd International Conference on World Wide Web. Seoul, Korea: ACM, pp. 833-838.

Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. 2013. "Comparing and Combining Sentiment Analysis Methods," Proceedings of the first ACM conference on Online social networks: ACM, pp. 27-38.
Group, T. W. 2009. "Guidelines for Temporal Expression Annotation for English for Tempeval 2010."
Hasanuzzaman, M., Dias, G., Ferrari, S., and Mathet, Y. 2016. "Iterative Tempowordnet," in The Semantic Web: Eswc 2016 Satellite Events, Heraklion, Crete, Greece, May 29 – June 2, 2016, Revised Selected Papers, H. Sack, G. Rizzo, N. Steinmetz, D. Mladenić, S. Auer and C. Lange (eds.). Cham: Springer International Publishing, pp. 40-45.
Kim, S.-M., and Hovy, E. 2004. "Determining the Sentiment of Opinions," Proceedings of the 20th international conference on Computational Linguistics: Association for Computational Linguistics, p. 1367.
Kolomiyets, O., Bethard, S., and Moens, M.-F. 2011. "Model-Portability Experiments for Textual Temporal Analysis," Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2: Association for Computational Linguistics, pp. 271-276.
Krippendorff, K. 1989. "Content Analysis," in International Encyclopedia of Communication, G.G. E. Barnouw, W. Schramm, T. L. Worth, & L. Gross (ed.). pp. 403-407.
Ling, X., and Weld, D. S. 2010. "Temporal Information Extraction," AAAI, pp. 1385-1390.
Liu, B. 2012. "Sentiment Analysis and Opinion Mining," Synthesis lectures on human language technologies (5:1), pp. 1-167.
Loughran, T., and McDonald, B. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," The Journal of Finance (66:1), pp. 35-65.
Madhoushi, Z., Hamdan, A. R., and Zainudin, S. 2015. "Sentiment Analysis Techniques in Recent Works," Science and Information Conference (SAI), 2015: IEEE, pp. 288-291.
Mani, I., and Wilson, G. 2000. "Robust Temporal Processing of News," Proceedings of the 38th annual meeting on Association for Computational Linguistics: Association for Computational Linguistics, pp. 69-76.
Miller, G. A. 1995. "Wordnet: A Lexical Database for English," Communications of the ACM (38:11), pp. 39-41.
Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. 1990. "Introduction to Wordnet: An on-Line Lexical Database," International journal of lexicography (3:4), pp. 235-244.
Ng, J.-P., Chen, Y., Kan, M.-Y., and Li, Z. 2014. "Exploiting Timelines to Enhance Multi-Document Summarization," ACL (1), pp. 923-933.
Nielsen, F. Å. 2011. "A New Anew: Evaluation of a Word List for Sentiment Analysis in Microblogs," arXiv preprint arXiv:1103.2903).
Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S. 2014. "Crisislex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises," ICWSM.
Olteanu., A. 2017. "Crisislex.Org." from http://crisislex.org/
Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," Foundations and trends in information retrieval (2:1-2), pp. 1-135.
Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2001. "Linguistic Inquiry and Word Count: Liwc 2001," Mahway: Lawrence Erlbaum Associates (71:2001), p. 2001.
Pustejovsky, J., Castano, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Katz, G., and Radev, D. R. 2003a. "Timeml: Robust Specification of Event and Temporal Expressions in Text," New directions in question answering (3), pp. 28-34.
Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., and Ferro, L. 2003b. "The Timebank Corpus," Corpus linguistics, p. 40.
Pustejovsky, J., Knippen, R., Littman, J., and Saurí, R. 2005. "Temporal and Event Information in Natural Language Text," Language resources and evaluation (39:2), pp. 123-164.

Read, J. 2005. "Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification," Proceedings of the ACL student research workshop: Association for Computational Linguistics, pp. 43-48.
Read, J., and Carroll, J. 2009. "Weakly Supervised Techniques for Domain-Independent Sentiment Classification," Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion: ACM, pp. 45-52.
Riloff, E., and Wiebe, J. 2003. "Learning Extraction Patterns for Subjective Expressions," Proceedings of the 2003 conference on Empirical methods in natural language processing: Association for Computational Linguistics, pp. 105-112.
Schilder, F., and Habel, C. 2001. "From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages," Proceedings of the workshop on Temporal and spatial information processing-Volume 13: Association for Computational Linguistics, p. 9.
Schwartz, H. A., Park, G., Sap, M., Weingarten, E., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., Berger, J., and Seligman, M. 2015. "Extracting Human Temporal Orientation from Facebook Language," Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 409-419.
SemEval. 2007-2017. "International Workshop on Semantic Evaluation." 2017
Stone, P. J., and Hunt, E. B. 1963. "A Computer Approach to Content Analysis: Studies Using the General Inquirer System," Proceedings of the May 21-23, 1963, spring joint computer conference: ACM, pp. 241-256.
Strapparava, C., and Valitutti, A. 2004. "Wordnet Affect: An Affective Extension of Wordnet," LREC: Citeseer, pp. 1083-1086.
Strötgen, J., and Gertz, M. 2010. "Heideltime: High Quality Rule-Based Extraction and Normalization of Temporal Expressions," Proceedings of the 5th International Workshop on Semantic Evaluation: Association for Computational Linguistics, pp. 321-324.
Strötgen, J., and Gertz, M. 2012. "Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards," LREC, pp. 3746-3753.
Strötgen, J., and Gertz, M. 2013. "Multilingual and Cross-Domain Temporal Tagging," Language Resources and Evaluation (47:2), pp. 269-298.
Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. 2011. "Lexicon-Based Methods for Sentiment Analysis," Computational linguistics (37:2), pp. 267-307.
Talukdar, P. P., Wijaya, D., and Mitchell, T. 2012. "Coupled Temporal Scoping of Relational Facts," Proceedings of the fifth ACM international conference on Web search and data mining: ACM, pp. 73-82.
Temnikova, I., Castillo, C., and Vieweg, S. 2015. "Emterms 1.0: A Terminological Resource for Crisis Tweets," ISCRAM 2015 proceedings of the 12th international conference on information systems for crisis response and management.
Thelwall, M., Buckley, K., and Paltoglou, G. 2011. "Sentiment in Twitter Events," Journal of the American Society for Information Science and Technology (62:2), pp. 406-418.
Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. 2010. "Sentiment Strength Detection in Short Informal Text," Journal of the American Society for Information Science and Technology (61:12), pp. 2544-2558.
Thelwall, M., Wouters, P., and Fry, J. 2008. "Information-Centered Research for Large-Scale Analyses of New Information Sources," Journal of the American Society for Information Science and Technology (59:9), pp. 1523-1527.
UzZaman, N., Llorens, H., Allen, J., Derczynski, L., Verhagen, M., and Pustejovsky, J. 2012. "Tempeval-3: Evaluating Events, Time Expressions, and Temporal Relations," arXiv preprint arXiv:1206.5333).
Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., and Pustejovsky, J. 2007. "Semeval-2007 Task 15: Tempeval Temporal Relation Identification," Proceedings of the 4th International Workshop on Semantic Evaluations: Association for Computational Linguistics, pp. 75-80.

Verhagen, M., Sauri, R., Caselli, T., and Pustejovsky, J. 2010. "Semeval-2010 Task 13: Tempeval-2," Proceedings of the 5th international workshop on semantic evaluation: Association for Computational Linguistics, pp. 57-62.
Wen, M., Zheng, Z., Jang, H., Xiang, G., and Rosé, C. P. 2013. "Extracting Events with Informal Temporal References in Personal Histories in Online Communities," ACL (2), pp. 836-842.
Wilson, T., Wiebe, J., and Hoffmann, P. 2005. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," Proceedings of the conference on human language technology and empirical methods in natural language processing: Association for Computational Linguistics, pp. 347-354.
Yang, C. C., Tang, X., Wong, Y., and Wei, C.-P. 2010. "Understanding Online Consumer Review Opinions with Sentiment Analysis Using Machine Learning," Pacific Asia Journal of the Association for Information Systems (2:3).