

Electronic Theses and Dissertations, 2004-2019

2018

Understanding Crisis Communication and Mobility Resilience during Disasters from Social Media

Kamol Roy

 Part of the [Civil Engineering Commons](#), and the [Transportation Engineering Commons](#)
Find similar works at: <https://stars.library.ucf.edu/etd>
University of Central Florida Libraries <http://library.ucf.edu>

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Roy, Kamol, "Understanding Crisis Communication and Mobility Resilience during Disasters from Social Media" (2018). *Electronic Theses and Dissertations, 2004-2019*. 6200.
<https://stars.library.ucf.edu/etd/6200>

**UNDERSTANDING CRISIS COMMUNICATION AND MOBILITY
RESILIENCE DURING DISASTERS FROM SOCIAL MEDIA**

by

KAMOL CHANDRA ROY
B.Sc. Bangladesh University of Engineering and Technology, 2014

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2018

Major Professor: Samiul Hasan

© 2018 Kamol Chandra Roy

ABSTRACT

Rapid communication during extreme events is one of the critical aspects of successful disaster management strategies. Due to their ubiquitous nature, social media platforms offer a unique opportunity for crisis communication. Moreover, social media usage on GPS enabled devices such as smartphones allow us to collect human movement data which can help understanding mobility during a disaster. This study leverages social media (Twitter) data to understand the effectiveness of social media-based communication and the resilience of human mobility during a disaster. This thesis has two major contributions. *First*, about 52.5 million tweets related to hurricane Sandy are analyzed to assess the effectiveness of social media communication during disasters and identify the contributing factors leading to effective crisis communication strategies. Effectiveness of a social media user is defined as the ratio of attention gained over the number of tweets posted. A model is developed to explain more effective users based on several relevant features. Results indicate that during a disaster event, only few social media users become highly effective in gaining attention. In addition, effectiveness does not depend on the frequency of tweeting activity only; instead it depends on the number of followers and friends, user category, bot score (controlled by a human or a machine), and activity patterns (predictability of activity frequency). *Second*, to quantify the impacts of an extreme event to human movements, we introduce the concept of mobility resilience which is defined as the ability of a mobility infrastructure system to manage shocks and return to a steady state in response to an extreme event. We present a method to detect extreme events from geo-located movement data and to measure mobility resilience and loss of resilience due to those events. Applying this method, we measure resilience metrics from geo-located social media data for multiple types of disasters occurred all over the world. Quantifying mobility resilience may help us to assess the higher-order socio-economic impacts of

extreme events and guide policies towards developing resilient infrastructures as well as a nation's overall disaster resilience.

Keywords: crisis communication; hurricane warning; evacuation; social media; Twitter; hurricane Sandy; disaster management; human mobility; resilience; geo-location data.

ACKNOWLEDGMENT

I would like to convey my heartiest gratitude to my honorable supervisor Dr. Samiul Hasan for his excellent supervision and constant support in this thesis. I would also like to acknowledge the support and encouragement from my family and friends.

TABLE OF CONTENT

LIST OF FIGURES	viii
LIST OF TABLES	x
CHAPTER 1: INTRODUCTION	1
1.1 Introduction	1
1.2 Thesis Contribution	5
1.3 The Objective of the Thesis	6
1.4 Thesis Organization.....	7
CHAPTER 2: LITERATURE REVIEW	8
CHAPTER 3: UNDERSTANDING CRISIS COMMUNICATION EFFECTIVENESS	13
3.1 Introduction and motivation	13
3.2 Data	14
3.3 Methods.....	15
3.3.1 Activity, Attention, and Efficiency Metrics in Twitter.....	15
3.3.2 Extraction of User Features	16
3.3.3 Contributing Features to Efficiency.....	17
3.4 Results	18
3.4.1 Distributions of User Features	18
3.4.2 Correlations between Activity and Attention	20
3.4.3 User Efficiency Analysis	22
3.4.4 User Attributes Contributing to Efficiency.....	28
3.5 Conclusions	35
CHAPTER 4: QUANTIFYING HUMAN MOBILITY RESILIENCE	38
4.1 Introduction	38
4.2 Data and Methodology	39
4.2.1 Extracting Location Time Series of a User.....	42
4.2.2 Displacement Metric.....	42
4.2.3 Extraction of Typical and Actual Displacements Time Series	43
4.2.4 Extreme Event Detection	44
4.2.5 Resilience Calculation	45
4.3 Results	46
4.4 Discussion	52
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS	54

REFERENCES 56

LIST OF FIGURES

Figure 3.1: (a) Activity, (b) Follower, (c) Followee Distributions (d) Bot Score and (e) Activity Entropy.....	19
Figure 3.2: Correlations between Activity and Attention in (a) Pre-Disaster Period (Oct 14, 2012 to Oct 22, 2012), (b) During Disaster (Oct 23, 2012 to Nov 1, 2012) Period and (c) Post Disaster Period (Nov 2, 2012 to Nov 11, 2012).....	21
Figure 3.3: Daily Efficiency Distribution for the Users Categorized by Active Days.	22
Figure 3.4: Average Daily Efficiency of the users categorized by efficiency (for hurricane related tweets).	24
Figure 3.5: Average Daily Efficiency of the users categorized by efficiency values (for Sandy specific tweets).....	24
Figure 3.6: Top 5 Efficient Users during Hurricane Declaration, Landfall and Overall (for hurricane related tweets).	26
Figure 3.7: Top 5 Efficient Users during Hurricane Declaration, Landfall and Overall (for Sandy specific tweets).....	27
Figure 3.8: Correlations between Efficiency and Different User Attributes (for hurricane related tweets).....	30
Figure 3.9: Relationship between Efficiency and Different User Attributes (for sandy specific tweets).....	31
Figure 4.1: Resilience and Resilience Loss Calculation. (a) Resilience Triangle (adopted from (Bruneau et al., 2003)), (b) Human Mobility Resilience (Decreased movement), (c) Human Mobility Resilience (Increased movement).	45

Figure 4.2: Resilience and Resilience Losses for Multiple Disasters. Note: DPU= Displacements Per User (Kilometer), TF= Trip Frequency 50

LIST OF TABLES

Table 3.1: Top 5 efficient users with user type and bot score during hurricane declaration and hurricane landfall. This analysis is based on the unfiltered (hurricane related tweets) data.	28
Table 3.2: Top 5 efficient users with user type and bot score during hurricane declaration and landfall. This analysis is based on the data which have the word sandy in the tweet.....	28
Table 3.3: Model Results for hurricane related tweets	33
Table 3.4: Model Results for Sandy specific tweets.....	34
Table 4.1: Data Description	40
Table 4.2: Comparison of Resilience, Resilience Loss and Recovery Time for Multiple types of Events Occurred in Different Location.....	51

CHAPTER 1: INTRODUCTION

1.1 Introduction

Extreme weather events have become common in recent decades (NELSON, 2013). Globally, natural disasters cause \$520 billion equivalent loss and are responsible for taking 26 million people below poverty line (The World Bank, 2016). Since 1980, the United States alone has spent more than \$1.5 trillion for managing 219 weather and climate related disasters which had overall damages exceeding \$1 billion (NOAA National Centers for Environmental Information (NCEI) U.S., 2018). Hurricanes along with other natural disasters in 2017 are expected to cause 135 billion US dollar insured cost (Munich RE, 2018). Effective disaster management plays a critical role in reducing the cost of a disaster with implications in its four phases including mitigation, preparedness, response and recovery operations. Households and communities require appropriate resources to meet different needs in these phases of disaster management (Comfort et al., 2004). To mitigate loss of lives and infrastructure damage, proper preparedness and organized response strategies are crucial. Information availability about the time and severity of an incident can greatly help disaster preparedness, response and recovery operations. Particularly for responding organizations, effective information sharing and coordination are critical (Bharosa et al., 2010; Yates and Paquette, 2011). Access to information enhances the efficiency of response actions and increases coordination throughout the network of responding organizations (Comfort et al., 2004).

Online social media platforms facilitate fast and easy exchange of information through sharing, discussion, and communication producing a huge amount of digital content (Huang et al., 2010). Social media data has been used to investigate many research topics such as human mobility (Hasan et al., 2013c; Hasan and Ukkusuri, 2014), transportation (Chen et al., 2017, 2014; Ni et al.,

2017; Rashidi et al., 2017; Zhang et al., 2018), tourism and hospitality (Leung et al., 2013), public health (Kass-Hout and Alhinnawi, 2013), disaster management (Huang and Xiao, 2015; Kryvasheyeu et al., 2016; Simon et al., 2015) and so on. Online social media can play a vital role in spreading timely updates about emergency and collecting feedback from the affected population. Emergency evacuation plan such as evacuation timing, mode and route choice depends on many different factors (Hasan et al., 2013a; Murray-Tuite and Wolshon, 2013; Sadri et al., 2014, 2013). Information from social media about situational awareness can influence shaping these decisions (Martín, Yago, Zhenlong Li, 2017).

Thus, a wide range of international, state, and local organizations have successfully used social media tools during disasters gaining broader interests among policy makers on how social media might be used to improve disaster response and recovery capabilities (Lindsay, 2011). For instance, during Hurricane Sandy, social media played an important role by sharing information, when the affected regions had limited access to traditional media (Kaufman, S., C. Qing, N. Levenson, 2012).

When using social media for information dissemination during disasters, it is critical to know what makes an information provider more effective. However, studying the effectiveness of social media users in disseminating information has been a challenging task. Such a study would require appropriate metrics applied over a large collection of disaster communication data. On the other hand, essential components of social communication such as human choices, disaster warning propagation and risk communication in large-scale social networks cannot be reproduced within the limits of typical social experiments.

Although social media data for disaster management has been used mainly for situational awareness and crisis communication that mostly covers temporal and contextual dimension, geo-tagged data offer collecting user needs, concerns and mobility traces spatially.

Human mobility analysis has drawn much attention in many research fields for its wide applications. Most of the studies have modeled mobility as probability distributions of the length of the traveled distance and the waiting time between any two displacements. Analyzing a wide range of data sets, studies have established that human mobility is not random rather it follows some specific patterns (Alessandretti et al., n.d.; Brockmann et al., 2006; Gonzalez et al., 2008; Jurdak et al., 2015). For instance, human mobility has been studied using large-scale trajectory datasets including bank notes (Brockmann et al., 2006), taxi data (Wang et al., 2015; Yao and Lin, 2016), GPS observations (J. Tang et al., 2015), Wi-Fi (Alessandretti et al., n.d.), cell phone call recordings (Deville et al., 2016; Song et al., 2010), and social media posts (Hasan et al., 2013c; Hasan and Ukkusuri, 2014; Rashidi et al., 2017). These studies have found that mobility follows power-laws (Beir?? et al., 2016; Brockmann et al., 2006; Deville et al., 2016; Gonzalez et al., 2008; Han et al., 2011; Hawelka et al., 2014; Noulas et al., 2012; Song et al., 2010; Vaca et al., 2014; Yao and Lin, 2016; Zhao et al., 2015b), log-normal (J. Tang et al., 2015; Wang et al., 2015), exponential distribution (Gallotti R, Bazzani A, 2016; Liang et al., 2012; Liu et al., 2015, 2014; Wu et al., 2014; Zhao et al., 2015a) or a combination of power-law and exponential distributions (Gallotti R, Bazzani A, 2016; Liu et al., 2015).

During extreme events, human mobility goes through a significant perturbation compared to regular periods. People are less likely to move the same way in emergency situations, such as a hurricane, typhoon, earthquake and other natural or manmade extreme events, as they do in normal

conditions. Understanding this perturbation will increase the effectiveness of disaster preparedness, information communication, reduce fatalities, and minimize economic losses (Wang and Taylor, 2016, 2014). Despite its importance, few studies have investigated human mobility under disasters. Although studies have investigated how individuals behave during an extreme event (Hasan et al., 2013b, 2011; Mesa-arango et al., 2013; Sadri et al., 2015, 2014, 2013), they are mainly based on post-disaster surveys with limited sample size. Based on these survey data, it is impossible to compare pre and post disaster human movements and measure mobility resilience at a system scale. Resilience is commonly used to indicate the ability of a system or entity to return to its normal state after a disruption due to a disaster event (Hosseini et al., 2016). To assess resilience, depending on the fields and events, both qualitative (Alliance, 2007; Kahan, Jerome H., Andrew C. Allen, n.d.; Speranza, Chinwe Ifejika, Urs Wiesmann, n.d.) and quantitative (Bruneau et al., 2003; McCallum et al., 2016; Nicholson, C. D., K. Barker, n.d.) approaches exist. While it has been widely studied for physical infrastructure systems, resilience of socio-economic systems is hard to quantify. Human mobility is a key factor to understand the impacts of disasters to our social and economic activities since socio-economic development is strongly associated with mobility (Pappalardo et al., 2015).

Thus, data unavailability is one of the main constraints of observing human movements during extreme events. While high resolution mobile phone calls, transit systems transactions, and GPS coordinates can provide us richer information on human mobility during disasters, these proprietary datasets are not widely available due to privacy concerns. Social media data can offer a promising direction in observing human movements during extreme events. A method that can quantitatively measure perturbations and recovery times will greatly impact disaster management

as well as in policy making towards building disaster resilient infrastructures, communities, and cities.

While disaster resilience has been studied in many fields, understanding mobility resilience under disasters is a relatively new topic. Donovan et. al. (Donovan and Work, 2017) have studied transportation system resilience for the New York City using taxi GPS data for multiple disasters. Recent studies (Qi and John E., 2014; Wang and Taylor, 2016, 2014) have shown that under disaster events human mobility goes through perturbation but still follows the same distributions similar to the ones in a steady state, and the shift in the center of mass and radius of gyration in a perturbed state are correlated with the steady state radius of gyration. Although, these studies have suggested that human mobility is somewhat resilient to disasters, a quantitative assessment of mobility resilience is still missing in the literature. Furthermore, these studies did not explore the expected correlations of mobility resilience across different types of extreme events.

1.2 Thesis Contribution

This thesis has made several contributions to disaster management. This study assesses social media-based communication efficiency in terms of attention gaining for activity in social media. We also investigate what are the contributing factors for efficiency. We develop models to identify the contributing factors and to predict the efficient users from the features. Adopting the contributing features for higher efficiency can lead to faster disaster communication by gaining more attention to the situation specific information. Another part of this thesis uncovers the human mobility resilience during multiple types of disasters. We present a method to detect extreme events from geo-located movement data and to measure mobility resilience and loss of resilience due to those events. Applying this method, we measure resilience metrics from geo-located social

media data for multiple types of disasters occurred all over the world. Quantifying mobility resilience may help us to assess the higher-order socio-economic impacts of extreme events and guide policies towards developing resilient infrastructures as well as a nation's overall disaster resilience.

1.3 The Objective of the Thesis

The main objective of this thesis is to apply social media data in disaster management. We focus on crisis communication during hurricane sandy in terms of attention gaining efficiency of the users. Our objective is to find out how activity frequency and activity pattern affect in attention gaining during a disaster. And what are the other factors that contribute to the efficiency in gaining attention during a disaster. To be specific, our objective is to answer the following research questions:

- Does a more active social media user gain more attention? What combinations of user activity will facilitate such attention in pre-disaster, during disaster and post-disaster periods? This study allows us to understand the correlation between activity and attention gained during these three phases of a disaster.
- How does user efficiency dynamics change over the pre, during and post disaster phases? We show that during disaster specially in hurricane declaration and landfall days users have higher average efficiency than that of pre-disaster and post disaster period.
- What are the factors contributing to user efficiency? How can efficient users be classified based on their activities and features? We present a model to classify efficient users highlighting the features contributing to user efficiency in disaster periods.

Another objective is to find human mobility resilience in response extreme events using location based social media data. To be specific we answer the following research questions:

- What is the definition of resilience in the context of human mobility? We define resilience for human mobility that can be calculated for geo-tagged data.
- How does human mobility resilience vary for different types of disaster? We calculate human mobility resilience for multiple types of disaster such as hurricane, earthquake, snowstorm etc.
- How does human mobility resilience vary in response to events of different intensities? We quantify human mobility resilience for different intensities of hurricane and earthquakes.

1.4 Thesis Organization

The rest of the thesis is organized as follows: Chapter 2 provides the literature review on social media usage on disaster management focusing on crisis communication and human mobility. Chapter 3 provides the data description, analysis, methodology and result to understand crisis communication efficiency during hurricane sandy. Chapter 4 describes the data description, methodology and result to quantify human mobility resilience during multiple types of disaster using social media. Chapter 5 presents the summary and conclusions of the thesis.

CHAPTER 2: LITERATURE REVIEW

Social media, the computer mediated technology is now one of the most integrated parts of our daily life. These technological advancements have transformed the view of disaster management professionals on disseminating information as well as interacting with the affected communities. Having a strong social network increases the likelihood of a person responding to a warning message (Aguirre et al., 1998). During crisis, warning message like evacuation decisions can be made anywhere and often with little advanced warning time (Murray-Tuite and Wolshon, 2013). Eye-witnessed information sources provide local and rapid updates during disaster and thus can be more helpful than official news for the decision makers (Palen et al., 2009; Shklovski et al., 2008).

Researchers have used social media in disasters from different perspectives (Kim and Hastak, 2018; Stieglitz et al., 2018). Studying 2013 Oklahoma tornado, it is shown that Twitter data can reveal relevant information as an additional data source for better understanding of individual behavior during a crisis (Ukkusuri et al., 2014). Visual analytics of microblog data can display public behavior in disaster events (Chae et al., 2014). Mobility patterns can be inferred from geo-tagged tweets (Hasan et al., 2013c; Hasan and Ukkusuri, 2014; Sadri et al., 2017a). Communities can be detected from user interactions on Twitter (Hasan et al., 2013c; Sadri et al., 2017a). Social media users can be used as social network sensors to increase disaster awareness (Kryvasheyev Y, Chen H, Moro E, Van Hentenryck P, 2015). Furthermore, social media data can be used to rapidly assess disaster damage, as it was shown that per capita damages were strongly correlated with per capita twitter activity during hurricane Sandy (Kryvasheyev et al., 2016).

Activities of social media users are greatly influenced by content production and sharing activities (Vaca et al., 2014). Most of the past studies focused on the popularity or propagation of the content in social media such as popular tweets (Hong et al., 2011; Mathioudakis et al., 2010), Flickr picture (Cha et al., 2009), YouTube video (Figueiredo et al., 2011), Twitter hashtag (Lehmann et al., 2012) etc. Scale free networks and affinity affect the propagation of information (Wu et al., 2004) but basic measures such as the raw number of social connections are not a good predictor for influence (Asur et al., 2011; Romero and Huberman, 2011). In addition to the graph properties of user networks, the popularity and influence of a twitter account depend on the personality and emotion of the human being behind that account (Quercia et al., 2011). Stai et al. (Stai et al., 2018) proposed an epidemic model to understand temporal dynamics of information diffusion in Twitter, explaining the burst like behaviors due to information diffusion (Myers and Leskovec, 2014). Vaca et al. (Vaca et al., 2014) observed that a combination of different type of social and content-producing activity is necessary to attract attention in social media. Using Sina-Weibo data during two hurricanes, this study (Dong et al., 2018) explores the information diffusion considering individual and network perspective. Analyzing the reposting behavior in Weibo.com during Yiliang earthquake, Li et. al have studied the propagation pattern of different types of information(Kim and Hastak, 2018). Kim et. al have analyzed the network characteristics of city of Baton Rouge Facebook page during 2016 Louisiana flood(Kim and Hastak, 2018). They have found higher information diffusion in Facebook than Twitter. A study(Kim et al., 2018) on storm Cindi using twitter data explores the role of four types of twitter users in emergency information diffusion. According to this study, news and weather agencies are the dominant twitter users as information sources whereas the public and organizations are the dominant twitter users as

information diffusers. Despite these efforts, what factors contribute to attract attention in social media during a disaster, remains an open question.

Another important aspect during disaster is human mobility because disaster limits the ability to move or create increased movement. Human mobility analysis has drawn much attention in many research fields for its wide applications. Most of the studies have modeled human mobility as probability distributions of the length of the traveled distance and the waiting time between any two displacements. Recently, human mobility has been studied using large-scale trajectory datasets including bank notes (Brockmann et al., 2006), taxi data (Wang et al., 2015; Yao and Lin, 2016), GPS observations (J. Tang et al., 2015), Wi-Fi (Alessandretti et al., n.d.), cell phone call recordings (Deville et al., 2016; Song et al., 2010) and social media posts (Hasan et al., 2013c; Hasan and Ukkusuri, 2014; Rashidi et al., 2017). These studies have found that mobility follows power-laws (Beir?? et al., 2016; Brockmann et al., 2006; Deville et al., 2016; Gonzalez et al., 2008; Han et al., 2011; Hawelka et al., 2014; Noulas et al., 2012; Song et al., 2010; Vaca et al., 2014; Yao and Lin, 2016; Zhao et al., 2015b) , log-normal (J. Tang et al., 2015; Wang et al., 2015), exponential distribution (Gallotti R, Bazzani A, 2016; Liang et al., 2012; Liu et al., 2015, 2014; Wu et al., 2014; Zhao et al., 2015a) or a combination of power-law and exponential distributions (Gallotti R, Bazzani A, 2016; Liu et al., 2015). Alessandretti et. al. (Alessandretti et al., n.d.) have found that both displacement and waiting time are best described by log-normal distribution and only for higher displacement and higher values of waiting time pareto distribution fitted better than log-normal. Exploration and preferential return were the two main principles for the individual mobility model developed by Song et al.(Song et al., 2010).

People are less likely to behave the same way in emergency situations such as a hurricane, typhoon, earthquake and other natural or manmade extreme events, as they do in normal conditions. Studies have investigated how individuals behave during an extreme event (Hasan et al., 2013b, 2011; Mesa-arango et al., 2013; Sadri et al., 2015, 2014, 2013). However, these studies are based on post-disaster surveys with limited sample size. Based on these survey data, it is impossible to compare pre and post disaster human movement patterns and measure mobility resilience at a system scale.

Data unavailability is one of the main constraints of observing human movements during extreme events. While high resolution mobile phone calls, transit systems transactions and GPS coordinates data can provide us richer information on human mobility during disasters, these proprietary datasets are not easily available due to privacy concerns. However, social media data can offer a promising direction in observing human movements during extreme events.

Increasing use of social media in disasters indicates its potential use as a communication and disaster management tool during extreme situations (Simon et al., 2015). Many studies have analyzed emerging social media data for understanding human behavior during disasters. Social network members have the potential to perform as early warning sensors and public sentiment sensing of social media posts can help detecting and locating disasters (Kryvasheyev Y, Chen H, Moro E, Van Hentenryck P, 2015). Evacuee behavior and evacuation compliance during disasters have been investigated using social media data (Arif Mohaimin Sadri; Satish V. Ukkusuri, Ph.D., M.ASCE; Pamela Murray-Tuite, Ph.D., M.ASCE; and Hugh Gladwin, 2014; Fry and Binner, 2015; Martín, Yago, Zhenlong Li, 2017; Sadri et al., 2017d, 2014). Thus, social media can play an

important role in collecting disaster information and help in making successful disaster management plans (Keim and Noji, 2010; Sadri et al., 2017a, 2017e; Z. Tang et al., 2015).

While disaster resilience has been studied in many fields, understanding mobility resilience under disasters is a relatively new topic. Donovan et. al. (Donovan and Work, 2017) have studied the transportation system resilience for the New York City using taxi GPS data for multiple disasters. Recent studies (Qi and John E., 2014; Wang and Taylor, 2016, 2014) have shown that under disaster events human mobility goes through perturbation but still follows the same distributions similar to the ones in a steady state, and the shift in the center of mass and radius of gyration in a perturbed state are correlated with the steady state radius of gyration. However, they did not find the expected correlations for some extreme cases which need further investigations. Although, these studies have suggested that human mobility has some resilience in disaster, they did not make any quantitative assessment of resilience such as the recovery time and the deviation from a steady state.

CHAPTER 3: UNDERSTANDING CRISIS COMMUNICATION EFFECTIVENESS

3.1 Introduction and motivation

Increasing use of social media in disasters will require a better understanding of the effectiveness of information spreading to an affected community. Finding the factors of information spreading is crucial for understanding the dynamics of social media systems. A better understanding of the underlying factors will provide insights into effective crisis communication strategies. Users more efficient in spreading information can play an important role during crisis, since user activities can draw a significant amount of attention to relevant topics/content from other users. Understanding the interplay among user activities, network properties and the attention received will help to identify the contributing factors in successful crisis communication in emergency situations (Sadri et al., 2017b, 2017c). Thus, understanding the influence of social media users has significant implications in a disaster management context.

Although influence of social media users has been studied in many different contexts, efficiency of information/awareness spreading in a disaster context still needs to be investigated. Previous studies have focused on the popularity of content instead of analyzing the effects of user behaviors on how other users respond to them (Vaca et al., 2014). Moreover, to the best of our knowledge, few studies have considered user categories and activity patterns while measuring the efficiency of information spreading in the context of disaster management. In social media dynamics, information diffusion creates sudden bursts of connections (e.g., friends or followers) by creating new edges or deleting existing edges (Myers and Leskovec, 2014). Similarly, during a disaster, information diffusion about situational awareness drives significant changes in the underlying social media connections of friends and followers. Such bursts in new followers may

happen due to common interest (textual similarity) (Myers and Leskovec, 2014) or attention (Vaca et al., 2014) to users or information source. In this study, using Twitter data, we analyze the efficiency of social media users in information spreading in the context of hurricane Sandy. We investigate user activity against the attention gained in pre-disaster, during (warning and response phase) and post-disaster periods. The efficiency of a user is defined as the ratio between attention gained over the number of tweets within a period. In this study, Twitter data before, during and after hurricane sandy have been analyzed to understand the factors contributing to the overall efficiency of a user in crisis communication. A model is also proposed to classify efficient users based on their attributes. This method has potential to be used to identify effective social media users during disasters for rapid communications.

3.2 Data

Hurricane Sandy, a late season post-tropical cyclone was the deadliest and most destructive hurricane of the 2012 Atlantic hurricane season. On October 20, Sandy's origin was primarily associated with a tropical wave that was assessed as a high potential for it to become a tropical cyclone within 48 hours (Blake et al., 2013). The hurricane was first classified and officially assigned its name as Sandy on October 22 (Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, 2015). After leaving a trail of damage over Jamaica, Cuba, and Bahamas, Sandy made its landfall on the United States at 23:30 UTC on 29 October 2012 near Brigantine, New Jersey. Sandy was responsible for 147 direct fatalities and damage in excess of \$50 billion, including 650,000 destroyed or damaged buildings (Blake et al., 2013). Sandy received a lot of media coverage both in traditional media and social media.

The dataset was collected from the publicly accessible data via doi:10.5061/dryad.15fv2 (DRYAD repository) (Kryvasheyev Y, Chen H, Moro E, Van Hentenryck P, 2015). Kryvasheyev et al. (Kryvasheyev Y, Chen H, Moro E, Van Hentenryck P, 2015) collected the dataset through an analytics company Topsy Labs who deals with Twitter data. The collected data contains tweets posted between October 15, 2012 and November 12, 2012. This duration covers the period before the formation of the hurricane to after the landfall in the United States. The dataset contains user id, timestamp, tweet text, tweet id, user followers count, user friends count, sentiment scores and locations. In total, there were 52,493,130 tweets from 13,745,659 unique twitter users.

3.3 Methods

3.3.1 Activity, Attention, and Efficiency Metrics in Twitter

We define *attention* as the number of new followers received and *activity* as the number of tweets or new followee added. For this study, tweet frequency is selected as an activity metric since this is the most frequent activity among all types of users; whereas followee addition is very low or zero for some organizational and personal users. A well-connected user (high initial followers in our study) has a large audience, thus a tweet posted by a well-connected user can reach to many users. But the existing followers may not be the targeted users during an emergency, thus not creating sudden burst in new connections (friends, followers etc.) as study(Myers and Leskovec, 2014) shows that information diffusion creates sudden burst in new connections. For that reason, we have used new follower gain as attention and existing followers as one of the factors. To measure the performance of the users in gaining attention, we use a metric called as *efficiency*. As shown in Equation (3.1), efficiency η of a user u for the time frame $(t_i \text{ to } t_j)$ is defined as the ratio between total attention received and total activity performed within that time frame.

$$\eta_u(t_i, t_j) = \frac{\sum_{k=t_i}^{t_j} att_k(u)}{\sum_{k=t_i}^{t_j} act_k(u)} \quad (3.1)$$

where $att_k(u)$ and $act_k(u)$ represent, in time period k by user u , attention gained and activities posted, respectively.

Although, equation similar to (3.1) are commonly used in fields like physics and economics, Vaca et al. (Vaca et al., 2014) used this term in a social media setting. Unlike most of the fields where efficiency is upper bounded to 1, it can take any value. Higher efficiency values indicate better engagement and higher influence in social media communication.

3.3.2 Extraction of User Features

From raw data, for all the tweets of each unique user activity frequency, initial follower count, initial followee count, total follower received, total followee added by the user and efficiency metrics were computed for a selected time interval. To measure the regularity of a user's activity approximate entropy of activities has been estimated. Approximate entropy is a statistical parameter that can quantify the predictability or regularity of a time series data. A repetitive pattern of fluctuation in a time series makes it more predictable than a time series without such patterns. Approximate entropy calculates the likelihood that similar patterns of observation will not found in the data in the subsequent observations. Thus, a higher value of approximate entropy implies less regularity and a smaller value indicates strong regularity (Kim et al., 2005). Approximate entropy has been used in many fields such as medical data (Srinivasan et al., 2007), finance (Pincus and Kalman, 2004), psychology (Pincus and Goldberger, 1994), complex system analysis (PINCUS, 1991) etc. Daily activity frequencies of a user for the whole analysis period were used as an input. Approximate entropy was best fitted as it has low computational demand, applicable

on small observation (points < 50) and can be applied in real-time. The detailed procedure of computing approximate entropy can be found in this study (Srinivasan et al., 2007). Furthermore, since a significant number of users are being operated autonomously (bot) (Ferrara et al., 2014), we have collected the bot score using [truthy botornot-python](#) API to evaluate whether a user account is controlled by human or machine (Davis et al., 2016).

3.3.3 Contributing Features to Efficiency

To find the linear relationship between different variables and the outcome variable, univariate and multivariate linear regressions were fitted with the extracted variables. The general form of such models is shown in Equations (3.2) -(3.4).

$$Y = \theta_0 + \theta_1 X + \varepsilon \quad (3.2)$$

$$Y = \theta_0 + \theta_1 X + \theta_2 X^2 + \theta_3 X^3 + \varepsilon \quad (3.3)$$

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \dots + \theta_k X_k + \varepsilon \quad (3.4)$$

Here Y is efficiency, treated as a dependent variable; X, X₁, X₂ etc. are the independent variables affecting efficiency; θ_0 is a constant term; and $\theta_1, \theta_2, \theta_3$ are the coefficients of the corresponding variables. While univariate linear regression describes the relationship of each independent variable with the dependent variable, a multiple linear regression model reveals the relationship of the combined effect of the explanatory variables. The independent variables in the best models are considered as the most influential and explanatory variables in determining efficiency. The best model is selected based on the adjusted R squared value.

Users are categorized based on their aggregate efficiency during the whole period. Besides understanding the effect of predictor variables in continuous change in efficiency, we estimate an

ordered logit model to understand the effect of the extracted features in predicting the category of the efficiency of a user. An ordered logit or proportional odd model was chosen as the outcome variable is ordered from low efficiency to high efficiency. This model gives the output as the probability or odd of falling an outcome in an efficiency category. The basic equation (Derr, 2013; Torres-reyna, 2012; Washington, S.P., Karlaftis, M.G. and Mannering, 2010) for interpreting this model is given in Equation (3.5).

$$\log \left[\frac{p_i}{1 - p_i} \right] = a_i + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_kx_k \quad (3.5)$$

where, p_i = probability of an outcome $\leq i$

a_i = intercept for outcome $\leq i$

b_1, b_2 , etc. are the co-efficient whereas x_1, x_2, x_3 are the explanatory independent variables. The best model is selected based on its AIC value.

3.4 Results

3.4.1 Distributions of User Features

This section describes the features collected for a user. Figure 3.1 shows the distributions of activity, followers followees, bot score, and activity entropy found in the data. Both X axis and Y axis are plotted in log scale for (a), (b), and (c). Counter cumulative probability (CCDF) is plotted in Y axis which represents the probability of a value x of being greater than the corresponding value in X axis. Both X axis and Y axis are plotted in normal scale for (d) and (e).

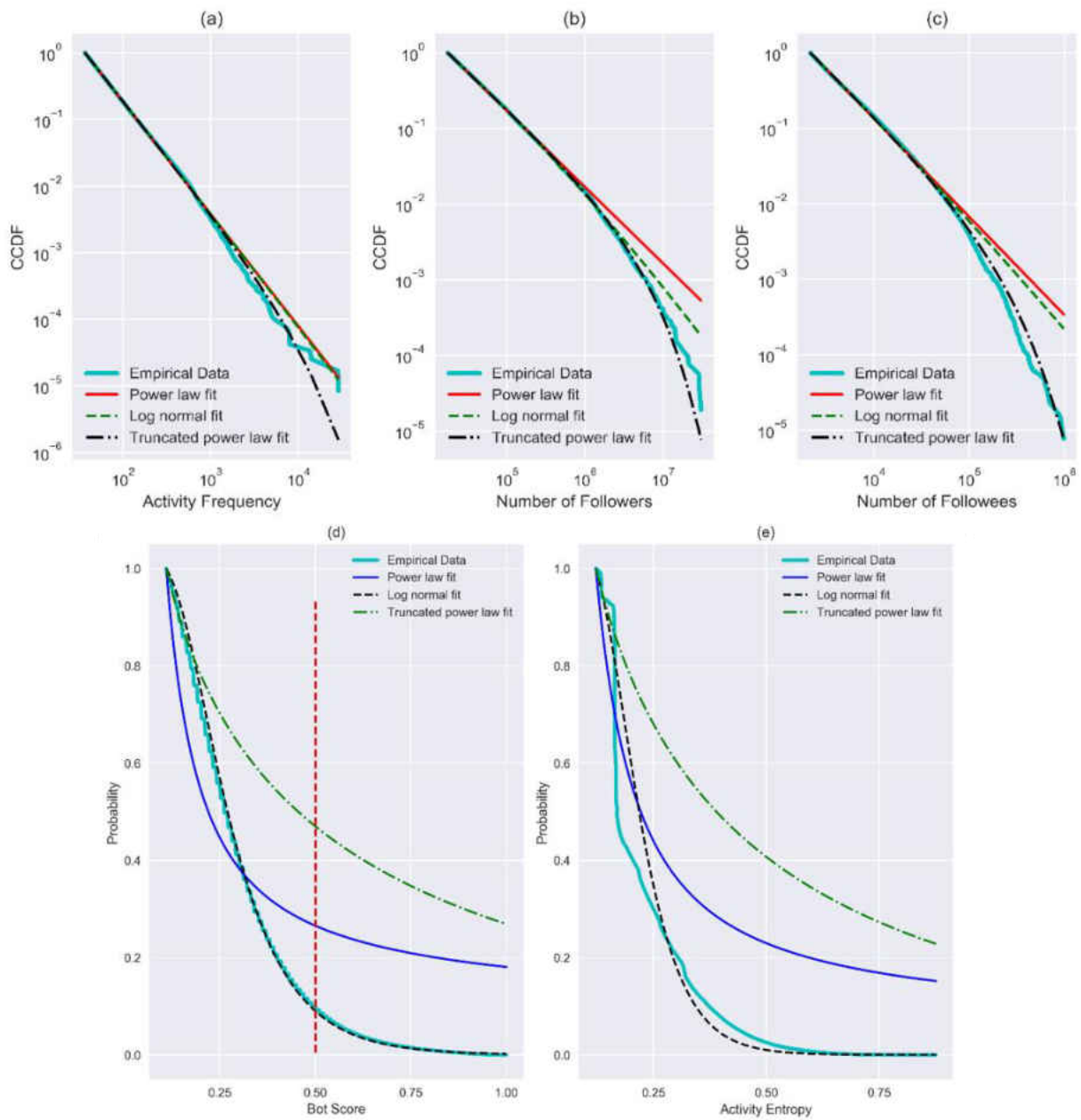


Figure 3.1: (a) Activity, (b) Follower, (c) Followee Distributions (d) Bot Score and (e) Activity Entropy.

Here the follower and followee counts are based on the counts when a user was first observed in the data set. The activity distribution is based on the total number of tweets observed during the

whole period in our dataset. Bot score and activity entropy distribution are plotted using values from 646,563 users. A bot score represents the likelihood of being a bot. An extreme value (0 or 1) represents more confidence of the bot-ness of the user. Higher the value, higher the likelihood of being a bot. In our study, we have used 0.5 as a threshold to separate bot-like behavior.

The empirical distributions of activity, initial followers and followees were best fitted to truncated power law among the fitted distributions shown in Figure 3.1(a), (b) and (c). For bot score and activity entropy (Figure 3.1 (d) and (e)), empirical distributions are best fitted to log-normal distribution. Log likelihood ratio tests were used to find the goodness of fit for the fitted power law, lognormal and truncated power law distributions.

3.4.2 Correlations between Activity and Attention

Activity frequency and followee added play roles in gaining attention. Attention gains may also vary over time and context. Figure 3.2 shows attention gains for different ranges of activities in pre, during and post disaster periods. Though the duration of pre (9 days), during (10 days) and post (10 days) periods disaster are almost same, followers received is the highest during disaster period compared to other two periods (compare the maximum values of the z scales in Figure 3.2).

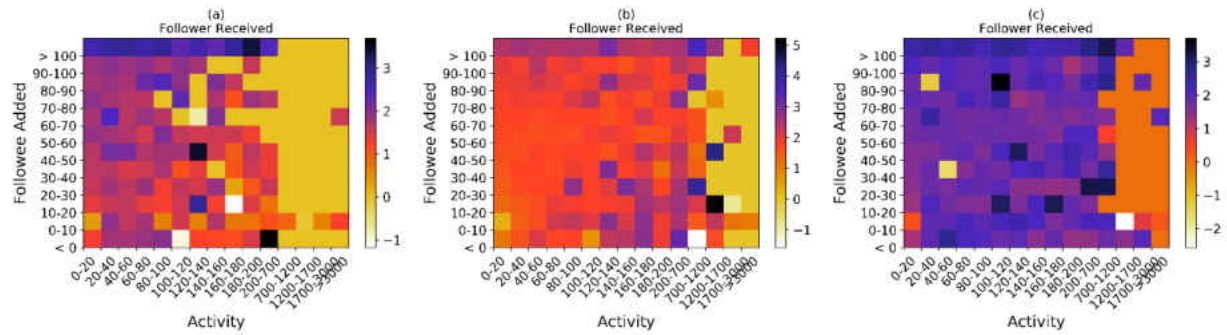


Figure 3.2: Correlations between Activity and Attention in (a) Pre-Disaster Period (Oct 14, 2012 to Oct 22, 2012), (b) During Disaster (Oct 23, 2012 to Nov 1, 2012) Period and (c) Post Disaster Period (Nov 2, 2012 to Nov 11, 2012).

The X-axis and Y-axis represent the range of number of tweets and number of followers, respectively. The Z axis shows the average number of followers received by the users falling in the corresponding x and y bin as color intensity. The Z axis values are shown in log scale.

In general, we do not observe any particular trend between activity frequency and gaining attention in the pre and post-disaster periods. Unlike pre and post disaster periods, during the disaster, higher activities tend to help gaining higher attention. Very high activity frequencies (activity>700) in pre and post disaster periods are not necessarily associated with a high number of followers received. During pre and post disaster periods, users with activity frequency less than 700 have received the highest number of followers (black rectangles in Figure 3.2); whereas during the disaster, the highest number of followers was gained for a user with activity frequency greater than 700. Follower added less than 100 has no impact in gaining attention but users adding followees greater than 100 have received higher attention during all the three phases. Another observation is that, the highest number of followers received in three phases occurred for the users with activities greater than 100. To study more in depth, we have studied user daily and aggregate efficiency based on different features and categories at different phases.

3.4.3 User Efficiency Analysis

Daily efficiency of a user is calculated by dividing the total follower gain by the total number of tweets of that day. Similarly, aggregate efficiency is calculated by dividing total daily follower gain by the total tweet in that period. Users are categorized based on their active days. Only the users who had at least one activity on each of the three periods (pre, during, post disaster) are selected in these categories. Figure 3.3 shows the daily efficiency distribution for the users categorized by their active days. It is found that a significant number of users have daily efficiency value equal to or less than zero. As the number of active day increases, the probability of having a user with efficiency less than or equal to zero decreases. It also shows that the probability of having efficiency less than or equal to zero is maximum for the users who were active less than 8 days. Although, the probability of having daily efficiency greater than 10 is low across all category of users, this probability increases as the number of active days for a user increases.

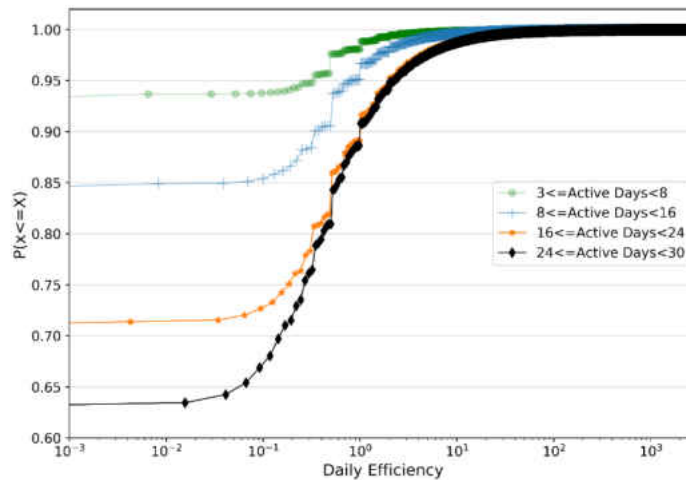


Figure 3.3: Daily Efficiency Distribution for the Users Categorized by Active Days.

X-axis shows the daily efficiency and Y-axis shows the cumulative probability which means the probability of being daily efficiency is equal or less than of the corresponding daily efficiency of X-axis. X-axis are plotted in log scale.

We further investigate, for each user, how daily efficiency varies over time. To find if there is any trend in the data, we have categorized users based on their overall efficiency values (i.e., ratio of sum of daily attention and activity measured over the whole observation period). We calculate average daily efficiency by taking the average of the daily efficiencies of the users for a particular category. For hurricane related tweets, average daily efficiency for the first two categories (overall efficiency less than or equal to zero) does not change that much (see the inset plot at figure 3.4). But daily efficiency for the highly efficient users provide an interesting insight. Users had higher values of average daily efficiency during hurricane declaration and landfall days (figure 3.4). The spikes on hurricane declaration and landfall days indicate that some users received higher attention for their activities on those days. But this trend shows a significant number of spikes even before the formation of Sandy and also long after its landfall. This indicates that some users might be gaining attention due to tweets unrelated to Sandy. To confirm, we analyze only Sandy related tweets (having 'sandy' within the text of the tweet) and found that efficiency was maximum just after the declaration day (October 23,2012) and decayed readily with a spike at landfall date (see Figure 3.5). It indicates that users were highly effective spreading the awareness about Sandy on the day after its declaration. We do not observe any efficiency curve before declaration because the term 'sandy' was not present before declaration.

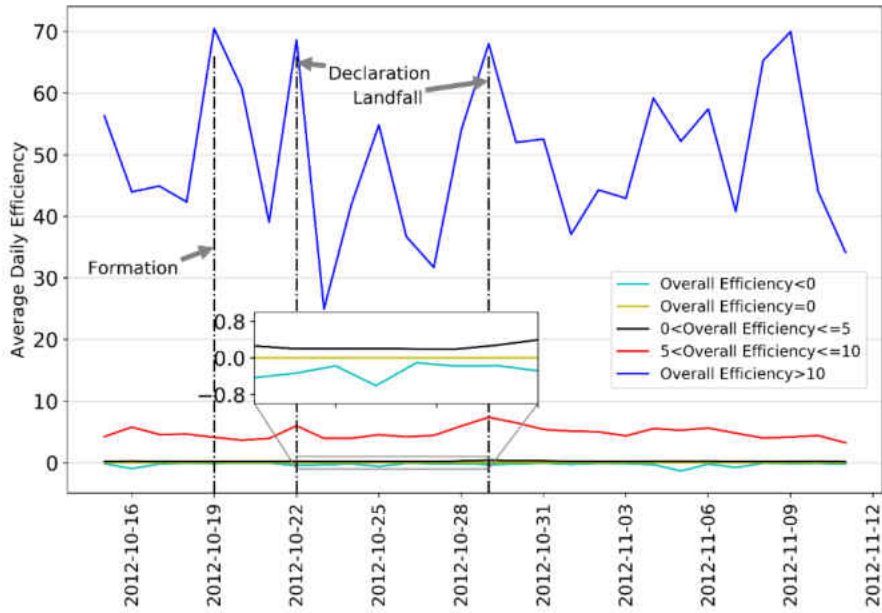


Figure 3.4: Average Daily Efficiency of the users categorized by efficiency (for hurricane related tweets).

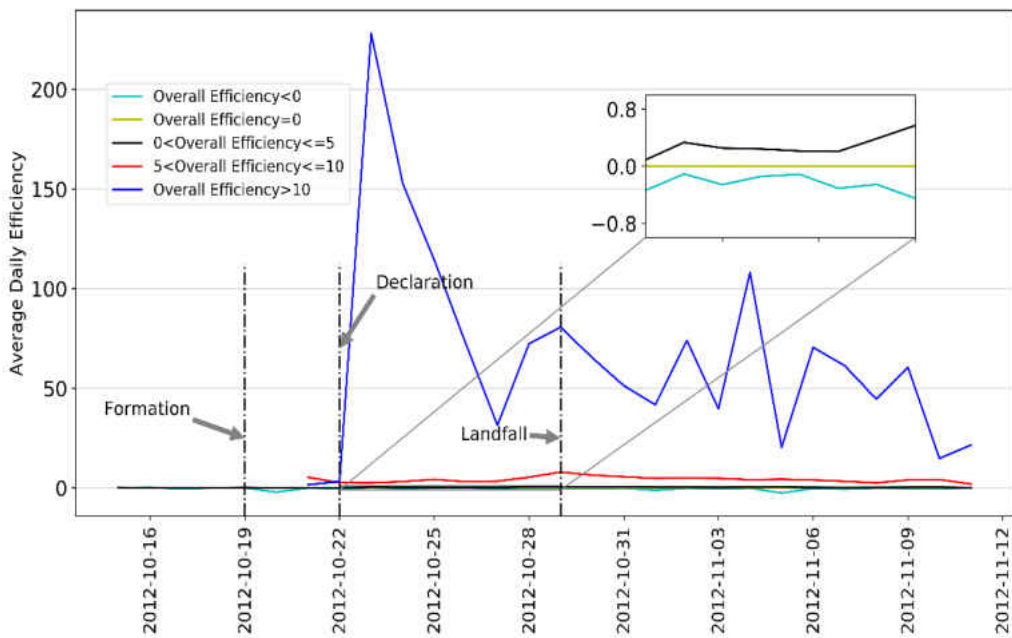


Figure 3.5: Average Daily Efficiency of the users categorized by efficiency values (for Sandy specific tweets).

Users are categorized by their overall efficiency values. To determine what type of users became highly effective spreading the awareness during hurricane declaration and landfall, we extract the top 5 efficient users for both hurricane related tweets and Sandy specific tweets. Figure 3.6 shows the daily efficiency values for the top 5 efficient users on hurricane Sandy declaration and landfall days and top 5 efficient users over the whole data coverage period. We find that majority of them are either political users or have no significant association with hurricane updates (see Table 3.1). However, analysis on Sandy specific tweets reveals significant spikes close to landfall day (see Figure 3.7). We find that these highly effective users are either storm update centers or weather reporter having close association with hurricane updates (Table 3.2). It highlights the importance of an appropriate filtering step when identifying highly effective users, specific to a disaster. Tweets collected for a general disaster context may contain ambiguous words (e.g., power, weather, recovery etc.) overlapping with other highly conversed contexts.

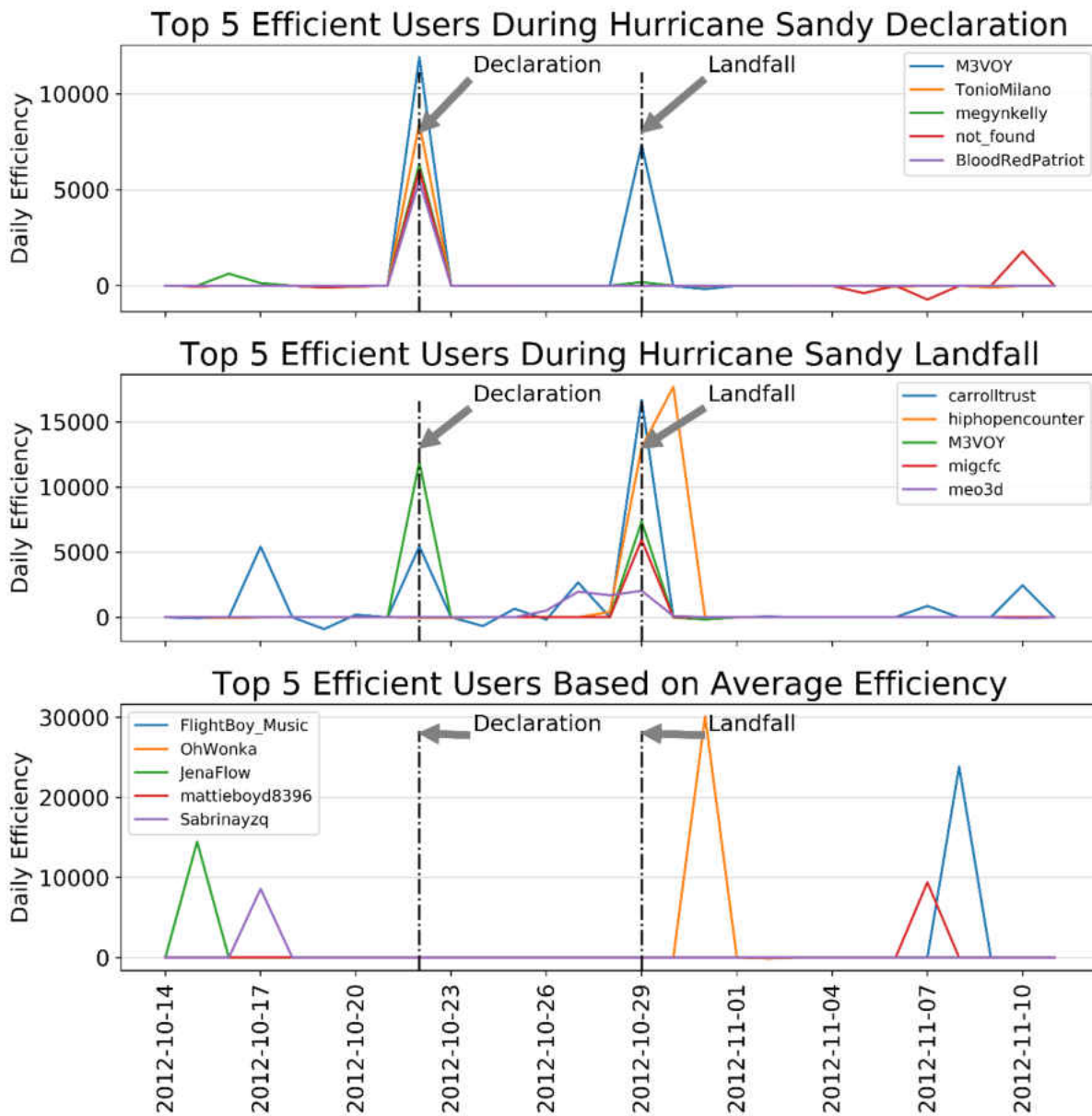


Figure 3.6: Top 5 Efficient Users during Hurricane Declaration, Landfall and Overall (for hurricane related tweets).

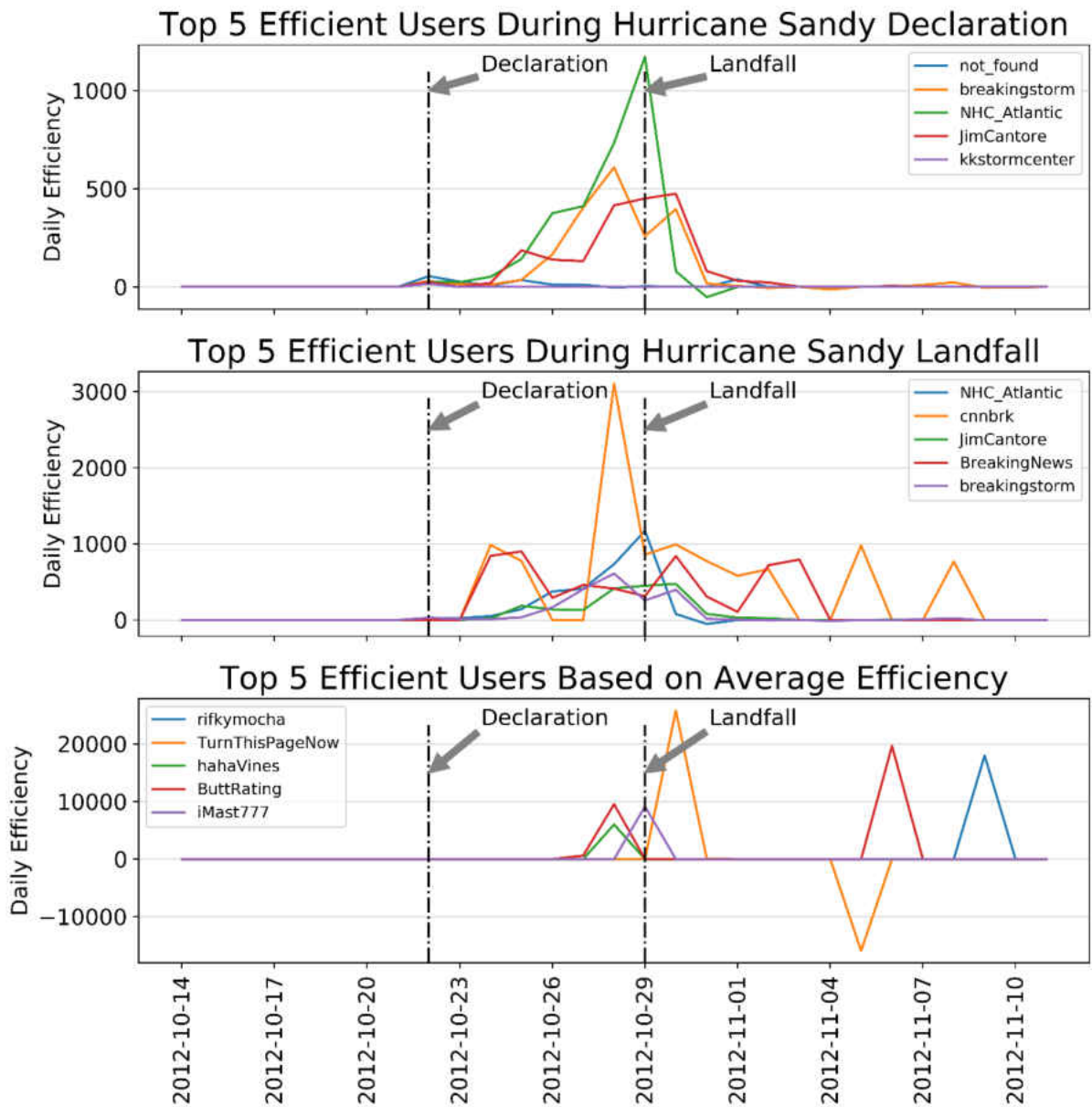


Figure 3.7: Top 5 Efficient Users during Hurricane Declaration, Landfall and Overall (for Sandy specific tweets)

Table 3.1: Top 5 efficient users with user type and bot score during hurricane declaration and hurricane landfall. This analysis is based on the unfiltered (hurricane related tweets) data.

Hurricane Declaration			Hurricane Landfall		
Screen Name	User Type	Bot Score	Screen Name	User Type	Bot Score
@M3VOY	Radio Operator	0.62	@carrolltrust	Organization	0.43
@TonioMilano	Personal	0.48	@hiphopencounter	Organization	0.76
@megynkelly	Anchor at NBC News	0.36	@migcfc	Personal	0.06
@BloodRedPatriot	Organization	0.73	@trio	Organization	0.55

Table 3.2: Top 5 efficient users with user type and bot score during hurricane declaration and landfall. This analysis is based on the data which have the word sandy in the tweet.

Hurricane Declaration			Hurricane Landfall		
Screen Name	User Type	Bot Score	Screen Name	User Type	Bot Score
Not Found	NA	NA	@NHC_Atlantic	Organization	0.43
@breakingstorm	Organization	0.66	@cnnbrk	Organization	0.51
@NHC_Atlantic	Organization	0.43	@Jimcantore	Broadcast Meteorologist	0.37
@Jimcantore	Broadcast Meteorologist	0.37	@BreakingNews	Organization	0.60
@kkstormcenter	Weather reporter	0.71	@breakingstorm	Organization	0.66

Note: Not found indicates that the user made tweets during Sandy but its screen name was not found when searched during our analysis.

3.4.4 User Attributes Contributing to Efficiency

To identify the contributing factors for an effective spreading of awareness, it is important to know the relationship between each feature and efficiency metric. Similar to the previous section, we analyze separately for hurricane related and Sandy specific tweets for understanding the factors in gaining attention in crisis communication. For hurricane related tweets, Figure 3.8 shows the

relationship between efficiency and each of the variables of followee add, initial follower and initial followee, considering two types of users: bot (bot score ≥ 0.5) and non-bot (bot score < 0.5). In addition, the relationship between efficiency and active days has been modeled considering user categories based on activity entropy. The result shows good correlation ($R^2 > 0.5$) between efficiency and initial follower for the non-bot users. From Figure 3.8, we find that efficiency is positively associated with all the variables. In all cases, R^2 values are lower for bot users compared to non-bot users. This reflects that efficiency of bots cannot be well predicted with a single variable. Similar associations have been found for Sandy specific tweets (see Figure 3.9).

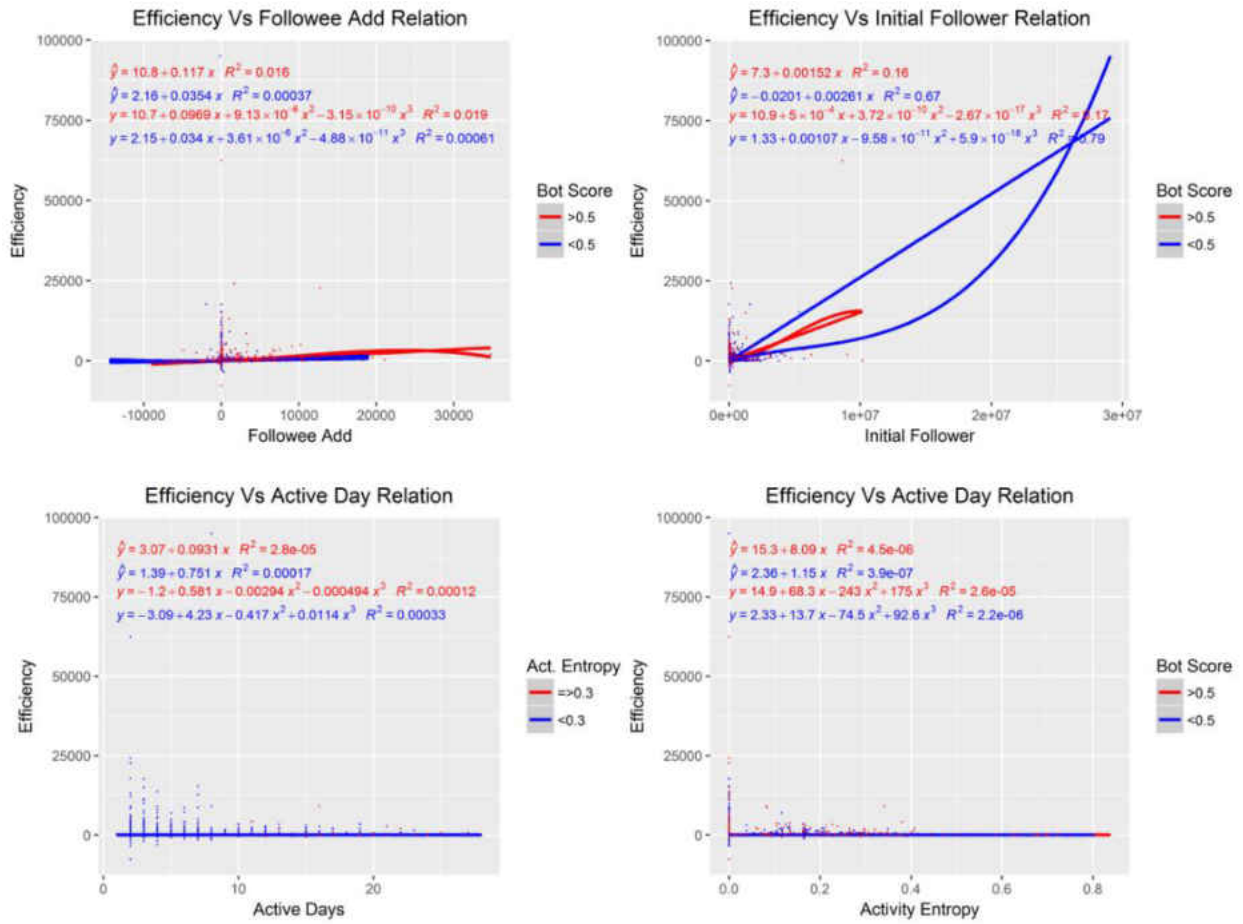


Figure 3.8: Correlations between Efficiency and Different User Attributes (for hurricane related tweets)

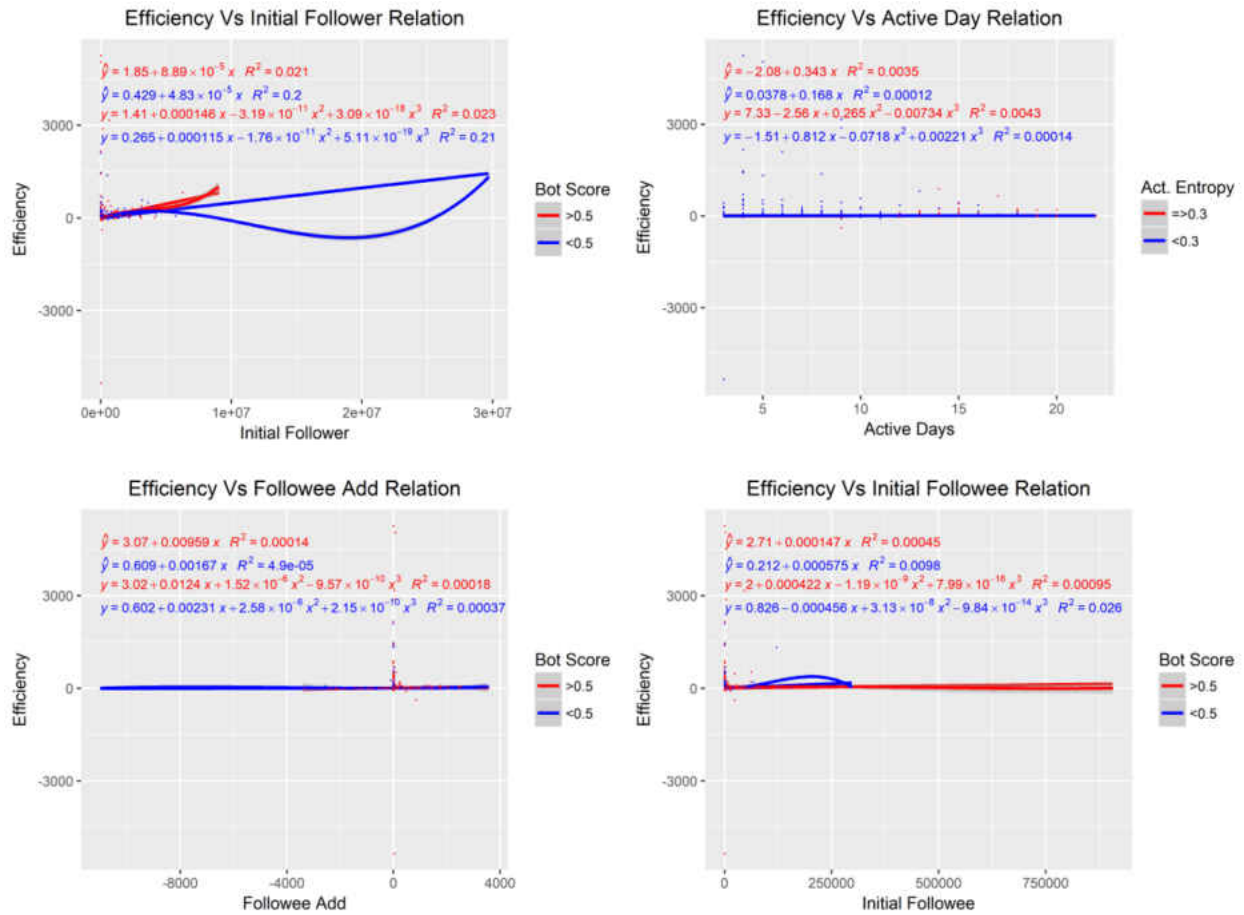


Figure 3.9: Relationship between Efficiency and Different User Attributes (for sandy specific tweets)

To determine the combined effects of the explanatory variables, we estimate a multivariate linear regression model. Table 3.3 presents the results of the model involving hurricane related tweets. All the variables are significant at 90 percent significance level, as each t statistics is greater than 1.65. It is found that efficiency increases with initial number of followers, bot score of bot users, initial number of followee of non-bot users, while decreases with total activity, initial number of followees, bot score of non-bot users. A negative coefficient for activity entropy implies that entropy values have negative correlation with efficiency and users having a predictable activity

pattern (lower entropy values) have higher efficiency values. The model estimated over Sandy specific tweets show similar results (see Table 3.4) except that total activity is not statistically significant in this case and bot score and activity entropy are positively correlated with efficiency.

From the regression analysis, we find how different user features influence a user's efficiency. However, while considering overall efficiency as an outcome, a minor change in efficiency does not provide any significant information about the user's performance in gaining attention. Thus, we have categorized efficiency into five classes: negative ($\text{efficiency} < 0$), zero ($\text{efficiency} = 0$), low ($0 < \text{efficiency} \leq 5$), moderate ($5 < \text{efficiency} \leq 10$) and high ($\text{efficiency} > 10$) as shown in Figure 3.4. The outcome variable, thus turned into an ordered categorical variable. In the hurricane related sample about 6%, 56%, 32%, 3%, and 3% of the users fall within the negative, zero, low, moderate and high efficiency category, respectively. We have estimated an ordered logit model using the same 582605 observations (see Table 3.3). All the parameters shown in the results are statistically significant at 99% significance level.

Table 3.3: Model Results for hurricane related tweets

Linear Regression Model			Ordered Logit Model			
Explanatory variable	Parameter Estimate	t statistic	Parameter Estimate	Odds Ratio Estimates		
				Point Estimate	95% Confidence Limits	Wald
Intercept 5	na		-5.0145			
Intercept 4	na		-4.2164			
Intercept 3	na		-1.1988			
Intercept 2	na		2.2357			
Constant	0.758	1.712	na			
Total activity	-0.039	-6.767	-0.0110	0.989	0.989	0.989
Followee add	0.098	77.845	0.00926	1.009	1.009	1.009
Initial follower	0.002	698.21	1.000	1.000	1.000	1.000
Initial followee	-0.004	-53.3	1.000	1.000	1.000	1.000
Active days	na		0.2951	1.343	1.340	1.347
Activity entropy	-36.754	-15.493	-1.9523	0.142	0.128	0.157
Bot score	11.622	9.793	-0.3974	0.672	0.648	0.697
Bot_Score*NonBotUser	-14.419	-10.041	na			
Initial_followee*NonBotUser	0.007	37.423	na			
Number of observations	582,605		582,605			
Adjusted R²	0.47					
AIC			1244359.6			

Note: na= not applicable, any variable included in one mode but not included in the other one is because we did not find it significant.

Table 3.4: Model Results for Sandy specific tweets

Linear Regression Model			Ordered Logit Model			
Explanatory variable	Parameter	t	Parameter Estimate	Odds Ratio Estimates		
	Estimate	statistic		Point Estimate	95% Confidence Limits	Wald
Intercept 5	na		-6.6001			
Intercept 4	na		-3.7514			
Intercept 3	na		4.6505			
Intercept 2	na		5.3600			
Constant	-.073	-.453	na			
Total activity	-0.001	-1.265	na	na	na	na
Followee add	0.003	5.411	0.00359	1.004	1.003	1.004
Initial follower	0.0001	231.54	0.000035	1.00	1.00	1.00
Initial followee	-6.85E-5	-6.219	-2.71E-06	1.000	1.000	1.000
Active days	na		-0.1312	0.877	0.852	0.903
Activity entropy	.704	1.940	1.2244	3.402	1.726	6.708
Bot score	1.053	3.735	2.115	8.289	5.227	13.145
Bot_Score*NonBotUser	.521	1.655	na			
Initial_followee*NonBotUser	-5.27E-5	-1.794	na			
Number of observations	15,792		15,792			
Adjusted R²	0.77					
AIC			7513.390			

Note: na= not applicable, any variable included in one mode but not included in the other one is because we did not find it significant

For the interpretation of this result, regarding total activity, a negative parameter estimate represents that if all other variables in the model remain constant, for an increase in the number of tweets a user is more likely to be in a lower level of efficiency. Similar to the results from the regression model, users with a predictable tweeting pattern (i.e., smaller entropy value) are more likely to be in a higher efficiency category. Moreover, we find that users with higher number of

active days and followee added are more likely to be in a higher category of efficiency. In contrast, users with higher total activity and bot score are less likely to be in higher category efficiency. An ordered Logit model estimated over the Sandy specific tweets shows similar association except that total activity is not statistically significant and activity entropy is positively correlated with efficiency (see Table 3.4). A higher number initial follower or initial followee does not result in higher or lower efficiency category for both hurricane related and sandy specific tweets.

3.5 Conclusions

In this study, we have analyzed twitter posts related to hurricane Sandy to understand the effectiveness of social media based communication during disasters. Effective crisis communication can ensure faster information dissemination to vulnerable communities who need timely information about disaster preparedness, evacuation warning, and recovery operations. To measure the effectiveness of a social media user in communicating information or awareness, we have estimated the efficiency of gaining attention within a specific time period as the ratio of follower gained over tweet frequency within the same time period. We consider that new follower gained represents attention received and tweet frequency represents activities made. As our data contains tweets from both pre and post-disaster periods, a comparison of user efficiencies in gaining attention among these periods has been possible.

Analyzing daily efficiencies in gaining attention, we have found that users had higher efficiency during the critical periods in hurricane Sandy such as declaration and landfall days. This indicates the potential of social media based crisis communication since higher attention to related information may help in providing situational awareness to vulnerable population. It might be the case that during Sandy some users' efficiency became abnormally high because a high number of

users started following them for hurricane related updates. These social media users could become one of the major sources of information for spreading hurricane awareness during future hurricanes.

During a disaster, general social media users seek information from other users for a timely update. When sharing information, some users gain more attention than the others. Thus, it's critical to understand what user features influence the process of gaining attention. For understanding the contributing features, we have estimated a regression and an ordered logit model considering overall efficiency and efficiency category, respectively as a dependent variable. We have found that higher activities relating to hurricanes are not necessarily associated with higher efficiencies. However, users with predictable tweeting patterns have gained higher efficiency values. We have also found that a higher bot score (typically associated with an organizational account) results in lower efficiencies. We have observed some differences on the effect of few user attributes on efficiency values for models estimated over general hurricane related tweets and Sandy specific tweets. User efficiencies in gaining attention for a crisis event are directly related with information spreading capacity of a system. A better understanding of the factors will provide insights on crisis communication both at organizational and individual levels. These insights will also help emergency agencies when using social media as a disaster communication tool.

Thus, our findings have significant importance in social media communication specially in disaster communication. For attaining high efficiency in spreading disaster related information, concerned organizational or personal accounts can plan their activity considering the factors which will maximize the chance to attain higher attention from the targeted population in social media. Also, prior to a major disaster event concerned authorities can select some efficient social media users

for disseminating information about situational awareness; a model based on user features could find the efficient users for this task.

However, our study has several limitations which can be improved in future. For example, bot scores were not collected during the period of hurricane Sandy, rather we have collected them at the time of our analysis. Bot scores could be different during hurricane Sandy than in the present when we have collected. We assume that all the tweets analyzed here were related to hurricane Sandy. User efficiency considering specific topics (e.g., evacuation) of a tweet should be analyzed in the future.

CHAPTER 4: QUANTIFYING HUMAN MOBILITY RESILIENCE

4.1 Introduction

Resilience is a broad concept, which is applied to many different fields to measure the ability to sustain an adverse situation. Previously, several concepts of resilience have been proposed. Hosseini et al. (Hosseini et al., 2016) have reviewed the methods of defining and quantifying resilience in various fields. Bruneau et al. (Bruneau et al., 2003) developed a framework for measuring resilience considering four dimensions: i) *robustness* reflecting the strength or ability of the system to reduce the damage; ii) *rapidity* representing the rate or speed of recovery; iii) *resourcefulness* reflecting the ability to apply materials and human resources by prioritizing goals when an event occurs; and iv) *redundancy* representing the capacity to achieve goals by prioritizing objectives to restrain loss and future disruptions. They have also proposed the following equation to measure resilience loss due to an earthquake:

$$RL = \int_{t_0}^{t_1} [100 - Q(t)]dt \quad (4.1)$$

where, RL denotes resilience loss, $Q(t)$ denotes a quality function at time t , and t_1 is the recovery time. This formula forms the basis of a resilience triangle. Although this metric was originally proposed for an earthquake, it can be applied to many other contexts (Hosseini et al., 2016).

However, measuring these resilience metrics, in a mobility context, has been difficult due to the lack of appropriate data over longer time periods. Geo-location data from social media can offer a solution to this problem. In this study, by analyzing user displacements from a pre-disaster period to a post-disaster one, we measure perturbation and recovery time for multiple types of disaster. To validate our results, we have used one-month of taxi data from the New York City recording

taxi movements before, during, and after hurricane Sandy. Quantifying resilience loss and recovery time from disruptions in response to an extreme event can help understanding the broader socio-economic impacts of disasters. Furthermore, these resilience metrics will help in making policy towards building resilient cities and communities.

This study makes several contributions. First, it defines the concept of mobility resilience and develops methods to detect extreme events in mobility data and to measure required metrics to measure resilience and resilience loss from movement data. Second, it applies the proposed method of measuring resilience to geo-located data collected from Twitter for multiple disasters. Thus, this paper shows that geo-located social media data can be effectively used to measure human mobility resilience to extreme events.

4.2 Data and Methodology

To measure mobility resilience, we have used geo-tagged tweets from several types of disaster (Table 4.1). The data sets have been collected from Dryad digital repositories <http://datadryad.org/resource/doi:10.5061/dryad.88354> (Wang Q, 2016), originally collected by Wang et. al. (Wang and Taylor, 2016) and <https://datadryad.org//resource/doi:10.5061/dryad.15fv2>, collected by Kryvasheyev et. al. (Kryvasheyev and Chen, n.d.).

Table 4.1: Data Description

Type	Disaster Name	Disaster Location	No. of Tweets	No. of Users
Hurricane	Sandy (all tweets)	USA	52,493,130	13,745,659
	Sandy (geo-tagged tweets)	USA	24,149,780	5,981,012
Earthquake	Bohol (Bohol)	Bohol, Philippines	114,606	7,942
	Iquique (Iquique)	Iquique, Chile	15,297	1,470
	Napa (Napa)	Napa, USA	38,019	1,850
Typhoon	Wipha (Tokyo)	Tokyo, Japan	849,173	73,451
	Halong (Okinawa)	Okinawa, Japan	166,325	5,124
	Kalmaegi (Calasiao)	Calasiao, Philippines	21,698	1,063
	Rammasun (Manila)	Manila, Philippines	408,760	27,753
Winter storm	Xaver (Norfolk)	Norfolk, Britain	115,018	8,498
	Xaver (Hamburg)	Hamburg, Germany	15,054	2,745
	Storm (Atlanta)	Atlanta, USA	157,179	15,783
Thunder storm	Storm (Phoenix)	Phoenix, USA	579,735	23,132
	Storm (Detroit)	Detroit, USA	765,353	15,949
	Storm (Baltimore)	Baltimore, USA	328,881	14,582
Wildfire	New South Wales (1)	New South Wales, Australia (1)	64,371	9,246
	New South Wales (2)	New South Wales, Australia (2)	34,157	4,147

To validate our approach of using social media data, we collected New York City taxi data which includes taxi movement for the period same as the hurricane Sandy twitter data. The data was collected from a repository hosted by New York City Taxi and Limousine Commission (http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml). In the data, each observation represents a trip and there were total 12,892,877 trips in the study period. Hurricane Sandy data have tweets from several places including USA, Canada, Mexico and other countries. For measuring resilience for a city or a state in response to hurricane Sandy, we have applied

appropriate location filters. For example, a trip can be made within the New York City or having only an origin or destination in it. Since displacements are calculated in six-hour periods, when calculating resilience for the New York City, if a location filter is applied, only the displacement within the New York City will be considered in a six-hour period. If a location filter is not applied, both displacements within the New York City and having origins or destinations at the New York City will be considered in a six-hour period. Except hurricane Sandy data, the rest of the data consist city-specific tweets where those cities were subject to a disruptive event. Thus, a location filter or constraint is not required for these cases.

In this study, we apply the concept of resilience for understanding human mobility under a disaster. Following the basic definition of resilience, we define *mobility resilience* as the ability of a mobility infrastructure system responsible for the movement of a population to manage shocks and return to a steady state in response to an extreme event. These events include a hurricane, earthquake, terrorist attack, winter storm, wildfire, flood, and others. We propose a simple method based on human movement data using *normalized per user displacement* as a key indicator of human mobility. Comparing the difference between per user displacements from typical displacements, the proposed method can detect a disruptive event from movement data and calculate the maximum deviation from normal conditions and the recovery time. Finally, applying the concept of resilience triangle, we estimate resilience and resilience loss for an event detected by the method. The proposed method can take any kind of movement data as inputs including coordinates from mobile phone call recordings, GPS observations, social media posts and many others. In this paper, we present our resilience analysis based on social media data from multiple types of disasters.

4.2.1 Extracting Location Time Series of a User

First, the coordinates of a user are sorted in an ascending order by timestamps. If there are not enough users for an hourly based analysis, we can divide each day in 4 periods such as 12 AM to 6 AM, 6 AM to 12 PM, 12 PM to 6PM and 6PM to 12AM. From the sorted time series, locations (i.e., latitude and longitude) of each user are extracted in six-hour interval for each day.

$$P_u^{d,t} = \{(x, y) | (x, y) \in (\text{latitude}, \text{longitude of a region})\} \quad (4.2)$$

where, $P_u^{d,t}$ denotes the set of locations of a user u in day d at period t $d \in (\text{days in the dataset}), t \in (\text{periods in a day}), u \in (\text{users in dataset})$

4.2.2 Displacement Metric

From the set of locations of a user, distances between two consecutive points are calculated using the Harvesine formula (C, n.d.) shown in Equation (4.3). For calculating displacements, a user must have at least two locations within a six-hour interval. Otherwise, the user is not considered in that interval.

$$C = 2r \times \sin^{-1} \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos \phi_1 \cos \phi_2 \sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right)} \right) \quad (4.3)$$

where r is radius of earth, ϕ is latitude and φ is longitude. Displacement between two consecutive points will be calculated for each user at every six-hour interval. The average of the displacements for an interval is calculated by dividing the sum of the displacements by the total number of users contributing to that displacements. Thus,

$$D^{d,t} = \left\{ \frac{\sum C^{d,t}}{\sum u^{d,t}} \right\} \quad (4.4)$$

where $D^{d,t}$ represents the average displacements at period t for day d . The term $\sum C^{d,t}$ indicates the summation of the displacements at period t for day d and the term $\sum u^{d,t}$ represents the total number of users contributing to these displacements within this period.

4.2.3 Extraction of Typical and Actual Displacements Time Series

The mobility dataset to be used for a resilience analysis should cover pre-disaster, disaster and post disaster periods. Using the average displacements value in the pre-disaster period, we can make four sets of typical values for the four periods considered in a day. These four typical values are calculated separately for weekdays and weekends.

$$D_{weekday}^t = \{D^{d,t} \text{ where, } d \in (\text{pre} - \text{disaster weekdays})\} \quad (4.5)$$

$$D_{weekend}^t = \{D^{d,t} \text{ where, } d \in (\text{pre} - \text{disaster weekend days})\} \quad (4.6)$$

where $D_{weekday}^t$ represents the set of displacements at period t considering only weekdays in the pre-disaster period. Similarly, $D_{weekend}^t$ represents the set of displacements at period t considering only weekends in the pre-disaster period. For instance, if we have 4 periods per day, and if we select first 7 days as a pre-disaster period, for each period, we have a set of 5 values of displacement for weekdays and a set of 2 values for weekends. The mean and standard deviation of these sets of displacement are used to compare the actual displacement at the corresponding periods of a day to check whether the displacement is typical or not. To capture this effect, we can compute standardized displacement, Z score, for each actual displacement using the equation given below:

$$Z^{d,t} = \left\{ \begin{array}{l} \frac{D^{d,t} - \text{mean of } D_{weekday}^t}{\text{standard deviation of } D_{weekday}^t} \text{ if } d \in (\text{week days}) \\ \text{else } \frac{D^{d,t} - \text{mean of } D_{weekend}^t}{\text{standard deviation of } D_{weekend}^t} \end{array} \right\} \quad (4.7)$$

where $Z^{d,t}$ represents the Z score at day d and period t . If d is a weekday, typical displacements for weekdays are used to compare; and if d is a weekend day, typical displacements for weekends are used.

4.2.4 Extreme Event Detection

An extreme event can disrupt human mobility by either increasing mobility or decreasing mobility. We consider two parameters for detecting an extreme event: a threshold z score α and the number of time intervals τ . The first parameter checks the amount of deviation from typical values and the second parameter checks how long this deviation persists.

$$Event_{d_i, t_p}^{d_j, t_q} = \left\{ Z^{d,t}: Z^{d,t} \leq \alpha_l \text{ and } \sum_{d_i, t_p}^{d_j, t_q} d, t \geq \tau \right\} \quad (4.8)$$

or,

$$Event_{d_i, t_p}^{d_j, t_q} = \left\{ Z^{d,t}: Z^{d,t} \geq \alpha_u \text{ and } \sum_{d_i, t_p}^{d_j, t_q} d, t \geq \tau \right\} \quad (4.9)$$

Equation (4.8) and (4.9) represent the event detection for decreased and increased mobility, respectively; where $Event_{d_i, t_p}^{d_j, t_q}$ represents an extreme event from day d_i period t_p to day d_j period t_q ; $d_i, d_j \in (\text{days in data set})$ and $t_p, t_q \in (\text{periods in a day})$; α_l, α_u represent the lower and upper threshold of Z score; and τ represents the threshold number of periods when Z score is above or below the threshold Z score. These parameters (α, τ) can be selected to identify shorter or longer extreme events depending on the type of a disaster and the area affected by it.

4.2.5 Resilience Calculation

Once an extreme event has been detected, maximum deviation and recovery time can be easily calculated. Bruneau et. al. (Bruneau et al., 2003) introduced a equation for calculating resilience loss as shown in Equation (4.1)

$$RL = \int_{t_0}^{t_1} [100 - Q(t)] dt$$

where, RL is the resilience loss which is the area (see Figure 4.1(a)) between the horizontal line from 100 and the curve $Q(t)$ for t_0 to t_1 which is the recovery period for any event.

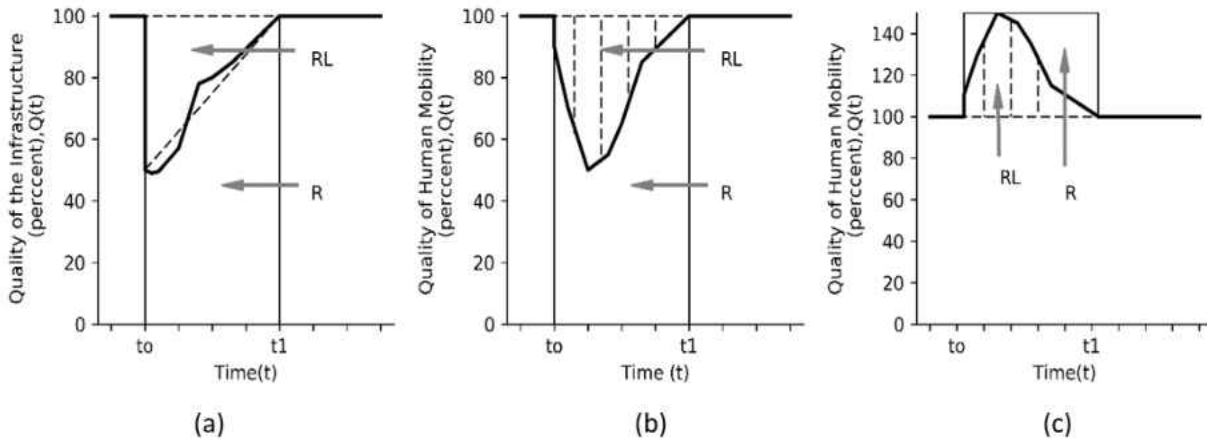


Figure 4.1: Resilience and Resilience Loss Calculation. (a) Resilience Triangle (adopted from (Bruneau et al., 2003)), (b) Human Mobility Resilience (Decreased movement), (c) Human Mobility Resilience (Increased movement).

A schematic representation of this equation (see Figure 4.1 a) is known as a resilience triangle. R and RL indicate resilience and resilience loss, respectively. From this triangle, the loss of resilience in any extreme event can be calculated as the area formed by the dashed lines and the vertical line (see Figure 4.1 a). Inspired from the resilience triangle, we represent the resilience by dividing this area into smaller trapezoids (see Figure 4.1b and 4.1c) having height equal to the increment of

time (six hours) considered in the analysis. This assumption is required since, unlike an idealized quality function, a real-world quality function indicating human mobility gradually drops from and improves to its typical values. Thus, assuming smaller trapezoids will minimize the loss in calculation.

In our analysis, we define quality as the ratio of actual displacements to typical displacements. If an actual displacement is equal to a typical displacement, the value quality function is 100 or the ratio is 1. The summation of the areas of all the small trapezoids is the resilience loss (indicated by RL in figure 4.1b and 4.1c). The residual area (indicated by R in Figure 4.1) represents the value of resilience during the recovery period. For increased mobility area considered in resilience calculation are defined by the maximum quality percentage/ratio (see Figure 4.1c).

4.3 Results

The approach to calculate resilience has been applied over location-based social datasets (see Table 4.2). During these events, we observe two types of responses in the mobility function which either significantly drops (decreased mobility) or significantly rises (increased mobility). To represent both types of events, two thresholds z scores (α values) have been used for detecting an extreme event. For decreased mobility cases, a threshold z score value of 40 percentile ($\alpha_l = 40$) and for increased mobility cases, a threshold z score of 90 percentile ($\alpha_u = 90$) have been chosen to detect an extreme event. However, when no event was detected with these thresholds, $\alpha_l = 60$ percentile have been chosen; this relaxes the lower threshold of z score. As the threshold duration of the extreme event when the z value is below α_l has been chosen as 7 time periods (i.e., $\tau =$

7 or 42 hours) and when the z value is above α_u has been chosen as 3 time periods (i.e., $\tau = 3$ or 18 hours).

Figure 4.2 shows the major steps in calculating resilience for three types of disasters namely: Each figure has three panels; the first panel shows the actual and typical values; the second panel shows the event detection by z score; and the third panel shows the resilience and resilience loss. Hurricane Sandy (Figure 4.2a), earthquake at Bohol (Figure 4.2b) and a thunder storm at Phoenix, Arizona (Figure 4.2c). Table 4.2 presents the results of resilience calculation for multiple types of disasters along with the threshold values used to detect the events. Events detected by 60 percentile thresholds are not comparable with the events detected by 40 percentile thresholds. The 40 percentile events are more severe than the 60 percentile events. Among 40 percentile events, the highest recovery time was found 144 hours for hurricane Sandy for the state of New York and the highest resilience loss was found 344.89 for earthquake Iquique. We have also calculated the ratio between resilience loss and resilience $\left(\frac{RL}{R}\right)$. The highest ratio of resilience loss over resilience has been found as 2.73 for the state of New York for hurricane Sandy. Among the 60 percentile events, the state of New Jersey during hurricane Sandy had the highest recovery time, resilience loss and resilience loss over resilience ratio. These metrics indicate the magnitude of impact of hurricane Sandy on the mobility systems of the states of New York and New Jersey.

In addition to Twitter data, we have used taxi trips data to calculate the resilience metrics. Figure 4.2d shows the resilience and recovery time for taxi movements in the New York City. For measuring resilience in taxi data, taxi trips have been used instead of the taxi trip distance. Most of the trips in taxi occurred between some frequently visited places and thus, the average traveled distances per trip were almost same for the disrupted days although there were significantly less

number trips in those days. For taxi trips, the maximum deviation at the landfall day is found as 0.052 which means only 5.2 percent of the typical trips occurred at the landfall day of hurricane Sandy; the recovery time is found 96 hours. A recent study (Donovan and Work, 2017) measuring transportation system resilience by taxi data using pace as a quality indicator found recovery time as 132 hours for hurricane Sandy. From Table 4.2, we can see that human mobility recovery time and resilience loss for New York city is 66 hours and 42.37, respectively. The two results between taxi resilience and human mobility resilience is not directly comparable because taxi is just one of the modes of human mobility.

During hurricane Sandy, among the states, the state of New York suffered the highest resilience loss followed by the states of New Jersey and Pennsylvania. For hurricane Sandy both recovery time and resilience loss are higher when a location constraint is not applied. Except hurricane Sandy data, typhoon, winter storm and rain storm data are location constrained. Thus, resilience losses for these events are lower compared to hurricane Sandy's unconstrained resilience loss. This finding is consistent with previous study (Qi and John E., 2014) that during these types of disasters, short trips are less affected compared to long trips. These events discussed above faced a significant amount of decrease in mobility from a typical mobility function.

However, in an earthquake, instead of a decreasing mobility function, we observe a significant increase in human mobility- probably due to the long-distance migration of people forced by severe infrastructure damages. Figure 4.2b shows the resilience calculation for an earthquake happened at Bohol, Philippines in 2013. The recovery time and resilience loss for this event are 54 hours and 162.31, respectively. Our method has detected one more event after around 3 days. This event may represent the increased mobility when displaced people returned to their

places as studies found that natural disaster like earthquake cause human migration. Table 4.2 shows the other earthquake resilience and recovery time results. Among the earthquakes analyzed in this study, Iquique had the highest deviation and resilience loss, 38.167 and 344.89, respectively and Napa had the lowest resilience loss and deviation. A study (Wang and Taylor, 2016) on the same data for measuring human mobility pattern found that although human mobility during most of typhoon, rainstorms, winter storms and Napa earthquake can be predicted by established patterns, mobility during earthquakes Bohol and Iquique cannot be predicted. Instead of decreased mobility, a significant increase in mobility with large resilience loss during these events may explain this result.

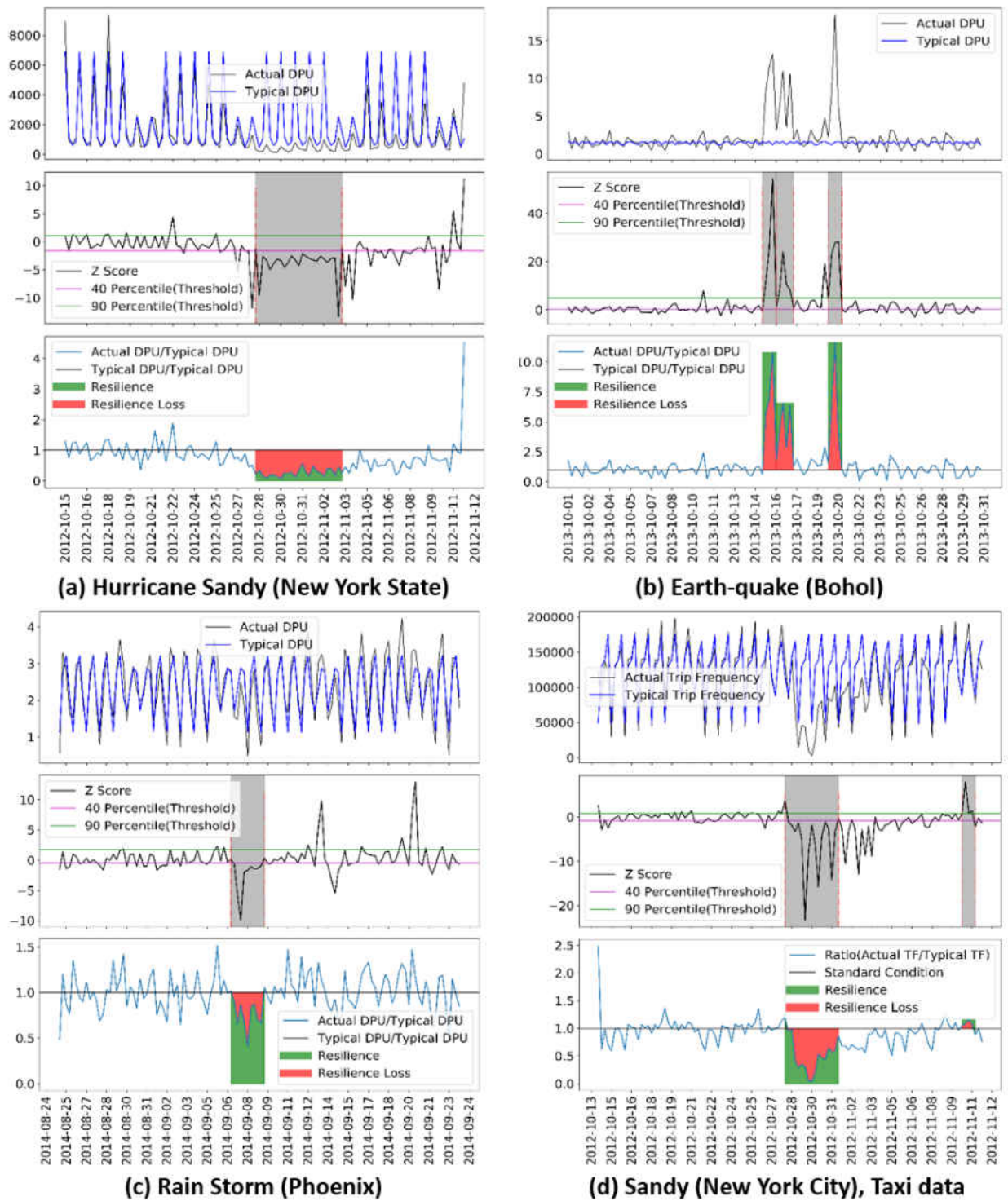


Figure 4.2: Resilience and Resilience Losses for Multiple Disasters. Note: DPU= Displacements Per User (Kilometer), TF= Trip Frequency

Table 4.2: Comparison of Resilience, Resilience Loss and Recovery Time for Multiple types of Events Occurred in Different Location

Disaster Name	Location	Threshold Z Score	Location Filter	Start Time	Recovery Time(hr.)	Max Deviation	Resilience (R)	Resilience Loss (RL)	Ratio (RL/R)	
Sandy	New York City	$\alpha_l=60$	Y	2012-10-26 00:00	132	0.540	106.73_0	19.260	0.181	
		$\alpha_l=40$	N	2012-10-28 12:00	66	0.010	17.620	42.370	2.404	
	New York State	$\alpha_l=40$	Y	2012-10-28 12:00	48	0.260	21.860	20.100	0.920	
		$\alpha_l=40$	N	2012-10-28 06:00	144	0.087	36.400	101.400	2.730	
	New Jersey State	$\alpha_l=60$	Y	2012-10-28 00:00	120	0.176	52.000	55.00	1.057	
		$\alpha_l=60$	N	2012-10-27 12:00	168	0.001	21.179	140.820	6.648	
				2012-11-06 00:00	48	0.018	8.907	33.090	3.715	
	Pennsylvania State	$\alpha_l=60$	Y	2012-10-28 00:00	120	0.180	58.930	49.060	0.833	
		$\alpha_l=60$	N	2012-10-26 06:00	144	0.003	12.600	125.390	9.949	
				2012-11-02 06:00	72	0.015	13.970	52.026	3.720	
	Earthquake	Bohol, Philippines	$\alpha_u=90$	NA	2013-10-15 00:00	54	9.330	120.47_0	162.310	1.340
					2013-10-19 18:00	24	11.035	64.956	115.680	1.780
Iquique, Chile		$\alpha_u=90$	NA	2014-04-02 18:00	48	38.167	519.05_8	344.890	0.664	
Napa, USA		$\alpha_u=90$	NA	2014-08-23 18:00	18	6.416	27.490	37.503	1.360	
Wild Fire	NSW1, Australia	$\alpha_u=90$	NA	2013-10-18 12:00	18	8.257	58.860	28.230	0.480	
		$\alpha_l=40$	NA	2013-10-19 06:00	48	0.188	22.440	19.550	0.870	
	NSW2, Australia	$\alpha_l=40,60$	NA				NO RL			

Disaster Name	Location	Threshold Z Score	Location Filter	Start Time	Recovery Time(hr.)	Max Deviation	Resilience (R)	Resilience Loss (RL)	Ratio (RL/R)
Winter Storm	Xaver, Norfolk, Britain	$\alpha_l=40$	NA	2013-12-02 12:00	48	0.339	25.370	16.629	0.655
	Xaver, Hamburg, Germani	$\alpha_l=40$	NA	2013-12-04 18:00	48	0.035	24.817	17.182	0.690
		$\alpha_u=90$	NA	2013-12-13 12:00	36	4.306	55.704	43.480	0.780
	Atlanta, USA	$\alpha_l=40$	NA	2014-01-28 12:00	54	0.261	20.450	27.545	1.346
Rain Storm	Phoenix, USA	$\alpha_l=40$	NA	2014-09-06 18:00	60	0.329	40.000	13.000	0.413
	Detroit, USA	$\alpha_l=40$	NA	Not Enough Pre-Disaster Data					
	Baltimore, USA	$\alpha_l=40,60$	NA	NO RL					
Typhoon	Wipha, Tokyo, Japan	$\alpha_l=40,60$	NA	NO RL					
	Halong, Okinawa, Japan	$\alpha_l=40$	NA	2014-07-29 06:00	96	0.616	74.000	10.000	0.135
	Kalmaegi, Philippines	$\alpha_l=40$	NA	2014-09-08 12:00	96	0.005	42.568	42.000	0.990
		$\alpha_l=40$		2014-09-23 12:00	54	0.003	24.188	23.811	0.980
	Rammasun, Philippines	$\alpha_l=40,60$	NA	NO RL					

Note: NA=Not Applicable, Y=Yes, N=No

4.4 Discussion

In this paper, we present a method to compute resilience metrics using geo-location data from social media. The proposed method can detect an extreme event from human movements, measure the recovery time and the maximum deviation from a steady state mobility indicator, and assess the values of resilience and resilience loss. Applying this method on multiple disaster data, we find that human movements within a geographic area (e.g., trips only within a city) is less affected

compared to all the movements associated with the area (e.g., trips from, to, and within the city). Disasters such as hurricane, typhoon, winter storm decrease human mobility and the amount of perturbation depends on the location and severity of the disaster. However, an earthquake increases human mobility causing a significant resilience loss. This is probably because an earthquake is unpredictable while for the other disasters people had warnings lasting over multiple days.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

Social media is a great tool for the individual users and the organizations to communicate, express views and concerns which is not possible through traditional media. Communication in social media is more dynamic as it allows two-way communication by allowing to be both content producer and consumer at the same time. These features of social media facilitate in disaster management in a unique and dynamic way during a disaster. But, content generated from some users/organizations get more attention than others. During a disaster, getting more attention to disaster related contents will lead to faster communications. In this thesis, we investigate the contributing factors to get more attention efficiency during hurricane sandy. If a user or organization adopt their activity and the other factors favorable to gain more attention, information spreading, or crisis communication is likely to be faster.

Another side of social media data is having the opportunity to collect location traces of the users. Especially during a disaster, that gives the opportunity to assess human mobility resilience which can indicate the overall disaster resilience of the region. In this thesis, we have used location based social media data to develop appropriate metrics to quantify human mobility resilience. This study uncovers that different types of disaster have different impact on human mobility depending on the intensity of the event. The findings of this study are very important for understanding the nature and amount of perturbation and the subsequent resilience loss in human mobility due to a disaster. Thus, it will help understanding the higher-order impacts of a disruptive event in human society and national economy. It can also help in policy making, as resilience assessment is critical for building a resilient transportation system. However, the proposed method has some limitations. It cannot detect events less than six hours long because a minimum period of six hours is chosen. Also, in a pre-disaster period, variations among weekdays and variations

between weekend days are not considered due to the lack of enough pre-disaster data. Movements of social media users may not represent well the actual population movement during a disaster.

Besides many opportunities, social media is subjected to several challenges due to its data volume and velocities. Further such data is also very unstructured and mixed with rumors, advertisements and uneducated opinions. Thus, extracting actionable information for the responders demands dynamic algorithms or systems to filter out the noises. Moreover, location traces can be collected only when a user post something about it. Hence mobility analysis using such data may not be representative to the actual mobility. Social media data is also prone to selection bias. Because it has different penetration rate for different geographic areas and users' groups may not represent the actual population proportion of the region.

Evacuation management is one of the major parts of disaster management. Existing studies use survey data to understand evacuation behavior during a disaster which is every costly, time consuming and often limited to small geographic area. Despite having the potential, social media data in this context remain underexplored. Future research direction can be how social media data can be used to understand individual and collective evacuation decision making behavior and how social media content can contribute to evacuation demand prediction.

REFERENCES

- Aguirre, B.E., Wenger, D., Vigo, G., 1998. A Test of the Emergent Norm Theory of Collective Behavior. *Sociol. Forum* 13, 301–320, Kluwer Academic Publishers–Plenum Publish.
- Alessandretti, L., Sapiezynski, P., Lehmann, S., Baronchelli, A., n.d. Multi-scale spatio-temporal analysis of human mobility. *PloS one* 12.2 e0171686.
- Alliance, R., 2007. Assessing resilience in Social-Ecological Systems - A workbook for scientists. *Transformation* 22, 1–53.
- Arif Mohaimin Sadri; Satish V. Ukkusuri, Ph.D., M.ASCE; Pamela Murray-Tuite, Ph.D., M.ASCE; and Hugh Gladwin, P.D., 2014. How to Evacuate: Model for Understanding the Routing Strategies during Hurricane Evacuation. *J. Transp. Eng.* 140, 61–69.
- Asur, S., Huberman, B.A., Szabo, G., Wang, C., 2011. Trends in Social Media: Persistence and Decay. *Proc. 5th Int. AAAI Conf. Weblogs Soc. Media* 434–437.
- Beir??, M.G., Panisson, A., Tizzoni, M., Cattuto, C., 2016. Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Sci.* 5.
- Bharosa, N., Lee, J., Janssen, M., 2010. Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Inf. Syst. Front.* 12, 49–65.
- Blake, E.S., Kimberlain, T.B., Berg, R.J., Cangia, Losi, J.P., Beven II, J.L., 2013. Tropical cyclone report Hurricane Sandy (AL182012) 22 – 29 October 2012. *Natl. Weather Serv. Natl. Hurric. Cent.* 1–157.
- Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. *Nature* 439, 462–465.
- Bruneau, M., Chang, S.E., Eguchi, R.T., Lee, G.C., O'Rourke, T.D., Reinhorn, A.M., Shinozuka, M., Tierney, K., Wallace, W.A., Von Winterfeldt, D., 2003. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthq. Spectra* 19, 733–752.
- C, R., n.d. The cosine-haversine formula. *Am. Math. Mon.* 1957; 64(1)38–40.
- Cha, M., Mislove, A., Gummadi, K.P., 2009. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. *Proc. 18th Int. Conf. World Wide Web* 721–730.
- Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D.S., 2014. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Comput. Graph.* 38, 51–60.
- Chen, Y., Frei, A., Mahmassani, H., 2014. From Personal Attitudes to Public Opinion :

- Information Diffusion in Social Networks towards Sustainable Transportation. *Transp. Res. Rec. J. Transp. Res. Board* 2430, 28–37.
- Chen, Y., Mahmassani, H.S., Frei, A., 2017. Incorporating social media in travel and activity choice models: conceptual framework and exploratory analysis. *Int. J. Urban Sci.* 0, 1–21. <https://doi.org/10.1080/12265934.2017.1331749>
- Comfort, L.K., Ko, K., Zagorecki, A., 2004. Coordination in Rapidly Evolving Disaster Response Systems. *Am. Behav. Sci.* 48, 295–313.
- Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F., 2016. BotOrNot: A System to Evaluate Social Bots. *arXiv* : 1602.009.
- Derr, B., 2013. Ordinal Response Modeling with the LOGISTIC Procedure. *SAS Glob. Forum* 1–20.
- Deville, P., Song, C., Eagle, N., Blondel, V.D., Barabási, A.-L., Wang, D., 2016. Scaling identity connects human mobility and social interactions. *Proc. Natl. Acad. Sci.* 113, 7047–7052.
- Dong, R., Li, L., Zhang, Q., Cai, G., 2018. Information Diffusion on Social Media During Natural Disasters. *IEEE Trans. Comput. Soc. Syst.* 5, 265–276.
- Donovan, B., Work, D.B., 2017. Empirically quantifying city-scale transportation system resilience to extreme events. *Transp. Res. Part C Emerg. Technol.* 79, 333–346.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A., 2014. The Rise of Social Bots. *arXiv Prepr. arXiv1407.5225* 1–11.
- Figueiredo, F., Benevenuto, F., Almeida, J.M., Fabr, F.F., Almeida, B.J.M., 2011. The tube over time: characterizing popularity growth of youtube videos. *Proc. fourth ACM Int. Conf. Web search data Min. - WSDM '11* 745–754.
- Fry, J., Binner, J.M., 2015. Elementary modelling and behavioural analysis for emergency evacuations using social media. *Eur. J. Oper. Res.* 249, 1014–1023.
- Gallotti R, Bazzani A, R.S., 2016. Towards a statistical physics of human mobility. *Int. J. Mod. Phys. C* 23(09):125.
- Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L., 2008. Understanding individual human mobility patterns 453.
- Han, X.P., Hao, Q., Wang, B.H., Zhou, T., 2011. Origin of the scaling law in human mobility: Hierarchy of traffic systems. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 83, 2–6.
- Hasan, S., Mesa-Arango, R., Ukkusuri, S., 2013a. A random-parameter hazard-based model to understand household evacuation timing behavior. *Transp. Res. Part C Emerg. Technol.* 27,

108–116.

- Hasan, S., Mesa-Arango, R., Ukkusuri, S., 2013b. A random-parameter hazard-based model to understand household evacuation timing behavior. *Transp. Res. Part C Emerg. Technol.* 27, 108–116.
- Hasan, S., Ukkusuri, S., Gladwin, H., Murray-Tuite, P., 2011. Behavioral Model to Understand Household-Level Hurricane Evacuation Decision Making. *J. Transp. Eng.* 137, 341–348.
- Hasan, S., Ukkusuri, S. V., 2014. Urban activity pattern classification using topic models from online geo-location data. *Transp. Res. Part C Emerg. Technol.* 44, 363–381.
- Hasan, S., Zhan, X., Ukkusuri, S. V., 2013c. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. *Proc. 2nd ACM SIGKDD Int. Work. Urban Comput. - UrbComp '13.*
- Hawelka, B., Sitko, I., Beinart, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* 41, 260–271.
- Hong, L., Dan, O., Davison, B.D., 2011. Predicting popular messages in Twitter. *Proc. 20th Int. Conf. World wide web* 57–58.
- Hosseini, S., Barker, K., Ramirez-Marquez, J.E., 2016. A review of definitions and measures of system resilience. *Reliab. Eng. Syst. Saf.* 145, 47–61.
- Huang, C.-M., Chan, E., Hyder, A. a, 2010. Web 2 . 0 and Internet Social Networking : A New tool for Disaster Management ? - Lessons from Taiwan. *BMC Med. Inform. Decis. Mak.* 10, 57.
- Huang, Q., Xiao, Y., 2015. Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery. *ISPRS Int. J. Geo-Information* 4, 1549–1568.
- Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D., 2015. Understanding human mobility from Twitter. *PLoS One* 10, 1–16.
- Kahan, Jerome H., Andrew C. Allen, and J.K.G., n.d. An operational framework for resilience. *J. Homel. Secur. Emerg. Manag.* 6.1 (2009).
- Kass-Hout, T.A., Alhinnawi, H., 2013. Social media in public health. *Br. Med. Bull.* 108, 5–24. <https://doi.org/10.1093/bmb/ldt028>
- Kaufman, S., C. Qing, N. Levenson, and M.H., 2012. Transportation during and after Hurricane Sandy.
- Keim, M.E., Noji, E., 2010. Emergent use of social media : A new age of opportunity for disaster

- resilience. *Am. J. Disaster Med.* 6, 47–54.
- Kim, J., Bae, J., Hastak, M., 2018. Emergency information diffusion on online social media during storm Cindy in US. *Int. J. Inf. Manage.* 40, 153–165.
- Kim, J., Hastak, M., 2018. Social network analysis: Characteristics of online social networks after a disaster. *Int. J. Inf. Manage.* 38, 86–96.
- Kim, W.-S., Yoon, Y.-Z., Bae, J.-H., Soh, K.-S., 2005. Nonlinear characteristics of heart rate time series: influence of three recumbent positions in patients with mild or severe coronary artery disease. *Physiol. Meas.* 26, 517–29.
- Kryvasheyev Y, Chen H, Moro E, Van Hentenryck P, C.M., 2015. Performance of Social Network Sensors During Hurricane Sandy. *PLoS one* 10.2 e0117288. 10.
- Kryvasheyev, Y., Chen, H., n.d. Performance of Social Network Sensors During Hurricane Sandy. *PLoS one* 10.2 e0117288.
- Kryvasheyev, Y., Chen, H., Obradovich, N., Moro, E., Hentenryck, P. Van, Fowler, J., Cebrian, M., 2016. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* 2.3 e1500779.
- Lehmann, J., Gonçalves, B., Ramasco, J.J., Cattuto, C., 2012. Dynamical Classes of Collective Attention in Twitter. *Proc. 21st Int. Conf. World Wide Web. ACM.*
- Leung, D., Law, R., van Hoof, H., Buhalis, D., 2013. Social Media in Tourism and Hospitality: A Literature Review. *J. Travel Tour. Mark.* 30, 3–22. <https://doi.org/10.1080/10548408.2013.750919>
- Liang, X., Zheng, X., Lv, W., Zhu, T., Xu, K., 2012. The scaling of human mobility by taxis is exponential. *Phys. A Stat. Mech. its Appl.* 391, 2135–2144.
- Lindsay, B.R., 2011. Social Media and Disasters: Current Uses, Future Options and Policy Considerations.
- Liu, H., Chen, Y.H., Lih, J.S., 2015. Crossover from exponential to power-law scaling for human mobility pattern in urban, suburban and rural areas. *Eur. Phys. J. B* 88, 1–7.
- Liu, Y., Sui, Z., Kang, C., Gao, Y., 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS One* 9.
- Martín, Yago, Zhenlong Li, and S.L.C., 2017. Leveraging Twitter to gauge evacuation compliance: spatiotemporal analysis of Hurricane Matthew. *PLoS One.* <https://doi.org/10.1371/journal.pone.0181701>
- Mathioudakis, M., Koudas, N., Marbach, P., 2010. Early online identification of attention

- gathering items in social media. Proc. third ACM Int. Conf. Web search data Min. - WSDM '10 301.
- McCallum, I., Liu, W., See, L., Mechler, R., Keating, A., Hochrainer-Stigler, S., Mochizuki, J., Fritz, S., Dugar, S., Arestegui, M., Szoenyi, M., Bayas, J.C.L., Burek, P., French, A., Moorthy, I., 2016. Technologies to Support Community Flood Disaster Risk Reduction. *Int. J. Disaster Risk Sci.* 7, 198–204.
- Mesa-arango, R., Hasan, S., Ukkusuri, S. V, Asce, A.M., Murray-tuite, P., 2013. Household-Level Model for Hurricane Evacuation Destination Type Choice Using Hurricane Ivan Data. *Nat. Hazards Rev.* 14, 11–20.
- Munich RE, 2018. Natural catastrophe review: Series of hurricanes makes 2017 year of highest insured losses ever [WWW Document]. URL <https://www.munichre.com/en/media-relations/publications/press-releases/2018/2018-01-04-press-release/index.html> (accessed 1.15.18).
- Murray-Tuite, P., Wolshon, B., 2013. Evacuation transportation modeling: An overview of research, development, and practice. *Transp. Res. Part C Emerg. Technol.* 27, 25–45.
- Myers, S.A., Leskovec, J., 2014. The bursty dynamics of the twitter information network, in: *Proceedings of the 23rd International Conference on World Wide Web.* pp. 913–924.
- NELSON, B., 2013. A calculated risk. *Nature* 495, 271–273.
- Ni, M., He, Q., Gao, J., 2017. Forecasting the Subway Passenger Flow under Event Occurrences with Social Media. *IEEE Trans. Intell. Transp. Syst.* 18, 1623–1632. <https://doi.org/10.1109/TITS.2016.2611644>
- Nicholson, C. D., K. Barker, and J.E.R.-M., n.d. Vulnerability analysis for resilience-based network preparedness. *Manuscr. Revis.*
- NOAA National Centers for Environmental Information (NCEI) U.S., 2018. Billion-Dollar Weather and Climate Disasters [WWW Document]. URL <https://www.ncdc.noaa.gov/billions/> (accessed 1.15.18).
- Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C., 2012. A tale of many cities: Universal patterns in human urban mobility. *PLoS One* 7.
- Palen, L., Vieweg, S., Liu, S.B., Hughes, A.L., 2009. Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event. *Soc. Sci. Comput. Rev.* 27, 467–480.
- Pappalardo, L., Pedreschi, D., Smoreda, Z., Giannotti, F., 2015. Using big data to study the link between human mobility and socio-economic development. *2015 IEEE Int. Conf. Big Data*

(Big Data) 871–878.

- Pincus, S., Goldberger, A., 1994. Physiological time-series analysis: What does regularity quantify? *Am. J. Physiol.* 266, H1643–H1656.
- Pincus, S., Kalman, R.E., 2004. Irregularity, volatility, risk, and financial market time series. *Proc. Natl. Acad. Sci.* 101, 13709–13714.
- PINCUS, S.M., 1991. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci.* 88.6 2297-2301.
- Qi, W., John E., T., 2014. Quantifying, Comparing Human Mobility Perturbation during Hurricane Sandy, Typhoon Wipha, Typhoon Haiyan. *Procedia Econ. Financ.* 18, 33–38.
- Quercia, D., Ellis, J., Capra, L., Crowcroft, J., 2011. In the mood being influential on twitter mood. *Proc. - 2011 IEEE Int. Conf. Privacy, Secur. Risk Trust IEEE Int. Conf. Soc. Comput. PASSAT/SocialCom 2011* 307–314.
- Rashidi, T.H., Abbasi, A., Maghrebi, M., Hasan, S., Waller, T.S., 2017. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transp. Res. Part C Emerg. Technol.* 75, 197–211.
- Romero, D.M., Huberman, B.A., 2011. Influence and Passivity in Social Media Background on Twitter. *WWW '11 Proc. 20th Int. Conf. companion World wide web.*
- Sadri, A.M., Hasan, S., Ukkusuri, S. V., 2017a. Joint Inference of User Community and Interest Patterns in Social Interaction Networks. *arXiv Prepr. arXiv1704.01706.*
- Sadri, A.M., Hasan, S., Ukkusuri, S. V., Cebrian, M., 2017b. Understanding Information Spreading in Social Media during Hurricane Sandy: User Activity and Network Properties. *arXiv Prepr. arXiv1706.03019.*
- Sadri, A.M., Hasan, S., Ukkusuri, S. V., Lopez, J.E.S., 2017c. Analyzing Social Interaction Networks from Twitter for Planned Special Events. *arXiv Prepr. arXiv1704.02489.*
- Sadri, A.M., Ukkusuri, S. V., Gladwin, H., 2017d. The Role of Social Networks and Information Sources on Hurricane Evacuation Decision Making. *Nat. Hazards Rev.* 04017005.
- Sadri, A.M., Ukkusuri, S. V., Gladwin, H., 2017e. Modeling joint evacuation decisions in social networks: The case of Hurricane Sandy. *J. Choice Model.* 1–11.
- Sadri, A.M., Ukkusuri, S. V., Murray-Tuite, P., 2013. A random parameter ordered probit model to understand the mobilization time during hurricane evacuation. *Transp. Res. Part C Emerg. Technol.* 32, 21–30.
- Sadri, A.M., Ukkusuri, S. V., Murray-Tuite, P., Gladwin, H., 2015. Hurricane Evacuation Routing

- Strategy from Miami Beach : Choice of Major Bridges. *Transp. Res. Rec.* 1–24.
- Sadri, A.M., Ukkusuri, S. V., Murray-Tuite, P., Gladwin, H., 2014. Analysis of hurricane evacuee mode choice behavior. *Transp. Res. Part C Emerg. Technol.* 48, 37–46.
- Shklovski, I., Palen, L., Sutton, J., 2008. Finding Community Through Information and Communication Technology During Disaster Events. 2008 ACM Conf. Comput. Support. Coop. Work 127–136.
- Simon, T., Goldberg, A., Adini, B., 2015. Socializing in emergencies - A review of the use of social media in emergency situations. *Int. J. Inf. Manage.* 35, 609–619.
- Song, C., Koren, T., Wang, P., Barabási, A.-L., 2010. Modelling the scaling properties of human mobility. *Nat. Phys.* 6, 818–823.
- Speranza, Chinwe Ifejika, Urs Wiesmann, and S.R., n.d. An indicator framework for assessing livelihood resilience in the context of social–ecological dynamics. *Glob. Environ. Chang.* 28 109-119.
- Srinivasan, V., Eswaran, C., Sriraam, N., 2007. Approximate Entropy-Based Epileptic EEG Detection Using Artificial Neural Networks 11, 288–295.
- Stai, E., Milaiou, E., Karyotis, V., Papavassiliou, S., 2018. Temporal Dynamics of Information Diffusion in Twitter: Modeling and Experimentation. *IEEE Trans. Comput. Soc. Syst.* 5, 256–264. <https://doi.org/10.1109/TCSS.2017.2784184>
- Stieglitz, S., Mirbabaie, M., Ross, B., Neuberger, C., 2018. Social media analytics--Challenges in topic discovery, data collection, and data preparation. *Int. J. Inf. Manage.* 39, 156–168.
- Tang, J., Liu, F., Wang, Y., Wang, H., 2015. Uncovering urban human mobility from large scale taxi GPS data. *Phys. A Stat. Mech. its Appl.* 438, 140–153.
- Tang, Z., Zhang, L., Xu, F., Vo, H., 2015. Examining the role of social media in California’s drought risk management in 2014. *Nat. Hazards* 79, 171–193.
- The World Bank, 2016. Natural Disasters Force 26 Million People into Poverty and Cost \$520bn in Losses Every Year, New World Bank Analysis Finds [WWW Document]. URL <http://www.worldbank.org/en/news/press-release/2016/11/14/natural-disasters-force-26-million-people-into-poverty-and-cost-520bn-in-losses-every-year-new-world-bank-analysis-finds> (accessed 6.15.17).
- Torres-reyna, O., 2012. Getting Started in Logit and Ordered Logit Regression Logit model. Princet. Univ. , URL <http://dss.princeton.edu/training/Logit.pdf>.
- Ukkusuri, S., Zhan, X., Sadri, A., Ye, Q., 2014. Use of Social Media Data to Explore Crisis Informatics. *Transp. Res. Rec. J. Transp. Res. Board* 2459, 110–118.

- Vaca, C., Aiello, L.M., Jaimes, A., Milano, P., 2014. Modeling Dynamics of Attention in Social Media with User Efficiency. *EPJ Data Sci.* 3.1 5.
- Wang, Q., Taylor, J.E., 2016. Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLoS One* 11, 1–14.
- Wang, Q., Taylor, J.E., 2014. Quantifying human mobility perturbation and resilience in hurricane sandy. *PLoS One* 9, 1–5.
- Wang Q, T.J., 2016. Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster [WWW Document]. Dryad Digit. Repos. URL <http://datadryad.org/resource/doi:10.5061/dryad.88354>
- Wang, W., Pan, L., Yuan, N., Zhang, S., Liu, D., 2015. A comparative analysis of intra-city human mobility by taxi. *Phys. A Stat. Mech. its Appl.* 420, 134–147.
- Washington, S.P., Karlaftis, M.G. and Mannering, F., 2010. *Statistical and Econometric Methods for Transportation Data Analysis*. CRC Press.
- Wu, F., Huberman, B.A., Adamic, L.A., Tyler, J.R., 2004. Information flow in social groups. *Physica A* 337, 327–335.
- Wu, L., Zhi, Y., Sui, Z., Liu, Y., 2014. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS One* 9.
- Yao, C.Z., Lin, J.N., 2016. A study of human mobility behavior dynamics: A perspective of a single vehicle with taxi. *Transp. Res. Part A Policy Pract.* 87, 51–58.
- Yates, D., Paquette, S., 2011. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *Int. J. Inf. Manage.* 31, 6–13.
- Zhang, Z., He, Q., Gao, J., Ni, M., 2018. A deep learning approach for detecting traffic accidents from social media data. *Transp. Res. Part C Emerg. Technol.* 86, 580–596. <https://doi.org/10.1016/j.trc.2017.11.027>
- Zhao, K., Chinnasamy, M.P., Tarkoma, S., 2015a. Automatic City Region Analysis for Urban Routing. 2015 IEEE Int. Conf. Data Min. Work. 1136–1142.
- Zhao, K., Musolesi, M., Hui, P., Rao, W., Tarkoma, S., 2015b. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Sci. Rep.* 5, 9136.