

DIFFUSION OF TWITTER MESSAGES ON DALLAS MASS SHOOTING:
PATTERNS AND FACTORS

by

Yahan Teng, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Geography
August 2017

Committee Members:

Yongmei Lu, Chair

Yihong Yuan

Alexander Savelyev

COPYRIGHT

by

Yahan Teng

2017

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgment. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Yahan Teng, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr. Lu and committee members, Dr. Yuan and Dr. Savelyev. In addition to academic advice, Dr. Lu always understands my situations, difficulties and needs, which I appreciate very much. I learned a lot and felt comfortable when working with her. Both of my committee members are very nice to work with. They are very supportive, always giving me timely feedbacks and wonderful ideas for me to continuously improve my work. I must say that I also feel encouraged when I see how they are devoted to their work.

I would like to thank Monica Medel, PhD student, for sharing the Twitter streaming data. I am also grateful to PhD student Guixing Wei for sharing his expertise in handling large dataset when I needed help. I would like to thank all the professors in the Department of Geography at Texas State University with whom I had an opportunity to take classes. What I learned from these classes prepared me for conducting research and writing the thesis. I want to thank all staff members in our department as well. Without their help and support, my study here wouldn't be such a pleasant experience.

Finally, I would like to thank my friends and family. Being abroad sometimes made me feel lonely. Friends' company and chatting with friends and family members online relieved the loneliness. Their support helped me overcome the difficulties during thesis writing. Especially, I want to thank my mom and dad, for their understanding and support all these years. They supported my decision to study abroad, and I'll never regret being a student at Texas State University.

TABLE OF CONTENTS

| | Page |
|--|-------------|
| ACKNOWLEDGEMENTS | iv |
| LIST OF TABLES | vii |
| LIST OF FIGURES | viii |
| ABSTRACT..... | ix |
| CHAPTER | |
| I. INTRODUCTION | 1 |
| II. BACKGROUND INFORMATION..... | 3 |
| 2.1 Twitter Data | 3 |
| 2.2 Dallas Mass Shooting | 4 |
| III. LITERATURE REVIEW | 6 |
| 3.1 Twitter on Public Response to Emergencies and Extreme Events | 6 |
| 3.2 Information Diffusion in Cyberspace | 8 |
| IV. RESEARCH QUESTIONS | 15 |
| V. DATA..... | 17 |
| VI. METHODOLOGY | 21 |
| 6.1 Descriptive Statistics, Visualization and Regression Model | 21 |
| 6.2 Hot Spot Analysis | 23 |
| 6.3 Logistic Regression..... | 25 |

| | |
|---|----|
| VII. RESULTS..... | 28 |
| 7.1 Temporal Pattern of Tweets' Volume..... | 28 |
| 7.2 Spatial Distribution of Tweets | 32 |
| 7.3 Hot Spot Analysis | 42 |
| 7.4 Logistic Regression Model | 47 |
| VIII. CONCLUSION AND DISCUSSION..... | 51 |
| REFERENCES | 55 |

LIST OF TABLES

| Table | Page |
|--|-------------|
| 1. Selected Data Fields Included in the Tweet Records..... | 19 |
| 2. Statistics of Tweets' Count at County Level | 23 |
| 3. Variables | 26 |
| 4. Statistics of First 20 Tweets | 34 |
| 5. Logistic Regression Results..... | 48 |

LIST OF FIGURES

| Figure | Page |
|---|-------------|
| 1. Study Area | 17 |
| 2. Accumulated Tweet Counts..... | 29 |
| 3. Tweet Counts Per Hour..... | 30 |
| 4. Location of Tweets | 33 |
| 5. Location of First 20 Tweets Tweeted after Occurrence of Dallas Mass Shooting | 33 |
| 6. Tweets in the First 2 Hours, 6 Hours, First Day, from July 7 th to July 31 st at County Level | 37 |
| 7. Tweets Mean Center Change | 38 |
| 8. Tweets' Count and Distance to Event..... | 40 |
| 9. Tweeting Time and Distance to Event..... | 41 |
| 10. Hot Spot Analysis for all Tweets | 43 |
| 11. Population Density Map | 45 |
| 12. Hot Spot Analysis for Tweets Standardized by Population..... | 45 |
| 13. Tweets Density Map | 46 |
| 14. Hot Spot Analysis for Tweets Standardized by Normal Tweets' Volume | 46 |

ABSTRACT

Through analyzing Twitter data on the Dallas mass shooting, the objective of this thesis research is to add to our understanding about the diffusion process of social media messages regarding mass-scale events, which may reflect public response to an event of such. Descriptive statistics, geographic visualization, hot spot analysis, model fitting and logistic regression are used in this study to examine the spatial and temporal patterns of message diffusion on Twitter and the factors associated with these patterns. We found that tweets' volume related to a mass-scale event grew very fast immediately following the occurrence of the event and decreased rapidly after a few hours, showing a negative exponential curve. Time lag to the event of interest and people's daily routine were the two main factors highly related to the volume of tweets. Physical distance is less apparent in information diffusion on social media not only because of minimum friction for online communication but also due to information source other than local witnesses, like news report. Most of the hot spots distributed in places near Dallas. Twitter users close to the event played an important role in message diffusion. These people post related information at the earliest stage and continuously share significantly larger amount of information than people from other places. Findings of this study help us understand the process of messages diffusion on Twitter, which may be used to predict the public's response on social media after emergencies or extreme events.

I. INTRODUCTION

Twitter is an online social networking service which allows people to send and read 140-character messages. Twitter had 317 million worldwide monthly active users in the 3rd quarter 2016 (Statistia 2016) which provide great quantity of study material for social science. Large quantity, real-time feature and high accessibility make Twitter data valuable and beneficial for geography studies and thus this data source has been widely used in studies of emergencies and extreme events. So far, studies related to terrorism on Twitter are mostly devoted to analyzing how terrorists promote terrorism on social media, and information of public reaction after terrorist attacks is overlooked (Klausen 2015; Dean et al. 2012).

Innovation diffusion, which means the origin and dissemination of cultural novelties, was first explored in geography by Torsten Hägerstrand. He argued that any visible cultural landscape, as study objects by geographers, was once innovation originating from somewhere and the aim of analyzing innovation diffusions is to gain understanding of distributional changes through time (Hägerstrand 1967). So far, there are studies about the information diffusion process on cyberspace (Takhteyev, Gruzd and Wellman 2012; Szüle et al. 2014; Li and Sakamoto 2015; etc.). However, when studying information diffusion process on social media related to different kinds of topics, the conclusions about information patterns and influencing factors cannot be universally applied (Eisenstein et al. 2014; Velde, Meijer and Homburg 2015).

Our study interests are information diffusion on social media about mass-scale events including terrorist attacks and other extreme events. In this research, using the Dallas mass shooting as an example, we studied message diffusion on Twitter. This

research may help us develop a better understanding of the general patterns of how the public reacts to mass-scale events through depicting their message tweeting on social media. This study may pave way for future investigations on public reaction on social media to a sudden and massive scale event. Because related studies were limited, our study started from asking general questions to investigate message spatial and temporal diffusion pattern and related factors. Specifically, we want to ask the following questions:

(1) How does the tweets' volume change through time?

(2) Did the volume of tweets follow distance decay?

(3) Did the spatial pattern of tweets' volume at county level change through time and were there any hot spots or cold spots?

(4) Were there any factors related to Twitter users' attributes and characteristics of tweets significantly related to the spatial distribution of tweets or the message diffusion process?

II. BACKGROUND INFORMATION

2.1 Twitter Data

Twitter is an online social networking service which allows people to send and read 140-character messages (known as “tweets”). Twitter gained popularity rapidly since its launch in 2006 and already had 317 million worldwide monthly active users in the 3rd quarter 2016 (Statistia 2016), in which 82 percent use Twitter on mobile devices (Twitter 2016). A tweet is a post made on the Twitter online message service (Merriam-Webster n.d.). Tweets may reflect many aspects of user’s life, including feeling, thoughts, opinions, interests, activities and so on, which can become abundant materials for social studies. By “retweeting” a tweet, Twitter users can easily share their tweets or tweets from other people to all their followers. When editing tweets, users may choose to report the location, indicating where the tweets were created. These “geo-tagged” tweets have latitude and longitude information or a bounding box reflecting geometrical boundary of the location where a tweet was posted. A geo-tagged tweet will be associated with precise location (latitude and longitude) when the user’s Twitter version is earlier than 6.26 for iOS or 5.55 for Android, or when the user chooses to toggle on the “Share precise location” option for later version (Twitter n.d.). Otherwise, the location information will be a bounding box (Twitter n.d.). Tweets, location information and other user profile information can be collected through Twitter Application Programming Interfaces (API) for free. Millions of tweets and their location information provide a wealth of material for geography research on many topics (Chew and Eysenbach 2010; Huang et al. 2015; Malleson and Andresen 2015; etc.).

2.2 Dallas Mass Shooting

A common definition of terrorism is the use of violence or the threat of violence, especially against civilians, in the pursuit of political goals (The American Heritage Dictionary of the English Language 2016). Although terrorism and terrorist attacks have long been a concern for American people and people all over the world, most of the related studies using social media were qualitative research focused on how terrorists and terrorist organizations tweeted jihadist ideas to public (Klausen 2015; Dean et al. 2012). So far, little is known about public response to these events on social media, and how the messages diffuse across cyberspace. This situation brings us the idea of conducting a case study, to look into the diffusion process of a special event using social media data with quantitative approaches. In this study, we are going to examine the diffusion of and the impacting factors for tweets regarding the Dallas mass shooting, an attack happened on July 7th 2016. A few minutes before 9 pm (Central Time), a man named Micah Xavier Johnson fired upon a group of police officers at Main Street and S. Lamar Street in Dallas, Texas during a peaceful protest, killing five officers and injuring nine. Two civilians were also wounded. Johnson was then killed by police with bomb in a building on the campus of El Centro College (Hacker 2016). A few days earlier, police killed black men Alton Sterling in Baton Rouge, Louisiana, and Philando Castile in Falcon Heights, Minnesota. It is said that Johnson was angry about police shootings involving black men and stated that he wanted to kill white people, especially white police officers. President Barack Obama delivered a speech after this attack and called the shooting a "vicious, calculated, despicable attack" and a "tremendous tragedy" (Mallin and Caplan 2016). Because no ties were found between Johnson and international terrorist or

domestic extremist groups (Kennedy 2016), whether this event could be considered as a terrorist attack raised a wave of discussion. While many people, media and law experts considered it as an act of domestic terrorism, many other thought it shouldn't be called "terrorism" (Bergen and Sterman 2016; John. T. Floyd Law 2016). No matter it's terrorism or not, because terrorism related shooting attacks are a subtype of mass shootings, findings from this case study will help us to gain better understanding of message diffusion process on social media related to terrorist attacks.

Beside the availability of data, we chose the Dallas mass shooting as example in this study was mainly because this tragedy was a big shock to the entire country and thus was of very broad impact and drew wide attention in social media, including Twitter. We expect the related tweet messages on Twitter to diffuse quickly across space and exist for an extended time period, so that we have an opportunity to examine the whole diffusion process of the related messages on Twitter.

III. LITERATURE REVIEW

3.1 Twitter on Public Response to Emergencies and Extreme Events

Large quantity, real-time feature and high accessibility make Twitter data valuable and beneficial for geography studies in many ways. Because tweets' contents are typed words or shared links reflecting user's feelings, thoughts, opinions, interests, activities and so on, Twitter data can be used in research reflecting public attitude towards many different kinds of topics, including climate change (Kirilenko and Stepchenkova 2014), depression (Yang and Mu 2015; Yang, Mu and Shen 2015), allergy (Gesualdo et al. 2015), crime (Malleon and Andresen 2015) and so on. Because of the real time nature of Twitter data, Twitter data especially stands out in collecting information of emergencies and extreme events with minimum cost of time and money (Chew and Eysenbach 2010; Nelson et al. 2015). During emergencies and extreme events, public response may change quickly. It would be too difficult for surveys including online surveys to capture and monitor dynamically the change of public attitude and the evolving of new knowledge due to several reasons. First, surveys data have time delay, thus this data source is not suitable for real time surveillance. Second, during emergencies or extreme events, most of the time people don't have enough prior knowledge about it and more information will be collected while an event develops for a better understanding about it (Takayasu et al. 2015). Surveys are always insufficient in acquiring and mining new knowledge and new ideas because of the limitation from pre-designed questions. Also, surveys may cost enormous money and efforts. Under these circumstances, real-time social media data, like Twitter data, become a good alternative data source. Twitter can capture dynamic status of public response with high temporal

resolution. Also, Twitter data is very suitable for daily management and surveillance of disasters and epidemic diseases for its low cost, as demonstrated by a fair amount of studies on detecting and monitoring events information and public reaction related to pandemics and public health (Chew and Eysenbach 2010; Widener and Li 2014; Signorini, Segre and Polgreen 2011), disaster management (Xiao, Huang and Wu 2015; Huang et al. 2015; Huang and Xiao 2015) and many other different subjects.

In the existing studies about the application of Twitter in public response of emergencies and extreme events, different research directions exist. Despite the data bias that some demographic groups (e.g. young adults, urban residents, etc.) tweet more than their peers (Smith and Brenner 2012), Twitter is still a valuable database in which information provided by millions of users could be extracted to answer many questions including that people perceive during an event. Thus, Twitter data in many studies were used as one type of secondary data to reveal people's feeling towards and cognition change after a specific event (Chew and Eysenbach 2010; Xiao, Huang and Wu 2015; Wu et al. 2014). However, compared to survey, there is no pre-designed question for people to answer on Twitter, and useful information was often identified and extracted later from millions of tweets. In order to make better use of Twitter data, the first group of studies were conducted to acquire information from tweets. Among those studies, some were devoted to extract, synthesize and present information (Cheng and Wicks 2014). Some developed methods and algorithms to interpret misspelled, abbreviated, incomplete words correctly (Han, Tsou and Clarke 2015). Others extracted sentiment information, toponyms information and many other information related to one topic with

efficiency (Chan, Vasardani and Winter 2014). Still others developed tools for collecting and presenting information, statistical analysis and management (Huang et al. 2015).

A second group of studies examined Twitter itself as a tool for providing and spreading information during an event. These studies aim at understanding the characteristics of tweets, finding out how, from whom and in what ways they are created, recreated, what kind of information are included, to gain better understanding of how information diffuses on social media and what affects the process. Twitter is a dynamic system where users highly interact and communicate with each other by creating, synthesizing and redistributing information (Starbird et al. 2010). During emergencies and extreme events, like a disaster, users who are affected or interested in the event become more active, producing more tweets than average and may actively volunteer to keep doing so for a longer time (Kim 2014; Starbird and Palen 2010). More people from afar receive information about the events of interest within a shorter time period. Because understanding how messages diffuse in cyberspace and what affects the process of message diffusion are closely related to this study, the next section of this thesis will discuss the related literature.

3.2 Information Diffusion in Cyberspace

As most people know, the term “innovation diffusion”, which means the origin and dissemination of cultural novelties was first explored in geography by Torsten Hägerstrand in his dissertation “Innovation Diffusion as a Spatial Process”, in which he argued that any visible cultural landscape, as study objects by geographers, was once innovation originated from somewhere, and the aim of analyzing innovation diffusions is

to gain understanding of distributional changes through time (Hägerstrand 1967). According to diffusion theory, innovation, time, communication channels and social system are the four main components that influence the innovation diffusion. This study highlighted the importance of analyzing spatial processes, and it has set up the foundation for the later developments in spatial diffusion theory and models. In the next few decades, there were some critiques about methods and approaches of the original work, discussions on the limitation of only considering external factors, and examination of the efficiency of applying Monte Carlo simulation (Meir 1982; Yapa 1974; Allaway et al. 1994; etc). There were new developments and applications in diffusion research as well. It is important to note that the research of diffusion process can be very broad, because the definition of “innovation” is not restricted to technological developments, but can be broad enough to include anything-new (Clark 1984). For example, Pred (1975) developed models to reveal how the locational patterns of major job-providing organizations and the inter-metropolitan circulation of specialized information interacted to influence the process of city-system development. Britain (2005) applied diffusion theory to describe the diffusion process of a number of linguistic innovations from the Southeast of England to a rural area to the northwest of East Anglia. Bjørkhaug and Blekesaune’s study (2013) explored that how neighborhood effect influenced the diffusion of organic farming in Norway.

Diffusion in cyberspace has similarities with innovation diffusion process in real world, but the ways how people connected with each other have differences. In the above-mentioned works, no matter what subject and its diffusion were studied, or what social economic factors were analyzed to explain a diffusion process, they all considered

that geographic proximity influences diffusion process. In other words, if a place is closer to the birthplace of innovation, chances are that the innovation is more likely to diffuse to this place. But the diffusion process in social media networks is very different from that in physical world. Because the way people connect to each other is very different on social media, the social media “neighbors” might be thousand miles away in real world. For example, on Twitter, users “tweet” to share information. By “following” other people, Twitter followers can read immediately everything tweeted by the people they follow, no matter how far they are from each other in physical world. Although in existing studies, there is a widely recognition that region, distance, national border and language barrier affect social ties online (Takhteyev, Gruzd and Wellman 2012; Szüle et al. 2014), the distance and some other factors on social media networks is different from physical world because the change of social system and communication channels (Lee, Agrawal and Rao 2015). Because physical distance does not significantly affect the speed of information diffusion online, the time needed for diffusion through physical distance becomes negligible and the spatial scope of potential influence increases greatly. Furthermore, since social media is not a closed space, the interaction between cyberspace and physical world may increase the complexity of what and how different factors may impact information diffusion on social media, as implied by some studies (Tafti, Zotti and Jank 2016; Eisenstein et al. 2014). The information people share and redistribute on Twitter may come from not only personal opinion or others’ tweets, but also external sources like television, internet, and newspapers and so on, which adds complexity to the diffusion process.

Research regarding information diffusion on social media in geography started not very long time ago, thus there are still questions need further discussion. First, there are a number of research aspects related to defining information diffusion on Twitter. In existing literature, the studies on the same social media diffusion research topic, an extreme event for example, may explore a variety of aspects with different research focuses and data. In some studies, tweets containing certain keywords (e.g. tweets with keyword “Tohoku earthquake”) or fitting certain criteria were treated indistinctively as valid data (Chew and Eysenbach 2010). Other studies looked into one or a few particular aspects, for example, retweeting behavior in an earthquake, or rumor in a bombing (Kim 2014; Lee et al. 2015). Studies of “retweeting” behavior are very popular, especially when the research topics are related to emergencies and extreme events (Xie et al. 2015; Li and Sakamoto 2015; Lee 2013). This is partly because retweeting is a fast way of accelerating information diffusion and thus is very important in spreading information during these events. Some studies also argued that the phenomenon of retweeting messages related to social issues, was analogous to the diffusion of an innovative idea (Lee 2013). This argument is understandable, because retweeting means that one user passes a message to his/her followers, just like innovation in real world passes from one person to one or a few others. Similar topics include rumor diffusion on social media, where a rumor once was created from one account, might spread to million others through re-sharing (Kim 2014). However, not all ideas diffuse this way. In many cases, information has multiple sources that appear at different times. Also, cyberspace and real world highly interact with each other so that it would be difficult to trace the process online. Think of an extreme example of how people learn and use popular online

language. A person might see a new word on Twitter for the first time, understand the meaning and practice in daily life. He may also unintentionally introduce the word to others during everyday conversation and use the word repeatedly. He never retweeted using the word but may include the word when he creates a new tweet. In this case, daily communication works as an external factor to help the diffusion process of the new word usage on Twitter.

Also, depending on different types of diffusion subjects, the diffusion time, extent, communication channels, its influencing factors and spatial-temporal dynamics are very different (Eisenstein et al. 2014; Velde, Meijer and Homburg 2015). For example, Eisenstein et al. (2014) modeled the usage of new words and phrases in tweets. Following a few phrases, they found that it took a few months or even more time for a linguistic change to diffuse to a couple of states or to the whole nation. But information related to emergencies and extreme events diffused very fast so that the most of related tweets may be collected within several hours or a few days after the happening of the event (Yoo et al. 2016; Sutton et al. 2015). In order to cover enough information as well as effectively process data analysis, we included three weeks' Twitter data in our case. The spatial extent of diffusion and the amount of tweets are also highly depending on the topic or theme of information. While diffusion of popular lexical change in social media did follow geographical proximity (Eisenstein et al. 2014), communication and information dissemination on the topic of "energy saving and climate change" seemed to be trapped in organizational loops and the related information were only shared by a group of people who were interested in the topic (Mohammadi et al. 2016). Even for a similar topic (e.g. natural hazard), depending on the scope of influence, spatial extent of

tweets dissemination may be very different (Starbird et al. 2010; Takayasu et al. 2015). Because we believe that the case we studied had nationwide impact, the study area included all the tweets we collected, which was southwest part of contiguous United States. For tweets related to different topics, the influencing factors, may vary largely, which are highly dependent on the topic of interest (Nagarajan et al. 2010). Considering emergencies and extreme events, a factor may affect the diffusion process of Twitter message related to one aspect, but not necessarily influence the Twitter message diffusion about another aspect. For example, the number of and decline in the speed of rumor diffusion in an earthquake was found to be related to the tweets correcting the rumor (Takayasu et al. 2015). Tweets' number is related to disaster impact and damage; it could also be related to community socioeconomic and demographic factors of a place (Xiao, Huang and Wu 2015). In existing literature, simple linear regression and multiple variable regression models are popular methods for studying the relationship between tweets' quantity and factors, or between one factor and other factors (Stieglitz and Xuan 2013; Oh, Agrawal and Rao 2013; Sutton et al. 2015).

So far, most of the terrorism related studies using social media were qualitative research focusing on how terrorists and terrorist organizations tweeted jihadist ideas to the public (e.g. Klausen 2015; Dean et al. 2012). Study on message diffusion as an indicator of public reaction after a terrorist attack is very limited. There are a few studies that looked at the mechanism of message diffusion and transmission on Twitter after a terrorist attack, as well as rumor diffusion after an attack (Oh, Agrawal and Rao 2013; Sutton 2015). In this study, taking the Dallas mass shooting as an example, our study interest is also the information diffusion process of social media messages on mass

shootings, to better understand terrorist shooting attacks. To be specific, we will study the temporal and spatial diffusion pattern of the related Twitter messages and try to identify the possible influencing factors.

IV. RESEARCH QUESTIONS

The objective of this study is to add to our understanding about the diffusion of Twitter messages on mass-scale events using the Dallas mass shooting as a case study. Our research began from asking general questions about message diffusion on social media, including speed, extent and the related factors. Findings of this study may help us to get an overall picture of how attack related messages diffuse on Twitter and to better comprehend the impact of such an event. In addition, it will help us understanding how people respond on social media after such attacks, and therefore may potentially help agencies to prepare for and manage any possible aftermath.

The first step of this research is to examine the spatial and temporal diffusion patterns of the selected Twitter messages. Considering that terrorist attacks are emergencies that affect people suddenly, and the relevant information are usually reported by news and bring up a lot of discussion, the spatial and temporal features might be similar to that of natural disasters, like an earthquake. In this study, by visualizing data and applying hot spot analysis, we examined the relationship between time, distance to the event and tweets' number, as well as interpret the hot spots and cold spots across space. We then investigated a series of factors which might influence the temporal and spatial patterns of the related tweets, including both attributes of tweets (such as source, time) and user's information (including friends' counts, follower counts and so on).

Specifically, the research questions are as follows:

- (1) How does the tweets' volume change through time?
- (2) Did the volume of tweets follow distance decay?

(3) Did the spatial pattern of tweets' volume at county level change through time and were there any hot spots or cold spots?

(4) Were there any factors related to Twitter users' attributes and characteristics of tweets significantly related to the spatial distribution of tweets or the message diffusion process?

Data and methodology for this study are discussed in the next two sections.

V. DATA

In this research, three types of data will be used for analyzing: Twitter data, the U.S. county shapefile and census data.

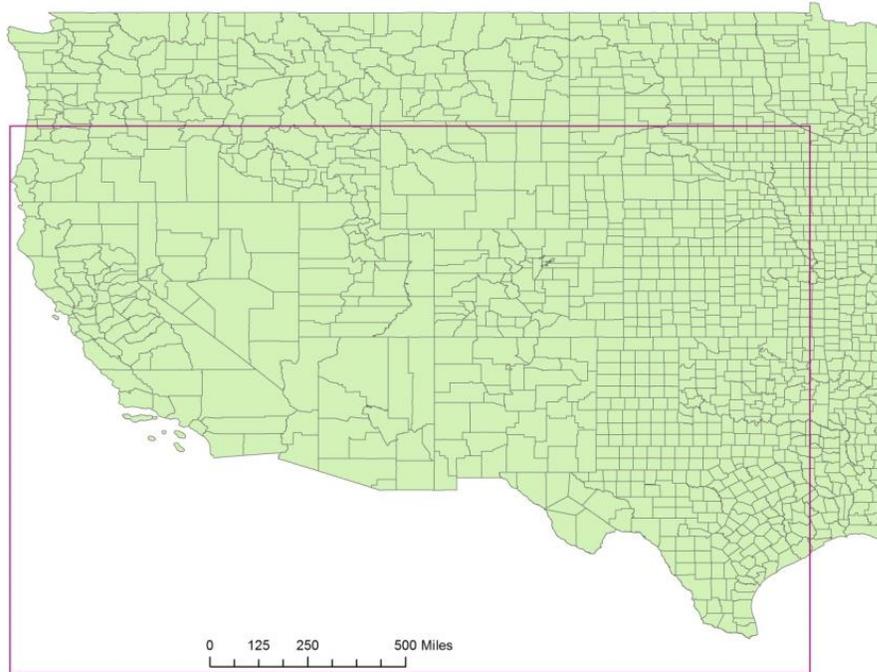


Figure 1. Study Area

Twitter data were collected using Twitter application programming interface (API) covering middle and west part of the continental USA as shown in the bounding box in Figure 1. We used Twitter data from July 7th 2016 to July 31st 2016, covering a total of 24 days including and after the day of the police mass-shooting event in Dallas. All the data are in JSON format where each Twitter message contains keys and values covering a variety of information including tweet id, tweets time, tweets content, all sorts of user background information and so on (Twitter n.d.). The data fields selected for this study are listed in Table 1 using one Twitter account example. When sending out the tweets, if the Twitter users choose to geo-tag their tweets, there would also be geographic

location information recorded in the data. Based on the precision of location a user chose, as well as Twitter service and cell phone devices, location information is recorded as either points or bounding boxes. Points contain latitude and longitude of points where the tweets were sent, while bounding boxes include geographic coordinates of four points which define the areas chosen by the user. According to literature, the geo-tagged tweets are less than 1% of all the tweets at a certain time and about 10 percent location information in those geo-tagged tweets are recorded as points (Mitchell et al. 2013). Although the proportion of tweets with geolocation information is small, it is believed that this set of tweets can still provide very useful information due to the quantity that as many as 500 million tweets are tweeted on Twitter each day.

In this study, we only examined tweets data related to the Dallas mass shooting and with location information. We used PyScriper (Kvlahos 2016) and wrote code in Python to process the data. We identified a series of keywords from tweets and news related to this event and use combinations of keywords for extracting the related tweets. The related keywords were grouped into four categories: those related to the shooter (“Micah Johnson”, “Micah Xavier Johnson”), those related to the victims (“police”, “cop”), those related to the action or consequence (“shoot”, “shot”, “attack”, “die”, “kill”, “death”), and those related to the place (“Dallas”). If tweets contained one keyword from “shooter”, or if tweet contained one keyword in “victims”, one in “action or consequence” and one in “place” at the same time, they would be considered as tweets related to this event and be included for further analyses. After searching through all the tweets in our database, a total number of 5353 tweets were selected. Using random sampling method, we manually examined about 10 percent of these tweets. It turned out

that in our manually-examined sample more than 98 percent of the tweets were valid information that was closely related to this mass-shooting event.

Table 1. Selected Data Fields Included in the Tweet Records

| Data fields | Example | Explanation |
|------------------------------------|--|---|
| created_at | Sat Jul 02 18:14:07 +0000 2016 | time the tweet created |
| source | \u003ca href=\"http://twitter.com/download/android\" rel=\"nofollow\"\u003eTwitter for Android\u003c/a\u003e | Whether the tweet is tweeted from android or apple or personal computer |
| Text | thanks for the shoutouts to the fire and police. You think if we put a ski mask on the city of Dallas will pay us? | content of tweet |
| user:verified | FALSE | whether the user has a verified badge |
| user:followers_count | 340 | number of followers |
| user:friends_count | 1137 | number of users this user is following |
| user:favourites_count | 139 | number of favorites this user has |
| user:statuses_count | 3479 | number of tweets this user has |
| user:lang | en | the user's selected default language |
| entities:hashtags | [] | text and indices of hashtags |
| entities:user_replies:screen_name | DwainPrice | user replied screen name |
| entities:user_mentions:screen_name | DwainPrice | user mentioned screen name |

We then assigned location information for each of these tweets. If the tweets contained point location information, the latitudes and longitudes would be taken as location information. If the tweets, however, had bounding box with diagonal line shorter than 80 kilometers, the latitudes and longitudes of the center points of the bounding box

was used to represent tweet location. We used 80 kilometers because the average area of the counties in the U.S. is 3128 km². Assuming the shape of county is similar to square, and then the diagonal line is about 80 kilometers. In this way, on average the error range of the points' location was controlled within a county. There were 4223 tweets whose location was generated this way, including 527 tweets with point location information. The length of diagonal line was calculated as spherical distance.

The U.S. county shapefile and census data were downloaded from the United States Census Bureau website. The county map with data the related tweets data joined to it was projected and analyzed in ArcGIS. Census 2010 population data for each county was extracted and used in the later analysis.

VI. METHODOLOGY

6.1 Descriptive Statistics, Visualization and Regression Model

To understand the relationship between time, distance and message quantity on Twitter, we applied descriptive statistics and visualization analyses.

Firstly, scatter graphs of the hourly volume of tweets and cumulative volume of tweets through time were drawn to show the public response and its change through time. After drawing the graphs, we found the relationship between time and hourly volume of tweets roughly followed an exponential curve. In order to testify and better describe the relationship, we used linear regression with logarithmic transformation to model the tweets' volume through time to describe the change more accurately. The formula is as follows:

$$\ln D = a + b_1 r_1 + b_2 r_2 \quad (1)$$

where D is the number of tweets every hour; r_1 is the time interval between the occurrence of the Dallas mass shooting and the time where the tweet was posted; $r_2 = \text{Min}(t_1, t_2)$, t_1 is the time interval between tweeting time (local time) and 12pm, while t_2 is the time interval between tweeting time and 6pm; a , b_1 and b_2 are constant terms. The variable r_2 designed in this model was following people's daily tweeting regularity. According to our experience, usually more tweets are created during people's lunch break and the hour immediately after work, which are around 12pm and 6pm. Adding variable r_2 allows us to capture the time decay effect of tweeting activities on the topic from these peak hours.

Secondly, we mapped out the tweets based on where they were posted to see the spatial distribution of tweets. We also joined the tweet point map to the county polygon

map to get tweets' counts in each county for the different time periods to see if the spatial pattern of tweets changes at different time phrases. After checking the histograms of count distributions of tweets across the counties, observing spatial distributions of tweets' counts and comparing the different classification methods, we chose to manually define class numbers and break values based on classification results from Geometrical Interval classification method. Geometrical Interval classification method is based on an algorithm that minimizes the square sum of elements per class and it works reasonably well on data that are not distributed normally. In order to make comparisons between different time phrases, people often use Quantile method, which assigns the same number of data values to each class, to avoid distorting spatial patterns when comparing multiple maps. However, in our case, the tweets are not normally distributed nor linearly distributed. Dallas and places near Dallas had much more tweets than other places that the maximum number of tweets tweeted in one county is 1117, while most of the counties in the study area had no tweet or just a few tweets. Statistics of tweets' counts and counties is shown in Table 2 below. After using Quantile classification to group equal number of counties in each class, we found that the result maps are misleading. Because most of counties had values of 0 or 1, equal number of features in each class was not guaranteed. After applying other methods, we found Natural Break method and others were not good enough to show the spatial differences either. The Geometrical Interval classification method was found to work better than all other classification methods when classifying similar values into same classes. So we defined break value for each class based on results from this method. We used fixed class values so that it would be easier to comparison between different time phases.

Table 2. Statistics of Tweets' Count at County Level

| Tweets' Count | Number of Counties |
|---------------|--------------------|
| 0 | 3026 |
| 1 | 64 |
| 2 | 29 |
| 3 | 12 |
| 4 | 14 |
| 5 | 8 |
| 6-10 | 22 |
| 11-20 | 13 |
| 21-50 | 15 |
| 51-1117 | 17 |

To further look at how distance affects the public response to the event through time, we drew a scatter graph of tweeting time and distance. We also created a histogram showing the number for tweets every 50 miles away from the event. The distance interval is 50 miles because: first, this is almost the average distance between two neighbor counties and thus county level difference can be reflected; also, 90 percent tweets have location information as bounding box which means considering data resolution, too small distance interval doesn't provide extra reliable information.

6.2 Hot Spot Analysis

In order to measure if there were any places having significantly larger number of tweets than other places, hot spot analysis in ArcGIS was conducted to reveal any hot spots or cold spots of tweets at county level. Hot spot analysis is based on local Getis-Ord G_i^* method developed by Ord and Getis in 2001 (Ord and Getis 2001). This statistic tests for local spatial autocorrelation in the presence of the global autocorrelation of

heterogeneous spatial data (Ord and Getis 2001). The Getis-Ord G_i^* local formula is as follows:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{x} \sum_{j=1}^n w_{i,j}}{s \sqrt{\frac{[n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2]}{n-1}}} \quad (2)$$

where x_j is the attribute value for feature j ; $w_{i,j}$ is the binary spatial weight matrix between feature i and j with a value of 1 if j is within defined distances from i ; n is the total number of features and :

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (3)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (4)$$

Significant large results are at with 90%, 95% and 99% confidence level. High value centers indicate hot spots, and low value centers indicate cold spots.

We firstly run hot spot analysis for spatial data of all the event related tweets at county level, to find out places where number of tweets are significantly greater than other places. But because larger population usually means larger volume of tweets and the hot and cold spots of tweets will easily reflect the hot and cold spots of population, we then run hot spot analysis for tweets' counts standardized by population. We also conducted hot spot analysis for the mass shooting related tweets after standardized by the overall tweets' volumes, or social media population. To calculate the tweets' volumes for each county, we used point tweets tweeted from July 1st to July 7th (Greenwich Mean Time). Normal tweets' volumes in each county represent population's general social media activities. Because characteristics of Twitter users are related to many social economic factors, social media population may work better as the standardization variable than the actual population.

6.3 Logistic Regression

After exploring the spatial and temporal patterns of message diffusion, we investigated if there were factors related to the spatial message diffusion process. As discussed in literature review section, based on different topics, the influencing factors for a diffusion process may be different. Related studies on terrorist attacks have studied the factors related to Twitter message retransmission and to the mechanisms of rumor diffusion through time with linear regression approaches (Oh, Agrawal and Rao 2013; Kim 2014; Sutton et al. 2015; Lee, Agrawal and Rao 2015). According to what they have found, during emergencies and extreme events, users affected by or interested in the event became more active, producing more tweets and sharing more information (Kim 2014). Number of followers and hashtag usage can improve the chances of message diffusion (Lee, Agrawal and Rao 2015). These factors may help to explain why certain areas had more tweets than other places. In this study, we wanted to testify if the factors mentioned above and some other factors are related to the spatial distribution of diffused messages. From results of hot spot analysis of tweets' volume at county level, we could find that some counties had significant large of small number of tweets and were identified as hot spots or cold spots. This formation of hot spots and cold spots may be related to a series of factors. So we chose to use logistic regression to look at the relationship between the dependent variable and a series of factors. The model is as follows:

$$\text{Ln}[Y/(1-Y)] = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n \quad (5)$$

where the dependent variable Y is if the tweet was in hot spots (yes=1, no=0); $x_1, x_2, x_3 \dots x_n$ are independent variables; $a, b_1, b_2, b_3 \dots b_n$ are constant terms.

In innovation diffusion theory, Hägerstrand argued that instead of considering situation at certain given times (t_{99} , t_{100} , etc), we must consider the situations' change between time t and $t+\Delta t$ (Hägerstrand 1967). Therefore, we included time as independent variables. For other independent variables, findings from existing literature were used to guide the selection of factors. For example, hashtag was tested in related studies and was found to be significant, and thus was included in our model. Some other variables were included in the model based on our filed knowledge of their relationship to tweet volume. For example, we included “verified account” as an independent variable because many of the verified accounts belong to news agencies that tweet a lot more than ordinary people do. The variables that were extracted from tweet message data fields and included in the model are listed in Table 3 below.

Table 3. Variables

| Variable | Explanation |
|----------------|--|
| language | the Twitter user's selected default language. English=1, else=0 |
| status count | number of tweets this Twitter user had |
| favorite count | number of favorites this Twitter user had |
| friend count | number of users this Twitter user was following |
| follower count | number of followers this Twitter user had |
| verified | whether the user had a verified badge. Yes=1, No=0 |
| source | source the tweet was tweeted from. Android=1, iPhone=2, else=0 |
| hashtag | if the tweet contained hashtag. Yes=1, No=0 |
| link | if the tweet contained link. Yes=1, No=0 |
| mention | if the tweet mentioned someone. Yes=1, No=0 |
| reply | if the tweet replied to someone. Yes=1, No=0 |
| three hours | if the tweet was tweeted within three hours after the event. Yes=1, No=0 |
| six hours | if the tweet was tweeted within six hours after the event. Yes=1, No=0 |
| day | if the tweet was tweeted during daytime (CST 6:00am-6:00pm). Yes=1, No=0 |
| weekday | if the tweet was tweeted during weekdays. Yes=1, No=0 |

The logistic regression analysis consisting of three models were conducted to explore the relationship between a series of variables and if a geo-tagged tweet was tweeted in a hot spot. In Model 1, only variables of user account information were considered. Variables included in this model are user default language, status count, favorite count, friend count, follower count, verified, and source. Model 2 added variables related to tweet attributes, which were hashtag, link, mention, reply and three hours. In Model 3 we added variables reflecting tweet posting time so that the relationship between posting time and if a tweet was in a hot spot was explored. When adding new variables to model 2 and model 3, change of relationship between existing independent variables and dependent variable, as well as relationship between dependent variable and new independent variables could be observed.

VII. RESULTS

7.1 Temporal Pattern of Tweets' Volume

Figure 2 and Figure 3 present statistics of tweet counts. The time label of first related tweet in our database was July 8th, 01:58:54 Greenwich Mean Time, which was July 7th 20:58:54 local time in Dallas (Central Standard Time). This tweet was posted right after the happening of the shooting. In order to facilitate data statistical process and analysis, we defined the 8 -9 pm, where the first tweet was tweeted, as the first hour. In both Figure 2 and Figure 3, the x axis is time. Although we collected Twitter data for more than three weeks, in these two graphs, we only showed tweet counts up to the first 72 hours. That was because there were 4387 pieces of related tweet messages posted in the first three days, which was about 82 percent of the overall tweets. In this way we improved the temporal resolution as much as possible while avoiding losing too much information. The y axis of Figure 2 is cumulative tweet counts by hour and the y axis of Figure 3 was counts of tweet per hour started from the first hour until the end of the third day.

From the Figure 2, we can tell that after the happening of the event, the tweet number grew very fast in the beginning. Tweet count reached 955 in the first 6 hours and 3393 in the first day. The number of tweets in the first day was more than all the other tweets in the next three weeks. Comparing to existing literature, the tweets' trend of the Dallas mass shooting is different from the tweets trend during the H1N1 Outbreak (Chew and Eysenbach 2010) and that related to climate change (Kirilenko and Stepchenkova 2014). Posting of Twitter messages related to the H1N1 outbreak in 2010 lasted for a few months, while the tweets about the climate change were tweeted by people who

concerned all year round. However, the tweet pattern of the Dallas mass shooting was very close to the tweets pattern during the Tohoku earthquake, which happened on 11 March 2011, in Japan (Kim 2014). During the Tohoku earthquake, the tweets grew rapidly in the first 12 hours after the earthquake and flattened out afterwards. From both our data and existing literature, we can conclude that the temporal pattern of tweets is highly related to the event's topic. Disasters and other emergencies may have rapid growth in tweets' volume, but how long this trend may last depends on the developments of the situation and the influence of the event. The case used in this study was known by many people in this country when it happened, thus many people shared related information and their feelings and opinions right after it. However, as time went by, although there was influence that can't be erased, less and less new information was discovered and people's interests shifted quickly.

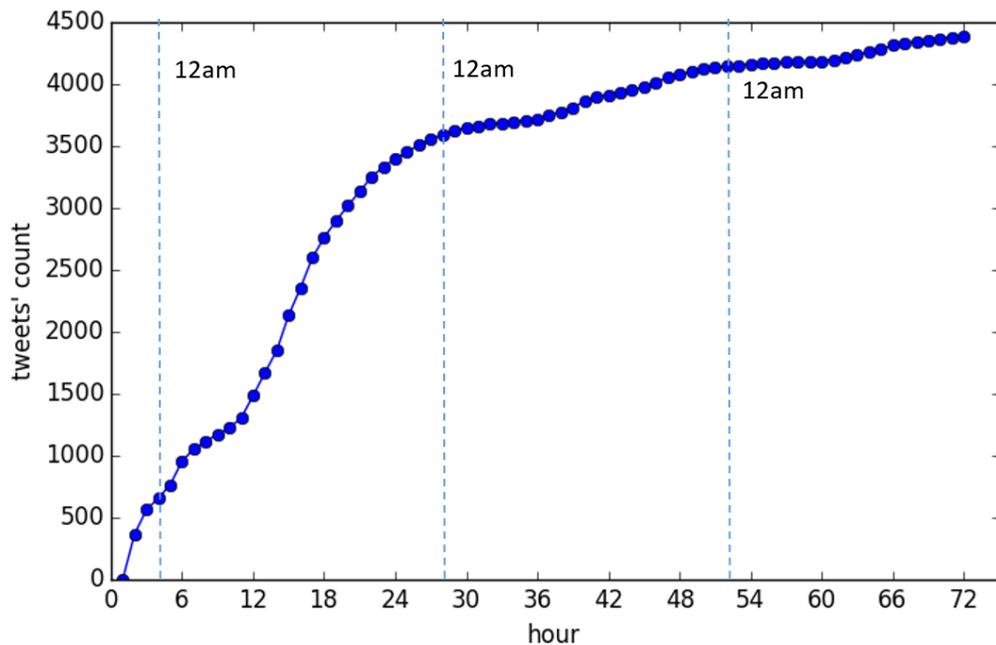


Figure 2. Accumulated Tweet Counts

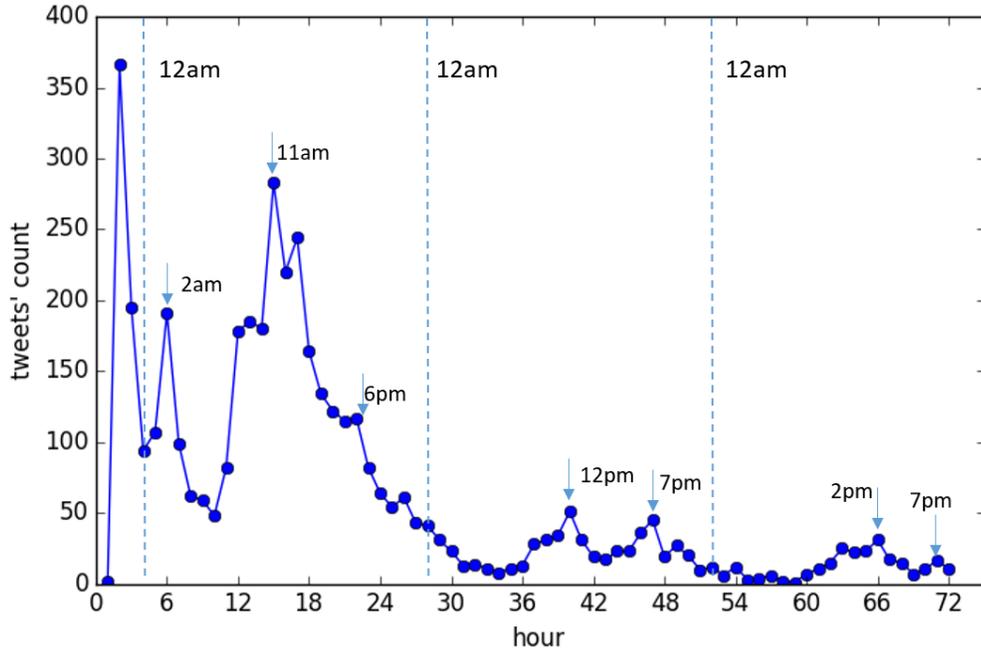


Figure 3. Tweet Counts Per Hour

Figure 3 clearly showed that the peak of tweets' counts per hour appeared in the hour right after the event with 366 tweets in total, followed by a drop in tweets' counts in general after 15 hours. Also, time cycle was observed within the time decay trend in general, which was also consistent with information diffusion on Twitter of emergency situations (Kim 2014; Yoo et al. 2016). This result can be explained by regular work and rest schedule of people. Except for the first few hours, we can observe a few peaks around the 6th, 15th, 22^{ed}, 40th, 47th, 63-66th and 71st hour. After converting to Central Time, these are about 2 am, 11 am, 6 pm, 12 pm, 7 pm, 11 am – 2 pm and 7 pm. It was not difficult to understand that rush hours of tweeting behavior are around 11-12 in the morning and 6-7 in the afternoon. These are the hours that most people have their lunch break and finish one day's work. People usually browse and share information on social media during these time periods.

The linear regression with logarithmic transformation revealed that all coefficients are significant at 99% confidence level and the adjusted R^2 is 0.82:

$$\ln D = 6.28 - 0.054r_1 - 0.221r_2.$$

This result indicated that the decay of tweets' volume followed two centers. First, the tweets' volume in next hour was about 0.95 times of the hour before. Second, for tweets' counts of two hours t_1 and t_2 , if the t_1 was one hour closer to 12 pm or 6 pm than t_2 , tweets' count in hour t_1 would be 1.25 times greater. However, according to the graph in Figure 3, the peak around 2 am, about 6 hours after the event seemed unusual. This violated the fact we observed that in general, there were fewer and fewer tweets each hour and people tweeted more in daytime than in nighttime. So we manually checked the content and user information of the tweets in the 6th hour and the 5th hour. There were 107 tweets related to the Dallas mass shooting tweeted in the 5th hour and 191 in the 6th hour. We found that in the 5th hour, there were two users that each tweeted 2 tweets and the other tweets were tweeted by different users. However, in the 6th hour, eight users each posted 2 messages, three users each posted 3 messages, two users each posted 4 messages, one tweeted 5 tweets and one tweeted 9 messages. They were in sum 47 tweets, which can explain 56 percent of the tweets' number larger than an hour before. In those 47 tweets, 22 of 47 or 47 percent contained a link. For all the tweets we filtered, 29 percent from the 5th hour and 33 percent from the 6th hour shared a link. Overall, more users in the 6th hour shared links than the 5th hour. When we further examined the links, we found that most of those links pointed to pictures or videos related to the Dallas mass shooting. Also, most of the pictures and videos were created by news agencies 2-3 hours ahead. We can't say for sure if those users were individual or public accounts. But we can

infer that the reason why in the 6th hour there were more tweets tweeted than in the 5th hour was mostly because there were more re-sharing behaviors by individual or public accounts who had interest in the event and were willing to share more information about it (Xie et al. 2015). The time of the peak was 6th hour, which was because of a few hours' time delay after the announcement of related news posted by many different news agencies and few individuals.

7.2 Spatial Distribution of Tweets

From the last section, we concluded that tweets of the Dallas mass shooting diffused rapidly through time. To figure out if distance would influence the diffusion process, we looked at the diffusion of tweets across space in different time periods. The tweets data used here were the 4223 pieces of tweets with geographical location information, as introduced in Data section. The locations of the tweets were mapped in Figure 4. In this map, it can be seen that the points are clustered near Dallas and a few big cities. Most of the counties with small number of population have none or a few points. We then explored the tweets in different time periods. The very first 20 tweets were posted after the occurrence of the Dallas mass shooting; they are displayed on the map in Figure 5 with number labels following time sequence. The points represent location of the tweets and longer length of bar indicates greater response speed (shorter time). Figure 6 showed counts of tweets tweeted in 2 hours, 6 hours, 1 day and all the tweets we collected at county level. The reason we chose 2 hours, 6 hours and 1 day was because tweets in the first 2 hours were tweeted by users aware of this event in the earliest stage, many more people get the information in 6 hours from news and other sources according

to analysis from last section, and 1 day covered a whole cycle of daily activities and more than 50 percent of tweets were tweeted in one day. We want to learn if there were any locational differences of tweets in these different times.

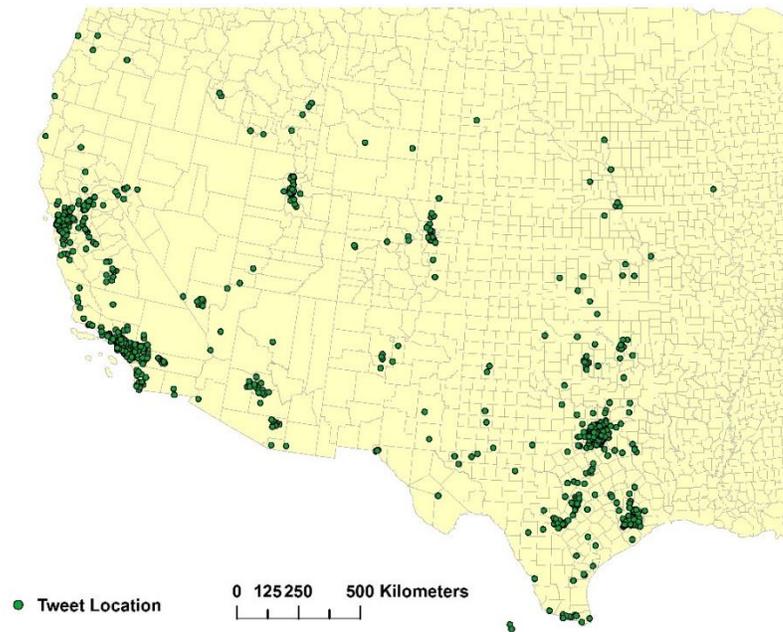


Figure 4. Location of Tweets

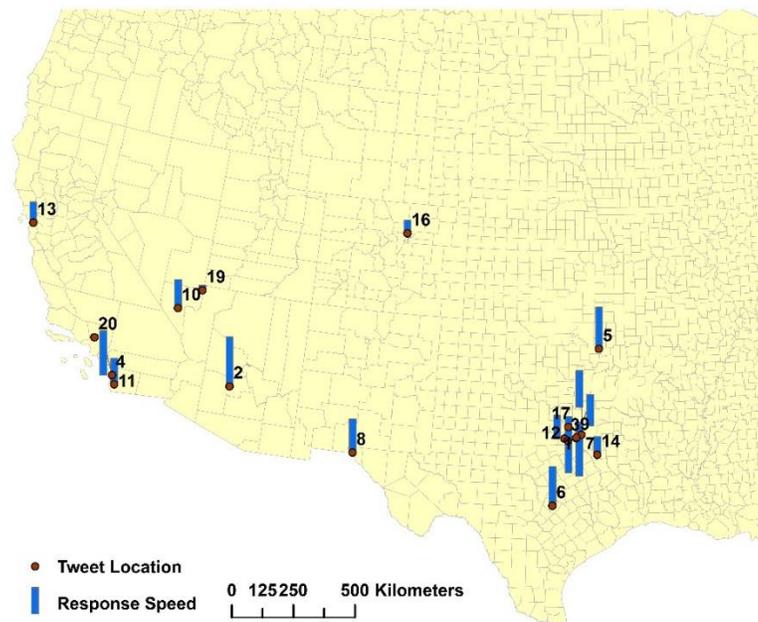


Figure 5. Location of First 20 Tweets Tweeted after Occurrence of Dallas Mass Shooting

Table 4. Statistics of First 20 Tweets

| No. | Distance from Event (km) | Time after First Tweet (minutes) | Content |
|-----|--------------------------|----------------------------------|--|
| 1 | 6 | 0 | Wtf THEYRE SHOOTING COPS IN DOWNTOWN DALLAS |
| 2 | 1427 | 1 | I think someone just shot a cop in Dallas! |
| 3 | 3 | 2 | The crowd of protestors seems to be growing as they march through Dallas. PoliceShooting NBCDFWNow NBC DFW https://t.co/Ng4H5XRt1 |
| 4 | 1909 | 3 | Just saw a cop get shot on Live TV in Dallas this is heart breaking the world we live in |
| 5 | 375 | 3 | JUST HAPPENED: 4 cops shot down in Dallas, TX during peaceful protest. |
| 6 | 288 | 5 | Woooo shots fired, officer down in Dallas. 2 dead cops in one day? |
| 7 | 29 | 5 | 3 cops shot in Dallas during protest. Fox News just accidentally streamed footage of their bodies |
| 8 | 910 | 5 | POLICE WERE SHOT & KILLED IN DALLAS, DURING PROTEST! |
| 9 | 6 | 5 | So I guess 4 cops got shot in Dallas, we saw the commotion in deep Ellum |
| 10 | 1708 | 6 | The protest in Dallas right now... 2 random police shot. YEAH. That'll fersure solve the problem. Disgusting. |
| 11 | 1895 | 6 | They shot a cop in Dallas |
| 12 | 45 | 8 | There's been shots fired at the Dallas protests tonight.... Not clear on if from Police or civilians |
| 13 | 2384 | 8 | Cop lying on ground lifeless in Dallas - video on Megan Kelly Fox News moments ago. Also apparently guy who may have shot him on ground. |
| 14 | 110 | 9 | Smh Dallas. Protesting a mans death then going to shoot at cops. Those cops didn't do it why take it out on them? |
| 15 | 6 | 9 | Having dinner in downtown Dallas and there's cops everywhere cause I shooting just happened, no biggie |
| 16 | 1076 | 9 | Cop shot in Dallas. Where is the outrage? Anybody? |
| 17 | 55 | 9 | Now we got people shooting cops in Dallas. #ClockIsTicking #2A |
| 18 | 32 | 10 | Bro I know Dallas people not out there really shooting police omg they T'd up |
| 19 | 1639 | 10 | @FOX5Vegas shots fired at cops Dallas TX Officers down |
| 20 | 2007 | 10 | #BREAKING: Shots fired during #Dallas protest over #police shootings. https://t.co/Extffe519B |

The 20 points in Figure 5 represent the location of the first 20 relevant tweets tweeted in study area. Among the 20 points, 11 of them were tweeted in Texas, 4 in California, 2 in Nevada, 1 in Arizona, 1 in Colorado and 1 in Oklahoma. Eight of them were less than 100 kilometers away from where the event happened. Although the very first one was near Dallas, the second tweeted within a minute shifted to Arizona, a thousand miles away. These indicate that after the Dallas mass shooting happened, everywhere in the study area, not limited to local places near Dallas, can possibly get information almost at the same time. However, within the study area, local people shared more information about this event.

Table 4 contains tweet's time order, distance to the event (with average resolution of 80 kilometers), time after the first tweet and tweet's content of the first 20 tweets. The 20 tweets were tweeted within 10 minutes, which was a short time. Combining with the distance information, we may also conclude that distance in physical world was not an important influential factor for spreading of information in this case.

Except for tweet No.3, all the rest of tweets were discussing about the mass shooting. After reading through all of them, there were a few things revealed. First, in the very early stage, there was limited information about this event. In those tweets, 7 of them mentioned one or at least one police officer was being shot; 2 of them guessed that there were 2 police officers got shot; 1 thought 3 got shot and 2 thought the number was 4. But actually, five officers and the perpetrator were killed, and nine others and two civilians were injured. Among those tweets, some of them indicated that the Twitter user was not very sure about what was actually happening. For example, tweet No.2 used

words “I think”, tweet No.9 used words “I guess” to describe what happened, and tweet No.12 said “not clear on if from Police or civilians”. Second, people who witnessed the crime took an active part in sharing information on Twitter in the first time. After comparing the user information, we found that No.1, No.9 and No. 15 tweets were tweeted by the same person. It seemed that people at the crime scene responded quickly. This user, in the first tweet, described what he or she witnessed, that police officers were shot in downtown Dallas. In the second tweet, the user gave more detailed information that he or she thought there were four victims and the location of what happened was near the Deep Ellum. In the third tweet, this user contributed to what happened next, that a lot more police officers came to the crime scene because of what just happened. Although the information the witness provided were not all correct, it helped others to get information about this event on Twitter. Third, the news agencies participated in news sharing quickly. In the No. 20 tweet, the user shared a link. This webpage link was a 230 words news titled “DALLAS SHOOTING: SNIPERS SHOOT 11 OFFICERS, KILLING 5 DURING PROTEST”. This tweet was posted 10 minutes after the mass shooting, which indicated that the news responded quickly. No. 4, No. 7 and No. 13 tweets mentioned that they found out information of this mass shooting from live television or videos from the crime scene. We thought this was something unique about this event, that this unexpected shooting event happened during a planned protest about which many news agencies had broadcasted live. Many people happened to witness the crime on TV, which made the news sharing process more quickly. However, most terrorist attacks happen unexpectedly and there may be no videos recording the entire process. Therefore,

caution should be taken when extending the findings from this research to other emergencies or general scenarios of extreme and unexpected situations.

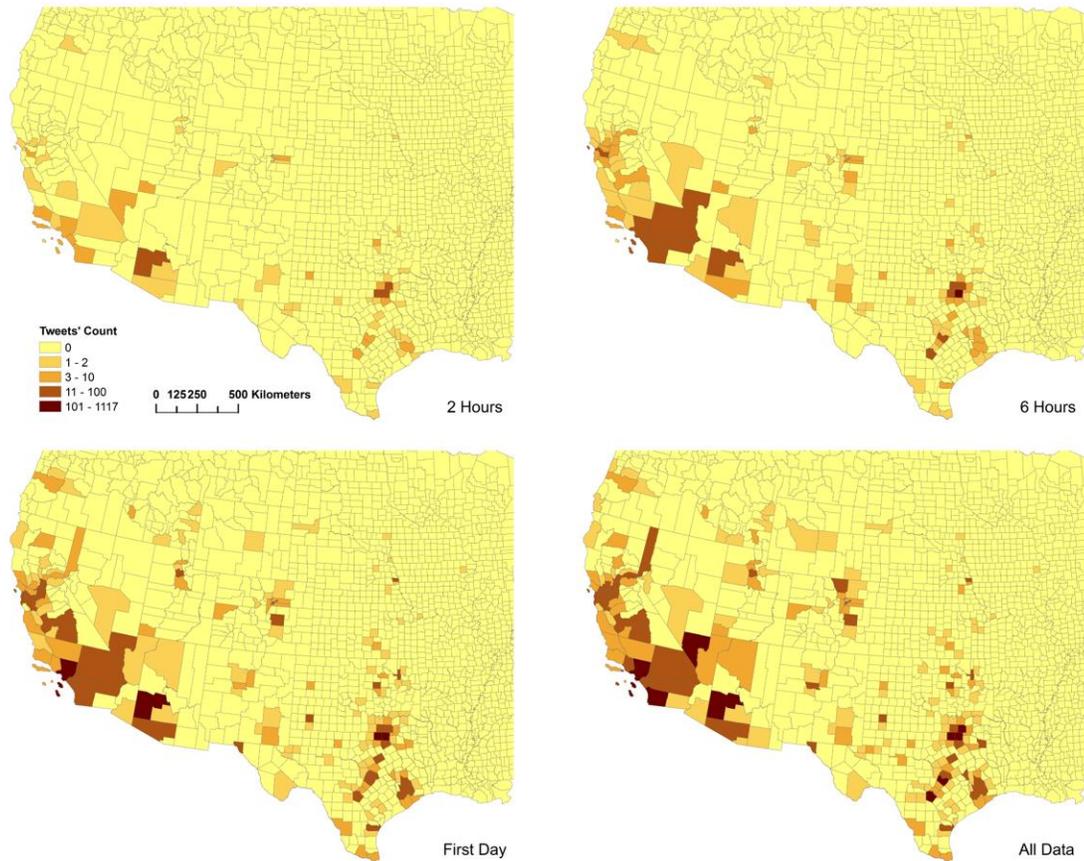


Figure 6. Tweets in the First 2 Hours, 6 Hours, First Day, from July 7th to July 31st at County Level

The four maps above in Figure 6 illustrate tweets' counts in each county in 2 hours, 6 hours, 1 day after the event and in the whole study period. In the first two hours, a lot of tweets tweeted from Texas, New Mexico, Colorado, Arizona, California and Nevada. A few states in the north didn't have any tweet yet. Dallas county had 93 tweets which was the maximum among all the counties. The few places which tweeted a lot about this event were Dallas and nearby places, Phoenix, Austin, San Antonio, Los Angeles, San Diego and Las Vegas. The later ones are big cities with large population.

After six hours, almost all the states within the study area had at least one tweet related to the Dallas mass shooting. After six hours, although the number of tweets kept increasing and tweets with location information continued on showing up in new places, the counties with more tweets remained the same. Dallas County had the maximum 1117 tweets, which was more than one fourth of the tweets included in our dataset. Tweets centered on the cities with larger population and nearby places of Dallas. It seemed that geographic location had limited impact on information diffusion speed related to the mass shooting, that almost all the places can receive information in a very short time, no matter how far they are from the scene.

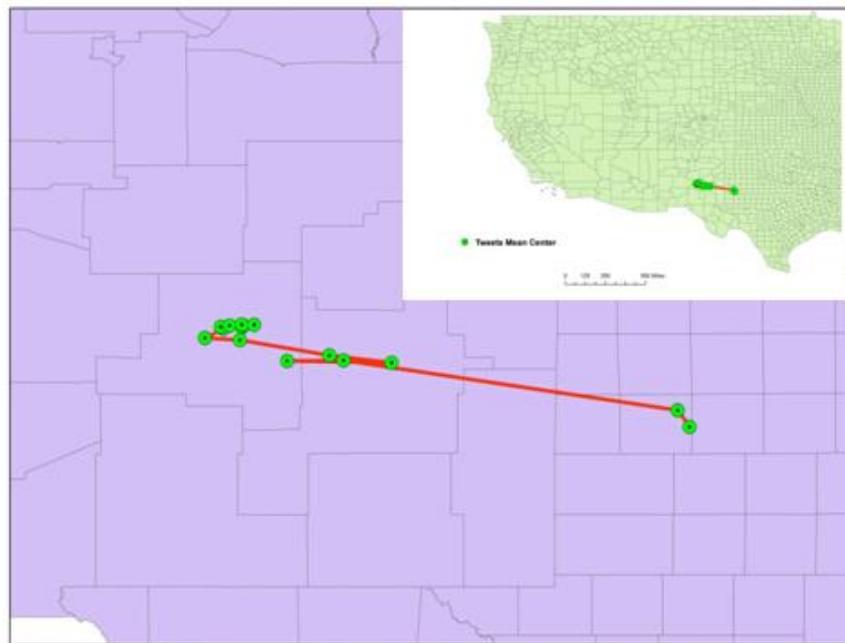


Figure 7. Tweets Mean Center Change

But in order to further understand if people's attention towards this event remained the same, we drew the mean center of the mass-shooting tweets in different times. In the map in Figure 7, points represent the mean center of tweets locations (average latitude and longitude) at 2 hours, 3 hours, 6 hours, 12 hours, 15 hours, 18

hours, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days and till July 31st. It started from the northwest of Texas, near Dallas. From the third hour to the sixth hour, the point moved towards northwest till New Mexico. In previous section, when discussing the tweets' volume and time relationship, we concluded that large amount of relevant news was produced and shared on Twitter at the sixth hour, leading to an unusual peak of tweets' volume. This may partially explain why the mean center shifted quickly towards northwest during this time period. In the beginning, witnesses of the mass shooting shared information mainly to Twitter users near Dallas due to their Twitter network. But news sharing on Twitter in the next few hours allowed users outside and Dallas and its nearby places gain information and tweet about this event. Because Dallas was located in the southeast of the study area, the mean center then moved to the northwest. After 6 hours, the location of mean center became quite stable and shifted around between two neighbor counties. As time passes, if the tweets' are created evenly across space, the mean center stays somewhere. But if the speeds of tweets creations of certain places are faster than other places, then the mean center will move towards places where people tweet more often. Thus from this map, we could learn that at the beginning, tweets were created relatively faster in places around Dallas than anywhere else. After 6 hours, the proportion of tweets started to become stable in both Texas and other states. People in Dallas and nearby places always had more interests and concerns about the Dallas mass shooting, the results above also showed that they had the fastest response in the earliest stage. Considering the information spreading process discussed at the beginning of this section, this phenomenon could be partially explained by the participation of public who witnessed the event in sharing relevant information on Twitter. People in Dallas who

witnessed the event shared information on Twitter right after the shooting, which became an important information source other than news. Because on Twitter, the tweets are shared to the followers, and normally people have more followers from nearby than in the distance, people from places near Dallas were likely to receive information first. Because this event happened to be recorded in the live news and also because of the high efficiency of news agencies, related information spreads fast on Twitter.

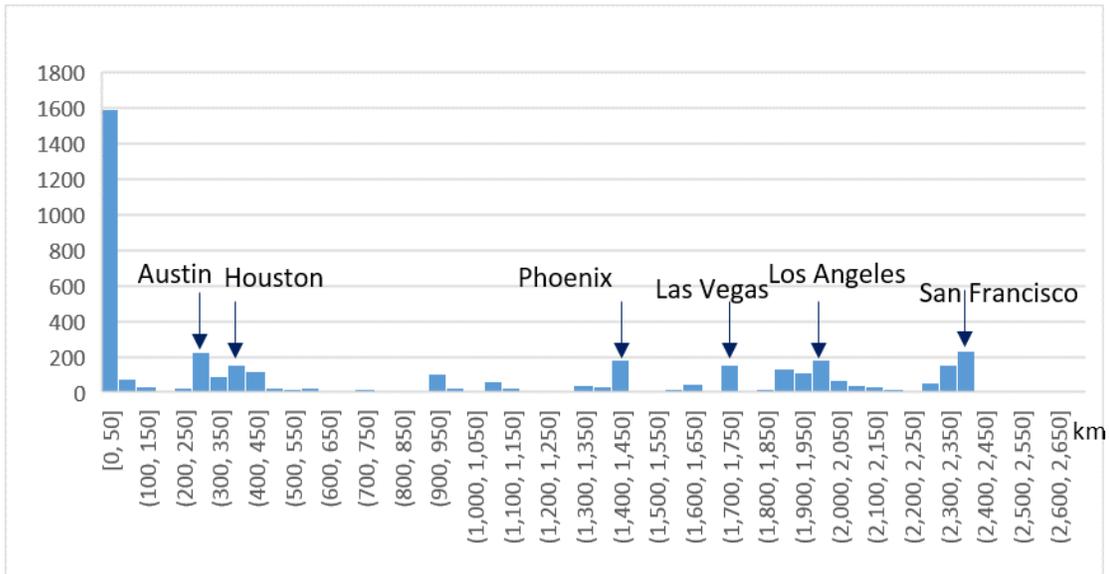


Figure 8. Tweets' Count and Distance to Event

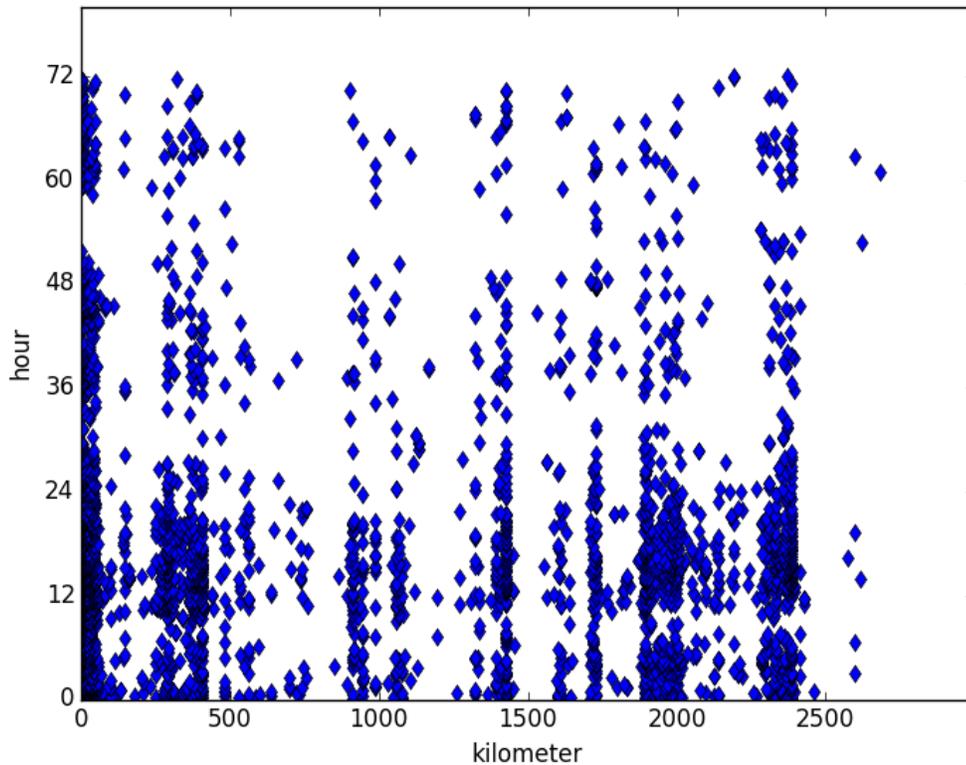


Figure 9. Tweeting Time and Distance to Event

In Figure 8, the relationship between distance and tweets' count is shown in a histogram. The x axis showed the distance of tweets where the reference latitude and longitude location was 32°46'46.4"N, 96°48'15.4"W, Main Street and S. Lamar Street of Dallas, where the shooting happened. Distance interval of 50 kilometers was used to make sure the resolution was good enough to capture the county level difference. Within 50 kilometers, there were 1595 pieces of tweets tweeted, where the number was much higher than other places. Although this event gained attention nationwide, local people tweeted much more than people from other places. There were a few peaks with about 200 tweets in about 300, 400, 1400, 1700, 2000, 2400 kilometers away. After measuring distances, we found that these distances correspond with the distance from Dallas to

Austin, Houston and San Antonio, Phoenix, Las Vegas, Los Angeles and San Francisco, all of which are cities with large population. Results were the same in Figure 9, a scatter diagram reflecting tweeting time and location relationship. Along the distance axis, density of points was the largest when distance was close to 0. There were a few distances having relatively larger amount of points. However, although population in Los Angeles was more than twice of population in San Antonio and was about four times as population in Austin (United States Census Bureau 2016), the volume of tweets was not larger, which indicated that people nearby posted more information on Twitter. In Figure 9, along the time axis, the density of points was larger when close to 0, and became smaller when rising to the top, indicating decreasing of tweets' volume when time passed.

7.3 Hot Spot Analysis

In order to measure if there were any places having significantly larger number of tweets than any other places, using hot spot analysis in ArcGIS, we then looked at if there were any hot spots or cold spots of tweets at county level. Based on the assumption that neighboring features have larger influence on the computations for a target feature than features that are far away, "INVERSE_DISTANCE" method was chosen. "Distance band or threshold distance" parameter was set up as 80 kilometers to make sure there was on average 8 neighboring features for each target feature and the result to be reliable (Esri n.d.).

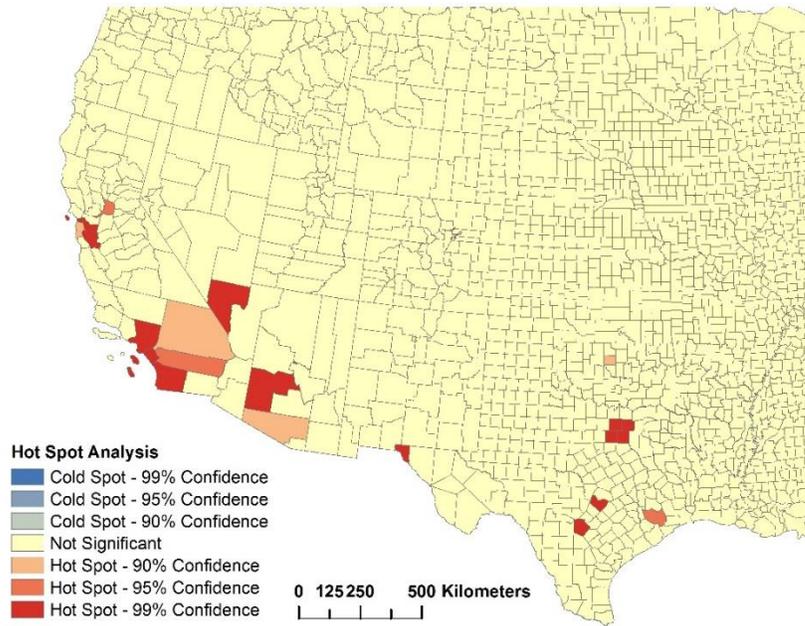


Figure 10. Hot Spot Analysis for all Tweets

Figure 10 shows the results of hot spot analysis for all the tweets. There were significant hot spots at 99% confidence level in counties around Dallas, San Antonio, Austin, El Paso, Phoenix, Las Vegas, San Diego, Los Angeles and San Francisco. Houston is significant at 95%. The result was very similar to Figure 8. All of them were cities with large population. The Result indicated that the amount of tweeting behavior in one place was positively related to the population. In other words, where there had more people, had more tweets. To eliminate influence by large number of population, we standardized tweets by population in each county (United States Census Bureau 2010). Figure 11 is population density map where each point represents 7,000 people. Northeast Texas, South Central Texas, West coast and a few places in the middle of the study area had large population density. And most of the other places had low population density. The uneven distribution of population caused many counties to have one or no tweet

related to the Dallas mass shooting, therefore the data quality was also low in those counties.

After we standardized tweets by population and run hot spot analysis again, the results became very different (See Figure 12). There were a few differences in this map. First, there were much more hot spots in the study area in this map than in Figure 10. This was because the standardized results eliminated influences of large population big cities. Second, most of the hot spots were located in Texas and east part of Texas (around Dallas, Austin, San Antonio and Houston). Again, this proved that people showed more interests when the location of the event was close to them. Third, comparing to the map in Figure 13, Los Angeles area was not a hot spot anymore, which mean that people in Los Angeles had average interest to this event compared to people from other places, and the hot spot in Figure 10 was purely because there were more people in Los Angeles. Finally, there seemed to be a few isolated hot spots in the north part of the study area. In fact, most of these hot spots are probably outliers from their immediate surroundings. Many of the counties forming these hot spots had probably one tweet about Dallas shooting in the county with small population and were surrounded by counties with no tweets on the event. This problem could be solved if there were more records in each county.

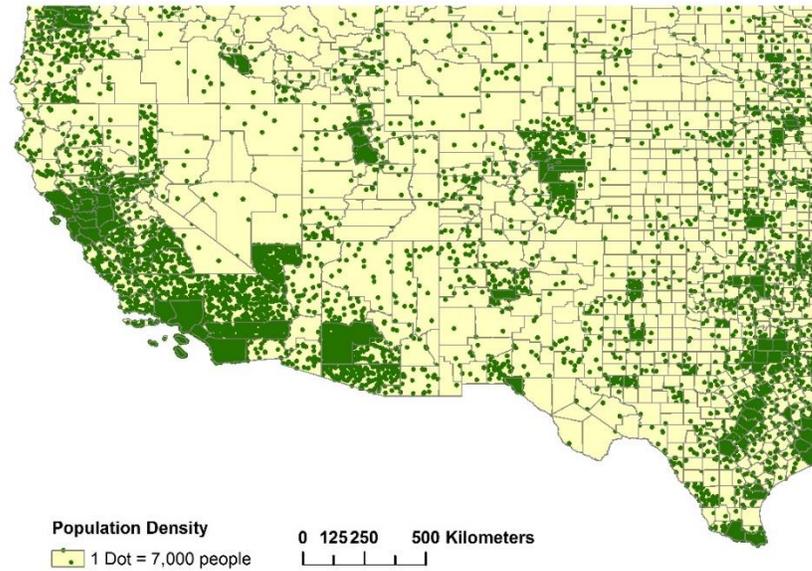


Figure 11. Population Density Map

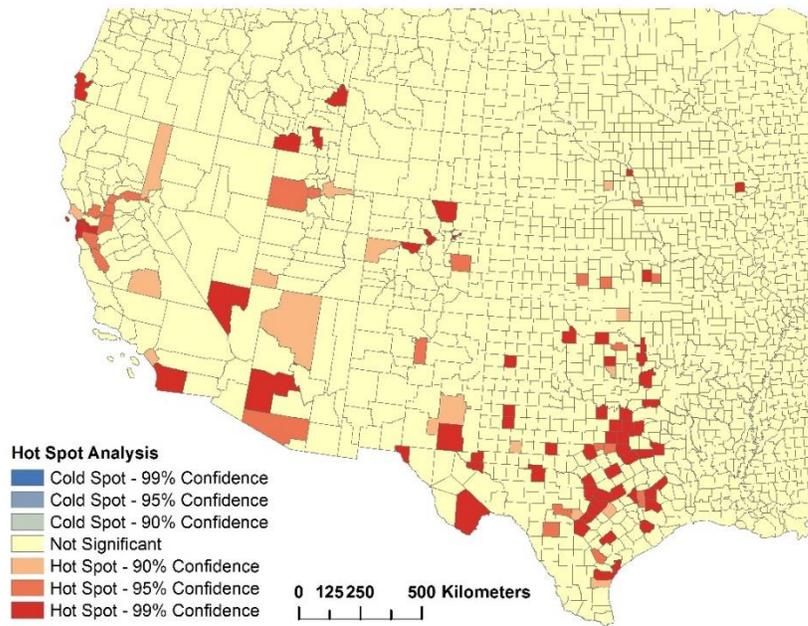


Figure 12. Hot Spot Analysis for Tweets Standardized by Population

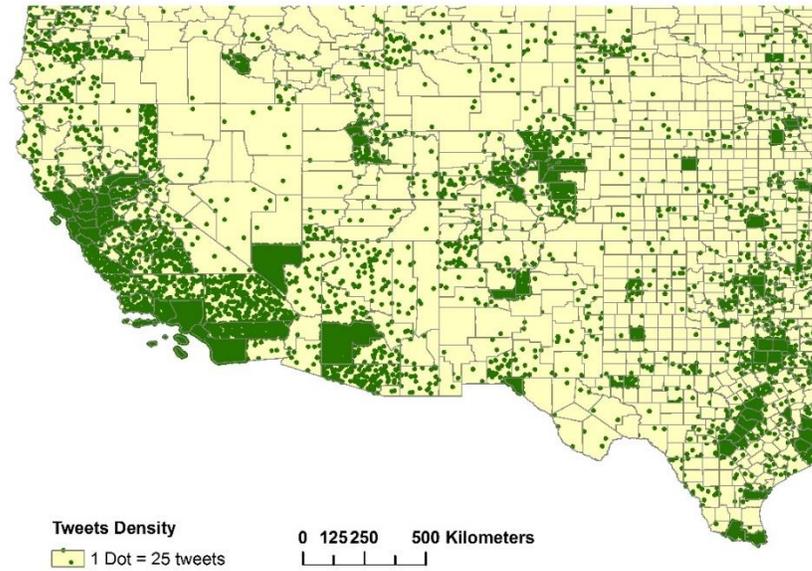


Figure 13. Tweets Density Map

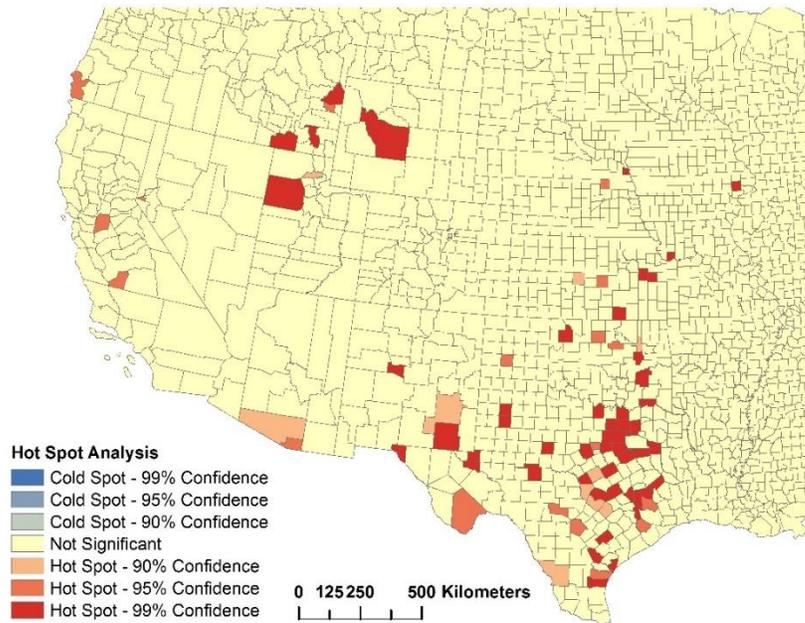


Figure 14. Hot Spot Analysis for Tweets Standardized by Normal Tweets' Volume

Figure 13 is a density map of the overall tweets' volume from our data set, which represents social media activity. We standardized the data on Dallas shooting by general

tweets' volume. The reference tweets used were from July 1st to July 7th (Greenwich Mean Time) and all of them had latitude and longitude information. The total number of tweets was about 600 thousand. We joined the point tweets layer with county map layer based on location to get the tweets' counts in each county. The tweets' counts were used as social media data for standardization. This was considering the characteristics of Twitter users. Population may be not the same as social media population. Although at first glance the two density maps were quite similar with each other except for a few places, for example, the Norton County in Kansas, the hot spot analysis results showed more differences when standardizing by normal tweets' volume comparing to standardizing by population. In Figure 14, comparing to Figure 12, most of the hot spots were concentrated in Texas. Places near San Diego, San Francisco, Phoenix and Las Vegas, however, were not hot spots any more. This indicated that in those places, the large number of tweets were caused by large number of Twitter users, and on average people in these places liked to tweet more. Since social media population was not the same as general population and the users' tweeting behaviors were not the same everywhere, normal tweets' volume should have better quality than population as reference data.

7.4 Logistic Regression Model

Table 5 listed odds ratio, 95% confidence interval and significance level of the variables in the three models. As more variables were included in the model, the coefficient of determination pseudo R^2 increased from 0.035 to 0.045, which meant that the dependent variable was explained more by independent variables. However, the

number was still small, indicating that there may be important related factors missing in the models.

Table 5. Logistic Regression Results

| Variable | Model 1 | Model 2 | Model 3 |
|--------------------------------|----------------------|----------------------|----------------------|
| | OR (95% CI) | OR (95% CI) | OR (95% CI) |
| Language | 0.26 (0.10, 0.68)*** | 0.29 (0.11, 0.76)** | 0.28 (0.11, 0.75)** |
| status count | 1.00 | 1.00 | 1.00 |
| favorite count | 1.00 | 1.00 | 1.00 |
| friend count | 1.00 | 1.00 | 1.00 |
| follower count | 1.00 | 1.00 | 1.00 |
| Verified | 4.06 (3.01, 5.48)*** | 3.75 (2.76, 5.09)*** | 3.61 (2.66, 4.91)*** |
| source | | | |
| <i>Android</i> | 1.04 (0.86, 1.26) | 1.05 (0.86, 1.27) | 1.05 (0.87, 1.27) |
| <i>iPhone</i> | 1.63 (1.39, 1.92)*** | 1.62 (1.37, 1.91)*** | 1.64 (1.39, 1.93)*** |
| Hashtag | | 1.15 (1.00, 1.32)** | 1.15 (1.00, 1.32)** |
| Link | | 1.28 (1.12, 1.47)*** | 1.29 (1.12, 1.48)*** |
| Mention | | 1.12 (0.93, 1.34) | 1.12 (0.93, 1.34) |
| Reply | | 0.70 (0.55, 0.89)*** | 0.71 (0.55, 0.90)*** |
| three hours | | 1.64 (1.34, 2.01)*** | 2.02 (1.51, 2.71)*** |
| six hours | | | 0.97 (0.74, 1.27) |
| Day | | | 1.26 (1.08, 1.48)*** |
| weekday | | | 0.88 (0.75, 1.05) |
| Pseudo R ² | 0.035 | 0.043 | 0.045 |
| N | 4219 | 4219 | 4219 |
| * p<0.1; ** p<0.05; *** p<0.01 | | | |

In Model 1, significant variables were language, verified account, and iPhone as data source. Comparing to users with other language, users who chose English as default language had 74% chance less likely to have posted their tweets in hot spots. Tweets in hot spots had three times greater chances to be tweeted from a verified user than tweets in other places. Because verified users usually indicate that their posts are of public interests, this result can be explained as local people share local news. Also, iPhone users who tweeted related messages were more likely to be in a hot spot than users using other

equipment except for Android and iPhone. However, because we didn't find literature discussing about the relationship between variable "language" and "source", further research is still needed to explain how this result is related to our study. Because most of the hot spots were in Texas, the result might just reflect the overlap of tweets' hot spots and certain language and electronic device's hot spots.

Model 2 included 5 more variables. Tweets with hashtag or link were more likely to be in a hot spot. If a tweet contains hashtag or link, it usually indicates that the tweet is sharing information. One reason why hot spots had greater number of tweets was because Twitter users in those places tweeted more about the event and a large proportion of tweets in those places were sharing information. Replied messages were less likely to be in a hot spot. This was because tweets replying someone were usually passively receiving information. The result proved that Twitter users in those places were message disseminator instead of message receiver. Tweets tweeted in three hours were more likely to be in hotspots. This was corresponding with what we have discussed in previous sections that in early stages, people who received information were mostly local people.

In model 3, odds ratio of variable "three hour" became larger after adding a few more variables. Tweets posted in three hours had two times chances to be in a hot spot comparing to others. However, tweets posted in six hours had no significant difference. According to analysis in previous sections, this may be because after three hours, a lot of news was produced and more and more people from other places received the information.

But overall speaking, although in all the three models, odd ratio of some variables were statistically significant, the logistic regression models were not good enough for

predicting results because of small R^2 . When adding more variable to raise R^2 of the model, chances are that the odds ratio of existing variables would change too. The results indicated misspecification of the model, which means the model missed some significant predictors.

VIII. CONCLUSION AND DISCUSSION

For the temporal pattern of volume of tweets, we found that tweets' number related to the Dallas mass shooting grew very fast soon after the happening of attack and decreased rapidly after a few hours following an exponential curve. After 3 days, the number of tweets posted everyday became stable. However, within the general time decay trend of tweets, time cycle was also observed. People tweeted more around 11am - 12 pm in the morning and 6 pm - 7 pm in the afternoon, which was consistent with people's regular daily work and rest schedule. Time interval to the event and people's daily routine were the two main factors that influenced the volume of tweets. The possible explanation was that this time was a few hours' behind the announcement of large number of related news from news agencies and individuals and this was caused by common re-sharing behaviors by individual or public accounts that had interest in the event. Geographic distance had impact on information diffusion speed, but the influence became smaller as time passes. In this case, people from faraway places got information within only a few minutes after the event. However, this did not mean that physical distance had no importance. In fact, in the very beginning, most of the tweets were located not far from the incident. Before a lot of news available on media, witness was an important source of information. Although the Dallas mass shooting was reported a lot by TV and online media, and drew great public attention, people near Dallas still tweeted more about this event than people from afar. Most of the hot spots distributed in places near Dallas from which a large number of tweets were created. These people posted information in the earliest stage and actively shared lots of information.

As a case study, the selection of the event may influence applicability of research conclusion. In this case, the shooting was accidentally broadcasted by some news agencies. This to some extent covered up the influence of physical distance in message diffusion on social media. In most of the emergencies and extreme events, because the event happens suddenly, witnesses or people affected always realize it in first time, and then spread related messages to others. However, the unexpected factor revealed the importance of external sources in message diffusion on Twitter. In our study, news online and from other sources significantly influenced the diffusion extent and speed. Before a lot of news published, most of information was collected and diffused in local places. The release of news changed the pattern that a lot more people from faraway places joined the information transmission and production process. The influence of these external information sources on information diffusion in social media must be further investigated by future studies.

There are some limitations to this study. First, compared to some other empirical analyses using Twitter data, the dataset in this study is not very large. We selected tweets for the related analyses using a strict set of keywords combination to ensure data accuracy and avoid unnecessary noise. We concerned more about data quality, rather than quantity. This was mainly because in this study, we would like to observe the message production in a few weeks. Because tweets' volume on a particular topic follows exponential decay through time, if filter criteria were not strict enough, impurities may take large proportion of daily tweets after a few days and may make the conclusion less reliable. But the flip side of this approach is related to that small numbers of tweets or zero tweet from some counties were identified as being related to the Dallas mass

shooting, which may have impacted the findings from the analyses. This problem could be solved in future studies by improving information extraction techniques including selecting more keywords. Second, the boundary of study area was close to where the event happened which made some of our conclusions less convincing. Also, in the logistic regression model, the pseudo R^2 in all of the three logistic regression models were small, indicating that the models had problem of misspecification. R^2 is usually considered as a measurement to reflect the variability in the response variables around the mean explained by the model. 3.5 percent to 4.5 percent are small numbers which mean that more than 90 percent variability were not captured by the model. The R^2 can become larger and the model might be improved when including more valid factors through further literature and empirical studies. However, according to the results we have, purely adding variables may not be good enough. Logistic model was probably not a very suitable approach and other models or methods should be considered.

For future studies, because this study aimed to understand general patterns of message diffusion on social media, more in-depth research should be conducted. For example, to study the temporal pattern, one may look at tweets in hourly, daily and weekly scales. Also, our results showed that local people tweeted more messages than people in other places and they actively produced and shared information instead of passively accept or follow tweets. Future study may look into details about the differences between tweeting behavior of local people and people from afar, including time, frequency, content and other aspects. Last, this study only examined spatial and temporal patterns of tweet quantity; future research may conduct sentiment analysis to

look at people's feelings about the event in different places and during different time periods after a mass scale event.

REFERENCES

- The American Heritage Dictionary of the English Language. 2016. "Terrorism." *The American Heritage Dictionary*. Accessed February 13, 2017.
<https://www.ahdictionary.com/word/search.html?q=terrorism&submit.x=40&submit.y=36>.
- Bergen, Peter and David Sterman. 2016. "Dallas: An act of domestic terrorism." *CNN: Breaking News, Latest News and Videos*. Accessed February 15, 2016.
<http://www.cnn.com/2016/07/08/opinions/dallas-domestic-terrorism-bergen/>.
- Cao, Nan, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu. 2012. "Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time." *Ieee Transactions On Visualization And Computer Graphics* 18, no. 12: 2649-2658.
- Chan, C.K., M. Vasardani, and S. Winter. 2014. "Leveraging Twitter to detect event names associated with a place." *Journal Of Spatial Science* 59, no. 1: 137-155.
- Cheng, T, and T Wicks. 2014. "Event Detection using Twitter: A Spatio-Temporal Approach." *Plos One* 9, no. 6: e97807. doi: 10.1371/journal.pone.0097807.
- Chew, Cynthia, and Eysenbach Gunther. 2010. "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak." *Plos One* 5, no. 11: e14118. doi: 10.1371/journal.pone.0014118.
- Clark, Gordon. 1984. "The Process of Component of Innovation Diffusion." In *Innovation diffusion : contemporary geographical approaches*, 6-21.: Norwich, England : Geo Books, c1984., 1984.

- Esri. n.d. "Hot Spot Analysis (Getis-Ord G_i^*)." *ArcGIS for Desktop*. Accessed February 19, 2017. <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm>.
- Gelernter, Judith, and Shilpa Balaji. 2013. "An algorithm for local geoparsing of microtext." *Geoinformatica* 1-33.
- Gesualdo, Francesco, Giovanni Stilo, Angelo D'Ambrosio, Emanuela Carloni, Elisabetta Pandolfi, Paola Velardi, Alessandro Fiocchi and Alberto E. Tozzi. 2015. "Can Twitter be a source of information on allergy? Correlation of pollen counts with tweets reporting symptoms of allergic rhinoconjunctivitis and names of antihistamine drugs." *Plos ONE* 10, no. 7: e0133706. doi: 10.1371/journal.pone.0133706.
- Hacker, Holly K., Corbett Smith and Marina Trahan Martinez. 2016. "El Centro College officials trace footsteps of Dallas police killer." *Dallas Morning News, The (TX)*, July 12.
- Han, Su Yeon, Ming-Hsiang Tsou, and Keith C. Clarke. 2015. "Do Global Cities Enable Global Views? Using Twitter to Quantify the Level of Geographical Awareness of U.S. Cities." *Plos ONE* 10, no. 7: 1-23.
- Herdagdelen, Amac, Wenyun Zuo, Alexander Gard-Murray, and Yaneer Bar-Yam. 2013. "An Exploration of Social Identity: The Geography and Politics of News-Sharing Communities in Twitter." *Complexity* 19, no. 2: 10-20.

- Huang, Qunying, Duido Cervone, Duangyang Jing and Chaoyi Chang. 2015. "DisasterMapper: A CyberGIS framework for disaster management using social media data." *Proceedings Of The 4Th ACM SIGSPATIAL International Workshop On Analytics For Big Geospatial Data, Bigspatial 2015 (Proceedings of the 4th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, BigSpatial)*. 2015: 1-6.
- Huang, Qunying, and David W.S. Wong. 2015. "Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data." *Annals Of The Association Of American Geographers* 105, no. 6: 1179-1197.
- Huang, Qunying., and Xiao Yu. 2015. "Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery." *ISPRS International Journal Of Geo-Information* 4, no. 3: 1549-1568.
- John. T. Floyd Law Firm, 2016. "Was the Dallas Shooting An Act of Domestic Terrorism?" *John T. Floyd Law Firm*. Accessed February 15, 2017.
<http://www.johntfloyd.com/dallas-shooting-act-domestic-terrorism/>.
- Jones, Taylor. 2015. "Toward a Description of African American Vernacular English Dialect Regions Using 'Black Twitter'." *American Speech* 90, no. 4: 403-440.
- Kennedy, Merrit and Tanya Ballard Brown. 2016. "What We Know About the Dallas Suspected Gunman." *NPR: National Public Radio*. Accessed February 15, 2017.
<http://www.npr.org/sections/thetwo-way/2016/07/08/485239295/what-we-know-about-the-dallas-suspected-gunman>.

- Kirilenko, Andrei P., and Svetlana O. Stepchenkova. 2014. "Public microblogging on climate change: One year of Twitter worldwide." *Global Environmental Change-Human And Policy Dimensions* 26, 171-182.
- Lee, Jaeung, Manish Agrawal and H.R. Rao. 2015. "Message diffusion through social network service: The case of rumor and non-rumor related tweets during Boston bombing 2013." *Information Systems Frontiers* 17, no. 5: 997-1005.
- Lin, Jie, and Robert G. Cromley. 2015. "Evaluating geo-located Twitter data as a control layer for areal interpolation of population." *Applied Geography* 58, 41-47.
- Longley, Paul A., and Muhammad Adnan. 2016. "Geo-temporal Twitter demographics." *International Journal Of Geographical Information Science* 30, no. 2: 369-389.
- Luini, Lorenzo P, Davide Cardellicchio, Fulvia Felletti, and Francesco S Marucci. 2015. "Socio-Spatial Intelligence: social media and spatial cognition for territorial behavioral analysis." *Cognitive Processing* 16 Suppl 1, 299-303.
- Malleson, Nick, and Martin A. Andresen. 2015. "The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns." *Cartography and Geographic Information Science* 42, no. 2: 112-121.
- Mallin, Alexander; David Caplan. 2016. "President Obama Describes Dallas Shootings As 'Vicious, Calculated, Despicable Attack'". *ABC News*. Accessed February 15, 2017. <http://abcnews.go.com/US/president-obama-receiving-regular-updates-dallas-shooting/story?id=40425624>.
- Mateos, Pablo, Paul A. Longley, and David O'Sullivan. 2011. "Ethnicity and Population Structure in Personal Naming Networks." *Plos ONE* 6, no. 9: 1-12.

- McConnell, Thomas H. 2007. *The Nature of Disease: Pathology for the Health Professions*. Baltimore, Mar.: Lippincott Williams & Wilkins.
- Merriam-Webster. n.d. "Definition of tweet." *Dictionary by Merriam-Webster*. Accessed February 13, 2017, <https://www.merriam-webster.com/dictionary/tweet>.
- Mitchell, L., M.R. Frank, K.D. Harris, P.S. Dodds, and C.M. Danforth. 2013. "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place." *Plos ONE* 8, no. 5: e64417. doi: 10.1371/journal.pone.0064417.
- Mocanu, Delia, Andea Baronchelli, Nicola Perra, Bruno Goncalves, Qian Zhang, and Alessandro Vespignani. 2013. "The Twitter of Babel: Mapping World Languages through Microblogging Platforms." *Plos One* 8, no. 4: e61981. doi: 10.1371/journal.pone.0061981
- Jonathan K., Nelson, Quinn Sterling, Swedberg Brian, Chu Wanghuan, and MacEachren Alan M. 2015. "Geovisual Analytics Approach to Exploring Public Political Discourse on Twitter." *ISPRS International Journal Of Geo-Information, Vol 4, Iss 1*, no. 1: 336-337.
- Oh, Onook, Manish Agrawal, and H. Raghav Rao. 2013. "Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises." *MIS Quarterly* 37, no. 2: 407-A7.
- Ord, J. Keith, and Arthur Getis. 2001. "Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation." *Journal of Regional Science* 41, no. 3: 411-432.

- Palmer, S. 2016. "Birds of a Feather: The Geographic Interconnection of Australian Universities on Twitter." *Journal of Applied Research In Higher Education* 8, no. 1: 88-100.
- Signorini, Alessio, Alberto Maria Segre, and Philip M. Polgreen. 2011. "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza a H1N1 Pandemic." *Plos ONE* 6, no. 5: 1-10.
- Sloan, Luke, Jeffrey Morgan, Pete Burnap, and Matthew Williams. 2015. "Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data." *Plos ONE* 10, no. 3: 1-20.
- Smith, A and J, Brenner. 2012. "Twitter Use 2012." *Pew Research Center*. Accessed June 25, 2017, <http://www.pewinternet.org/2012/05/31/twitter-use-2012/>.
- Starbird, Kate, Leysia Palen, Amanda L. Hughes and Sarah Vieweg. 2010. "Chatter on The Red: What hazards threat reveals about the social life of microblogged information." *Proceedings Of The ACM Conference On Computer Supported Cooperative Work, CSCW* no. 2010 ACM Conference on Computer Supported Cooperative Work, CSCW 2010: 241-250.
- Starbird, Kate, and Leysia Palen. 2010. "Pass it on?: Retweeting in mass emergency." *ISCRAM 2010 - 7Th International Conference On Information Systems For Crisis Response And Management: Defining Crisis Management 3.0, Proceedings*.
- Starbird, Kate, and Leysia Palen. 2011. "'Voluntweeters': Self-organizing by digital volunteers in times of crisis." *Conference On Human Factors In Computing Systems – Proceedings*.

- Statista. 2016. "Number of monthly active Twitter users worldwide from 1st quarter 2010 to 3rd quarter 2016 (in millions)." *Statista - The Portal for Statistics*. Accessed February 13, 2017, <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
- Sutton, Jeannette, C. Ben Gibson, Emma S. Spiro, Cedar League, Sean M. Fitzhugh, and Carter T. Butts. 2015. "What it Takes to Get Passed On: Message Content, Style, and Structure as Predictors of Retransmission in the Boston Marathon Bombing Response." *Plos ONE* 10, no. 8: 1-20.
- Szule, Janos, Daniel Kondor, Laszlo Dobos, Istvan Csabai, and Gabor Vattay. 2014. "Lost in the City: Revisiting Milgram's Experiment in the Age of Social Networks." *Plos One* 9, no. 11: e111973. doi: 10.1371/journal.pone.0111973.
- Tafti, Ali, Ryan Zotti, and Wolfgang Jank. 2016. "Real-Time Diffusion of Information on Twitter and the Financial Markets." *Plos ONE* 11, no. 8: 1-16.
- Takhteyev, Yuri, Anatoliy Gruzd, and Barry Wellman. 2012. "Geography of Twitter networks." *Social Networks* 34, no. 1: 73-81.
- Twitter. 2016. "Twitter Usage/ Company facts." *Twitter Company*. Accessed February 13, 2017. <https://about.twitter.com/company>.
- Twitter. n.d. "FAQs about adding location to your Tweets." *Twitter Help Center*. Accessed February 16, 2017. <https://support.twitter.com/articles/78525>.
- Twitter. n.d. "About verified accounts." *Twitter Help Center*. Accessed February 19, 2017. <https://support.twitter.com/groups/31-twitter-basics/topics/111-features/articles/119135-about-verified-accounts>.

- Twitter. n.d. "Tweets." *Twitter Developer Documentation*. Accessed June 20, 2017.
<https://dev.twitter.com/overview/api/tweets>.
- United States Census Bureau. 2016. "Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More, Ranked by July 1, 2015 Population: April 1, 2010 to July 1, 2015 - United States -- Places of 50,000+ Population." *United States Census Bureau*. Accessed February 19, 2017.
<https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>.
- United States Census Bureau. "2010 Census Data." *United States Census Bureau*. Accessed February 19, 2017. <http://www.census.gov/2010census/data/>.
- Van de Velde, Bob, Albert Meijer, and Vincent Homburg. 2015. "Police message diffusion on Twitter: analysing the reach of social media communications." *Behaviour & Information Technology* 34, no. 1: 4-16.
- Widener, Michael J., and Wenwen Li. 2014. "Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US." *Applied Geography* 54, 189-197.
- Wu, Yingcai, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. 2014. "OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media." *Ieee Transactions On Visualization And Computer Graphics* 20, no. 12: 1763-1772.
- Xiao, Yu, Qunying Huang, and Kai Wu. 2015. "Understanding social media data for disaster management." *Natural Hazards* 79, no. 3: 1663-1679.
- Yang, Wei, and Lan Mu. 2015. "GIS analysis of depression among Twitter users." *Applied Geography* 60, 217-223.

- Yang, Wei, Lan Mu, and Ye Shen. 2015. "Effect of climate and seasonality on depressed mood among Twitter users." *Applied Geography* 63, 184-191.
- You, Quangzen, and John Krumm. 2014. "Transit tomography using probabilistic time geography: planning routes without a road map." *Journal Of Location Based Services* 8, no. 4: 211-228.