Spatiotemporal Variation in Emotional Responses to 2017 Terrorist Attacks in London Using Twitter Data

Ruixue Wang
SPATIOTEMPORAL VARIATION IN EMOTIONAL RESPONSES TO 2017 TERRORIST ATTACKS IN LONDON USING TWITTER DATA

by

RUIXUE WANG

Under the Direction of Dajun Dai, PhD

ABSTRACT

Terrorist attacks have a significant impact on human lives. This study examined emotional responses after the terrorist attacks in London in March and June of 2017, respectively. This research extracted tweets related to the two attacks by developing a Python tool interacting with the Twitter Application Program Interface (API). The tweets were analyzed for its negative emotion expression such as sadness. This study then analyzed these negative tweets using the space-time permutation model in SatScan and assessed their variation in space and time. Results suggested two significant clusters of negative tweets after the first attack. These clusters located in the metropolitan area of London and between Manchester and Liverpool within ten days of the attack. The findings may contribute to quick surveillance of emotional responses on the Twitter users.

INDEX WORDS: Terrorist attacks, Emotional responses, Negative tweets, Space-time.
SPATIOTEMPORAL VARIATION IN EMOTIONAL RESPONSES TO 2017 TERRORIST ATTACKS IN LONDON USING TWITTER DATA

by

RUIXUE WANG

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in the College of Arts and Sciences

Georgia State University

2018
SPATIOTEMPORAL VARIATION IN EMOTIONAL RESPONSES TO 2017 TERRORIST ATTACKS IN LONDON USING TWITTER DATA

by

RUIXUE WANG

Committee Chair:  Dajun Dai

Committee:  Jeremy Diem  Lawrence Kiage

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2018
DEDICATION

I dedicate this thesis to my parents who have been a great source of inspiration and support. This thesis is also dedicated to my husband who encouraged me to build my motivation towards the world of GIS.
ACKNOWLEDGEMENTS

I would like to convey my gratefulness to my advisor Dr. Dajun Dai, because this work would not have been completed without his help. I sincerely thank Dr. Daniel Deocampo for his guidance and encouragement in writing a proposal in the Research Method class. I also would like to express my gratitude towards my committee members, Dr. Diem and Dr. Kiage, for their suggestions and help.

My thanks and appreciations also go to my colleagues and people who have willingly helped me out with their abilities.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... V

LIST OF TABLES ......................................................................................................................... VIII

LIST OF FIGURES ....................................................................................................................... IX

LIST OF ABBREVIATIONS .......................................................................................................... X

1 INTRODUCTION .................................................................................................................... 1

1.1 Background ......................................................................................................................... 1

1.2 Literature Review ............................................................................................................... 4

1.2.1 Application of Social Media ....................................................................................... 4

1.2.2 Use of Geospatial Data in Social Media Studies ....................................................... 5

1.2.3 General Approach of Studying Emotional Data ....................................................... 7

1.2.4 Application of Space-time Analysis .......................................................................... 9

1.2.5 Importance of Studying Negative Emotional Data in Mental Health .............. 11

1.2.6 Rural/Urban Differences in Social Media Adoption and Mental Health .. 12

1.3 Research Question and Objectives ................................................................................. 14

1.4 Significance of this Study ................................................................................................. 14

2 METHODOLOGY .................................................................................................................. 16

2.1 Study Area ......................................................................................................................... 16

2.2 Methods ............................................................................................................................ 17

2.2.1 Data Collection ........................................................................................................... 17
2.2.2 Data Processing

3 RESULTS

4 DISCUSSION AND CONCLUSION

4.1 Tweets, Social Network Usage, and Spoken Language

4.2 Mental Health and Human Behaviors

4.3 Mental Health and Rural/Urban Classification

4.4 Social Media Attention to Various Attacks

4.5 Limitations

REFERENCES
LIST OF TABLES

Table 1. Samples of negative tweets ................................................................. 20
Table 2. Rural/urban reclassification in U.K .................................................. 22
Table 3. Number of geotagged tweets in U.K ............................................... 27
Table 4. Summary of geotagged tweets in March (r=75 km) ......................... 30
Table 5. Description of the first cluster in March ....................................... 30
Table 6. Description of the second cluster in March .................................. 31
Table 7. Correlation of negative tweet and rural/urban classification ............ 34
Table 8. Official language and prevailing social media platforms for top 5 countries having active Twitter accounts .................................................. 35
LIST OF FIGURES

Figure 1. Research workflow ........................................................................................................17
Figure 2. A sample of python script ............................................................................................18
Figure 3. Sample tweets retrieved from Twitter ........................................................................19
Figure 4. Process of calculating generic time .............................................................................21
Figure 5. Distribution of geotagged tweets in March ................................................................24
Figure 6. Distribution of geotagged tweets in June ....................................................................25
Figure 7. Daily number of all relevant tweets in the world ........................................................25
Figure 8. Daily number of geotagged tweets in the world ........................................................26
Figure 9. Geotagged tweets in U.K. ............................................................................................28
Figure 10. Clusters of the attack in March ................................................................................29
Figure 11. Tweets and rural/urban classification in U.K. ..........................................................32
Figure 12. Tweets within each classification .............................................................................33
Figure 13. Number of tweets related to the attack in U.K. and Syria .........................................37
LIST OF ABBREVIATIONS

US: United States
EU: European Union.
GIS: Geographic Information System.
API: Application Programming Interface.
POS: Part of Speech.
SVM: Support Vector Machine.
IDW: Inverse Distance Weighting.
KDE: Kernel Density Estimation.
PTSS: Post Traumatic Stress Syndrome.
1 INTRODUCTION

1.1 Background

Terrorist attacks affect human lives tremendously, both physically and mentally. Terrorist attacks may be attributed to the foreign terrorist organization (international terrorism) or domestic movements exposed to extremist ideologies of a political, religious, social or environmental nature (Federal Bureau of Investigation, 2018). Terrorism has become a global issue, especially after the September 11 attack in the United States (US) in 2001 (Goldman 2011). In the year of 2016, there was a total of 11,072 terrorist attacks worldwide, accounting for 30 attacks daily. More than 34,676 people died due to terrorism in 2016, while the number was 9,362 in 2006 (Roser, Nagdy et al. 2013).

As terrorism expanded to other countries, Europe has become the target with increasing concerns for the people living there (Barros, Proenca et al. 2007, Draca, Machin et al. 2011, Drakos and Muller 2011, Salerno 2017). As one of the member states in European Union (EU), the United Kingdom (U.K.) had the highest number of terrorist attacks, with a total of 76 separate terrorist attacks in 2016, while the number was 23 in France and 5 in Germany (Statista 2017). Simcox stated that the terrorist attacks in the U.K. would continue since over twenty-three thousand terror suspects lived in the country and government couldn’t control the significant amount of domestic threats (Simcox 2017). The detrimental impact that terrorist attacks brought to people were long-term. A study of London bombing terrorist attack in 2005 has suggested that the pregnant woman who lived in northern England and had been exposed to media coverage of terrorist attacks had infants with lower mean birth weight (Nugent, Khashan et al. 2011). Terrorist attacks also had influences on people’s behaviors. Stecklov and others observed people’s driving behaviors in Israel after the terrorist attack. They found, though, traffic volume declined in both peak and off-
peak hour a few days after terrorist attacks, traffic fatalities increased (Stecklov and Goldstein 2004).

In recent years, many studies have proven that terrorist attacks have threatened public mental health (Boscarino, Galea et al. 2002, Whalley and Brewin 2007, Gruebner, Sykora et al. 2016). Those studies suggested that there was a significant increase of anxiety/depression among a specific group (Boscarino, Galea et al. 2002, Whalley and Brewin 2007, Gruebner, Sykora et al. 2016, Thoresen, Jensen et al. 2016). Besides, studying the negative responses could help to identify the potential threats, which would be meaningful to disaster relief processes and emergency management. Nevertheless, few of the studies have investigated the factors that may associate with the negative responses both in space and time.

Twitter plays an important role in people’s daily lives in the United Kingdom. As a social media platform, Twitter provides people an opportunity to interact with others more quickly and get engaged in social events more easily (Katkina 2015). There were 30% of Internet users used Twitter in U.K. (Dutton, Blank et al. 2013), and these users were younger, wealthier, and better educated than the off-line British population (Blank 2017). The contents that the user post contain the comments, including feelings and attitudes of the user. By extracting the feelings within the contents, Twitter data can be used to monitor the post-disaster situation (Bai and Yu 2016), detect real-time traffic incident (Gu, Qian et al. 2016), and predict match outcomes (Schumaker, Jarmoszko et al. 2016).

Twitter adoption varies between countries, and many factors can affect the Twitter usage. Studies have suggested that age and the income, as well as the access to Internet or an Internet-enabled device, are the factors that may have impact on people’s engagement in Twitter in the United Kingdom (Pearce and Rice 2013, Hargittai 2015, Blank and Lutz 2017), while the
socioeconomic status, gender, age, and Internet skills are considered as the factors in the US (Hargittai 2015). Moreover, the trust in the internet may contribute to user’s internet activities. Blank and et al. examined user’s trust in the internet in Britain between 2003 and 2009 by comparing two reported results. They found that Britain users’ trusts in the internet were influenced by the general attitudes toward technology, and older people tend to be more skepticism about the high technology (Blank and Dutton 2012). Although these factors may affect Twitter data representation, collecting and analyzing the collective Twitter data within a specific area can, at least, represent the general attitudes toward social events of Twitter users in that particular area.

Since terrorist attacks frequently happened in U.K. (Engene 2007), understanding the collective emotional responses toward terrorist attacks became urgent and important. It is also necessary to know how the negative responses vary in a geographic area. It had been proven in many studies that negative feelings, including depression, anxiety, and stress, were harmful to human (Rodrigues, Lopes et al. 2017, Herbert, Hesse et al. 2018, Kanchanatawan, Sirivichayakul et al. 2018, Miskowiak, Larsen et al. 2018). Depressive emotions and stress had a significant influence on humans’ appetite and body weight, as well as the accuracy of their decision-making skills (Wei, Li et al. 2018). Being trapped in depressive feelings for a long time could result in a mood disorder, or even worse, lead to a psychotic disorder (Herbert, Hesse et al. 2018, Kim and Na 2018). Negative emotion acts as a motivating factor underlying politicians’ support (Rico, Guinjoan et al. 2017), and it could result in crimes, which would, on the other hand, generate more negative emotions (Guang, Feng et al. 2017). Therefore, studying the negative feelings within the emotional responses toward terrorist attack is important.
1.2 Literature Review

1.2.1 Application of Social Media

Traditionally, news related to a terrorist attack is covered by mainstream media, such as the TV news channels, newspapers, and radios. The coverage may be biased depending on a media’s interest, and do not necessarily reflect the view or feeling of the people. Studies aiming on the emotional response to terrorism were mostly survey-based, which required a large number of participants and were time-consuming (Bleich, Gelkopf et al. 2006, Fischer, Postmes et al. 2011). However, the rapid development of information and the advent of new technology have significantly reshaped the ways of how each of us connects to the world. Building an information-sharing network provides us a better way to get involved in the world (Panagiotopoulos, Barnett et al. 2016). Twitter (Dorsey 2006), as well as Facebook (Zuckberg, Saverin et al. 2004) and Instagram (Systrom and Krieger 2010), have increasingly become the popular social media platforms, in addition to the traditional media. Unlike TV and newspaper, digital social media have more potential customers, a broader range of spread, and more importantly, a fastest way of spreading messages. With a large amount of monthly active users—Twitter (330 million), Instagram (800 million), and Facebook (2.2 billion)—social media can spread messages all over the world within an hour (Statista 2017).

Twitter, as one of the online networking platforms, has approximately one-quarter of adult users in America and about 320 million users worldwide (Ranney, Chang et al. 2016). Tweets, the texts that people post on their twitter accounts, contain information for us to capture and study. The rich information released by Twitter makes tweets the most favorite resource for social media data acquisition (Smock, Black et al. 2012, Lin, Lachlan et al. 2016, Panagiotopoulos, Barnett et al. 2016). Twitter also allows users to attach location information to tweets. A location-based
tweet, also known as geotagged tweet, contains the latitude and longitude of the digital device where the tweet was sent, such as a mobile phone, a tablet, or a computer. These devices can automatically record the location information of users when they mark their locations on Twitter.

1.2.2 Use of Geospatial Data in Social Media Studies

The location-based tweet allows us to track its geolocation information, which can be accessed with programming languages (Landwehr, Wei et al. 2016). Twitter provides Application Programming Interface (API) for public access, which offers multiple ways to collect tweets based on the needs of users. Twitter REST API can retrieve tweets within the past seven days, while, the Streaming API can be used to capture the streaming tweets. In order to protect the privacy of Twitter users, limitations existed in both two APIs, all standard (free) Twitter APIs no longer provide any access to old tweets beyond past seven days (https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets). In other words, the maximum timeframe for Twitter REST API to work is within the past one week, and the Twitter Stream API can access the most recent data, usually within one day. Twitter also sets the rate limit of APIs, which defines how many times the APIs can be used in an hour. The detailed description and tips for avoiding being rate limited can be found elsewhere (https://developer.twitter.com/en/docs/basics/rate-limiting).

The research that aims to study the behaviors or investigate the public’s reactions to an event at different phases would be more likely to use REST API to gather relevant tweets for (Massey, Leader et al. 2016). However, studies focusing on detecting real-time incident or forming an online information-sharing system would use Streaming API, instead of REST API (Padrez, Ungar et al. 2016, Uchida, Kosugi et al. 2016). With both APIs, tweets containing geolocation information could be extracted. As the Twitter user guide denotes, there are two classes of
geographical metadata: tweet location and account location. The tweet location usually associates with the geo-tagged tweets, while the account location is based on the ‘home’ location provided by the user in their profile. However, using the tweet location would provide more precise location data as it contains the point coordinates where the tweet is from (Developer 2018).

Analyzing the event-related geotagged tweets offers a lot of geographic insights and provides a target area for actions. It could not only help to reveal the place of the incident but also imply more when communicating risks to the public (Landwehr, Wei et al. 2016, Ranney, Chang et al. 2016). Event-related geotagged tweets posted within the period would contain rich information associating with the target event, including the location, moving path, and side effects it brought. The information regarding the situation of the disaster could be updated through analyzing related tweets, and government and emergency responders could give emergency alerts and promptly respond to the disaster (Dong, Halem et al. 2013, Wang, Wang et al. 2016, Wang, Ye et al. 2016).

Studying geotagged data can help us understand human behaviors. Researchers identified that there was a correlation between social connection and users’ check-in locations, and such correlation could represent the users’ geolocation information (Ma, Sandberg et al. 2017). A similar conclusion has been drawn in Shao et al.’s research, which studied travel tours by analyzing the check-in locations of tourists (Shao, Zhang et al. 2017). In this study, scientist detected the tourists’ spatial-temporal behaviors after analyzing the social media data and then extracted the locations to form the city’s tourism districts. It is believed that location-based data has made contributions to the fields of urban planning and tourism (Shao, Zhang et al. 2017). Geotagged data also presents a way of visualizing the community of a specific group (Zhao, Sui et al. 2017) and helps to allocate management resources in emergent situations (Chen, Elmes et al. 2016). Since
the pattern that tweets present is spatiotemporal (Wang, Ye et al. 2016), researchers would like to restrict their studies to a specific timeframe and explore research questions both in space and time.

Whereas geotagged tweets are the most favorite studying objects, previous studies indicated that less than 1% tweets were geotagged (Sloan, Morgan et al. 2013). Reasons may associate with the location service on Twitter and the demographic characteristic of Twitter users. The location service on Twitter is off by default, and users need to opt-in. Otherwise, the message they post won’t contain any location information (Twitter 2018). Tasse et al. suggested that people chose not to share their locations in order to protect privacy (Tasse, Liu et al. 2017). What’s more, adding a location to the message was a different process than adding a tag (Tang, Lin et al. 2010). Studies of the demographic characteristic of Twitter users have suggested that males are more likely to geotag their tweets than females, and the age for those who use geotagging was 0.82 years older than those who don’t (Sloan and Morgan 2015). Sloan et al. have also concluded, though small differences have existed between users who use geo-service or not, those who use geotagged tweets don’t represent the whole Twitter population (Sloan and Morgan 2015).

1.2.3 General Approach of Studying Emotional Data

Posted tweets also contain users’ opinions and feelings regarding specific events, and these become a good source for mining the semantic data behind social media. The analysis of semantic data, also known as sentiment analysis, is an ongoing field of research in text mining (Medhat, Hassan et al. 2014). Its initial use was made to analyze sentiment based on texts such as letters, emails and so on. Topics, emotions, feelings and any other features related to opinions are extracted to detect the semantic hidden within the texts. With the advent of new technologies, such as online applications, microblogging, and social networks, these fields of studies gain more attention from the scientists (Harsh and Dhiren 2016). However, the development of technology also brings a lot
of challenges. The free access to tweets creates an unpredictably large data set which makes sentiment analysis complicated. Also, as the abuse use of symbols, slangs, and emotions in a tweet, it has much noise information that requires pre-processing before we initialize the sentiment analysis, making the whole process challenging.

The general approach for sentiment analysis is based on knowledge-based techniques, which calculates the average aggregate of the semantic orientation of the words and its phase (Kaur, Kumar et al. 2015). However, due to the use of informal language and the length limitation of tweets, summarizing the polarity of words cannot accurately reflect the sentiment of the topic-based tweets. In addition to that, lowering the uppercase letters, removing URL and user reference, removing digits, removing stop words, removing repeated letters, tokenizing, detecting Part-Of-Speech (POS) tags and lemmatizing should be done prior to extracting feelings within the tweet texts (Ficamos and Liu 2016). After that, tweets containing a bag of words for features/topics classification and extraction would be kept. Supervised machine learning, which requires a classifier to extract the features and feelings, can be used to detect the sentiment in tweets. There are two data sets in supervised machine learning. The training set is used to train the classifier based on their selected features with the available algorithms, such as the Support Vector Machine (SVM), the Naïve Bayes and the Decision Trees. The test data set is used to evaluate the performance of the classifier (Kaur, Kumar et al. 2015, Khan, Qamar et al. 2016, Khan, Qamar et al. 2016).

Despite the complexity of processing, sentiment analysis has been successfully applied in social media studies. It has been widely used in studying emergency response, identifying the sentiments from online social networks and understanding how users’ sentiments change accordingly in space and in time (Neppalli, Caragea et al. 2017). By observing the increase or
decrease of the frequently used words on social media or exploring the sudden changes regarding the keywords positivity or negativity, sentiment analysis can detect an event or make a prediction (Ji, Cao et al. 2016, Paltoglou 2016, Schumaker, Jarmoszko et al. 2016). It can also help us to understand the public’s awareness of diseases as well as their stigmas at early phases (Kim, Jeong et al. 2016, Oscar, Fox et al. 2017, Salas-Zarate, Medina-Moreira et al. 2017). However, as mentioned in other papers, the accuracy of the classifiers is limited and needs to be improved in the future work (Go, Bhayani et al. 2009, Kaur, Kumar et al. 2015, Giatsoglou, Vozalis et al. 2017). Also, the classifiers can only use the non-emotion feature to determine the sentiment, which could be a limitation when classifying the emotion-embedded social media data (Go, Bhayani et al. 2009).

1.2.4 Application of Space-time Analysis

The space-time analysis in SatScan (Kulldorff 1997) can be used to pinpoint any geographic variation of activities and identify the peaks/troughs of an event. Researchers used this method to identify the spatial clusters of infectious and chronic diseases, as well as disease vectors and risk factors (Stelling, Yih et al. 2010, Sherman, Henry et al. 2014). The space-time analysis used a cylindrical window with a circular (or elliptic) geographical base and a specific height to identify the potential clusters (Kulldorff 1997). Cluster detection is a fundamental approach for studies aiming to explore the relationship between the objects and their geographic locations. Azage investigated the spatiotemporal clustering and seasonal variability of childhood diarrhea in northwest Ethiopia using space-time scan statistics. They found that the childhood diarrhea had both spatial-temporal variation and seasonal pattern (Azage, Kumie et al. 2015). The same method was used in another research to study the geographic cluster of resistant Escherichia coli within the regional network and the changes in clustering over time (Park, O'Brien et al. 2016).
Space-time analysis also provides multiple statistical models for various types of research in addition to detecting clusters. The space-time permutation model and the Bernoulli model are the most frequently used in space-time studies. The space-time permutation model requires less information and is more straightforward. Only with the case data, the model can identify the number of the observed cases in a cluster by using the spatial location and time for each case, and there is no further information of the control data required (Kulldorff M, Heffernan R et al. 2005). The model would gradually move a cylindrical window across space and time to examine the number of observed and expected cases in the cylinder at a location. A cluster in a particular geographic area was detected during a specific timeframe, suggesting that this area had a higher proportion of cases during this timeframe compared to the remaining areas. Due to the less requirement of the data, the model has been used to identify the cluster of negative emotions in the terrorist attack in Paris (Gruebner, Sykora et al. 2016). Similar research had been done to identify the anomalous diameter distribution of disease in Pennsylvania, USA (Rubin and MacFarlane 2008), detect the location of emergency events (Jasso, Hodkgiss et al. 2009), or monitor terrorist events in their early stage (Li, Jiang et al. 2010). The Bernoulli model, contrary to the space-time permutation model, requires more on the testing dataset, which is divided as cases and non-cases representing by a 0/1 variable, denoting as the cases and controls to serve for the purpose of the study. The model also requires the data containing not only the location but also the time (Kulldorff 1997). Gu and his colleagues employed the Bernoulli model to identify qualified patients for breast cancer clinical trials from free-text medical reports, and the results showed that the Bernoulli model was easier to implement and performed better than other techniques (Gu, Kallas et al. 2013). The same method has also been used in Agustian and his colleagues’ research to detect localized disease clusters (Agustian, Rodd et al. 2012).
Aside from the space-time analysis, a wide range of spatial analysis methods have been developed to accommodate the social media studies. Geographic Information System (GIS) techniques have been used for mining the data behind the social media in recent years (Smock, Black et al. 2012, Dong, Halem et al. 2013, Simon, Goldberg et al. 2014, Ukkusuri, Zhan et al. 2014, Shekhar and Setty 2015, Lin, Lachlan et al. 2016, Martin and Schuurman 2017). Based on the needs of visualize geotagged data, Kernel Density Estimation (KDE) and Inverse Distance Weighting (IDW) in GIS are the two frequently adopted methods (Shook, Leetaru et al. 2012, Spitzberg, Tsou et al. 2012, Li, Goodchild et al. 2013, Tsou, Yang et al. 2013, Padmanabhan, Wang et al. 2014, Han, Tsou et al. 2015). The KDE estimates the density of the sample data based on a specific feature, which can be employed to investigate the relationship between social media content and its geolocation. Researchers used KDE to generate the kernel density map to study the 2012 U.S. Presidential Election (Tsou, Yang et al. 2013), and this method can also be employed to visualize the tweets density while studying the spatiotemporal pattern of Twitter (Li, Goodchild et al. 2013). The IDW can be used to predict the value of an unmeasured the location, and this can help to interpolate the value of the nearby location.

1.2.5 Importance of Studying Negative Emotional Data in Mental Health

Tweets containing the humans’ feelings toward an event would cluster in an area. The negative feeling is an important study object for examining the relationship between human behaviors and their surroundings. This negative feeling, different from positive or neutral feeling, could reveal more severe problems, disclosing the impact of surroundings on human beings’ mental health. Understanding Twitter users’ emotional responses to terrorist attacks, whether they feel threatened, angry, fearful, or overwhelmed, can guide researchers to study the global impact of terrorism. Simon and others studied the tweets related to the terrorist attack in the Westgate
Mall in Kenya, and they found information tweeted by the public had facilitated a greater awareness of the situation (Simon, Goldberg et al. 2014). Feedback toward terrorist attacks was collected once individuals or organizations posted their reactions. This feedback could allow us to examine the variation of negative response based on its location, identify the clusters of negative sentiment, and evaluate how it could be affected by the incident location.

Although many studies have used the emotional data to investigate the health impact of emergency events (Rico, Guinjoan et al. 2017, Herbert, Hesse et al. 2018, Ripper, Boyes et al. 2018), few have examined how both geographical and temporal distance affects the distribution of emotions. In addition to distance, the pattern of emotions would also be affected by human behaviors and area’s economic situation, which should be taken into consideration when analyzing the spatial pattern.

1.2.6 Rural/Urban Differences in Social Media Adoption and Mental Health

People from rural/urban communities may have various attitudes toward social media. Eric and et al. studied how the rural communities used modern technologies in the US. They found rural users were younger than the urban users, and they were more attempted to set their profiles to private than the urban users (Gilbert, Karahalios et al. 2010). What’s more, Christopher and others indicated that there was a reciprocal relationship between social networks and social trust in rural communities, while the relationship was linear in the urban communities (Beaudoin and Thorson 2004). In other words, urban residents can be more easily getting engaged in social media, as well as to pay more attention to a social event.

Many studies have suggested that terrorist attacks have impacts on people’s mental health, resulting depressions and anxieties (Paykel, Abbott et al. 2000, Boscarino, Galea et al. 2002, Philo, Parr et al. 2003, Whalley and Brewin 2007, Thirthalli, Reddy et al. 2017). Terrorist attacks, as well
as crimes, are more likely to occur in urban areas than in rural areas (Sameem and Sylwester 2018). The urban dwellers in Britain, compared with the rural dwellers, are at the high risk of being a target (Hinkkainen and Pickering 2018). In this case, if a terrorist attack happened, the urban dwellers in U.K. would have more depressive feelings.

People’s mental health varies in urban/rural areas. It has been proven that the actual different dimensions of rural/urban spaces—physical, demographic, economic, social and cultural—impact the mental health of rural/urban dwellers (Philo, Parr et al. 2003). Historical studies indicated that adults living in urban areas were thought to be at a higher risk of depression and general psychotic disorder (Dohrenwend 1975, Barquero, Muñoz et al. 1982, Neff 1983). A further study has been conducted on people that were older than 75 years in Britain, and the result suggested that higher population density was associated with increased depression and anxiety in older people (Barquero, Muñoz et al. 1982). Since most of the urban areas in U.K. are highly populated than the rural ones (Suprageography 2016), the older adults in urban areas are expected to have more depressions and anxieties than the ones in rural areas in U.K. However, when the same study was conducted in China, it had a contrary conclusion. Wang and others indicated that older adults living in urban areas in China would have less depressive symptoms than the ones living in rural areas, because the physical and social environments were better in cities (Wang, Chen et al. 2018). Also, Chavez and others identified there was little difference in mental health if the study object were young adults (Chavez, Kelleher et al. 2018).

Since many studies had different conclusions than the previous studies (Allan, Williamson et al. 2017, Chavez, Kelleher et al. 2018, Wang, Chen et al. 2018), the mental health difference in rural/urban areas should be investigated case by case. What’s more, the variations in demographics
and social media adoption in rural/urban areas may also need to be considered when using social media data to investigate the mental health impact of the terrorist attack.

1.3 Research Question and Objectives

Using the two terrorist attacks of London, the United Kingdom in 2017, this thesis raises a research question—how Twitter users’ emotional responses toward terrorist attacks vary in geographic space and time. To answer this question, this study assumes that the location of the tweet represents the residential location of the users. This study has three objectives: first, extract geotagged tweets with Twitter API; second, after collecting the data, use space-time analysis to identify the clusters of negative tweets and examine the variation in space and time; third, after identified the cluster of negative, use Chi-square test to investigate factors that could be associated with such variation in rural and urban differences.

1.4 Significance of this Study

Previous studies focused on using terrorism-related emotional data to examine how terrorist attack may affect people’s life (Bleich, Gelkopf et al. 2006, Salib and Cortina-Borja 2009, Fischer, Postmes et al. 2011, Solana, Nogueira et al. 2017). However, few of them have focused on investigating the distribution pattern of emotional responses toward terrorisms or examining the factors associated with the distribution in space and time. This study used geographical methods to explore how the 2017 terrorist attacks in London affected people’s emotional response both in space and time. It also discussed the factors that may associate with the spatiotemporal pattern of the negative tweets. This study hypothesized that Twitter users’ responses varied in the United Kingdom. Urban areas would have the most significant amount of related responses, as well as the negative responses, indicating users living in these areas would have a greater concern
than the rest areas. Concerning the high frequency of terrorist attacks in Europe, this study could be used as a guide for post-disaster trauma counseling and management.
2 METHODOLOGY

2.1 Study Area

The study area of this thesis was the United Kingdom, which consists of England, Wales, Scotland, and Northern Ireland. The U.K. has been the favorite frequent target of terrorism in Europe. London, as the capital city, had experienced more than ten reported terrorist attacks in the 21st century. Four of them happened in the year of 2017. On March 22nd, 2017, a driving car ran into the pedestrians on Westminster Bridge in London and caused 40 people to be injured and 6 deaths. On June 3rd, 2017, a van with three attackers was driven into pedestrians on London Bridge, and the terrorists stabbed people before the policemen shot them dead. This attack left 48 people injured and 11 people dead, including the three attackers.

To examine the difference in public emotional reactions to these two attacks in both time and space, and to identify the cluster of negative sentiment that would be involved with locations, this study used the geotagged tweets that fell inside the U.K. as the study objects. The contents of the tweets and the location associated with each related tweet were collected. Worldwide geotagged tweets were also collected as a comparison of these in U.K.
2.2 Methods

2.2.1 Data Collection

This study was divided into two phases: data collection and data analysis. A workflow was provided above (Figure 1). A Twitter account was required for retrieving tweets and managing Twitter applications on the Twitter Apps platform, which can be accessed at https://apps.twitter.com/. Within each Twitter application, a consumer key (API key) and an access token were assigned to users for retrieving tweets with Twitter APIs. Both the API key and the access token were required for each independent request. This study adopted the Tweepy API, which provided access to the entire Twitter RESTful API method (Tweepy 2018). Geotagged tweets containing event-related information were intentionally collected by two python programming scripts (Figure 2). This study used two keywords, i.e., “London attack” and “U.K. Parliament attack,” and then made two independent requests with each keyword. For the attack that happened in June, “London Bridge attack” was set as the keywords to retrieve relevant tweets.
The designed scripts would return all geotagged tweets that contained the specific keyword without considering its upper/lower case and word order.

```python
import tweepy
from tweepy import OAuthHandler
import datetime
import codecs
import json

twitter_api_key = "wv4i4hgoawAwT21eH2lAbS"
twitter_api_secret = "ur0Bx99P9v3OEITff3FU0q9JNDup1Rz7EFAJ50yeAmp51Le585"
twitter_access_token = "z761456714-YjRgdi:b0vA7nNeOsgpcIKUXUqge56CE41jXK88"
twitter_access_token_secret = "y7V08XR8tSTvOYom1DbdrcrrsOq2DaeJvDqSoCw49b"

auth = OAuthHandler(twitter_api_key, twitter_api_secret)
auth.set_access_token(twitter_access_token, twitter_access_token_secret)

api = tweepy.API(auth)

file1 = open("./coordinates_bridge_attack_0622.txt", "w")
file2 = codecs.open("./json_tweet_bridge_attack_0622.txt", "w", "utf-8")
file3 = open("./textonly_bridge_attack_0622.txt", "w")

startDT = datetime.date(2017, 3, 29, 0, 0, 0)
endDT = datetime.date(2017, 4, 12, 0, 0)
max_id = -1
maxTweets = 17000
word="London Bridge attack"
i=0

while i < maxTweets:
    if max_id<=0:
        results = api.search(q=word, count=100)
    else:
        results = api.search(q=word, count=100, max_id=Str(max_id))
    max_id = results[-1].id
    for result in results:
        i=i+1
        if result.created_at > endDT or result.created_at < startDT:
            continue
        coords = result.place
        print coords
        if coords is not None:
            lon = coords.bounding_box.coordinates[0][0][0]
            lat = coords.bounding_box.coordinates[0][0][1]
            print lon, lat, result.created_at, result.text
            file1.write(str(lon) + " \"\"" + str(lat) + \\
"\"
            file2.write(str(lon) + "\"\"" + str(lat) + "\"\"
            file3.write(json.dumps(result._json))

file1.close()
file2.close()
file3.close()```

**Figure 2. A sample of python script**

A filter was set in the scripts as the study aimed to collect English-written tweets containing geolocation information. When a full request was made, it would attempt to return the exact location associating with each relevant tweet. Coordinates of the geotagged tweets were extracted and later would be mapped on ArcMap. Three requests were made continuously during 3/22/2017 to 4/12/2017 and 6/5/2017 to 6/21/2017. A total number of 2,183 geotagged tweets (March: 1,382, June: 801) were collected within the timeframes worldwide. A challenge this study faced was that there was an overwhelming number of tweets immediately following each attack. Twitter set a rate limit on how many times the Twitter API can be used per user access token, either 15 calls every
15 minutes or 180 calls every 15 minutes (https://developer.twitter.com/en/docs/basics/rate-limiting.html). Therefore, this research was limited due to missing geotagged tweets at the beginning of each collection period. In total, there were 1,056 geotagged tweets in U.K.

2.2.2 Data Processing

Retrieved tweets (Figure 3) contained the longitude, latitude, date, time, and the texts. As the local time varied from country to country, this study used the Coordinated Universal Time (UTC) to unify all the time zone differences. Duplicate tweets were removed if they occurred at the same coordinates and time with the same content. However, tweets with same coordinates and contents, but different time, would all be kept due to the unknown purposes behind these Twitter accounts and the small number of the available dataset.

This research then extracted tweets containing negative emotions. A typical negative tweet showed fear, sadness, anger, shock, confusion, disgust, shame, among others (Sykora, Jackson et al. 2013, Gruebner, Sykora et al. 2016). In this study, negative tweets were considered as those that delivered severe concerns regarding anything that related to the attack. As shown in the literature review, using text mining, such as machine learning, to conduct the sentiment analysis can help to understand the contents. However, these methods require a large amount of data for
training and testing. Since the number of the geotagged tweets was not as large as expected, this study manually scrutinized each tweet to assess its negative emotion. To minimize personal bias, all categorized tweets were verified later by a peer. Below (Table 1) listed several negative tweets.

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>Contents</th>
</tr>
</thead>
</table>
| -0.297251 | 51.685439| the last leg's response to the London attack is f—brilliant ????
| -0.775435 | 51.279904| so sad watching the vigil in London today. our thoughts &amp; prayers are with everyone who was affected by yesterday’s attack
| -0.15191  | 51.410792| r.i.p ????. i had tears in my eyes walking past where the attack was london you are beautiful ??londonstrong
| -3.0812071| 51.549936| i think they are sensible &amp; very necessary to stop a tory government exploiting brexit to attack people's rights…

To identify the clusters of negative tweets, this study used the space-time permutation model in SatScan to evaluate negative geotagged tweets (cases) collected within the two timeframes from March 22\textsuperscript{nd}, 2017 to April 12\textsuperscript{th}, 2017 and June 5\textsuperscript{th}, 2017 to June 21\textsuperscript{st}, 2017. The space-time permutation model only required case data, including each case’s geographic coordinates and time. The special nature of social media data was that the entire background population who were using tweets was unknown. Therefore, the space-time permutation model, which required only cases, was more appropriate compared to the Bernoulli or the Poisson models.

To reflect the temporal dimension, this study assigned a numerical value to each tweet corresponding to the number of hours this tweet was posted since the date of an attack. Moreover, the posting time of each tweet had been cautiously converted to generic time (Figure 4) using the equation below (Equation 1). The equation set the time 0:00 of March 22\textsuperscript{nd} as the starting point to measure how far each tweet was away from that point in timescale.
Figure 4. Process of calculating generic time

Generic time = DD*24 + Hour + 1  (1)

Where DD represented the day difference, measuring how many days it had been since the attack happened, and the hour represented the actual hour of the posting time.

To understand the nature of the location where tweets were sent, this study also reclassified the urban/rural areas in England, Wales, Scotland, and Northern Ireland. The original classification was conducted in those countries separately based on census Output Area (OA) (Northern Ireland Statistics and Research Agency 2015, Statistics 2016, Statistics 2017), and each had a unique rural/urban classification. After examining the existing differences among the classifications in those countries, this study consolidated all the classes into six categories (Table 2).
<table>
<thead>
<tr>
<th>Original Classification</th>
<th>Reclassification (6 fold)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belfast city, Derry City, Urban major conurbation.</td>
<td>Class 1: Large Urban</td>
<td>Settlements over 125,000.</td>
</tr>
<tr>
<td>Large Town, Medium Town, Urban minor conurbation, Urban City and Town.</td>
<td>Class 2: Other Urban</td>
<td>Settlements between 10,000 and 125,000.</td>
</tr>
<tr>
<td>Small Town, Urban City and Town in a sparse setting, Rural Town and Fringe.</td>
<td>Class 3: Accessible Small Town</td>
<td>Settlements between 3,000 and 10,000.</td>
</tr>
<tr>
<td>Intermedia settlements, Rural Town and Fringe in a sparse setting, Rural village.</td>
<td>Class 4: Remote small Town</td>
<td>Settlements between 3,000 and 10,000. Beyond a 30-minute drive time of a settlement of 10,000 or more.</td>
</tr>
<tr>
<td>Village, Rural village in a sparse setting.</td>
<td>Class 5: Accessible Rural</td>
<td>Settlement below 10,000 and within a 30-minute drive time of a settlement of 10,000 or more.</td>
</tr>
<tr>
<td>Open Countryside and small village, Rural hamlets and isolate dwelling, Rural hamlets and isolate dwellings in a sparse setting.</td>
<td>Class 6: Remote Rural</td>
<td>Settlement below 10,000 and beyond a 30-minute drive time of a settlement of 10,000 or more.</td>
</tr>
</tbody>
</table>
This study also used Chi-square test (Angresti 2007) to examine whether the number of the negative tweets correlated to an area’s urban/rural classification. Each geotagged tweet was denoted as 1/0 to indicate whether it was negative/non-negative. Also, the class that the tweet belonged to was represented by 1, and the rest were represented by 0. The Chi-square test would investigate the relationship between all the negative tweets and each rural/urban classification, and then give the correlation coefficient and the significance of this correlation.
3 RESULTS

A total of 2,183 geotagged tweets worldwide (March: 1,382, June: 801) had been collected during the study period. The worldwide distribution maps (Figures 5&6) suggested that tweets focused mainly in North America and Europe, and sporadically presented in the rest of the world. The United Kingdom, as the victim of the attacks, had the most intensive tweets. The Eastern United States had more relevant tweets clustered compared to the west coast. This was possibly resulted from the use of English keywords and countries with English as the official language. Australia was an exception where only a few messages were tweeted.

Figure 5. Distribution of geotagged tweets in March
Figure 6. Distribution of geotagged tweets in June

Figure 7. Daily number of all relevant tweets in the world

Note: Data is possibly missing on the first day for both worldwide tweets and geotagged tweets due to Twitter API rate limitation and overwhelming capacity of tweets (day 1, 2…, refers to the date from the incident).
The number of both global geotagged tweets and all relevant tweets (both geotagged and non-geotagged) grew rapidly at the beginning of the incident, usually reached the highest number on the first day of the event, and later gradually decreased with time until it turned to 0 (Figures 7& 8). In general, after two weeks most tweets were retweets of previous messages and new tweets became rare. The overall global pattern can be interpreted as the public’s attention reached to the peak by the sudden occurrence of the terrorist attack on the first day, and it decreased for the subsequent days until it disappeared. There were some specific days where the number bounced back but never exceeded the peak value on the first day.

![Geotagged Tweets](image)

*Figure 8. Daily number of geotagged tweets in the world*

Although the overall pattern for the daily number of geotagged tweets was similar to that of all tweets, it fluctuated on some days that needed to be explained in more detail (Figure 8). The increased number of geotagged tweets from Day 1 to Day 3 in March and from Day 1 to Day 2 in June was possibly resulted from the rate limit error set by Twitter’s server and thus may affect the total number of retrieved tweets within that periods. There were also two timeframes where the
number of geotagged tweets had small fluctuations, such as Day 7- Day 8 and Day 10 – Day 11. As the investigation of the terrorism was moving forward through the time, updates related to the event, such as the attacker’s motivation, and the total number of injuries or the fatalities could affect the daily number of geotagged tweets. However, such unstable fluctuations haven’t been found on the number of all relevant tweets (Figure 7). Comparison of the two terrorist attacks indicated more tweets in March than that in June, suggesting attention to terrorist attacks faded when they repeated three months apart.

Table 3. Number of geotagged tweets in U.K.

<table>
<thead>
<tr>
<th>Events</th>
<th># of geotagged tweets</th>
<th>Unique locations</th>
<th># of negative tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>March</td>
<td>717</td>
<td>222</td>
<td>100</td>
</tr>
<tr>
<td>June</td>
<td>339</td>
<td>74</td>
<td>28</td>
</tr>
</tbody>
</table>

There was a total number of 1,056 geotagged tweets in U.K., of which 100 negative tweets were created in March and 28 negative tweets in June (Table 3). Among these tweets, 222 unique locations corresponded to the tweets in March, and 74 individual locations corresponded to those in June. In other words, many geotagged tweets were sent from the same location. It was also noticeable that the number of geotagged tweets, as well as the negative tweets, was higher in March than these in June.
Figure 9. Geotagged tweets in U.K.

Given the map above (Figure 9), the majority of the tweets were sent from England, and most of them were clustered in the southeast area, which is close to the city of London where the two attacks happened. The rest were scattered in Northern Ireland, Scotland, and Wales.
Figure 10. Clusters of the attack in March

*Note: The #1 cluster was made up of repeated points*

A map containing the information regarding all the significant clusters was given at the end of the analysis (Figure 10). When the search window had a radius of 75 km, all the clusters
were in England. The first cluster, as well as the second cluster, was close to the metropolitan areas of London and between Manchester and Birmingham, respectively.

<table>
<thead>
<tr>
<th>Table 4. Summary of geotagged tweets in March (r=75 km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study period</td>
</tr>
<tr>
<td>0 to 527</td>
</tr>
<tr>
<td>Number of locations</td>
</tr>
<tr>
<td>222</td>
</tr>
<tr>
<td>Total number of cases</td>
</tr>
<tr>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. Description of the first cluster in March</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinates / radius</td>
</tr>
<tr>
<td>(51.527592 N, 0.336055 W) / 0 km</td>
</tr>
<tr>
<td>Timeframe</td>
</tr>
<tr>
<td>260 to 260</td>
</tr>
<tr>
<td>Number of cases</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>Expected cases</td>
</tr>
<tr>
<td>9.61</td>
</tr>
<tr>
<td>Observed / expected</td>
</tr>
<tr>
<td>3.23</td>
</tr>
<tr>
<td>Test Statistic</td>
</tr>
<tr>
<td>17.674800</td>
</tr>
<tr>
<td>P-value</td>
</tr>
<tr>
<td>0.000000000078</td>
</tr>
</tbody>
</table>

After running the space-time permutation model, the result (Table 4) indicated that there was a total of 717 geotagged tweets created in March, and 222 unique locations were associated with these tweets. The first most likely cluster (Table 5) took place on April 1\textsuperscript{st}, 2017 from 19:00 to 20:00, ten days after the attack. A closer examination of the tweets suggested that these were from the same twitter account, even though the time stamps associated with these tweets were different. It is unclear why the user repeated tweeting the same message. This study considered this as a false positive cluster given the tweets’ relatively narrow timeframes from the same user.
Table 6. Description of the second cluster in March

<table>
<thead>
<tr>
<th>Coordinates / radius</th>
<th>(52.948941N, 2.226804 W) / 45.73 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeframe</td>
<td>86 to 98</td>
</tr>
<tr>
<td>Number of cases</td>
<td>6</td>
</tr>
<tr>
<td>Expected cases</td>
<td>0.72</td>
</tr>
<tr>
<td>Observed / expected</td>
<td>8.33</td>
</tr>
<tr>
<td>Test Statistic</td>
<td>7.584542</td>
</tr>
<tr>
<td>P-value</td>
<td>0.023</td>
</tr>
</tbody>
</table>

The second cluster was a circular area with a center of 52.948941N, 2.226804 W, and a radius of 45.73 km (Table 6). Tweets that fell inside the cluster were created between March 25th, 2017 at 15:00 and March 26th, 2017 at 1:00. This second most likely cluster occurred three days after the attack within a circle with a 45.73 km radius. It included six negative cases, while only 0.72 were expected. The clustering of the negative feeling of the users covers a massive area. To examine how sensitive the result was, this study alternated the search window size to 36 km and 100 km. SatScan consistently reported a significant cluster of negative tweets within the region between Manchester and Birmingham. When the same operation was applied to the tweets generated in June, the result was not as significant as that in March, so that this study wouldn’t report it here.
Overlying the rural/urban reclassification map and the tweets in the U.K. (Figure 11), it can be found that most of the geotagged tweets were created in urban areas although a small portion fell into the rural area. England had the most tweets, and the suburb area of London had the highest
density of tweets, which was denoted as large urban area or other urban areas. Northern Ireland, Scotland, and Wales did not have as many tweets as England.

Figure 12. Tweets within each classification

The histogram (Figure 12) showed that the class of large urban area had the largest number of tweets (both negative and non-negative ones) among the six classes. Most areas had more non-negative tweets than the negative ones. However, areas categorized as Class 2, i.e., other urban areas, had more negative tweets (count: 15) than the non-negative (count: 12). The class of accessible rural area (Class 5) had the least of both negative and non-negative tweets.
Table 7. Correlation of negative tweet and rural/urban classification

<table>
<thead>
<tr>
<th>Classification</th>
<th>Correlation coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.029</td>
<td>0.435</td>
</tr>
<tr>
<td>2</td>
<td>0.035</td>
<td>0.344</td>
</tr>
<tr>
<td>3</td>
<td>0.09</td>
<td>0.802</td>
</tr>
<tr>
<td>4</td>
<td>0.031</td>
<td>0.409</td>
</tr>
<tr>
<td>5</td>
<td>0.006</td>
<td>0.88</td>
</tr>
<tr>
<td>6</td>
<td>0.057</td>
<td>0.13</td>
</tr>
</tbody>
</table>

To verify whether the negative sentiment of each tweet is associated with the area’s rural/urban classification, this study ran the Chi-square test. As indicated in Table 7, the correlation coefficients—0.029, 0.035, 0.09, 0.031, 0.006, and 0.057—for all testing groups were positive. The P-values, i.e., 0.435, 0.344, 0.802, 0.409, 0.88, and 0.13, for all the classes were larger than 0.05, indicating there was no significant correlation between the negative sentiment and being rural or urban.
4 DISCUSSION AND CONCLUSION

4.1 Tweets, Social Network Usage, and Spoken Language

Clustering of the collected tweets in North America and Europe was mainly attributed to the selected English keywords and the countries speaking English. As shown in Table 7 updated in May 2016, the top five countries that have the most number of active Twitter users in leading markets were: United States, India, Indonesia, Japan, and China (Statista 2018). Some non-English spoken countries, such as Japan, China, although they have a large number of active Twitter users, have their own predominant social media platform rather than Twitter. That may explain why there were few tweets existing in those countries.

Table 8. Official language and prevailing social media platforms for top 5 countries having active Twitter accounts

<table>
<thead>
<tr>
<th>Country</th>
<th>Language (Official)</th>
<th>Popular social media platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>English</td>
<td>Facebook, Twitter, etc.</td>
</tr>
<tr>
<td>India</td>
<td>English</td>
<td>Facebook, YouTube, etc.</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Indonesian</td>
<td>YouTube, Facebook, etc.</td>
</tr>
<tr>
<td>Japan</td>
<td>Japanese</td>
<td>Line, Twitter, etc.</td>
</tr>
<tr>
<td>China</td>
<td>Chinese</td>
<td>WeChat, Weibo, etc.</td>
</tr>
</tbody>
</table>

4.2 Mental Health and Human Behaviors

This study identified that most of the negative tweets clustered in the metropolitan areas (Figure 10), indicating that Twitter users in these areas would tweet more event-related messages and have greater concerns than the rest areas. The SatScan program consistently reported the two significant clusters near London and Liverpool, indicating the cluster of fear/sadness had been detected in these areas (Gruebner, Sykora et al. 2016). In this case, Twitter users in these areas have more negative responses regarding the attacks than the other areas.
The negative sentiment triggered by terrorist attacks may result in the abuse of social media (Marino, Gini et al. 2018). Several studies have suggested that there was a strong association between problematic social media use and mental health problems, such as depressive, fear and sadness (Moreau, Laconi et al. 2015, Miskowiak, Larsen et al. 2018). That may explain why there were a lot of identical tweets coming from the same Twitter account. A study suggested that the Post Traumatic Stress Syndrome (PTSS) associating with the terrorist attacks had a long-term mental impact on parents, because they felt guilty for exposing their children to trauma and had to be cautious for the dangers that may affect their children (Thoresen, Jensen et al. 2016).

4.3 Mental Health and Rural/Urban Classification

This thesis did not find a significant correlation between the clustering of negative tweets and rural/urban classification. The locations of the two terrorist attacks were both in the city of London. When the terrorist attacks happened, the clustering of messages and news related to the attack was mainly concentrated locally in London and spread out quickly.

The negative tweets were mostly sent from the urban areas in England, and the two most likely clusters also existed in these areas. Nevertheless, the Chi-square test indicated there was no significant correlation between the negative tweets and the rural/urban settings. As discussed before, the urban dwellers in the UK intended to have more depressions than the rural dwellers (Barquero, Muñoz et al. 1982). These urban civilians were at the high risk of being the target for a terrorist attack (Hinkkainen and Pickering 2018). Twitter users in urban areas would have more negative responses to terrorist attack than those in rural areas. This explained why this study identified more negative tweets clustered in the urban areas. However, the variation of social attention may also be associated with the Twitter adoption and user demographics (Philo, Parr et
al. 2003). In this case, while an area’s urban/rural classification may affect its residents’ attention on a social event, the creation of negative tweets won’t be necessarily affected by this setting.

4.4 Social Media Attention to Various Attacks

Findings from this study suggested that different countries have various reactions to terrorist attacks around the world. This research found that terrorist attacks in developed countries attracted much more attention than those in developing countries in English-speaking regions. To further confirm this, this study analyzed the social media attention on another attack in Syria. On the 16th of April 2017, a suicide car bomber killed and injured at least 100 evacuees as they waited to leave their towns in Syria. This study collected both geotagged tweets and all other tweets related to this attack, containing the keywords “evacuee attack” and “rashidin attack,” between April 17th and May 1st. Compared to the attacks in U.K. (the total number of tweets was 71,096 in March and 34,112 in June), there were only 1,247 relevant tweets within the first day following the attack in Syria, and most were tweeting the news report without indicating any emotions.
(Figure 13). Among these tweets, only one tweet was geotagged. However, it was possible that using Arabic keywords may have collected a larger number of tweets. But the comparison here suggested that attacks in developing countries were unable to draw as much attention as the ones occurring in developed countries. Since Han and others suggested that “Twitter users living in heavily populated areas would have the more geographical awareness of other cities that were far away than those living in less populated areas,” that awareness may be more associated with big cities (Han, Tsou et al. 2015). In other words, if a terrorist attack happened in a big city, instead of one in a small city, the Twitter users living in a populated area would likely pay more attention than the ones living in a less populated area.

4.5 Limitations

This study may be limited in several aspects. First, it excluded related tweets that didn’t have the location information or the selected keywords. Several attempts have been made to retrieve tweets without keywords, yet all the efforts failed due to the rate limit set by Twitter. Second, most tweets were not geotagged. The distribution of negative feelings may be skewed if these non-geotagged tweets would have been geographically located. Third, manually labeling each tweet as negative or non-negative is subjective, and the classification may be changed if the results had been verified by others. The negative sentiment analysis would vary due to the differences in labeling tweets, and the manual label method could only be applied to a small number of study objects. Fourth, the results would vary if the search criteria were modified to collect data from another social network or in another language. The variability and unpredictability of human behaviors, as well as the differences between countries, were also factors that need further research. Last but not the least, due to the large differences on demographics of Twitter users in different countries, many studies have suggested that Twitter
data is not suitable for any research where representativeness is important (Sloan and Morgan 2015, Sloan, Morgan et al. 2015, Blank 2017, Blank and Lutz 2017). Findings in this study therefore are typically applied only to Twitter users.

Understanding the concentrated emotional reactions in real-time after a disaster is critical to provide guidance for the provision of mental health services (Whalley and Brewin 2007, Gruebner, Sykora et al. 2016). This study, in general, made a contribution in this regard by identifying the clusters of emotional reactions in space and time based on recorded geotagged social media data. The results indicated that the terrorist attacks, especially the one in March, had significant impacts on public’s emotional responses. The two attacks both occurred in England, making this place to generate the most tweets. The negative tweets were mostly generated in the urban area, with one cluster being detected three days after the attack, and the other cluster being detected ten days after the attack. Although the negative tweets were associated with urban areas, there was no significant correlation between negative tweets and the area’s urban/rural setting. This study used GIS in social media analysis to identify the emotional impact after disasters. Future research would need to find a solution using a bounding box, not keywords, to retrieve data. A bounding box can help to restrict the target area based on the need of the study, and all tweets sent from that area could be collected. This is helpful when the study site is a small specific area. Researchers also need to come up with a way to bypass the Twitter limit error to fully use Twitter API.
REFERENCES


