A Visual Insight Into Tweeting Activity Before and During Natural Disasters: Case Study of Hurricane Harvey

Shadi Maleki, Texas State University, USA Milad Mohammadalizadehkorde, Texas State University, USA

ABSTRACT

Big data provided by social media has been increasingly used in various fields of research including disaster studies and emergency management. Effective data visualization plays a central role in generating meaningful insight from big data. However, big data visualization has been a challenge due to the high complexity and high dimensionality of it. The purpose of this study is to examine how the number and spatial distribution of tweets changed on the day Hurricane Harvey made landfall near Houston, Texas. For this purpose, this study analyzed the change in tweeting activity between the Friday of Hurricane Harvey and a typical Friday before the event.

KEYWORDS

Cartography, Data Mining, Disaster Analysis, Geospatial Analysis, GIS, Hurricane Harvey, Tweet, Tweeting Activity, Twitter

INTRODUCTION

The increasing occurrence of natural disasters and widespread use of social media have increased the scholarly interest to examine the role that social media could potentially play in disaster resilience (Wang et al., 2019). Social media data have been used by researchers in different academic fields (Mocanu et al., 2013), including Geography and Geographic Information Science (GIS) (Wang and Ye, 2019). In particular, the availability of geotagged tweets has opened an important frontier for geographic analysis (Thome et al., 2016; Yuan, & Medel, 2016).

Twitter is one of the most highly used microblogging services, with millions of monthly active users across the world. Microblogging in Twitter is a form of communication initially composed of 140 characters and recently raised to 280 for all users. The number of twitter users has been increasing significantly over the last few years. According to the Twitter First Quarter of 2020 Earnings Report, monetizable daily active users (mDAU) who accessed Twitter daily are 24% higher than the first quarter of 2019 (Twitter, 2020). The mDAU count is 166 million, including 133 million international users and 33 million U.S based users.

Through Twitter, it is possible to obtain information regarding users' profiles, their situations, feelings, and experiences, as it often provides immediate and on-location updates (Kireyev, Palen & Anderson, 2009). The scale of information and data provided by Twitter is incomparable to the traditional methods where data is collected primarily by surveys making Twitter a fascinating tool of data collection for research. In particular, Twitter has become a powerful data collection tool during

DOI: 10.4018/IJAGR.2021070102

This article published as an Open Access Article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

natural disasters (Wang and Ye, 2018), helping environmental managers and planners to identify populations at risk and prepare responses (Houston et al., 2015).

When a disaster occurs, time becomes a determinative factor, and being virtually connected on social media can be an immediate communication solution. Twitter has been revealed a vital source for people affected by a natural disaster to collect real-time information, report about their situation, and requesting help (Olteanu, Vieweg, & Castillo, 2015). For example, during Hurricane Irene, Twitter played a critically vital role as a communication channel (Mandel et al., 2012; Takahashi et al., 2015), adding to the value of Twitter data in disaster research (Li et al., 2015). Many people tweeted to provide or obtain information about the ongoing situation in their communities and to communicate with their friends and families during crisis time (Bird, Ling, & Haynes, 2012; Landwehr & Carley, 2014). Coupling of this information and GIS represents a unique opportunity for decision-makers and planners to improve risk identification and emergency responses (Singleton and Arribas-Bel, 2019; Wang and Ye, 2019).

In disaster studies, social media data has been explored through four dimensions of space, time, content, and network (Wang and Ye, 2018). Among these dimensions, space and content have been explored more extensively than other dimensions (Ye and Wei, 2019). A recent review of 94 papers showed that more than half of the examined articles were focused on content analysis and the rest focused respectively on space, time, and then network analysis (Wang and Ye, 2018).

Despite the recent popularity of social media data in research, obtaining meaningful insight from big data can be challenging in the frame of geography (Singleton and Arribas-Bel, 2019; Crampton et al., 2013). One of the challenges is to detect the pattern of change before and during a disaster. By visualizing the spatial distribution of tweets, we examine how tweeting patterns changed on Friday, August 25, the day when Hurricane Harvey occurred, compared to a typical Friday before Hurricane Harvey in Houston Metropolitan Area (HMA). Considering that previous studies (ex. Mandel et al., 2012) found that, in general, during a natural disaster the frequency of tweeting increases compared to the past, we hypothesized that there was an increase in tweeting activity on the day that the hurricane hit the HMA. We will use a visualization technique to test this hypothesis and gain further insights.

BACKGROUND

Big data provided by social media has been increasingly used in various fields of research (Goodchild and Glennon, 2010; Roche et al., 2013; Palen et al., 2010). As introduced by Laney (2001) and highlighted by Wang et al. (2015) "Big data are high volume, high velocity and wide variety datasets that require new forms of processing to enable enhanced process optimization, insight discovery, and decision making" (Wang et al., 2015, p.33). The process of obtaining insight from big data is also named "data science" since "big data" is arguably a technological matter without necessarily providing any insight (Singleton and Arribas-Bel, 2019).

The use of big data in geographic research has provided geographers with the opportunity to build improved models of human-environment relationships over space and time (Singleton and Arribas-Bel, 2019). The interdisciplinary nature of geography, development of critical GIS and quantitative geography contribute to the idea that geographers are well prepared to "combine technical practice, quantitative methods, and critical scholarship" (Graham and Shelton, 2013, p. 3) to turn big data into insight (Singleton and Arribas-Bel, 2019). Although the modifier "big" is difficult to define due to its constant change throughout the time (Graham and Shelton, 2013), utilizing big data in geographic research has obtained popularity in recent decades.

The proliferation of big data usage in research is in part because of significant progress in big data storage due to the quiet recent advancement in hardware technology (Keim, 2002; Graham and Shelton, 2013). While this flourishing data production is getting bigger and faster, big data provided by social media has been increasingly used in various fields of research, including disaster studies, emergency management (Goodchild and Glennon, 2010; Roche et al., 2013; Palen et al., 2010), and

geographic studies (Graham and Shelton, 2013). It is supposed that big data can empower geographers to build "better models of human relationships and activities over space and time" if geographers are actively engaged with data science (Singleton and Arribas-Bel, 2019, p.2).

A few concerns arise regarding the use of big data in research. One concern is that big data may undermine some valuable knowledge, not encapsulated within the big data framework (Graham and Shelton, 2013). Also, big data is not representative of the entire population, excluding people who do not use or have access to social media. Further, data science methods may neglect some information (ex. location) in favor of simplicity (Singleton and Arribas-Bel, 2019).

Despite all the limitations, big data has been particularly effective in disaster studies and emergency management in recent decades (Martin et al., 2020; Graham and Shelton, 2013; Goodchild and Glennon, 2010; Roche et al., 2013; Palen et al., 2010). During natural disasters, people use social media to contact friends and families to exchange information about their situation and make sure they are safe (Bird, Ling, and Haynes, 2012; Landwehr and Carley, 2014; Wang et al., 2019). Many people post messages about their unsafe situation and ask for help (Acar and Muraki, 2011).

Twitter, in particular, has played a significant role in improving situational awareness in emergency circumstances during the 2011 flood in Thailand (Kongthon et al., 2012) and in the 2009 Oklahoma grassfires and Red River floods (Vieweg et al., 2010). Not only has Twitter been used as an official source for hazards-related information (Chatfield, Scholl, and Brajawidagda, 2013), it has also become a space where users willingly share their experiences with family, friends, and the broader community. In a 2015 study, Dailey and Starbird assessed the human reaction to the B.P Deep Water Horizon Oil Spill, flowing for 87 days starting on April 10, 2010, by defining how people used Twitter to extract meaning from the usage of oil dispersant. In another study, Jung and Uejio (2017) examined the relationship between heat-related themes in different cities around the United States to identify whether the number of tweets containing heat/AC keyword changes with the rising temperature. Another example is given by the work of Ye and Wei (2019), where Twitter is used to get insight into how El Niño is discussed in the United States.

While social media could be used as a source of information in all phases of natural disaster management, many studies have focused on disaster response, leaving a gap in pre and post-disaster phases (Wang and Ye, 2018). Various studies found that social media is significantly used to contact friends and families to get information about their situation during natural disasters (Bird, Ling, and Haynes, 2012; Landwehr and Carley, 2014; Wang et al., 2019). Many people in disaster-hit areas post messages about their unsafe situation, while people in other areas mostly write to let their family and friends know that they are safe (Acar and Muraki, 2011). The focus on the disaster response phase is probably caused by the capture of streaming content during a disaster in addition to data sparsity that can lead to a non-reliable analytical approach (Wang and Ye, 2018).

Researchers have introduced different tools and perspectives for analyzing disaster-related tweets. For example, Kumar et al. (2011) introduced the TweetTracker tool, which can help first responders and relief organizations to track locations of people in need of help during disasters. In another example, Ashktorab et al. (2014) introduced a data mining tool called "Tweetdr," which can be used to cluster, classify, and extract information from tweets. They validated their approach with tweets collected from 12 crises about infrastructural damage, damage type, and casualties in the United States since 2006. Mandel et al. (2012) examined 65,000 tweets for demographic differences through training a sentiment classifier. This study found that the peak in the number of tweets matched the time when Hurricane Irene hit the land. Results were shown using a temporal frequency diagram and differentiated by lines and colors.

Effective data visualization plays a central role in generating meaningful insight from big data. In several studies, the exact position of a given series of tweets is plotted on a map with a highlighted limitation in detecting spatial patterns (Wang and Ye, 2018). Big data visualization has been a challenge due to the high complexity and high dimensionality of it. Moreover, visualizing a large number of points can lead to a phenomenon known as overplotting, which can become an overwhelming

cognitive process for users (Wang et al.,2015). Overplotting occurs when a large number of points overlap and eventually create a homogenized surface, which makes it visually difficult to detect any pattern (Dang and Wilkinson, 2010).

Conventional data visualization and less known methods of display are summarized by Wang et al. (2015) as (1) The known methods consist of using tables, histograms, scatter plots, different types of charts, timelines, Venn diagrams, data flow diagrams, entity-relationship diagrams, etc. and (2) Less known methods such as parallel coordinates, treemap, cone tree, and semantic network. Different ways can be used to improve the visualization of patterns and correlation, but they are not always applicable. Standard data reduction techniques such as sampling and filtering can be used to, for example, exclude outliers and obtain a reduced dataset. Another approach is to manage the transparency of data points in a way that each point becomes slightly transparent. Another method to address this issue is to slightly offset the position of data points to avoid the overlapping problem. However, managing the transparency and offsetting techniques are more effective when working with smaller datasets but not as much practical use with more massive datasets (Poorthuis and Zook, 2015).

Another approach is to represent density using graduated symbols within an area (Dang, Wilkinson, and Anand, 2010). For example, counts can be shown within political and geographic regions by sizing symbols. However, this method cannot completely resolve the problem of overplotting as symbols can still cover each other if their size exceeds the distance with their neighboring symbols. Furthermore, this method can affect perception and visual interpretation (Ware, 2008). A similar problem exists with using choropleth maps to represent the magnitude (Keim, 2002). Another way is to divide a space into partitions and change the size of barriers to show counts of data in each partition (Gastner and Newman, 2004). This technique can effectively avoid the overplotting problem, but on the other hand, can cause distortions in location as well as in perception (Dang, Wilkinson, and Anand, 2010).

Heatmaps and Kernel density methods are also used to address the overlapping problem within large datasets by generating a density surface using color or intensity gradient (Poorthuis and Zook, 2015). This approach is more effective to represent a continuum, assuming that the nature of data is continuous such as natural processes. Nevertheless, heatmaps could also be used to aggregate points to larger areas such as political or geographic divisions. The problem with political or geographic boundaries is that the unequal area may falsify the results as it creates a higher chance of having more points inside their borders. Thus, it is recommended to aggregate points to equal-area geometric shapes such as rectangle or hexagons to minimize this problem (Poorthuis and Zook, 2015).

Kernel density estimation (KDE) is one of the methods used to study high volume social media data during Hurricane Sandy (Guan and Chen, 2014). Density-based clustering is another method used by Wang et al. (2015) to study Chinese social media called Weibo in the case of the 2012 Beijing Rainstorm (Wang and Ye, 2018). Any high volume of activity could be a direct effect of areas with a larger population, which is addressed by Dual KDE and excluding the population impact in Wang et al. (2016). Multiple topics can be extracted from the content of tweets before, during, and after a natural disaster. For example, Huang and Xiao (2015) assessed the social response topic during Hurricane Sandy.

Among the most effective methods used to avoid the overlapping problem within large datasets is generating heatmaps (Poorthuis and Zook, 2015). Kernel density method can be considered a type of heatmap which represents the smoothed density using color or intensity gradient. This approach is most commonly used to represent a continuum, assuming that the nature of data is continuous such as natural processes. A more appropriate approach for using heatmaps in big data analysis is to aggregate each point to a larger area such as political or geographic divisions, or the area could be other geometric shapes such as rectangles or circles. The problem with political or geographic boundaries is the unequal area, which may falsify the results as it creates a higher chance of having more points inside their borders. To avoid this problem, Poorthuis and Zook (2015) have suggested to aggregate points to equal-area geometric shapes such as rectangle or hexagons.

METHODS

This study aims to discover how the tweeting frequency changed on Friday, August 25, 2017, the day that Hurrican Harvey made landfall in HMA compared to the past three Fridays in the same month. Previous studies often found an increased tweeting activity on the day of a natural disaster occurrence. Thus, this study is formed based on the hypothesis that the tweeting frequency is higher on Friday, August 25, compared to the tweeting activity on the same day in the past. Using a visualization method, this study aims to examine whether this hypothesis is valid for this study.

Study Area

The study area consists of Houston and The Woodlands-Sugar Land, also called the Houston Metropolitan area or Greater Houston (datausa.io). This area is composed of nine counties: Austin, Brazoria, Chambers, Montgomery, Galveston, Waller, Fort Bend, Harris, and Liberty (Figure 1). The largest city in this area is Houston, which is the fourth most populous city in the United States following New York, Los Angeles, and Chicago (houstontx.gov). Houston has also been the most populated city in Texas since 1930. The Houston area has been subject to one of the highest numeric population change from April 2010 to July 2019, with 10.7% population gain (Census.gov). The estimated population reached 2,320,286 in July 2019 (Census.gov). Houston's shallow topography and its geographic location in proximity to the Gulf of Mexico make this place highly vulnerable to hurricanes, storms, and floods.

Data

The data used in this study consists of a sample of tweets collected through Twitter streaming API for August 2017. The dataset includes only geotagged tweets, and it has further been reduced to the coordinates falling in the HMA. We did not use any keywords to filter the data because this study is interested in looking for a change in the frequency of tweeting activity in general. This study focuses mainly on time and space, aiming to provide a comprehensive point of view regarding the tweeting activity in an ordinary Friday before Hurricane Harvey and Friday, August 25, when the hurricane landed in HMA. Table 1 shows the number of tweets on four Fridays of August 2017. The tweets of the first three Fridays (August 4, 11, and 18) are used to estimate the tweeting activity on a typical Friday of August 2017.

Analysis

As discussed in the background section, aggregating Twitter points to a regular rectangular (fishnet) or hexagonal frame can help address the problem of overlapping points to some extent. Such a frame can address the Modifiable Areal Unit Problem (MAUP) to some extent as well (Poorthuis and Zook, 2015). In addition to rectangles and hexagons, other geometric shapes such as circles and triangles are used to create a grid (Birch, Oom, and Beecham, 2007; Carr, Olsen, and White, 1992). The triangular grid is not as popular as others because of the position of triangles with two different orientations (Sahr et al., 2003). Hexagonal grids have higher visual accuracy compared to circular or rectangular grids, and they provide a more precise representation of spatial patterns (Birch, Oom, and Beecham, 2007; Scott, 1988). Figure 2 shows the plotted tweet points for each Friday (August 4, 11, 18, 25) before the application of a fishnet, thus showing the problem of overplotting (Figure 2).

Therefore, we generated a hexagonal grid in ArcGIS to aggregate up from the original point pattern. After an exploratory process to identify an appropriate cell size, we found that 25 mi² (64.7 km²) is a suitable size for obtaining meaningful visualization. Therefore, a hexagonal grid of the latter size was generated, and the tweet points for each of the four Fridays were aggregated up to the hexagonal grid by joining each Friday's attribute table to the attribute table of the hexagonal grid. Using the function of Field Calculator in ArcMap, the difference of tweet numbers between the Friday of Hurricane Harvey (August 25, 2017) and those of the three Fridays before the hurricane

International Journal of Applied Geospatial Research

Volume 12 • Issue 3 • July-September 2021

Figure 1. Study area composed of nine counties in Texas



Table 1. Number of tweets in 4 Fridays in August 2017 (August 25 is the day of Hurricane Harvey)

August 4	August 11	August 18	August 25
5031	4848	4872	4526

day were calculated (Equation 1). Figure 3 shows this difference based on the median. To obtain more significant results, we took a step further to normalize these results:

Difference = $n_4 - Median(n_1, n_2, n_3)$

where:

 $n_4 = \text{count of tweets on August 25}$ $n_{1,2,3} = \text{count of tweets respectively on August 4, 11, and 18}$

Normalization Process

In the previous steps, after addressing the overplotting problem by aggregating the points to a hexagonal grid, we calculated and mapped the difference in the spatial distribution of tweet frequency on the day that Hurricane Harvey hit HMA compared to a typical Friday before this event (Figure 3). To obtain a more significant representative ratio, the results needed to be normalized. We used the median as a measure of central tendency to normalize the results (Equation 2). The median is preferred in comparison to the mean because the median is considered a more accurate measure as the mean can be more affected by the outliers in the dataset, thus skewing the results:



Figure 2. Georeferenced tweets in four Fridays of August 2017 in HMA. This picture represents the problem of overplotting.

Normalization = Difference / Median (n_1, n_2, n_3)

where:

 n_{123} = count of tweets respectively on August 4, 11, and 18

Another way commonly used is to normalize the data by population. This method is used when the spatial distribution of data is a reflection of the population (Wang and Ye, 2018). This study took a different approach due to the presence of imposed boundaries (hexagon), known as *fiat* regions by Smith and Varzi (2000). Imposed boundaries are often used in GIS to provide a form of sampling for underlying social reality, potentially misleading in multiple ways (O'sullivan & Unwin, 2014). Hexagons and other types of geometrical shapes bear little relationships with the distribution of the population (O'sullivan & Unwin, 2014). A study area divided by hexagons will not match the population pattern, which is often represented by conventional boundaries such as census tracts or block groups. Also, population information available based on these conventional geographic units does not specify the spatial distribution. Thus, estimating the population of each hexagon and focusing on the ratio between the number of tweets and the population may lead to exclusion, or under/overrepresentation of populations (Steiger et al., 2015).

RESULTS

This study examined how the tweeting activity changed between Friday, August 25, 2017, the day that Hurricane Harvey made landfall in HMA, and the median of tweeting activity on the three Fridays

Figure 3. The classified difference in the spatial distribution of tweets (before normalization) between August 25, 2017 and a typical Friday in the same month and year



before the Hurricane occurrence. The median of the past three Fridays in August of the same year was used as an estimation of tweeting activity on a typical Friday before the hurricane occurrence.

To avoid overplotting, we aggregated the tweet points to a hexagonal grid, thus obtained the number of tweets in each hexagon. The use of the grid allowed us to obtain the count of tweets in each hexagon and improve the visualization. The difference of tweet counts on August 25 (Hurricane Harvey occurrence) and the three Fridays before it was calculated for each hexagon and normalized by the median.

The results showed that 449 hexagons out of 4249 experienced a change in the spatial distribution of tweets. The resulting ratio has a midpoint of 0, which indicates that there was no change in the number of tweets between August 25 and a typical Friday before the hurricane occurrence. Values larger than one show an increase in the number of tweets on August 25 compared to the expected number of tweets on a typical Friday before that day. Values lower than one are indicative of fewer tweets on the day of the hurricane than expected. Overall, 6.9% of hexagons showed a decreasing trend in tweeting activity versus 3.6%, which indicated an increase in twitting activity on August 25, 2017. Figure 4 visualizes the change in the spatial pattern of tweeting with most of the hexagons showing a decline in tweeting on the day of the hurricane occurrence.

In Figure 5, the spatial pattern of change is shown in the main cities in the HMA. As expected, Houston shows the highest positive change, thus increase, in tweeting activity. Other cities showing an increase are respectively Sugar Land, Montgomery, Northwest Harris, and Fulshear-Simonton. The number of tweets decreased in many cities on the day Hurricane Harvey made landfall in the HMA. We will discuss a few reasons that may have caused this trend in the spatial pattern of tweeting in the discussion section.

Figure 6 shows the median of tweet counts of the first three Fridays of August 2017 classified into three groups (low, medium, and high). This classification is meant to show the level of confidence



Figure 4. The figure shows the difference in the spatial distribution of tweets (after normalization based on the median) between August 25, 2017, and typical Fridays in the same month and year

of the results in each hexagon. For example, the blue hexagons indicate that the number of tweets used to calculate the median was less than 50. The yellow hexagons indicate that the median was calculated based on tweet counts between 50 and 100. The red hexagons refer to more than 100 tweets used to calculate the median. Where a larger number of tweets are available results are more accurate. On the other hand, where a lower number of tweets are used to obtain the median results may not be as accurate.

DISCUSSION AND CONCLUSION

This study examined whether the tweeting activity increased on the day that Hurricane Harvey hit the HMA compared to the past. It is important to examine the pre-disaster conditions to improve our understanding of disaster data and its limitations (Wang and Ye, 2018). Unlike to the findings of some previous studies that reported an increase in tweeting during natural disasters (Mandel et al., 2012; Vieweg et al., 2010; Blanford et al., 2014) this study found that the tweeting frequency decreased in most of the areas in HMA on August 25, 2017. This difference may be partly due to the fact that our study did not use any keywords to filter the data, while other studies used filtered datasets. For this study, we did not intentionally use any keywords to obtain a more comprehensive understanding of the change in the pattern of tweeting before and during Hurricane Harvey occurrence.

The highest activity rates were concentrated in the city of Houston and few towns around it. Factors such as ease of access to the internet, population, and digital divide can cause different activity patterns. When a natural disaster occurs, people may leave their places and move to other towns to find a safe refuge. During storms and hurricanes, a power outage may occur, potentially affecting people's access to social media. However, according to Mandel et al. (2012), a power outage cannot effectively impact social media usage since smartphones have access to the internet even if they are

Volume 12 • Issue 3 • July-September 2021



Figure 5. The difference in tweeting activity between August 25^{TH} , 2017, and a typical Friday before August 25 (when Hurricane Harvey hit this area). This map is made based on the median value and shows the name of the main cities in the Houston Metropolitan area.

Figure 6. The classification: less than 50, larger than 50 and less than 100, and larger than 100, respectively indicate to low, medium, and high reliability of the results. The figure shows the confidence level of results calculated based on the median value.



not connected to a residential Wi-Fi. It is important to consider that many people may avoid using their phones to write on social media to save the battery of their phones for a longer time in times of emergency.

Overall, this study builds upon previous studies to leverage visualization techniques to explore the spatial distribution of tweeting activity during a natural disaster. This study contributed to the previous knowledge by examining the difference in the spatial distribution of tweets, not only during but also before, a natural disaster. Comparing users' tweeting attitudes at different times, before and during a natural disaster, can benefit emergency management and planning communities in preparing responses.

Future studies are needed to explore the reasons behind the changes in tweeting behavior during disasters (Mihunov et al., 2020). Also, it would be interesting to examine the tweeting activity in the aftermath of Hurricane Harvey. Examining the socio-demographic characteristics of users is important for a more comprehensive understanding of the factors that may cause a change in tweeting activity. Previous findings indicated that people with a higher level of education tend to tweet more (Li, Goodchild, and Xu, 2013). Age can also influence tweeting behavior. Younger individuals produce more tweets (Huang and Xiao, 2015). Geographic location is another influencing factor. It seems that people in more physically vulnerable communities tend to be more socially aware of the impacts of a disaster. These and many other factors are interesting aspects that need to be further examined in future studies.

Limitations of the Study

Tweets are gathered on first three Fridays of August 2017 estimate the frequency of tweeting on a typical Friday before the Friday that Hurricane Harvey hit HMA (August 25, 2017). We used the median of the three Fridays in August of the same year to define a typical Friday. This seemed a reasonable choice, as all the tweets are gathered on the same day and month. However, using a larger dataset may provide a more accurate estimation of how the tweeting activity looks like on a typical Friday.

Another limitation of this study is known as "the small number problem" (Dean and Dixon, 1951), meaning that areas with a small number of tweets show highly varying ratios, which may cause misinterpretation of the results. The small number problem occurs because of the sparse population in some areas, which leads to a lower number of tweets, which makes making the final ratio less reliable. For example, in this study, the median of tweets ranges from 1 to about 500, meaning that in some of the hexagons, there are only one or few tweets, and in some, there are a larger number of tweets. Therefore, a ratio based on only a few tweets is less reliable than the same ratio based on 500 tweets.

To show this problem, we created a confidence interval for each odds ratio to obtain an indicator of reliability (Figure 6). A larger number means higher confidence in the results. Thus, the median values are classified as less than 50, larger than 50, and less than 100, and larger than 100 showing respectively low, medium, and high reliability of the ratio. As Figure 7 illustrates, only a few hexagons contained more than 100 tweets. These hexagons are within the city limits of Houston, West University Place, and Sugar Land. To increase the accuracy of the results, using a larger dataset is recommended.

Besides using a larger dataset, future studies may explore other methods of normalization to explore the difference. Also, we suspect that using different sizes of hexagons may lead to a change in the spatial pattern. The size of the hexagon used in this study (25 mi²) was decided after exploring larger and smaller sizes. Our goal was to identify a hexagon size that best represents the spatial pattern and allows a meaningful visualization. Future studies may examine other methods to identify other shapes and sizes of grids to explore how the results may be different. Conducting content analysis may also shed some light on the change in the tweeting activity before and during Hurricane Harvey. Using other methods of research, including surveys and interviews, would also help to identify what other factors should be considered in order to improve our understanding of the results.

Figure 7. The results for areas within the city limits of Houston, West University Place, and Sugar Land have higher confidence compared to other places



ACKNOWLEDGMENT

We thank Dr. Ronald Hagelman III, Dr. Alexander Savelyev and Dr. Nathaniel Dede-Bamfo for their thoughts and insights.

REFERENCES

Acar, A., & Muraki, Y. (2011). Twitter for crisis communication: Lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities*, 7(3), 392–402. doi:10.1504/IJWBC.2011.041206

Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). *Tweedr: Mining twitter to inform disaster response*. Paper presented at the Iscram.

Birch, C. P., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment, and simulation in ecology. *Ecological Modelling*, 206(3-4), 347–359. doi:10.1016/j.ecolmodel.2007.03.041

Bird, D., Ling, M., & Haynes, K. (2012). Flooding Facebook - the use of social media during the Queensland and Victorian floods. *Australian Journal of Emergency Management*, 27(1), 27-33. Retrieved from http://libproxy. txstate.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-84861837365&site=eds-live&scope=site

Blanford, J. I., Bernhardt, J., Savelyev, A., Wong-Parodi, G., Carleton, A. M., Titley, D. W., & MacEachren, A. M. (2014, May). Tweeting and tornadoes. ISCRAM.

Carr, D. B., Olsen, A. R., & White, D. (1992). Hexagon mosaic maps for display of univariate and bivariate geographical data. *Cartography and Geographic Information Systems*, 19(4), 228–236. doi:10.1559/152304092783721231

Chatfield, A. T., Scholl, H. J. J., & Brajawidagda, U. (2013). Tsunami early warnings via twitter in government: Net-savvy citizens' co-production of time-critical public information services. *Government Information Quarterly*, 30(4), 377–386. doi:10.1016/j.giq.2013.05.021

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: Situating "big data" and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2), 130–139. doi:10.1080/15230406.2013.777137

Dailey, D., & Starbird, K. (2015). It's raining dispersants: Collective sensemaking of complex information in crisis contexts. *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, 155-158. doi:10.1145/2685553.2698995

Dang, T. N., Wilkinson, L., & Anand, A. (2010). Stacking graphic elements to avoid over-plotting. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1044–1052. doi:10.1109/TVCG.2010.197 PMID:20975142

Dean, R. B., & Dixon, W. J. (1951). Simplified statistics for small numbers of observations. *Analytical Chemistry*, 23(4), 636–638. doi:10.1021/ac60052a025

Gastner, M. T., & Newman, M. E. (2004). Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America*, 101(20), 7499–7504. doi:10.1073/pnas.0400280101 PMID:15136719

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth*, *3*(3), 231–241. doi:10.1080/17538941003759255

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, *3*(3), 255–261. doi:10.1177/2043820613513121

Guan, X., & Chen, C. (2014). Using social media data to understand and assess disasters. *Natural Hazards*, 74(2), 837–850. doi:10.1007/s11069-014-1217-1

Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., & McElderry, J. A. (2015). Social media and disasters: A functional framework for social media use in disaster planning, response, and research. *Disasters*, *39*(1), 1–22. doi:10.1111/disa.12092 PMID:25243593

Huang, Q., & Xiao, Y. (2015). Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information*, 4(3), 1549–1568. doi:10.3390/ijgi4031549

Jung, J., & Uejio, C. K. (2017). Social media responses to heat waves. *International Journal of Biometeorology*, 61(7), 1247–1260. doi:10.1007/s00484-016-1302-0 PMID:28078449

Volume 12 • Issue 3 • July-September 2021

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1–8. doi:10.1109/2945.981847

Kireyev, K., Palen, L., & Anderson, K. (2009). *Applications of topics models to analysis of disasterrelated twitter data*. Paper presented at the NIPS Workshop on Applications for Topic Models: Text and Beyond.

Kongthon, A., Haruechaiyasak, C., Pailai, J., & Kongyoung, S. (2012). The role of twitter during a natural disaster: Case study of 2011 Thai flood. *Technology Management for Emerging Technologies (PICMET)*, 2012 Proceedings of PICMET'12, 2227-2232.

Kumar, S., Barbier, G., Abbasi, M. A., & Liu, H. (2011, July). TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief. ICWSM.

Landwehr, P. M., & Carley, K. M. (2014). Social media in disaster relief. In Data mining and knowledge discovery for big data. Springer.

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. *META Group Research Note,* 6(70), 1.

Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., . . . Tapia, A. H. (2015, May). Twitter Mining for Disaster Response: A Domain Adaptation Approach. ISCRAM.

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, 40(2), 61–77. doi:10.1080/15230406.2013.777139

Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., & Rodrigue, J. (2012, June). A demographic analysis of online sentiment during hurricane irene. In *Proceedings of the Second Workshop on Language in Social Media* (pp. 27-36). Association for Computational Linguistics.

Martín, Y., Cutter, S. L., Li, Z., Emrich, C. T., & Mitchell, J. T. (2020). Using geotagged tweets to track population movements to and from Puerto Rico after Hurricane Maria. *Population and Environment*, 42(1), 1–24. doi:10.1007/s11111-020-00338-6

Mihunov, V. V., Lam, N. S., Zou, L., Wang, Z., & Wang, K. (2020). Use of Twitter in disaster rescue: Lessons learned from Hurricane Harvey. *International Journal of Digital Earth*, *13*(12), 1–13. doi:10.1080/17538947 .2020.1729879

Mocanu, D., Baronchelli, A., Perra, N., Gonalves, B., Zhang, Q., & Vespignani, A. (2013). The twitter of babel: Mapping world languages through microblogging platforms. *PLoS One*, *8*(4), e61981. doi:10.1371/journal. pone.0061981 PMID:23637940

O'sullivan, D., & Unwin, D. (2014). Geographic information analysis. John Wiley & Sons.

Olteanu, A., Vieweg, S., & Castillo, C. (2015, February). What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 994-1009). ACM. doi:10.1145/2675133.2675242

Palen, L., Starbird, K., Vieweg, S., & Hughes, A. (2010). Twitter-based information distribution during the 2009 Red River Valley flood threat. *Bulletin of the Association for Information Science and Technology*, *36*(5), 13–17. doi:10.1002/bult.2010.1720360505

Poorthuis, A., & Zook, M. (2015). Small stories in big data: Gaining insights from large spatial point pattern datasets. *Cityscape (Washington, D.C.)*, 17(1), 151.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2013). GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal*, 78(1), 21–40. doi:10.1007/s10708-011-9423-9 PMID:32214617

Sahr, K., White, D., & Kimerling, A. J. (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, *30*(2), 121–134. doi:10.1559/152304003100011090

Scott, D. W. (1988). A note on choice of bivariate histogram bin shape. Journal of Official Statistics, 4(1), 47.

Singleton, A., & Arribas-Bel, D. (2019). Geographic data science. *Geographical Analysis*, gean.12194. doi:10.1111/gean.12194

Smith, B., & Varzi, A. C. (2000). Fiat and bona fide boundaries. *Philosophical and Phenomenological Research*, 401-420.

Steiger, E., Westerholt, R., Resch, B., & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with U.K. census data. *Computers, Environment and Urban Systems*, 54, 255–265. doi:10.1016/j.compenvurbsys.2015.09.007

Takahashi, B., Tandoc, E. C. Jr, & Carmichael, C. (2015). Communicating on twitter during a disaster: An analysis of tweets during typhoon Haiyan in the Philippines. *Computers in Human Behavior*, 50, 392–398. doi:10.1016/j.chb.2015.04.020

Thomee, B., Shamma, D. A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., & Li, L. (2016). YFCC100M: The new data in multimedia research. *Communications of the ACM*, *59*(2), 64–73. doi:10.1145/2812802

Twitter, Inc. (2020). *Quarterly results*. Retrieved from https://investor.twitterinc.com/financial-information/ quarterly-results/default.aspx

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 10791088. doi:10.1145/1753326.1753486

Wang, L., Wang, G., & Alexander, C. A. (2015). Big data and visualization: Methods, challenges, and technology progress. *Digital Technologies*, *1*(1), 33–38.

Wang, Y., Wang, T., Ye, X., Zhu, J., & Lee, J. (2016). Using social media for emergency response and urban sustainability: A case study of the 2012 Beijing rainstorm. *Sustainability*, 8(1), 25. doi:10.3390/su8010025

Wang, Z., Lam, N. S., Obradovich, N., & Ye, X. (2019). Are vulnerable communities digitally left behind in social responses to natural disasters? An evidence from Hurricane Sandy with Twitter data. *Applied Geography (Sevenoaks, England)*, *108*, 1–8. doi:10.1016/j.apgeog.2019.05.001

Wang, Z., & Ye, X. (2018). Social media analytics for natural disaster management. *International Journal of Geographical Information Science*, 32(1), 49–72. doi:10.1080/13658816.2017.1367003

Wang, Z., & Ye, X. (2018). Social media analytics for natural disaster management. *International Journal of Geographical Information Science*, 32(1), 49–72. doi:10.1080/13658816.2017.1367003

Wang, Z., & Ye, X. (2019). Space, time, and situational awareness in natural hazards: A case study of Hurricane Sandy with social media data. *Cartography and Geographic Information Science*, *46*(4), 334–346. doi:10.108 0/15230406.2018.1483740

Wang, Z., Ye, X., & Tsou, M. H. (2016). Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Natural Hazards*, 83(1), 523–540.

Ware, C. (2008). Visual Thinking: for Design. Morgan Kaufmann Series in Interactive Technologies.

Ye, X., & Wei, X. (2019). A Multi-Dimensional Analysis of El Niño on Twitter: Spatial, Social, Temporal, and Semantic Perspectives. *ISPRS International Journal of Geo-Information*, 8(10), 436. doi:10.3390/ijgi8100436

Yuan, Y., & Medel, M. (2016). Characterizing international travel behavior from geotagged photos: A case study of Flickr. *PloS One*, *11*(5), e0154885.

Milad Mohammadalizadehkorde is a Ph.D. student in Geographic Information Science (GIS) in the Department of Geography at Texas State University.

Shadi Maleki is a Ph.D. candidate in Geographic Information Science (GIS) in the Department of Geography at Texas State University. Her dissertation research examines the neighborhood effect on children's independent mobility and access to opportunities.