Computers & Graphics 38 (2014) 51–60

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on Visual Analytics

Public behavior response analysis in disaster events utilizing visual analytics of microblog data

Junghoon Chae [a], Dennis Thom [b], Yun Jang [c,*], SungYe Kim [d], Thomas Ertl [b], David S. Ebert [a]

[a] Purdue University, USA
[b] University of Stuttgart, Germany
[c] Sejong University, South Korea
[d] Intel Corporation, USA

ARTICLE INFO

Article history:
Received 22 June 2013
Received in revised form
7 October 2013
Accepted 8 October 2013
Available online 21 October 2013

Keywords:
Visual analytics
Social media analysis
Disaster management

ABSTRACT

Analysis of public behavior plays an important role in crisis management, disaster response, and evacuation planning. Unfortunately, collecting relevant data can be costly and finding meaningful information for analysis is challenging. A growing number of Location-based Social Network services provides time-stamped, geo-located data that opens new opportunities and solutions to a wide range of challenges. Such spatiotemporal data has substantial potential to increase situational awareness of local events and improve both planning and investigation. However, the large volume of unstructured social media data hinders exploration and examination. To analyze such social media data, our system provides the analysts with an interactive visual spatiotemporal analysis and spatial decision support environment that assists in evacuation planning and disaster management. We demonstrate how to improve investigation by analyzing the extracted public behavior responses from social media before, during and after natural disasters, such as hurricanes and tornadoes.

1. Introduction

For emergency and disaster management, analysis of public behavior, such as how people prepare and respond to disasters, is important for evacuation planning. As social media has played a pervasive role in the way people think, act, and react to the world (more than 40 million Americans use social media web sites multiple times a day [1]), social media is changing the way people communicate not only in their daily lives, but also during abnormal events, such as natural disasters. In emergency situations, people even seek social confirmation before acting in response to a situation, where they interact with others to confirm information and develop a better informed view of the risk [2]. A study commissioned by the American Red Cross, found that roughly half of the respondents would mention emergencies and events on their social media channels, and more than two-thirds agree that response agencies should regularly monitor postings on their websites [3]. Moreover, a growing number of people are using Location-based Social Network services, such as microblogs, where they create time-stamped, geo-located data and share this information about their immediate surroundings using smart phones with GPS. Such spatiotemporal data has great potential for enhancing situational awareness during crisis situations and providing insight into the evolving event, the public response, and potential courses of action.

For public behavior analysis in disasters, however, finding meaningful information from social media is challenging. It is almost impossible to perform a straightforward qualitative analysis of the data, since the volume of the data exceeds the boundaries of human evaluation capabilities and normal computing performance. Even though we could extract certain information from the data, it is not always easy to determine whether the analysis result of the extracted information is meaningful and helpful. Thus, there is a need for advanced tools to handle such big data and aid in examining the results in order to understand situations and glean investigative insights. Given the incomplete, complex, context-dependent information, a human in this analysis and decision-making loop is crucial. Therefore, a visual analytics approach offers great potential through interactive, scalable, and verifiable techniques, helping analysts to extract, isolate, and examine the results interactively. In this paper, we present an interactive visual analytics approach for spatiotemporal microblog data analysis to improve emergency management, disaster preparedness, and evacuation planning. We demonstrate the ability to identify spatiotemporal differences in patterns between emergency and normal situations, and analyze spatial relationships among spatial distributions of microblog users, locations of multiple types of infrastructure, and severe weather conditions.

* Corresponding author. Tel.: +82 1025483148.
E-mail address: jangy@sejong.edu (Y. Jang).

0097-8493/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2013.10.008

Furthermore, we show how both spatiotemporal microblog and disaster event data can help the analysts to understand and examine emergent situations, and evaluate courses of action.

This study is performed using Twitter messages called Tweets, as Twitter has been the most popular microblog service in the United States. In this paper, we extend our previous work [4] with additional features of our system and examine their capabilities with several expanded examples in Section 4.2. We also add a discussion section for comparisons and analysis of the case studies.

Our system evaluates visual analytics of spatiotemporal distribution of Tweets to identify public behavior patterns during natural disasters. The main features of our approach are as follows:

- *Spatial analysis and decision support*: The system provides effective analysis for exploring and examining the spatial distribution of Twitter users and supporting spatial decision-making using a large volume of geo-located Tweets and multiple types of supplementary information during specific time periods (i.e., disaster events).
- *Temporal pattern analysis*: Our visualization system enables the analysts to analyze the temporal distribution of the number of Twitter users posting Tweets in a given location and time.
- *Spatiotemporal visualization*: We provide a visualization that allows the analysts to simultaneously analyze both aspects: space and time in a single view.

We first review previous work in Section 2 and describe our interactive analysis system in Section 3. We present analysis results for two natural disaster events in Section 4 and discussion in Section 5. Finally conclusion and future directions are presented in Section 6.

2. Related work

In recent years social media data has become a popular topic in a range of application domains. Several researchers have proposed and presented systems for social media analysis and important studies covering the use of social media during crisis events have been conducted. Most recent analysis environments for crisis-related social media exploration and visualization are from MacEachren et al. [5], Marcus et al. [6], and Thom et al. [7]. Their systems combine traditional spatial and geographic visualizations with means for automated location discovery, trend and outlier search, anomaly and event discovery, large scale text aggregation and highly interactive geovisual exploration. Approaches putting less focus on visualizations and more on fully automated data mining mechanisms have been proposed by Sakaki et al. [8] that use Kalman and Particle Filters to detect the location of earthquakes and typhoons based on Twitter. Various techniques for spatiotemporal data analysis and anomaly detection using visualization or machine learning techniques have been proposed by Andrienko et al. [9], Lee and Sumiya [10], and Pozdnoukhov and Kaiser [11]. Twitcident from Abel et al. [12] provides a web-based framework to search and filter crisis-related Tweets. Using the Netherlands emergency broadcast system, Twitcident automatically reacts on reported incidents and collects related information from Twitter based on semantic enrichment. In all these system the focus is primarily on individual messages and aggregated message volumes and how insight can be generated by understanding their content. In contrast, our system investigates a more user focused approach that tries to identify the whereabouts and movements of people in order to understand mass behavior.

Researchers have also examined the usage of Twitter during incidents and disasters. Terpstra et al. [13] investigate more than 90k Twitter messages that were sent during and after a storm hit the Belgium *Pukkelpop* musicfestival in 2011. They categorize Tweets into warnings about the severe weather conditions, rumors and self organization of relief measures. They show that valuable information for crisis response and decision support can be gathered from the messages. Vieweg et al. [14] investigate the differences in reaction to different crisis events. For their study they investigate eyewitness reports in Twitter from people that were affected by Oklahoma Grassfires in April 2009 and Red River Floods in March and April 2009. Their research also demonstrates the high value that the extraction of meaningful comments from crisis-related communication can have to generate insights. Furthermore, Heverin et al. [15] demonstrate that Twitter can also be a useful source of information for smaller events as they investigate the reaction to a shooting of four police officers and the subsequent search for the suspect that took place in the Seattle-Tacoma area. Based on the collection and categorization of 6000 messages they are able to show that citizens use the service to communicate and seek information related to the incident.

In this paper we also present a case study on crisis-related information gathered from Twitter data. However, in contrast to the discussed studies that harvest information directly out of the content of the messages, our method is primarily based on observing movement patterns and identifying local hotspots in order to learn about the effects of the crisis and the performance of evacuation measures.

3. Problem statement and interactive analysis process design

Analysis of public behavior, such as how people prepare and respond to disasters, plays an important role in crisis management, disaster response, and evacuation planning. Recently, social media becomes popular and people utilize it for communications not only in their daily lives, but also in abnormal disastrous situations. Thus, Location-based Social Networks services offer a new opportunity for enhancing situational awareness during disaster events. Unfortunately, collecting relevant data can be costly and finding meaningful information from the huge volume of social media data is very challenging. Therefore, there is a need for an advanced tool to analyze such massive ("big") streaming data and aid in examining the analysis results to better understand situations more efficiently.

Our proposed visual analytics approach provides multiple analysis methods: spatial analysis, spatial decision support, temporal pattern analysis, abnormal topic analysis, and interactive spatiotemporal visualization as shown in Fig. 2. In our system, all methods are tightly integrated based on a user-centered design in order to enhance the ability to analyze huge social media data (Fig. 2 (A–C)). Our Tweet collection component obtains real-time Tweets using the Twitter API—to collect about 2.2 million geo-tagged Tweets within the United States per day. In general for spatial analysis, the required accuracy of the geocoordinate depends upon the required level of location granularity. The data, however, is generated by very reliable GPS and software. We can be reasonably certain about the data accuracy as illustrated in [16]. For the temporal accuracy of Tweets, we use the time when each Tweet is created. Therefore, it is highly accurate if the time setting of the device posting a Tweet is correct. This large volume of data is stored in our database in order to maintain and track the history of the Twitter stream. Our system allows the analysts to query Tweets with a specific area and time span condition (Fig. 2(A)). The initially selected spatiotemporal context of Tweets can be represented by two different analytics components: spatial analysis and spatiotemporal visualization. Spatial analysis allows the analysts to examine the overall distribution of Twitter users and discover hotspots where relatively more Twitter users post Tweets. The analysts are able to add supplementary information (infrastructure locations, tornado paths) on top of current information representing outcomes in order to better understand events and increase situational awareness (Fig. 2(B)). Furthermore, the analysts can select a sub-region within the initial area, so that he can analyze the temporal patterns of the number of Twitter users and extract abnormal topics from the text messages in the selected region (Fig. 2 (C)). In addition, our interactive spatiotemporal visual analytics provides a single view representation for the analysis of both aspects: spatial and temporal characteristics of Tweets at the same time.

4. Spatiotemporal analysis

In this work, we present a visual analytics approach to handle the vast amount of microblog data such as Twitter messages, provide interactive spatiotemporal analysis, and enable the use of multiple types of supplementary spatial infrastructure information for spatial decision support. Analysts select an initial spatiotemporal context of Tweets to be represented in the visualization to serve as a basis for analysis. They can also perform the interactive spatiotemporal queries that load the relevant datasets from a larger database.

4.1. Spatial analysis

As mentioned in Section 1, social media embedding geolocation information into the data is extremely useful in analyzing location-based public behaviors. Such spatial analysis, therefore, is important in order to manage and prepare plans for disaster and emergency situations.

In late October in 2012, a massive hurricane, Sandy, devastated Northeastern United States [17]. Due to the severeness of the hurricane, on October 28th in 2012, the New York City Authorities ordered residents to leave some low-lying areas—the mandatory evacuation zones (red color) are shown in Fig. 9 (Right). We investigate an area of Manhattan, since the area is the most populated and severely damaged. Through the map view in our system, analysts navigate to the Manhattan area in New York City and filter Tweets posted within the area. Initially we tried to reveal public movement flows during the disaster event, but the movement patterns were too complicated to find meaningful flows due to movement randomness and the visual clutter of the flows. Then, we examined the spatial distribution of the users for specific time frames. Based on our experiments, a geospatial heatmap was useful for an overview of the spatial distribution and for trend approximation. We utilize a divergent color scheme to generate the heatmap, where saturated colors are used for the data distribution to avoid any confusion from the color scheme from the desaturated colormap of the background map. Analysts can specify a threshold range to emphasize hotspots, where the upper bound is mapped to a red color and the lower bound to a yellow color. Additionally, the blue color is mapped by the analysts to the value of the overall distribution of Twitter users. In Fig. 1, we show three heatmaps of spatial user-based Tweet distribution from 12:00 PM to 4:00 PM on October 14th (Left), 21st (Center), and 28th (Right). In this work, we use the number of Twitter users instead of the number of Tweets for the heatmap generations to properly reflect the flow of evacuation unbiased by personal Tweet activity or behavior of individual users, since some enthusiastic Twitter users generate a large number of Tweets at the same location during a short time period (more than 20 Tweets per hour). The heatmaps in Fig. 1 (Left and Center) represent normal situations of Twitter user distribution in the Manhattan area, and the heatmap (Right) shows the situation right after the evacuation order that was announced at 10:30 AM on October 28th, 2012. This standard heatmap visualization allows analysts to explore the spatial pattern of Twitter users for any specified time period. In Section 4.2, we will provide further analysis for the spatial decision support.

Hurricane Sandy damaged not only New York City, but also the entire eastern coast area of New Jersey. Most cities in the area also announced evacuation orders on October 28th, 2012. The distribution of Twitter users in the area from Atlantic City to the upper eastern shore area for two different dates are shown in Fig. 3. The heatmaps in Fig. 3 (Left) represent the previous normal situation of Twitter user distribution on October 24th and the heatmap (Right) shows the post distribution after Sandy passed over the area on October 31st. As shown in the result, many hotspots are gone or diminished. This situation shows that the number of Twitter users had significantly decreased after the hurricane damaged the area. In fact, a huge number of homes were damaged or destroyed and a couple of million households lost power because of Hurricane Sandy [18]. In disaster management this type of visualization can support analysts estimating which areas were highly damaged and even which areas still need reconstruction.

4.2. Spatial decision support

In Section 4.1, we introduced our spatial analysis to explore the Twitter user distribution. In addition to the analysis, our system allows the analysts to utilize supplementary information in order to support understanding of the situations and decision-making in disaster management. The spatial characteristics together with heterogeneous information can assist in disaster management and migrating hazards where the problems have spatial components [19]. The supplementary information can be various types of infrastructures (i.e., school, park, supermarket, and shelter), as well as spatial information of disaster events (i.e., hurricane path and damage area of a tornado). In this section, we describe how our system supports spatial decision-making by correlating such spatial information with location-based microblog data.

4.2.1. Infrastructure data

During a natural disaster event, such as Hurricane Sandy, analysts would assume that many people might want to go to the supermarket before staying or evacuating, but they would need supporting evidence before making appropriate decisions and plans. With our system support, the analysts can simply overlay the locations of large supermarkets on the heatmap of the Twitter user distribution. The infrastructure locations are indicated by standard symbols [20] as shown on the right side of Fig. 1. A relatively large number of people immediately went to supermarkets nearby the evacuation area, instead of the emergency shelter as shown in Fig. 1 (Right). However, October 28th was Sunday and many people generally would go for grocery shopping on Saturday or Sunday; therefore, the analysts might need to verify whether the heatmap shown in the figure is a normal periodic situation. The analysts can investigate new Twitter user distributions for different time frames by simply manipulating the time context. In Fig. 1 (Left and Center), we show two distributions for one and two weeks before the disaster period respectively. Here, we see that the hotspot locations are very different from the ones for October 28th shown in Fig. 1 (Right). For further analysis, we can explore another popular Sunday location—large parks—by superimposing the locations on each heatmap. As shown in Fig. 1 (Left and Center), many hotspots

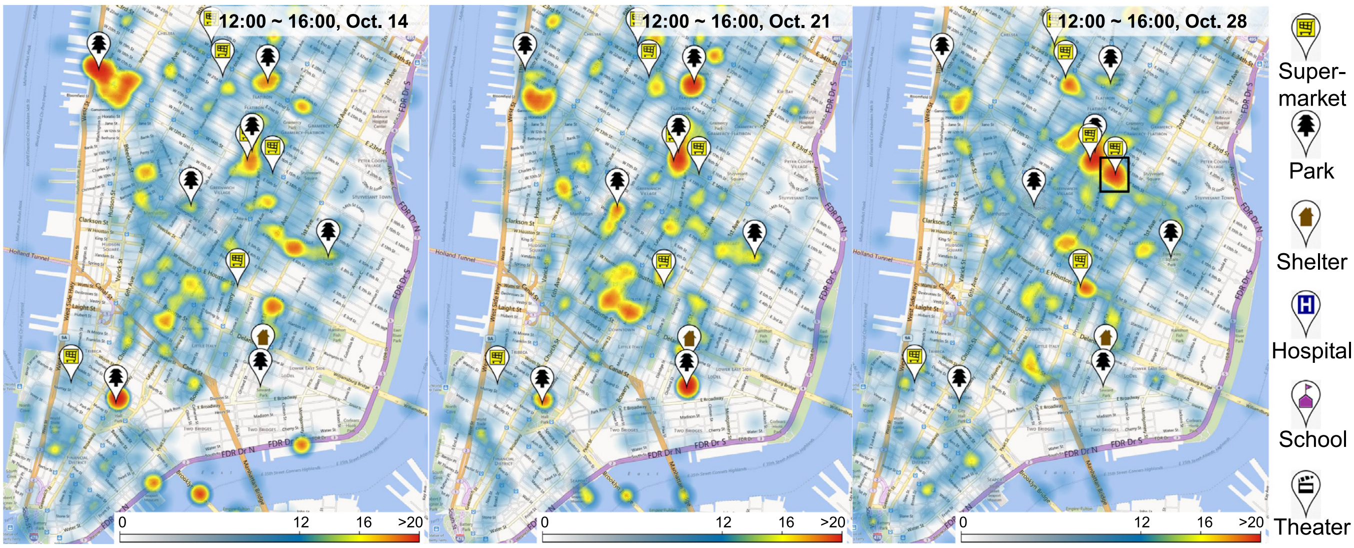

Fig. 1. Spatial user-based Tweet distribution in the Manhattan area in New York City during 4 h right after the evacuation order (from 12:00 PM to 4:00 PM on October 28th, 2012 (Right)). Previous distribution of Tweets on 14th (Left) and 21st (Center).

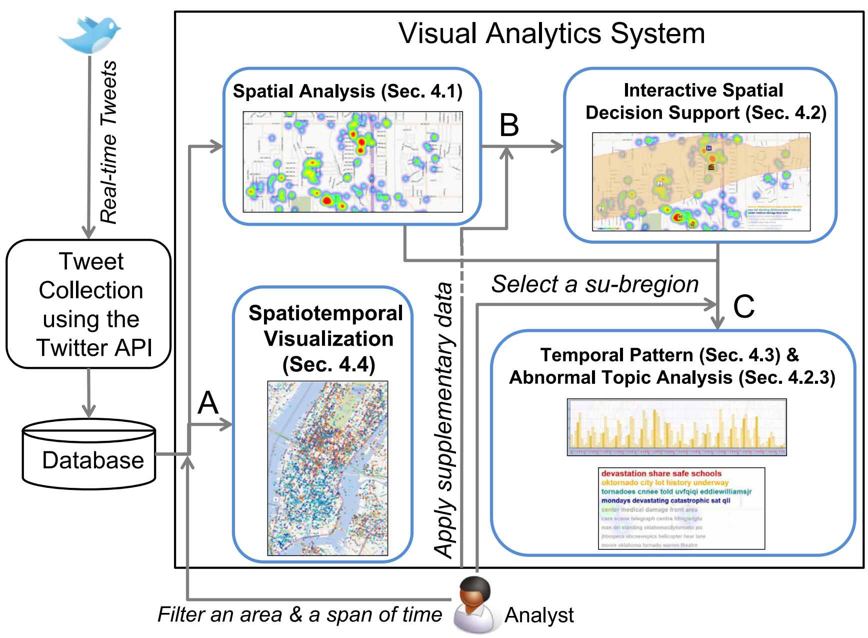

Fig. 2. Overview of our interactive analysis scheme for public behavior analysis using social media data.

overlap with the park areas in normal situations. Therefore, we can conclude that the situation on October 28th is an unusual non-periodic pattern.

4.2.2. Disaster event data

In Section 4.2.1, we explained how the infrastructure data help the analysts to understand and examine the emergent situations. During severe weather conditions, people tend to be sensitive to the dynamic variance of the weather conditions. Relationship analysis, therefore, between the public responses and the spatio-temporal pattern of the severe weather is important. Our system overlays geographic information of disaster events, for example, center positions and tracks of a hurricane, and damaged areas by a tornado, in order to provide further analysis. Two case studies are presented as follows:

Track of hurricane: Fig. 4(1) and (2) shows the southeastern coast areas of the United States, whereas, Fig. 4 (3), (4), and (5) shows the northeastern coast areas. In the figures the distributions of Twitter users for each consecutive date, from October 26th to 30th, 2012, are presented using the heatmap visualizations. We use the number of Twitter users who posted Twitter messages containing one of the following keywords: *hurricane*, *storm*, and *sandy* in order to analyze Tweets that are highly related to Hurricane Sandy. Note that Hurricane Sandy reached the south-eastern Florida coast on October 26th and passed, then, over the northeastern coast on October 30th, 2012 [17]. As shown in Fig. 4, our system is able to overlay the track of the hurricane on the map. The blue pins and the blue lines represent the center locations of the hurricane and its path respectively.

Twitter users also actively respond to the severe weather conditions. In Fig. 4, we indicate that the distribution pattern of Twitter users had dynamically varied along the track of the hurricane center locations. When Sandy moved to the south-eastern coast on October 26th, there were bursts on eastern Florida's coast (Fig. 4(1)). Next day, the bursts disappeared,

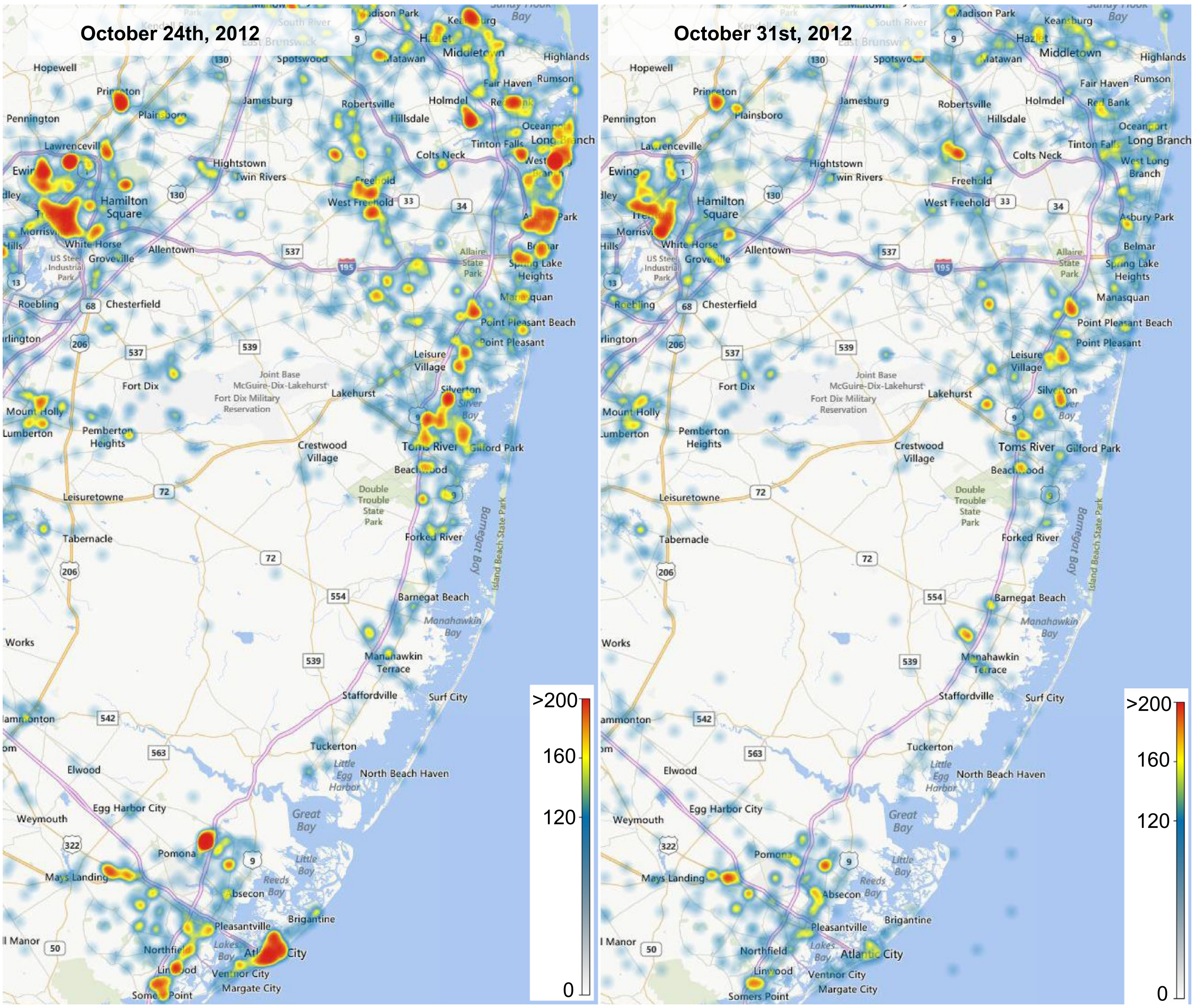

Fig. 3. Twitter user distribution on the eastern coast area in New Jersey, after the hurricane passed over the area on October 31st (Right). Previous distribution on October 24th is shown on the Left.

because Sandy moved toward the northeast away from the east coast of United States (Fig. 4(2)). Sandy kept moving toward a few hundred miles southeast of North Carolina on October 28th (Fig. 4(3)). In the next day, the hurricane's track bent toward the north and the hurricane made landfall at night in the northeast of Atlantic City (Fig. 4(4)). Throughout the days, Twitter users were actively reacting to Hurricane Sandy' arrival in a wide range of areas. After the landfall, the storm turned toward the northwest and was gradually weakened. The big outbreaks were diminished on October 30th as shown in Fig. 4(5). As shown in the figures, we can see how Twitter users reacted according to the spatiotemporal pattern of the severe weather conditions in the social media domain.

Damage area from a Tornado: An extremely strong Tornado passed through the city of Moore in southern metropolitan Oklahoma City [21] in the afternoon on May 20th, 2013. The larger than one-mile-wide tornado damaged the city with a wind speed of more than 200 mph. Fig. 5 shows the damaged part of the city. The tornado entered the area at about 3:16 PM and exited the area after about 10 min. We visualize the distribution of Twitter users on the map during 24 h, from May 20th 4:00 PM to 21st 4:00 PM. We also overlay an approximate extent of tornado damage (transparent orange color) and locations of multiple infrastructures, such as schools, hospitals, and supermarkets, on the map view. Since the tornado suddenly happened and disappeared, we were not able to find significantly abnormal patterns before and during the event. After the disaster event, however, many Twitter users moved toward some specific areas: two elementary schools, a medical center, a theater, and two large supermarkets. The two elementary schools, the medical center, and the theater were located within the highly damaged area and they were severely destroyed. Also many people were hurt and died in these infrastructures. The increased number of Twitter users was probably due to the fact that many people went to these places in order to rescue the victims [22]. Moreover, people might have gone to supermarkets to obtain indispensable things. In Fig. 5(1), the heatmap shows a normal situation of Twitter user distribution in the same area. The distribution is very different from the situation after the tornado hit the area. This example demonstrates how our visual analytics system enables the analysts to analyze public responses using spatial disaster data and infrastructure data for disaster management.

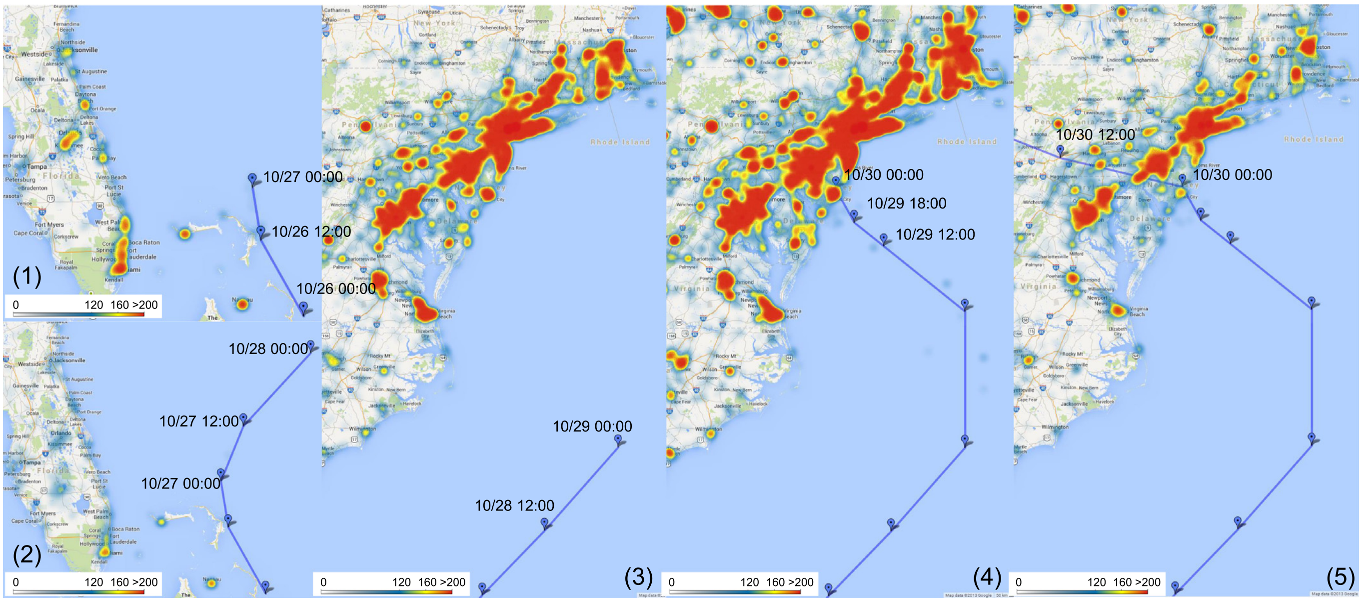

Fig. 4. Distribution of Twitter users of each consecutive date (October 26–30, 2012), who post hurricane related Tweets on the southeastern (1 and 2) and northeastern coast (3, 4, and 5) area of the United States. We can see the variance of Twitter user reactions along the track of the hurricane center locations.

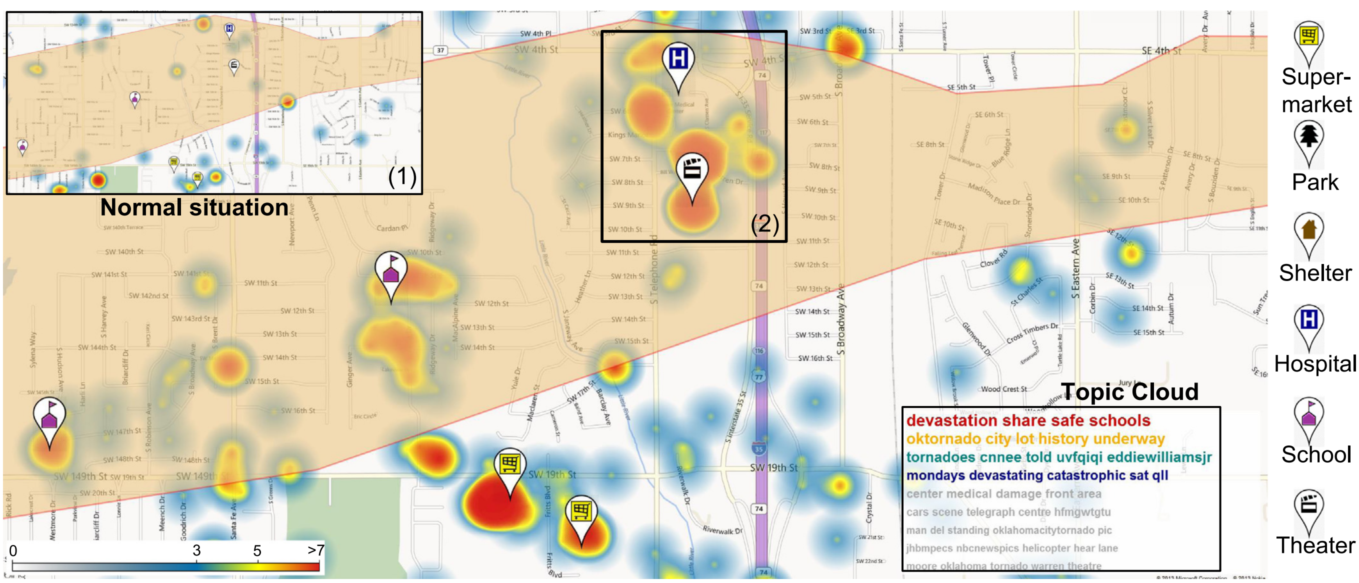

Fig. 5. Spatial pattern of Twitter users during 24 h in the city of Moore after damages from a strong tornado. Relatively many people moved to severely damaged areas after the disaster. This situation is much different from the previous normal situation (1). We selected a specific region (2) that includes severely damaged areas in order to extract topics (3) from Tweets within the selected area.

4.2.3. Abnormal topic analysis

Our system also provides analysts with abnormal topic examination within the microblog data. Each Twitter message provides not only spatiotemporal properties, but also textual contents. The text messages are also important to understand and examine the emergent situations. Our system allows the analysts to extract major topics from many Tweets posted within a specific area using Latent Dirichlet Allocation (LDA) [23]. We also employ, then, a Seasonal-Trend Decomposition procedure based on Loess smoothing (STL) [24] to identify unusual topics within the selected area. For each extracted topic of the LDA topic modeling, our algorithm retrieves messages associated with the topic and then generates a time series consisting of daily message counts from their timestamps. The time series can be considered as the sum of three components: a trend component, a seasonal component, and a remainder. Under normal conditions, the remainder will be identically distributed Gaussian white noise, while a large value of the remainder indicates substantial variation in the time series. Thus, we can utilize the remainder values to implement control chart methods detecting anomalous outliers within the topic time series. We have chosen to utilize a seven day moving average of the remainder values to calculate the z-scores. Note that we use the z-score as the abnormality score in this work. If the z-score is higher than 2, events can be considered as abnormal within a 95% confidence interval. The details of these techniques are described in the previous work [25]. We select a sub area in Fig. 5(2) that includes severely damaged areas: the selected region (black rectangle) on the map. The extracted topics, which are ordered based on their abnormalities, are displayed as Topic Clouds at the bottom-right corner (Fig. 5(3)) on the map. The topic cloud is enlarged and shown in Fig. 6. In this case study, most topics are related to the disaster event. However, the last topic—*moore*, *oklahoma*, *tornado*, *warren*, *theatre*, has a relatively low abnormality although they seem related to the disaster event, because tornadoes frequently occur in the area. Fig. 7 shows an abnormality graph for the first topic in Fig. 6. The abnormality score for the topic had significantly increased when the tornado hit the region on May 20th (Marked region). As shown in Fig. 7, the abnormality score (6.75) is much higher than the average abnormality score(0.42); therefore, the analysis of the microblog data provides a statistically significant difference during this severe weather condition.

4.3. Temporal pattern analysis

In the previous sections, we presented the spatial analysis of social media and spatial decision support. In this section, we

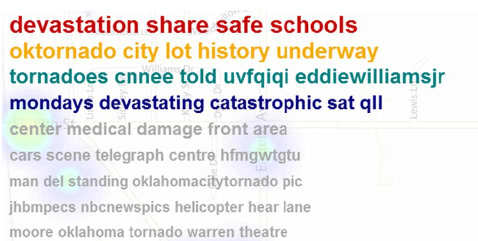

Fig. 6. Topic cloud: Topics from Tweets within the selected area in Fig. 5 (2) are ordered by their abnormality scores.

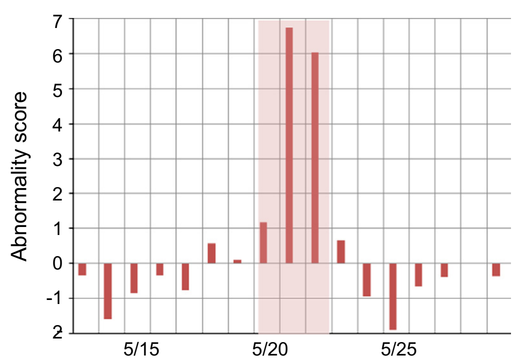

Fig. 7. Abnormality of the first topic in Fig. 6. The abnormality score of the topic had significantly increased when the tornado hit the region on May 20th (Marked region).

demonstrate analysis of the relationships between the temporal patterns of the number of Twitter users and certain public situational behaviors: how many people go where and how different is it from the previous situations? Analysis of temporal trends and relationships between data values across space and time provides underlying insights and improves situational awareness [26,27].

After selecting the initial spatiotemporal context of Tweets as a basis for the analysis, the analysts can explore the temporal patterns of the number of Twitter users who posted Tweets within the spatial boundary using the bar chart as shown in Fig. 8. The values of each bar are the number of users in 4 h intervals and represent data two weeks before and after the selected date. Once a mouse cursor hovers over one of the bars in the graph, every bar that corresponds to that time period, is highlighted in dark yellow color as shown in Fig. 8. As previously mentioned, the heatmap in the figure shows the Twitter user density distribution from 12:00 PM to 4:00 PM on October 28th, right after the announcement of the evacuation order. We select a hotspot that includes one of the supermarket locations: the selected region (black rectangle) on the map in Fig. 1 (Right). We can indicate that the number of Twitter users (red rectangle in Fig. 8) in the corresponding time period is higher than for the same time period from other dates (October 14th, 21st and November 4th, 5th) by 35% more from the average. Moreover, there is another interesting finding—the number of people during each of the following time frame (4:00–8:00 PM) on the dates from the previous weeks are higher than the number of people in the selected time frame. This is because many shoppers were lining up at stores and emptied the shelves to prepare for Hurricane Sandy. Some actual Twitter messages posted in the area are following: "*The line at Trader Joes is unbelievable*…" and "*There is amazing line here*…". Furthermore, since October 29th, the number of people has significantly decreased because most residents left the area before the arrival of the hurricane. The increase in the number of people after one week reflects that some people came back to the area.

4.4. Spatiotemporal visualization

There is abundant research published on the topic of spatiotemporal data visualization. Still, exploration of time-referenced geographic data is still a challenging issue [28]. We introduce a modest visualization that enables analysts to analyze both aspects: space and time in a single view. Each Tweet is independent and

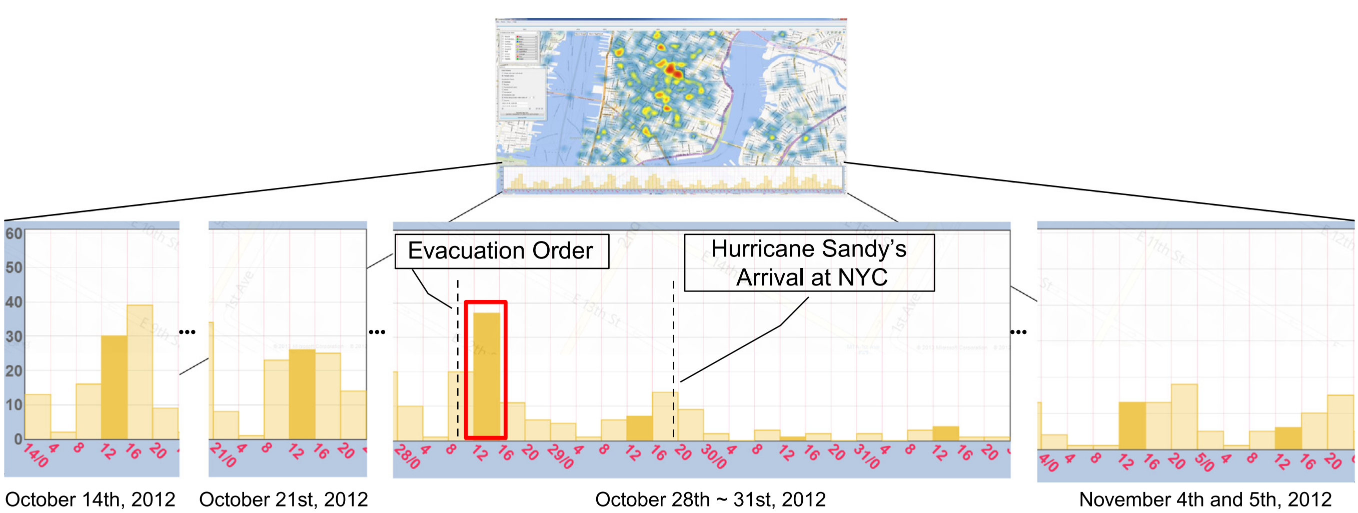

Fig. 8. Temporal analysis for public behaviors during the disaster event, Sandy. Top shows our entire system view. The bar chart (Bottom) for the number of Twitter users within the selected region including a supermarket in Fig. 1 (Right) in 4 h intervals is shown. We see that many people went to the supermarket right after the evacuation order.

contains multiple properties, such as location, time, the number of re-Tweet, etc. In this study, therefore, we utilize a glyph-based visualization to depict both location and time aspects of the independent data record using two visual features. As shown in Fig. 9 (Left), each hexagon corresponding to a Tweet represents the spatial and temporal information where the center of each hexagon is the location of each Tweet and the color represents its posting time. In other words, space and time properties are encoded in a single visualization to harness the spatial analysis features of human visual perception [29]. In Fig. 9 (Left), the hexagons with blue (12 PM–6 PM) or green (6 PM–12 AM)) color correspond to Tweets published on October 29th, 2012 and ones with orange or red color correspond to Tweets posted on the following day after the hurricane. New York City announced the evacuation of Zone A (red color) in Fig. 9 (Right); residents in Zone A faced the highest risk of flooding, whereas, Zone B (yellow color) and Zone C (green color) are moderate and low respectively. In the visual representation, analysts can indicate overall spatiotemporal patterns of people and their movements during the disaster event —many people still remained at home one day after the mandatory evacuation order, but most people left home on the following day as the hurricane damaged the city.

5. Discussion and evaluation

In this work we found out that the public responses to disaster events in social media streams are different according to the disaster event types. Hurricane Sandy had a long time duration—more than one week, and affected a wide range of areas. Therefore, there were many reactions in the potential damage area before the hurricane impacted the area. However, no or significantly less hotspots were found right after the hurricane passed over the area. This was because the hurricane severely affected the areas—communication facility damage and power outages occurred in the area. Moreover, we found out that unusual post-event situations in the Twitter user distribution continued for a certain time period from a couple of days to more than one week as shown in Figs. 4 and 8. The analysts could estimate how long it took for the reconstructions in the areas.

Regarding the tornado case, we intended to find abnormal patterns in the Twitter user distribution before and during the disaster event but there was no unusual patterns in the area. In contrast to the hurricane, the tornado generally affected the areas relatively shortly, for example, a few minutes to an hour. The abrupt natural disaster did not strongly influence the social media

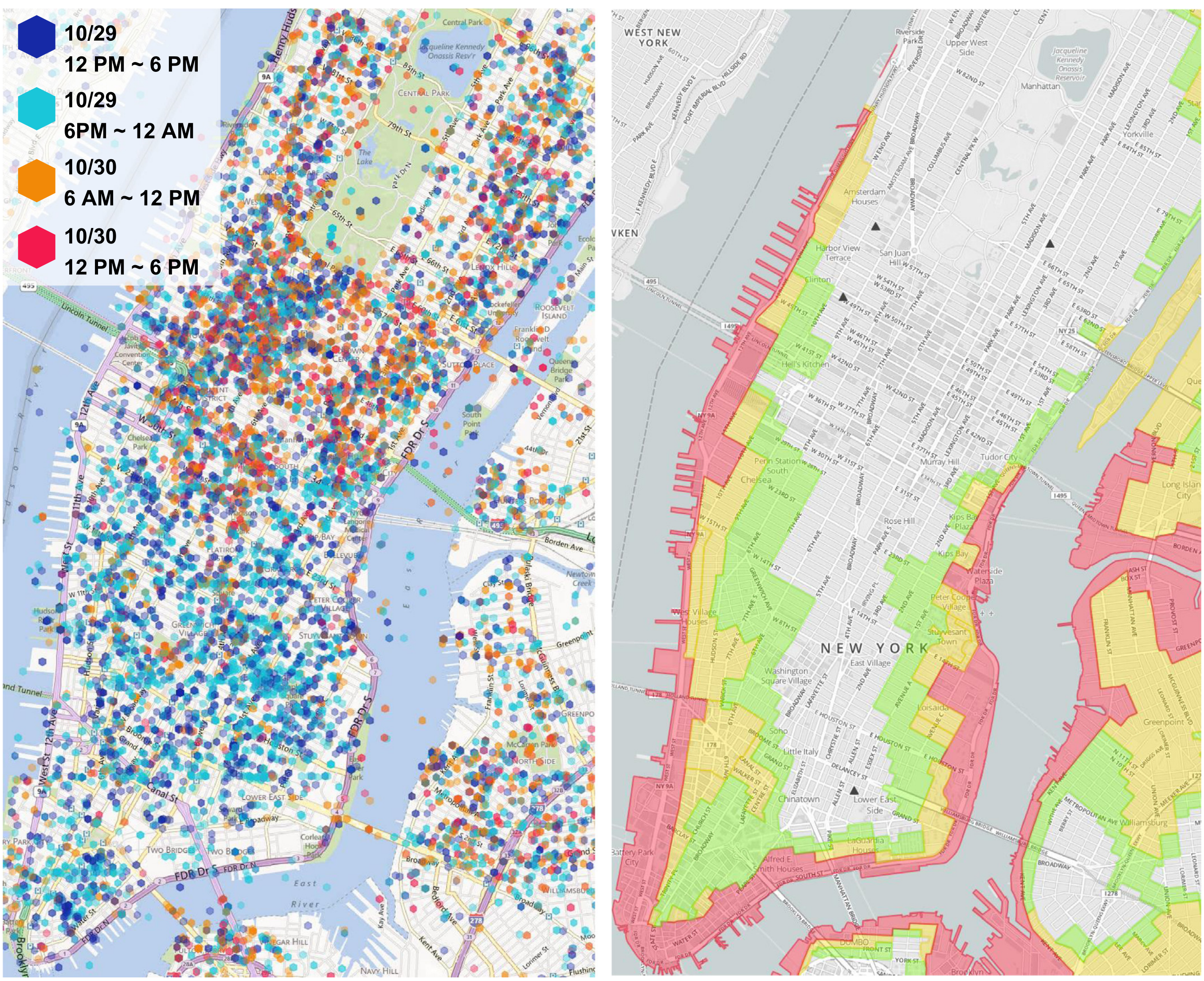

Fig. 9. Visualization for spatiotemporal social media data (Left). A hexagon represents the spatial (position) and temporal (color) information of a Tweet. Hurricane evacuation map [30] (Right). Residents in Zone A (red) faced the highest risk of flooding, Zone B (yellow) and Zone C (green) are moderate and low respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

stream before and even during the event. However, as shown in Fig. 5, we were able to find many hotspots within the damaged areas after the tornado passed. In fact, the tornado damaged some small areas (i.e., a couple of miles wide), in contrast to the wide range of damaged areas for the hurricane case. This indicated that communication facilities were still available and many people were interested in the disaster, similar to the hurricane. Thus, our social media analysis could support the analysts to make plans and manage for the emergent situations according to the types of the disasters.

The above cases demonstrate how our system supports spatial decision making through evaluation of varying-density population area to determine changes in behavior, movement, and increase overall situational assessment. This increased spatial activity and behavioral understanding provides rapid situational assessment and provides insight into evolving situational needs to provide appropriate resource allocation and other courses of action (e.g., traffic rerouting, crowd control).

We requested informal feedback for the usability of our system from users within our universities, and received useful and positive comments and suggestions. They were interested in the findings of the abnormal situations during the disaster events in Sections 4.1 and 4.3. They also noted that the use of the infrastructure symbols on the heatmaps improved the legibility of the Twitter user distributions in Fig. 1 and they suggested a visualization for the deviations between multiple heatmaps in order to show the differences clearly, which we plan to develop in the future.

6. Conclusions and future work

In this work we presented a visual analytics system for public behavior analysis and response planning in disaster events using social media data. We proposed multiple visualizations of spatiotemporal analysis for disaster management and evacuation planning. For the spatial decision support, we demonstrated an analytical scheme by combining multiple spatial data sources. Our temporal analysis enables analysts to verify and examine abnormal situations. Moreover, we demonstrated an integrated visualization that allows spatial and temporal aspects within a single view. We have still some limitations with these techniques including the potential occlusion issues in the spatiotemporal visualization. For future work, we will investigate the flow of public movement before and after disasters and the analysis for recovering from disasters and crises. We also plan to design the glyphs with varied sizes adapting to the zoom level in the spatiotemporal visualization. In addition, we will conduct a user evaluation for the usability and effectiveness of the geospatial visual support, and the impact of interactive spatiotemporal visual analytics using social media data on disaster management.

Acknowledgment

This work was partially funded by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0003. Jang's work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A1011170). We would like to thank the reviewers for their valuable suggestions and comments, which helped to improve the presentation of this work.

References

[1] Webster T. The social habit – frequent social networkers, Retrieved March 3, 2013. ⟨http://www.edisonresearch.com/home/archives/2010/06/the_social_habit_frequent_social_networkers_in_america.php⟩; 2010.
[2] Committee on Public Response to Alerts and Warnings Using Social Media: Current Knowledge and Research Gaps; Computer on Science and Technology Board; Division on Engineering and Physical Sciences; National Research Council. Public Response to Alerts and Warnings Using Social Media: Report of a Workshop on Current Knowledge and Research Gaps. The National Academies Press; 2013.
[3] American Red Cross. Social media in disasters and emergencies. Retrieved March 3, 2013, ⟨http://i.dell.com/sites/content/shared-content/campaigns/en/Documents/red-cross-survey-social-media-in-disasters-aug-2010.pdf⟩; 2010.
[4] Chae J, Thom D, Jang Y, Kim SY, Ertl T, Ebert D. Visual analytics of microblog data for public behavior analysis in disaster events. In: EuroVis workshop on visual analytics; 2013. p. 67–71.
[5] MacEachren A, Jaiswal A, Robinson A, Pezanowski S, Savelyev A, Mitra P, et al. Senseplace2: geotwitter analytics support for situational awareness. In: 2011 IEEE conference on visual analytics science and technology (VAST); 2011. p. 181–90.
[6] Marcus A, Bernstein MS, Badar O, Karger DR, Madden S, Miller RC. Twitinfo: aggregating and visualizing microblogs for event exploration. In: Proceedings of the 2011 annual conference on human factors in computing systems. ACM; 2011. p. 227–36.
[7] Thom D, Bosch H, Koch S, Woerner M, Ertl T. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In: IEEE pacific visualization symposium (PacificVis); 2012. p. 41–8.
[8] Sakaki T, Okazaki M, Matsuo Y. Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. WWW '10. ACM; 2010. p. 851–60.
[9] Andrienko N, Andrienko G, Gatalsky P. Exploring changes in census time series with interactive dynamic maps and graphics. Comput Stat 2001;16(3):417–33.
[10] Lee R, Sumiya K. Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In: Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks. LBSN '10. ACM; 2010. p. 1–10.
[11] Pozdnoukhov A, Kaiser C. Space-time dynamics of topics in streaming text. In: Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks. LBSN '11. New York, NY, USA: ACM; 2011. p. 1–8.
[12] Abel F, Hauff C, Houben GJ, Stronkman R, Tao K. Semantics+ filtering+ search= twitcident. exploring information in social web streams. In: Proceedings of the 23rd ACM conference on hypertext and social media. ACM; 2012. p. 285–94.
[13] Terpstra T, de Vries A, Stronkman R, Paradies G. Towards a realtime twitter analysis during crises for operational crisis management. In: Proceedings of the 9th international ISCRAM conference; 2012.
[14] Vieweg S, Hughes AL, Starbird K, Palen L. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: Proceedings of the 28th international conference on Human factors in computing systems. ACM; 2010. p. 1079–88.
[15] Heverin T, Zach L. Microblogging for crisis communication: examination of twitter use in response to a 2009 violent crisis in the seattle-tacoma, Washington area. In: Proceedings of the 7th international ISCRAM conference–Seattle, vol. 1; 2010.
[16] Twitter. The geography of tweets. Retrieved August 24, 2013, ⟨https://blog.twitter.com/2013/geography-tweets-3⟩; 2013.
[17] Wikipedia. Hurricane sany. Retrieved April 30, 2013, ⟨http://en.wikipedia.org/wiki/Hurricane_Sandy⟩; 2012.
[18] Wikipedia. Effects of hurricane Sandy in New Jersey. Retrieved June 11, 2013, ⟨http://en.wikipedia.org/wiki/Effects_of_Hurricane_Sandy_in_New_Jersey⟩; 2012.
[19] Andrienko G, Andrienko N, Jankowski P, Keim D, Kraak M, MacEachren A, et al. Geovisual analytics for spatial decision support: setting the research agenda. Int J Geogr Inf Sci 2007;21(8):839–57.
[20] Robinson A, Roth R, Blanford J, Pezanowski S, MacEachren A. Developing map symbol standards through an iterative collaboration process. Environ Plan B: Plan Des 2012;39(6):1034–48.
[21] Wikipedia. 2013 moore tornado. Retrieved June 12, 2013, ⟨http://en.wikipedia.org/wiki/2013_Moore_tornado⟩; 2013.
[22] Twitchy. Convoy of hope: Glenn beck arrives in okla. With two truckloads of food, water and diapers. Retrieved August 26, 2013, ⟨http://twitchy.com/2013/05/21/convoy-of-hope-glenn-beck-arrives-in-okla-with-two-truckloads-of-food-water-and-diapers/⟩; 2013.
[23] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993–1022.
[24] Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: a seasonal-trend decomposition procedure based on loess. J Off Stat 1990;6(1):3–73 (with discussion).
[25] Chae J, Thom D, Bosch H, Jang Y, Maciejewski R, Ebert D, et al. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In: IEEE conference on visual analytics science and technology; 2012. p. 143–52.
[26] Maciejewski R, Hafen R, Rudolph S, Larew S, Mitchell M, Cleveland W, et al. Forecasting hotspots – a predictive analytics approach. IEEE Trans Visual Comput Graph April;17(4):440–53.

[27] Malik A, Maciejewski R, Hodgess E, Ebert D. Describing temporal correlation spatially in a visual analytics environment. In: 2011 44th Hawaii international conference on system Sciences (HICSS). p. 1–8.

[28] Andrienko N, Andrienko G, Gatalsky P. Exploratory spatio-temporal visualization: an analytical review. J Vis Lang Comput 2003;14(6):503–41 (Visual Data Mining).

[29] Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol 1980;12(1):97–136.

[30] New York City. Hurricane evacuation map. Retrieved March 3, 2013, ⟨http://gis.nyc.gov/oem/he/map.htm⟩; 2012.