EuroVis Workshop on Visual Analytics (2013)
M. Pohl and H. Schumann (Editors)

# Visual Analytics of Microblog Data for Public Behavior Analysis in Disaster Events

Junghoon Chae[†1], Dennis Thom[‡2], Yun Jang[§3], SungYe Kim[¶4], Thomas Ertl[‡2] and David S. Ebert[†1]

[1]Purdue University, USA, [2]University of Stuttgart, Germany,
[3]Sejong University, South Korea, [4]Intel Corporation, USA

**Abstract**

*In disaster management, analysis of public behavior plays an important role for evacuation planning. Unfortunately, finding meaningful information for analysis is challenging and collecting relevant data can be very costly. However, the growing dataset of Location-based Social Networks services with its time-stamped, geo-located data offers a new opportunity. Such spatiotemporal data has substantial potential to increase the situational awareness of local events and provide for both planning and investigation. In this paper, we present a visual analytics tool that provides users with interactive social media data analysis and investigation in order to help evacuation planning, analysis, and response. We demonstrate how to improve investigation by analyzing the extracted public behavior responses before and after the evacuation order during the natural disaster event, such as Hurricane Sandy.*

Categories and Subject Descriptors (according to ACM CCS): Information Interfaces and Presentation [H.5.2]: User Interfaces—GUI; Information Storage and Retrieval [H.3.3]: Information Search and Retrieval—Information filtering, relevance feedback

## 1. Introduction

For emergency and disaster management, analysis of public behavior, such as how people will respond and prepare for disasters, is important for evacuation planning. As social media have 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 [Web10]), social media are changing the way people communicate not only in daily use, but also during abnormal events, such as natural disasters. In the emergency situations, people even generally seek social confirmation before acting in response to the situation, where they interact with others to confirm information and develop a view of the risks [oPRtAKoS*13]. 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 [Ame10]. A growing number of people are using *Location-based Social Network* services 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 benefits for enhancing situational awareness during crisis situations and providing insights into the evolving event, public response, and potential course of action.

For public behavior analysis in disasters, however, finding meaningful information from social media is challenging, since data volume has increased beyond the capabilities of manual evaluations. Even though we could extract certain information from this dataset, 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 even aid in examining the results in order to understand the situations and glean investigative insights. Given the incomplete, complex,

† e-mail: {jchae|ebertd}@purdue.edu
‡ e-mail: {dennis.thom|thomas.ertl}@vis.uni-stuttgart.de
§ e-mail: jangy@sejong.edu
¶ e-mail: sungye.kim@intel.com



DOI: 10.2312/PE.EuroVAST.EuroVA13.067-071

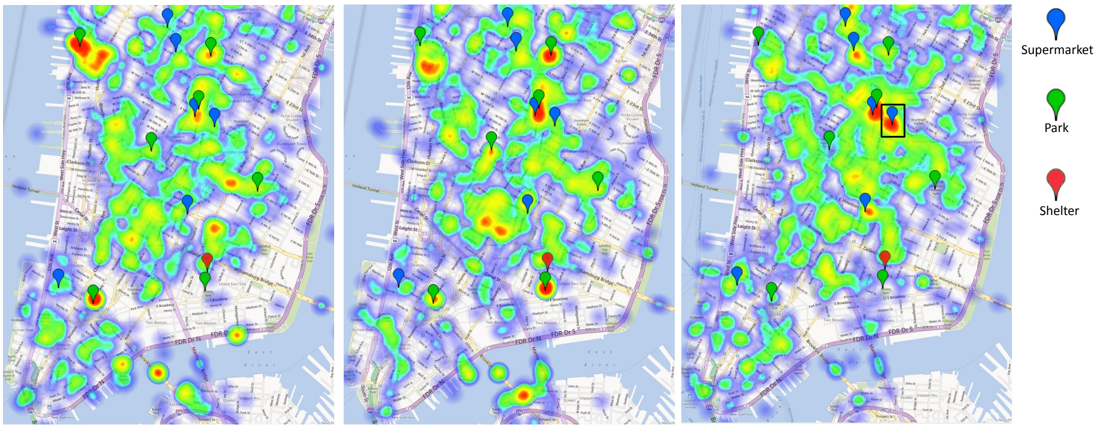


**Figure 1:** *Spatial user-based Tweet distribution during four hours right after evacuation order on October 28th (Right). Previous distribution on 14th (Left) and 21st (Center).*

context dependent information, a human in the analysis and decision-making loop is crucial. Therefore, a visual analytics approach offers great potential. In this paper, we present an interactive visual analytics tool 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 between location-based public behavior and locations of multiple types of infrastructures.

This study is performed using Twitter messages called Tweets that were published before, during, and after *Hurricane Sandy*. For this study, 1,740,114 Tweets published by 92,683 people are used. We propose an approach that visualizes spatial and temporal distribution of Tweets to identify public behavior patterns during the disaster. The main features of our approach are:

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

## 2. Related Work

In recent research, social media have become a popular and influential data source in many domains. Especially, analysis of *Location-based Social Networks* can be essential for situational awareness in disaster management [VHSP10, SOM10, TSdVP12].

Sarah Vieweg et al. [VHSP10] focused on communications for identifying features of Tweets generated during emergencies. MacEachren et al. [MJR*11] demonstrated a visualization system that denotes the message density of actual or textually inferred Twitter message locations. Their work also has shown that social media can be a potential source for crisis management. Bosch et al. [BTW*11] developed a scalable system, ScatterBlogs, enabling analysts to find quantitative information and detect abnormal event within a large set of geo-located microblog messages. Sakaki et al. [SOM10] introduced a natural disaster alert system using Twitter users as virtual sensors. Thom et al. [TBK*12] showed a spatiotemporal anomaly overview based on a streaming enabled clustering approach in order to generate a spatially and temporally explorable term map of large amounts of microblog messages as an entry point for closer examination. Chae et al. [CTB*ct] proposed the combination of Latent Dirichlet Allocation and Seasonal-Trend Decomposition based on locally-weighted regression for an ad-hoc analysis of a user selected set of messages regarding the topical distribution of messages and the abnormal presence of topics. Due to this characteristic, the system provides an iterative analysis loop for qualitative analysis and drill down operations.

## 3. Spatiotemporal Analysis

Since many social media channels provide time-referenced geographic data, traditional techniques for spatiotemporal zooming and filtering can now be applied to explore social media data. However, as the volume of the data exceed the boundaries of human evaluation capabilities and even normal computing performance, it is almost impossible to perform a straightforward qualitative analysis of the data. The tasks of examining and determining whether the extracted result is meaningful, are still challenging. In order to address these issues, traditional visualization methods have to be enhanced with interactive, scalable and verifiable techniques,

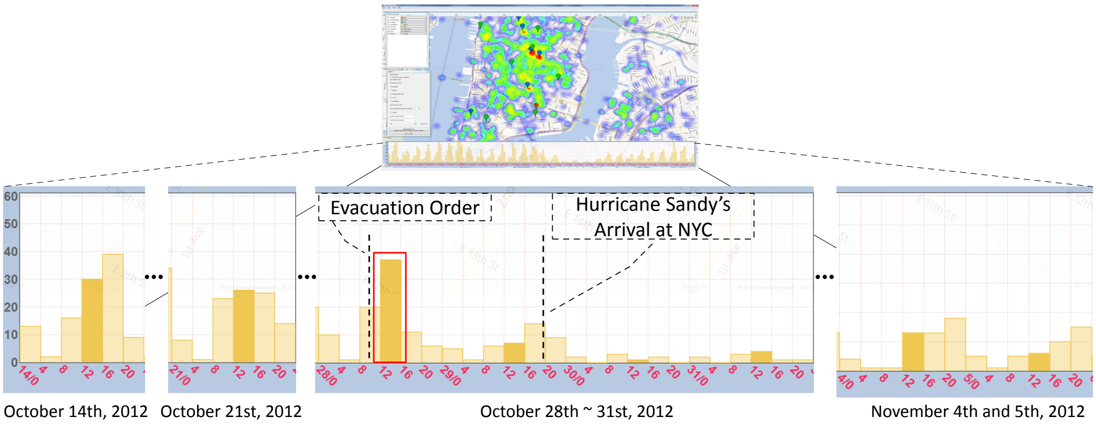


**Figure 2:** *Temporal analysis for public behavior during disaster event. Our entire system view (Top). The bar chart (Bottom) for the number of Twitter users within the selected region including a supermarket in Figure 1 (Right) in four hours intervals is showing some interesting time frames of the entire plot on the system. We see that many people go to supermarket right after the evacuation order.*

helping users extract, isolate, and examine the results interactively. We present a visual analytics tool that handles vast amount of microblog data, provides interactive spatiotemporal analysis, and allows combinational analysis on multiple spatial datasets for spatial decision supporting. Users select an initial spatiotemporal context of Tweets to be represented in the visualization serving as a basis for analysis. They can also perform spatiotemporal queries that load the relevant data set from a larger database.

### 3.1. Spatial Analysis and Spatial Decision Support

As mentioned in Section 1, social media, which enable users to embed geo-location information into the data, can be significantly useful in analyzing location-based public behavior. Such spatial analysis is important in order to manage and prepare disasters and emergency situations. The spatial characteristics together with heterogeneous information can assist in disaster management and migrating hazards where the problems have spatial components [AAJ*07]. In this section, we describe how our system supports spatial decision-making by combining multiple spatial data sources: location-based microblog data and spatial infrastructure data.

In this work, we focus on one recent disaster event, *Hurricane Sandy* [Wik12]. New York City Authorities ordered residents to leave some low-lying areas—the mandatory evacuation zones (red color) are shown in Figure 3 (Right). We focus on the area of Manhattan that is the most populated area and experienced severe damage. Through the map view of our system, users move to the Manhattan area in New York City and filtered Tweets that are posted within this area appear in the view during two weeks before and after October 28th. In Figure 1 (Right), we show a heatmap of spatial user-based Tweet distribution from noon to 4:00 PM on October 28th, right after the evacuation order that was announced at about 10:30 AM. To properly reflect the flow of evacuation unbiased by personal Tweet activity or behavior of individual users, we use the number of Twitter users instead of the number of Tweets, since some enthusiastic users generate a large number of Tweets at the same location (more than 20 messages per hour).

Investigating and making decisions using only the heatmap without any supplement information are demanding and complicated tasks. Our system, therefore, allows users to apply spatial information of various types of infrastructure (e.g., transportation centers, routes, school, venue, and business locations). In this case, analysts can assume that many people might go to the supermarket before staying or evacuating, but they would need supporting evidence and results before making decisions and plans. Through our system, the analysts can simply overlay the locations of big supermarkets (blue pins) on the user distribution heatmap. As shown in Figure 1 (Right), a relatively high number of people immediately went to supermarkets nearby the evacuation area, instead of the emergency shelter (red pin).

However, October 28th was Sunday and many people generally would go grocery shopping on Saturday or Sunday; therefore, analysts might need to verify whether the result shown on the map is a normal periodic situation. The analyst can obtain new results for different time frames by simply manipulating the time context. In Figure 1 (Left and Center), we show two new results of one and two weeks before the same time period. Here, we see that the hotspot locations in the results are very different from the ones for October 28th shown in Figure 1 (Right). For further analysis, we can explore another popular Sunday location—large parks—by superimposing the locations of the large parks (green pins) on each map. As shown in Figure 1 (Left and Center), many hotspots overlap with the park areas. This confirms the unusual, non-periodic event pattern on October 28th. This system can support the analyst to understand the unusual event

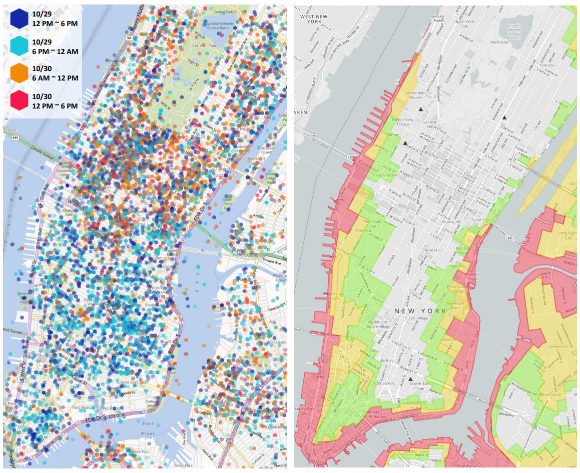


**Figure 3:** *Visualization for spatiotemporal social media data (Left). A hexagon represents the spatial (position) and temporal (color) information of a Tweet. Hurricane evacuation map [New12] (Right).*

patterns of movement and plan resource allocation accordingly for such emergency events.

### 3.2. Temporal Pattern Analysis

In Section 3.1, we presented spatial analysis of social media and spatial decision support using multiple data sources in order to reveal where and why a number of people move. In this section, we 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 previous situations? Analysis of temporal trends and relationships between data values across space and time provides underlying insights and improves situational awareness [MHR*il, MMHEan].

After selecting the initial spatiotemporal context of Tweets as a basis for the analysis, the analyst can explore the temporal patterns of the number of Twitter users who posted Tweet messages within the spatial boundary using the bar chart as shown in Figure 2. The values of each bar are the number of users in four hours 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 Figure 2. As previously mentioned, the heatmap in this Figure shows the Twitter user density from noon 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 Figure 1 (Right). We can indicate that the number of people (red rectangle in Figure 2) in the corresponding time period is higher than for the same time period from other dates (October 14th, 21st and November 4th, 5th). Moreover, there is another interesting finding—the number of people during each of the following time frame (4:00 $\sim$ 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*. Furthermore, since October 29th, the number of people has significantly decreased, because most residents left before the arrival of the hurricane. The increase of the number of people after one week reflects that a portion of people came back home and even shows when the stores reopened.

### 3.3. Spatiotemporal Visualization

There is abundant research work published on the topic of spatiotemporal data visualization. Exploration of time-referenced geographic data is still a challenging issue [AAG03]. We introduce a modest visualization that enables users to analyze both aspects: space and time using a single view. As shown in Figure 3 (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 features of human visual perception [TG80]. In Figure 3 (Left), the hexagon with blue (12 PM $\sim$ 6 PM) or green (6 PM $\sim$ 12 AM) ) color correspond to Tweets published on October 29th 2012 and the others with orange or red color correspond to Tweets posted the following day after the hurricane. New York City announced the evacuation of Zone A (red color) in Figure 3 (Right); residents in Zone A faced the highest risk of flooding, Zone B (yellow color) and Zone C (green color) are moderate and low respectively. In the visualization, analysts can realize 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 the following day as the hurricane damaged the city.

## 4. Conclusions

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 spatial decision support, we demonstrated an analytical scheme by combining multiple spatial data sources. Our temporal analysis enables users 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 focus on the flow of public movement before and after disasters and analysis for recovering from disasters and crises.

## References


[AAG03] ANDRIENKO N., ANDRIENKO G., GATALSKY P.: Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing 14*, 6 (2003), 503 – 541. <ce:title>Visual Data Mining</ce:title>. URL: http://www.sciencedirect.com/science/article/pii/S1045926X03000466, doi:10.1016/S1045-926X(03)00046-6. 4

[AAJ*07] ANDRIENKO G., ANDRIENKO N., JANKOWSKI P., KEIM D., KRAAK M., MACEACHREN A., WROBEL S.: Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science 21*, 8 (2007), 839–857. URL: http://www.tandfonline.com/doi/abs/10.1080/13658810701349011, arXiv:http://www.tandfonline.com/doi/pdf/10.1080/13658810701349011, doi:10.1080/13658810701349011. 3

[Ame10] 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. 1

[BTW*11] BOSCH H., THOM D., WORNER M., KOCH S., PUTTMANN E., JACKLE D., ERTL T.: Scatterblogs: Geo-spatial document analysis. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (2011), pp. 309 –310. doi:10.1109/VAST.2011.6102488. 2

[CTB*ct] CHAE J., THOM D., BOSCH H., JANG Y., MACIEJEWSKI R., EBERT D., ERTL T.: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (Oct.), pp. 143–152. 2

[MHR*il] MACIEJEWSKI R., HAFEN R., RUDOLPH S., LAREW S., MITCHELL M., CLEVELAND W., EBERT D.: Forecasting hotspots - a predictive analytics approach. *Visualization and Computer Graphics, IEEE Transactions on 17*, 4 (April), 440–453. doi:10.1109/TVCG.2010.82. 4

[MJR*11] MACEACHREN A., JAISWAL A., ROBINSON A., PEZANOWSKI S., SAVELYEV A., MITRA P., ZHANG X., BLANFORD J.: Senseplace2: Geotwitter analytics support for situational awareness. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (oct. 2011), pp. 181 –190. 2

[MMHEan] MALIK A., MACIEJEWSKI R., HODGESS E., EBERT D.: Describing temporal correlation spatially in a visual analytics environment. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on* (Jan.), pp. 1–8. doi:10.1109/HICSS.2011.144. 4

[New12] NEW YORK CITY: Hurricane evacuation map. Retrieved March 3, 2013, http://gis.nyc.gov/oem/he/map.htm, 2012. 4

[oPRtAKoS*13] ON PUBLIC RESPONSE TO ALERTS C., KNOWLEDGE W. U. S. M. C., ON SCIENCE R. G. C., ON ENGINEERING T. B. D., COUNCIL P. S. N. R.: *Public Response to Alerts and Warnings Using Social Media: Report of a Workshop on Current Knowledge and Research Gaps*. The National Academies Press, 2013. URL: http://www.nap.edu/openbook.php?record_id=15853. 1

[SOM10] 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* (New York, NY, USA, 2010), WWW '10, ACM, pp. 851–860. URL: http://doi.acm.org/10.1145/1772690.1772777, doi:10.1145/1772690.1772777. 2

[TBK*12] 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). 2

[TG80] TREISMAN A. M., GELADE G.: A feature-integration theory of attention. *Cognitive Psychology 12*, 1 (1980), 97 – 136. URL: http://www.sciencedirect.com/science/article/pii/0010028580900055, doi:10.1016/0010-0285(80)90005-5. 4

[TSdVP12] TERPSTRA T., STRONKMAN R., DE VRIES A., PARADIES G.: Towards a realtime twitter analysis during crises for operational crisis management. In *Proceedings of the 9th International ISCRAM Conference. Vancouver, Canada* (2012). 2

[VHSP10] VIEWEG S., HUGHES A. L., 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* (2010), ACM, pp. 1079–1088. 2

[Web10] 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. 1

[Wik12] WIKIPEDIA: Hurricane sany. Retrieved April 30, 2013, http://en.wikipedia.org/wiki/Hurricane_Sandy, 2012. 3