Observing Community Resiliency in Social Media

Robert M. Patton, Chad A. Steed, Chris G. Stahl, and Jim N. Treadwell

Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN, USA, 37831
{pattonrm,steedca,stahlcg,treadwelljn}@ornl.gov

Abstract. In spite of social media's lack of structural integrity, accuracy, and reduced noise with respect to other forms of communication, it plays an increasingly vital role in the observation of societal actions before, during, and after significant events. In October 2012, Hurricane Sandy making landfall on the northeastern coasts of the United States demonstrated this role. This work provides a preliminary view into how social media could be used to monitor and gauge community resilience to such natural disasters. We observe, evaluate, and visualize how Twitter data evolves over time before, during, and after a natural disaster such as Hurricane Sandy and what opportunities there may be to leverage social media for situational awareness and emergency response.

Keywords: social media, temporal analysis, community resilience.

1 Introduction

Originally developed for entertainment purposes, social media systems have rapidly evolved to provide valuable benefits for such areas as business intelligence, national security, and disaster management. In combination with the development and adoption of smart phone technology, people have become mobile sensors providing "eyes" and "ears" as events unfold. Leveraging this capability provides significant advantages for situational awareness.

Unfortunately, this technology contains significant noise and error in the data as well as the inability to position the "sensors" in critical areas. Consequently, leveraging this capability can be quite challenging depending on the application purposes. Research into this issue has opened up several opportunities to overcome this problem. The work described here is a preliminary investigation into resolving the data noise issue as it relates to disaster management.

This work narrowly focuses on resolving a specific problem: identifying and characterizing the community resilience to natural disasters relating to physical infrastructure and social behavior using information obtained via Twitter. Community resilience is a measure of a community's ability to prepare, respond, withstand, and recover from disasters both natural and man-made. In the unfortunate event of a disaster, communities that have low resilience will suffer significant loss of life and damage to critical infrastructure. In addition, communities with low resilience require longer periods of time to recover to pre-disaster operations and quality of life. Communities with higher resilience suffer fewer losses and return more quickly to

B. Murgante et al. (Eds.): ICCSA 2013, Part V, LNCS 7975, pp. 491–501, 2013.

pre-disaster operation and quality of life. Unfortunately, when a disaster occurs, the effects of any weakness in a community's resiliency will be amplified. Any ability for disaster management personnel and leadership to observe in real-time the effects of a disaster on a community provides a significant advantage and opportunity to respond more quickly. Consequently, the work described here is focuses on utilizing social media, specifically the Twitter platform, as a means of providing a real-time view into the impacts of a disaster on the community.

2 Related Works

Investigating the impacts of natural disasters on society is not a new research area. There are many works relating to the identification of different impacts as well as different studies that have been performed [1][11][12]. Many of these works and studies, however, have identified impacts that are difficult to measure quantitatively either before, during, or after the event. In addition, some works have relied on remote sensing techniques for assessing and monitoring disaster situations [4][9][13]. For our work, the goal is to develop an approach that would provide a more quantitative view of the event impact as well as be founded on field reporting about the event.

More recently, research has shifted into utilizing social media for disaster management. In [5], a system for crowdsourcing the collection of information for disaster management is described. Leveraging social networks, information is received and exchanged between volunteers in the field and disaster management command and control (DMCC). The DMCC can request the volunteers to visit specific locations and report on what is seen and heard, thereby making the DMCC much more adaptive to the circumstances. Unlike the work of [5] which relies on some structured data (e.g., automated geo-tagging of the volunteer's mobile device) to reduce the noise, the work of [10] uses natural language processing and data mining to extract situational awareness from Twitter data. Utilizing Twitter's location based search API, tweets are collected for specific areas and then processed and mined to provide a view into the current situation of a specific area. Results are then displayed as a combination of tag clouds and map-based user interfaces.

The primary problem with the two previous works is that they are very granular in detail as it relates to situational awareness. In disaster management, there are various "layers" of detail that allow disaster response personnel to view the event at different levels of abstraction. The two previous works provide little to no layers of abstraction. In the work of [2], Twitter data is processed in increasingly higher levels of abstraction such that individual tweets are eventually grouped into areas such as Health, Transport, Communication, and Lodging. While their work provides a critical connection of tweets to higher-level abstractions, what is lacking is a visualization or user interface to support a "top-down" view of the different layers of abstraction. The work described here attempts to address this deficit.

Others have also investigated the impact of events from news reports, but in different ways. In the work of [8], information from blogs was analyzed with respect to the actual sales of a book. The authors discovered that there was a direct relationship between blog chatter prior to a book release and the actual volume of book sales after the book release. The more people talked about a book on their blogs prior to its release, then the higher the volume of sales for the book after its release. The authors show clearly that blog chatter is a good indicator of the impact that a book release will create on sales volume. In the work of [6-7], the relationship between financial news and stock prices was investigated. The foundation of these works is the premise that financial news can have either a positive, negative, or natural impact on the price of a stock, and that the time lag between the news and the stock price was minimal, if any. The authors observed that, in fact, some of the financial news could provide indications as to the direction of the stock price (up, down, unchanged). However, their results are based on market simulations using real data. Regardless, their work provides evidence that events and event impacts can be monitored and gauged using news reports.

Finally, the work described here is an extension of our previous work in this area. In [14], news media was investigated in order to better understand various impacts of a disaster on society. In that work, the data source tended to be less noisy in terms of the language that was used. News media tend to be more factual and use more structured language as well as correct spellings. In contrast, the Twitter data uses abbreviated or incorrect syntax and spelling, and can be less factual and more biased to the individual needs and wants of a person. As addressed in [2] work, a bottom up approach was developed to identify trends in a larger population of people based on the tweets that depict individual wants and needs. The approach described here supports visualizing the higher-level trends and enables the ability to dive into details for specific areas of concern.

3 Approach

The primary focus is to detect, analyze, and characterize community resiliency metrics from events impacting society as observed in Twitter data. Inspiration for this work partially originated from the concept of decomposing a time-series into sub-series as discussed in [15], where time-series are decomposed into three sub-series referred to as seasonal, trend, and noise. The goal is to find complete or partial periodic patterns in a time-series with trends. Their work provides an approach to observe both short and long-term periodic patterns. They demonstrate their approach on atmospheric CO2 levels as well as stock prices. Figure 1 shows the conceptual view of our approach.

The original data begins as a time-series of all of the word and phrase counts that are observed in the Twitter fire hose. Next, a "textual prism" is created that is comprised of a set of taxonomies that describe the words and phrases related to the topics of interest by the user. The taxonomies are defined by the user and applied to the

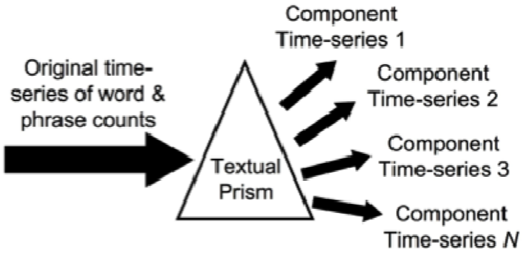

Fig. 1.

original time-series in order to produce a component time-series. This component time-series shows how the words and phrases of the taxonomy change over time.

As discussed previously, there have been investigations into the various impacts caused by natural disasters. For this work, the research performed by [3] provides the foundation for the impacts to be investigated. This work was chosen for its simplicity and extensibility. In [3], impacts can be observed in 4 areas: Technical, Organizational, Social, and Economic. Technical refers to the infrastructure (roads, bridges, power grid, water systems, etc.) of a society. Organizational refers to areas such as service crews that maintain or respond to the Technical aspects. Social refers to areas related to housing, shelters, provisions for human needs, etc. Economic refers to the impacts on the economy. This work focuses on the Technical and Social impacts of a natural disaster.

To begin, a set of taxonomies were developed for both the Technical and Social categories as shown in Table 1. These taxonomies were developed manually by analyzing news reports from the time period of August 25, 2005 to September 5, 2005. This is the time period when Hurricane Katrina made landfall in the Gulf Coast of the United States. This natural disaster created significant and widespread damage and resulted in extensive impacts on U.S. society. The news reports during this time period were clustered. The clusters were then analyzed to determine the most popular words and phrases that were used to describe specific conditions. These words and phrases were then categorized according to the framework defined in [3] as shown in Table 1.

As a specific example, the Shelters taxonomy consists of word and phrases such as: taking cover, taking shelter, seek refuge, shelters. The Movement taxonomy consists of words and phrases such as: mandatory evacuation, dawn curfew, flights canceled, etc. The Power taxonomy consists of phrases such as: downed power lines, backup power, emergency generator, and emergency power. The Roads taxonomy consists of phrases such as: alternate route, traffic backups, evacuation route, major streets, and interstates.

Table 1.

Technical	Social
Communications	Crime
Fuel	Death
Light	Health Hazards
Outage	Medical
Power	Movement
Rail	Shelter
Roads	Water (Health)
Structures	
Water (Infrastructure)	

After creating these taxonomies, they are then applied to the Twitter data as a form of "textual prism" to create component time-series as described in Figure 1. Every word and phrase observed from these categories is counted each day. For this particular investigation, the data was focused on the Hurricane Sandy event, and was collected using the Twitter location-based search to identify tweets in the geographic area of impact. When visualizing the volume of tweets for each taxonomy on the same chart, trends in numerically larger scales can obscure significant trends in smaller scales. To account for the volume of tweets for a particular taxonomy, the total counts of each day for each category was converted to be the percentage of total tweets for that taxonomy across the entire time that tweets were collected. For example, the Shelters taxonomy may have a total of 10,000 tweets over a 10-day period while the Medical taxonomy may have a total of 50 tweets over the same time period. On day 1, the Shelters total tweet count may have been 1,000 while the Medical may have been 5. For each taxonomy, the number of tweets on day 1 is 10% of the total number of tweets over the 10-day period, respectively. Thus, the data is normalized to a scale of 0 to 100.

4 Results

The results from the approach described previously are shown in Figures 2 through 5. Figures 2 and 3 show the trends of the Social taxonomies of Table 1. Figures 4 and 5 show the trends of the Technical taxonomies. Figures 2 and 4 show the total volume of tweets collected for the respective taxonomies, while Figures 3 and 5 show the normalized tweet counts, respectively.

As Hurricane Sandy makes landfall in New Jersey around midnight of October 29th, the trends in nearly every area begin moving upward. Of particular note, the Shelter and Movement taxonomies begin trending upward considerably and prior to the upward trends Death and Medical taxonomies. Unfortunately, most of the tweets in the Shelter and Movement taxonomies are not particularly useful. Many of the tweets are either originating from news agencies, retweets from news agencies, or tweets that contain information that could more easily be obtained from news articles. Furthermore, many of the tweets are simply wishing or praying for people to stay safe and seek shelter. Table 2 shows some example tweets during that time period.

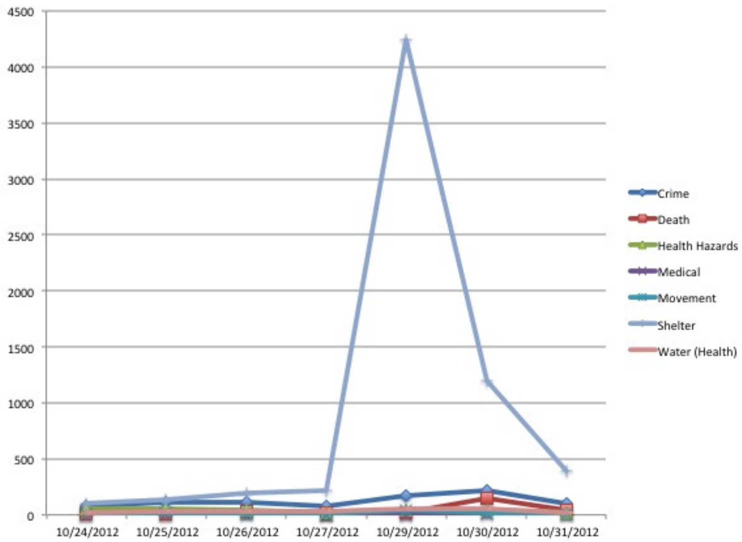

Fig. 2. Total number of tweets for Social taxonomies

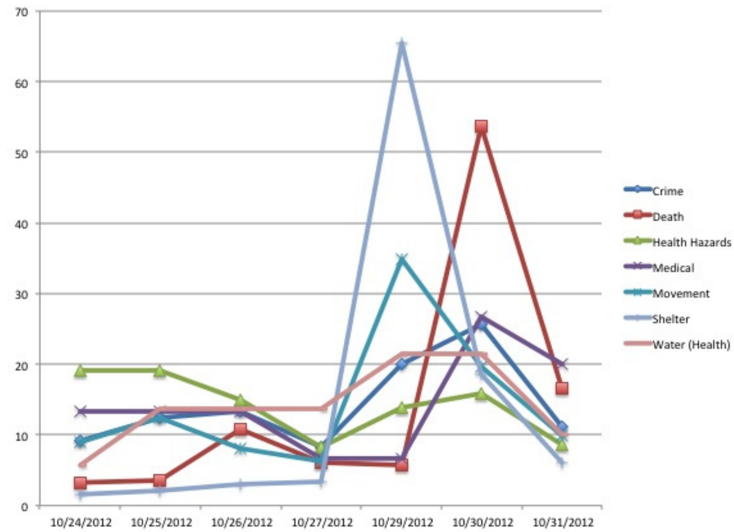

Fig. 3. Normalized number of tweets for Social taxonomies

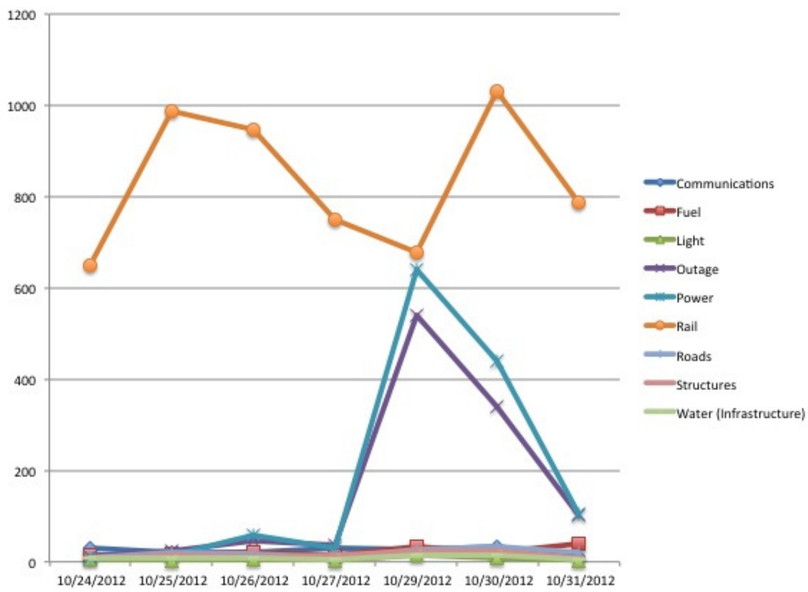

Fig. 4. Total number of tweets for Technical taxonomies

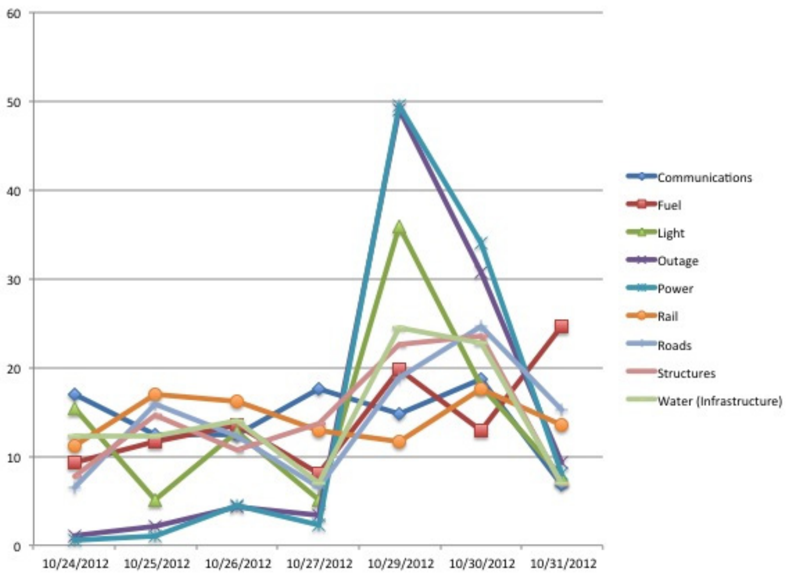

Fig. 5. Normalized number of tweets for Technical taxonomies

Table 2.

rt @sapendeleds: philadelphia to open shelters at 2 pm tomorrow for evac of flood-prone areas of city.
rt @corybooker: animal lovers: our jfk center shelter will have facilities to care for people displaced by storm who also have pets.
those without power: as of 11:30 pm the active living center in jim barnett park will be an emergency shelter for those without power.
all's mostly quiet in #cherrydale. still no power. headed to bed. stay safe all! #fb
please stay safe people who are on the east coast! #prayforeastcoast
… no we are staying put. as of now it's not mandatory in our town. plus we would have to find a dog friendly shelter :/

Table 3 shows example tweets from the Death and Medical taxonomies. Like the Shelter and Movement taxonomies, the tweets in Table 3 are mostly news oriented, retweet of news, or contain information more easily obtained from other sources. However, in two example tweets that departs from this pattern. One tweet is requesting information on where to receive medical attention in New York City (NYC), while another tweet states that a specific Twitter user needs medical attention. In the latter example, one would hope that 911 were called to assist the person if, in fact, it was an emergency and not a sarcastic remark.

Table 4 shows example tweets from the Power, Light, and Fuel taxonomies. While still consisting of tweets originating from news agencies, more of the tweets from these taxonomies consisted of eyewitness accounts such as downed power lines and trees and shortages at gas stations. Many of these tweets also tended to express more sentiment such as fear, anxiety, and relief.

These results highlight the discrepancy in the language that is used by news agencies and individuals who use Twitter. News agencies tend to express more factual information, while tweets from individuals tend to describe more specific details or express more emotion. In addition, the news agencies often include http links for followers to obtain more information, while the individuals tend to produce tweets directly to a specific follower.

Table 3.

death toll continues to rise after sandy steamrolls northeast http://t.co/2opfeg2b
rt @stormchaser4850: developing: update: ap - death toll from hurricane #sandy is now at least 33 via @abc7chicago
if ppl are in need of medical attention in nyc right now what options are available if any? #sandy cc. @hurricanehackrs
@<Twitter ID> needs medical attention!
sandy death toll climbs; millions remain without power. thousands of homes are flooded others destroyed.

Table 4.

current power outages in monmouth and ocean counties http://t.co/j4zai70n via @asburyparkpress
power lines down and trees in nb
and the valley view power outages begin
backup generator just literally blew up
light pole fell and smashed a parked car on my street. #sandy is real!
people are lining up at gas stations where there isn't even gas just waiting for gas to come this is scary.
wtf all these gas stations are packed ????
just went on the road ..so dark no houses have light in them long lines at gas stations omg ??

In regard to trends, each taxonomy experiences an increasing trend as Hurricane Sandy approaches landfall and through the initial hours of landfall, but then quickly start declining. One exception of particular notice is the Fuel taxonomy as shown in Figure 5. While the other taxonomies are declining, the Fuel taxonomy makes a significant increase on the last day of our collection. Analyzing the individual tweets from this taxonomy reveals that a number of the tweets are from news agencies that are reporting shortages at gas stations. Further, there are an increasing number of tweets by individuals expressing frustration at the gas shortages and long lines at gas stations. For example, one tweet stated “people are lining up at gas stations where there isn't even gas just waiting for gas to come this is scary.” Another tweet stated, “what gas stations are open and are not packed with hundreds of people. i highly doubt any.” In the days and weeks after Hurricane Sandy, this gas shortage becomes a significant problem resulting in near riots and citizens venting their anger in various ways. Consequently, there is an opportunity here to supplement this preliminary work with sentiment analysis as well as additional taxonomies to capture social unrest in order to further capture the dynamics of the community as an event transpires.

5 Summary

This work is a preliminary investigation into the use of community resilience taxonomies applied to Twitter data. Visualizing these taxonomies over time provides insight into how communities respond to events in a variety of different ways. Our approach helps curb the effects of noise and errors in the Twitter data and provides a “top-down” visual analytics approach to leveraging this data for situational awareness purposes.

Additional work will incorporate sentiment analysis and additional taxonomies to capture more detail in the data as well as additional filters to further reduce noise and errors in the data. Furthermore, additional visualization involving parallel coordinates and a more sophisticated user interface to enable the user to thoroughly analyze the data will be developed.

Acknowledgements. Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy. LDRD #6427.

Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285; managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR2225. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

1. Alexander, D.: The Study of Natural Disasters, 1977-97: Some Reflections on a Changing Field of Knowledge. Disasters 21(4), 284–304 (1997)
2. Birregah, B., Top, T., Perez, C., Chatelet, E., Matta, N., Lemercier, M., Snoussi, H.: Multi-layer Crisis Mapping: A Social Media-Based Approach. In: 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), pp. 379–384 (2012)
3. Bruneau, M., et al.: A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. Earthquake Spectra 19(4), 733–752 (2003)
4. Chi, T., et al.: Research on information system for natural disaster monitoring and assessment. In: Proc. of the 2003 IEEE International Geoscience and Remote Sensing Symposium (2003)
5. Chu, E.T., Chen, Y.-L., Lin, J.-Y., Liu, J.W.-S.: Crowdsourcing support system for disaster surveillance and response. In: 2012 15th International Symposium on Wireless Personal Multimedia Communications (WPMC), pp. 21–25 (2012)
6. Fung, G.P.C., et al.: Stock prediction: Integrating text mining approach using real-time news. In: Proc. of the 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, pp. 395–403 (2003)
7. Fung, G.P.C., et al.: The predicting power of textual information on financial markets. IEEE Intelligent Informatics Bulletin 5(1) (June 2005)
8. Gruhl, D., et al.: The predictive power of online chatter. In: Proceeding of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, Chicago, Illinois, pp. 78–87 (2005)
9. Hussain, M., et al.: Emerging geo-information technologies (GIT) for natural disaster management in Pakistan: an overview. In: Proc. of the 2nd International Conference on Recent Advances in Space Technologies (2005)
10. Jie, et al.: Using Social Media to Enhance Emergency Situation Awareness. Intelligent Systems 27(6), 52–59 (2012)

11. Kreimer, A., Arnold, M., Carlin, A.: Building Safer Cities: The Future of Disaster Risk. World Bank Publications (2003)
12. Little, R.G.: Toward more robust infrastructure: observations on improving the resilience and reliability of critical systems. In: Proc. of the 36th Annual Hawaii International Conference on System Sciences (2003)
13. Llinas, J.: Information fusion for natural and man-made disasters. In: Proc. of the 5th International Conference on Information Fusion, pp. 570–576 (2002)
14. Patton, R.M., et al.: Discovery, analysis, and characteristics of event impacts. In: 2008 11th International Conference on Information Fusion, pp. 1–8 (2008)
15. Yu, J.X., Ng, M.K., Huang, J.Z.: Patterns discovery based on time-series decomposition. In: Cheung, D., Williams, G.J., Li, Q. (eds.) PAKDD 2001. LNCS (LNAI), vol. 2035, p. 336. Springer, Heidelberg (2001)