

# “Superstorm” Sandy’s Social Media Surge in Twitter: A Three-Dimensional Analysis

Craig A. Ramseyer, J.A. Rackley, A.W. Black, J.A. Knox

University of Georgia, Department of Geography



The University of Georgia



## Background

Post-tropical storm Sandy made U.S. landfall near Brigantine, New Jersey with the center of circulation coming onshore on October 29, 2012 at approximately 2330 UTC with winds estimated at 70 kt and a minimum pressure of 945 mb.

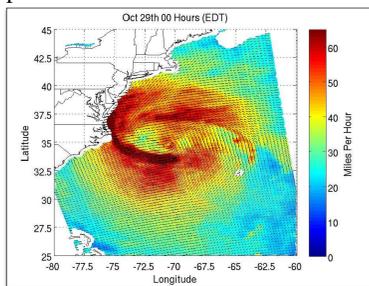


Image credit: ISRO/NASA/JPL-Caltech

Earlier in day, Sandy lost most of its tropical characteristics as it moved over cooler waters and interacted with a cold air mass located over the eastern U.S. The National Hurricane Center (NHC) suggests that Sandy became extratropical by 2100 UTC while the center of circulation was still offshore. Sandy posed unprecedented challenges in how watches and warnings were to be issued. Knowing Sandy would likely be a post-tropical cyclone at landfall in the Northeast U.S., the NWS issued high wind watches and warnings over 60 h before Sandy made landfall instead of the NHC issuing Hurricane watches/warnings (Blake et al. 2013).

Microblogging has become a popular communication tool with millions of users sharing opinions on different aspects of life and current events. Twitter can be used as a corpus for sentiment analysis and opinion mining (Pak and Paroubek 2010). Twitter provides several million daily potential data points, and has been used in scientific research across most disciplines (e.g. Bollen et al. 2010, Shelton et al. 2014). This study utilizes Twitter data to analyze how the general public reacted to Sandy in real time and what terminology they used to talk about Sandy. In addition, a basic sentiment analysis was performed by using OpinionFinder 2.0 (Riloff and Wiebe 2003).

## Data & Methods

- Tweets were acquired from GNIP for the period 25 October – 3 November. Tweets that contained either/or the terms, “Sandy” and “superstorm” were retained. Only those tweets that were geo-tagged were used. Re-tweets were not included to reduce non-original sentiment.
- Tweets were grouped by hour in order to analyze the temporal evolution of certain keywords
- Tweets were parsed to remove punctuation.
- OpinionFinder2.0 was used to determine the sentiment of each word. The sentiment for each tweet was determined based on the sum of sentiment for each tweet.
- Tweets were aggregated to census tract level.
- Sentiment for each tract was determined using the sum of the total tweet sentiment.

## Analysis

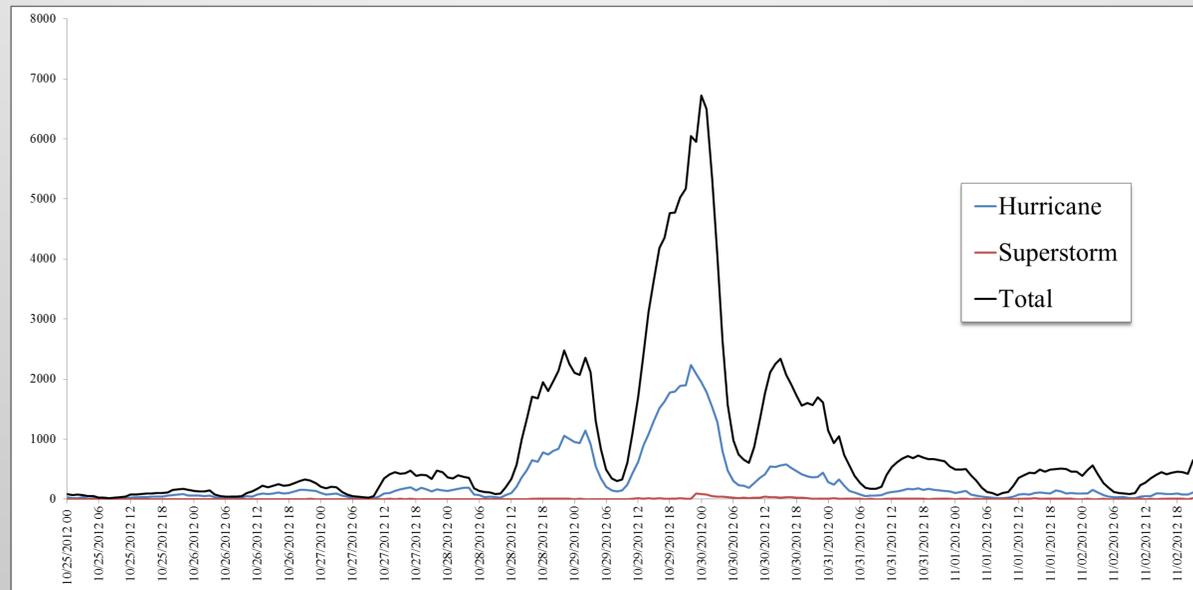


Fig. 1. Shows the temporal evolution in the number of tweets containing the search terms (black), the number of tweets referencing Sandy as a hurricane (blue), and the number of tweets referencing Sandy as a “superstorm”. The authors note the low frequency of the term “superstorm”, as well as the peak in Tweets is coincident with the center of circulation making landfall.

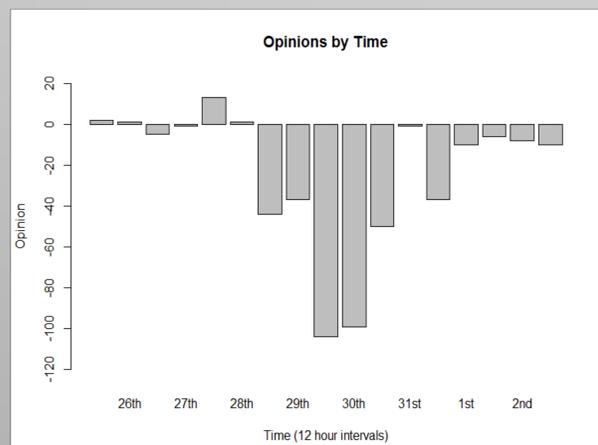


Fig. 2. Shows the temporal evolution of sentiment over the study period. Prior to landfall, sentiment appears to be neutral or slightly positive. However, as conditions began to deteriorate and during the peak intensity of exposure to the storm, sentiment became largely negative.

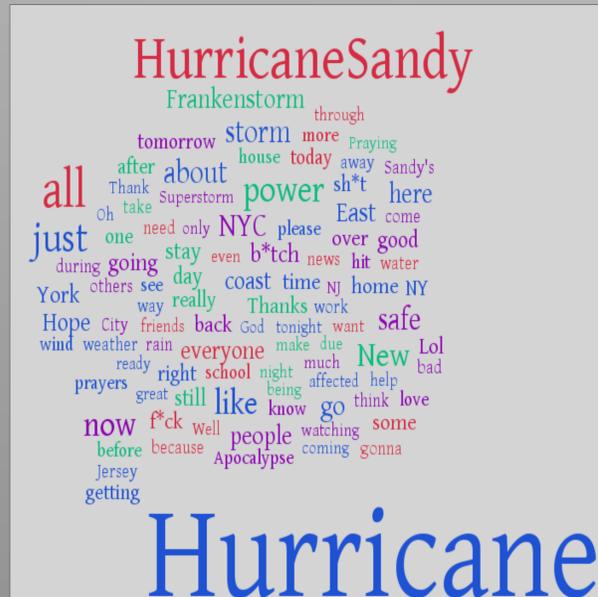


Fig. 4. Word cloud of top words in Sandy tweets.

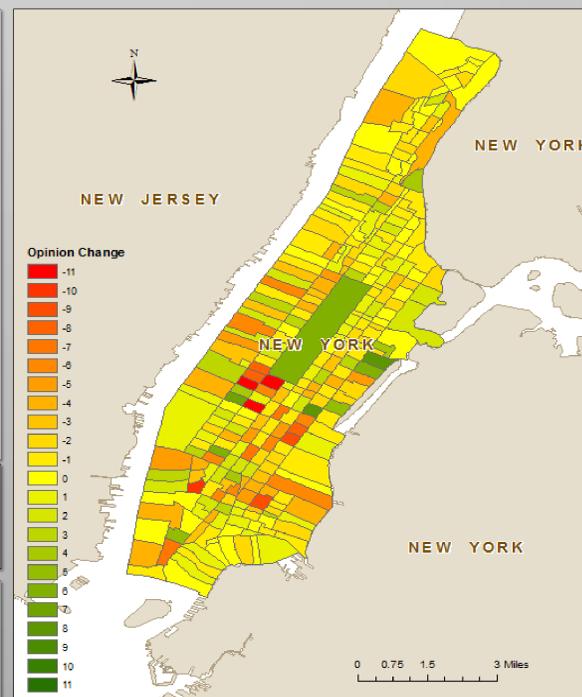


Fig. 3. Manhattan census tract spatial scale analysis of changing sentiment over the before and after peak exposure. The authors note the greatest changes occurred in lower Manhattan.



Fig. 5. A sample of Hurricane Sandy tweets that expressed confusion.

## Conclusions

- While many in the media referred to the storm as Superstorm Sandy, social media was largely referring to Sandy as a hurricane even after it was deemed post-tropical by the NHC. While Sandy was not a hurricane at landfall, the authors suggest that it didn’t change the public perception on social media. This mischaracterization of Sandy may have led to heightened awareness and preparedness for “hurricane-like” impacts.
- The number of Tweets per hour peaked during the same hour the center of circulation made landfall. While the pattern exhibits a diurnal trend observed in Twitter data, the authors suggest it may also be a response to peak exposure.
- Sentiment became highly negative as Sandy made landfall. However, sentiment before landfall was neutral or slightly positive. This suggests that tweets before exposure to Sandy were more lighthearted and may signal skepticism of the forecasted impacts. 24 h prior to exposure in the Manhattan area, tweets become more negatively skewed. The ratio of positive-to-negative became most negative as Manhattan began experiencing the direct impacts of Sandy.
- Analysis of social media text by systems such as OpinionFinder 2.0 can be difficult due to the over classification of neutral text. Sentiment is difficult to capture with current linguistic analysis packages due to the frequent use of slang and curse words. The authors suggest that these data still need to be considered as they can provide insight on sentiment and opinion.

## Future Work

This is a preliminary study and the authors expect to further explore the potential of social media data by developing a dictionary of confusion keywords and phrases based on social media text. The authors will build on the current sentiment analysis literature to include slang terminology and abbreviations. In addition, emoticons would be explored as additional sources of sentiment. Further analysis would include the spatial and temporal evolution of confusion as Sandy made its extratropical transition and as high wind watches and warnings were issued.

## References

- Blake, E.S., T.B. Kimberlain, R.J. Berg, J.P. Cangialosi, and J.L. Beven II. 2013. National Hurricane Center Tropical Cyclone Report for Hurricane Sandy (AL182012). 157 pp. Available at NOAA: nhc.noaa.gov/data/tcr/AL182012\_Sandy.pdf.
- Bollen, J., H. Mao, and X. Zeng, 2010: Twitter mood predicts the stock market. *J. Comput. Sci.* 2, 1-8.
- Pak, A. and P. Paroubek, 2010: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the Seventh conference on International Language Resources and Evaluation LREC'10*, Valletta, Malta, European Language Resources Association ELRA.
- Riloff, E. and J. Wiebe, 2003: Learning Extraction Patterns for Subjective Expressions. *Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*. ACL SIGDAT. 105-112.
- Shelton, T., A. Poorthuis, M. Graham, M. Zook, 2014: Mapping the Data Shadows of Hurricane Sandy: Uncovering the Sociospatial Dimensions of ‘Big Data’. *Geoforum* (forthcoming). Available at SSRN: ssn.com/abstract=2379966.

Contact: Craig A. Ramseyer, University of Georgia, Climatology Research Laboratory, ramseyca@uga.edu