Lessons from Harvey: Crisis Informatics for Urban Resilience
# Table of Contents

2 Executive Summary  
3 Introduction  
4 Data  
5 Results  
11 Discussion  
12 Conclusion  
13 Appendix A  
14 References
Social media systems and crowdsourced data sites were incredibly active during Hurricane Harvey. Residents, first responders and officials all turned to these systems to impart information and make calls for assistance. These systems will likely continue to hold a central informational and communication role in future disasters. Analyzing the trends and information that come from these sources in real-time could aid the recovery process and help public agencies, first responders and researchers more quickly assess damages during and immediately after a disaster. This report demonstrates some of the work that could be done with this type of information. It suggests that public agencies work closely with social media companies and the operators of crowdsourced sites to install plans to incorporate these systems into the disaster recovery process.

Methodology

Social media, in particular Twitter, has been used to evaluate the spatial and temporal aspects of disaster and emergency situations, including terrorist attacks, earthquakes and floods. The study evaluates the efficacy of parcel-level FEMA damage estimates using Twitter-sourced damage estimates and water rescue records from CrowdSource Rescue (CSR), a crowdsourcing platform set up during Hurricane Harvey, to proxy damaged structures.

Key Findings

- Immediate damage estimates based on FEMA models can miss areas of heavy impact. Augmenting initial models with real-time analysis of social media and crowdsourced information can help identify overlooked areas. Twitter-sourced estimates were virtually available as people tweeted distress signals, of these parcel-level damage estimates, 46 percent were not captured by FEMA estimates.
- Social media accounts have become an essential tool for both information and calls for assistance during storms and other crises. This role will only increase.
- In order to take advantage of this information, public agencies should proactively create structures and plans to include it in recovery analysis.
- Social media and emergency crowdsourced sites can support normal channels of emergency communication, such as the 911 system, which can become overwhelmed during large disaster events.
- Data agreements between cities, corporations and research entities are a valuable asset to save time and create robust and accurate damage assessments.
Introduction

The most frequently occurring and damaging natural disasters in the world are floods. Houston’s three major floods in three consecutive years demonstrate the impact these disasters can have on a region. Clearly, flooding events and hurricanes are among the biggest risks to Houston’s future. In the wake of Hurricane Harvey, the Houston region has worked tirelessly to rethink how its prepares for and mitigates against the impacts of these disasters. However, the region has not yet effectively leveraged several key datasets that are available through social media and crowdsourcing platforms. Working jointly, researchers, public agencies and social media companies have the opportunity to use crisis informatics—the process of analyzing data and knowledge collected from a variety of sources during a disaster—to inform and improve response to these events.

Crisis informatics can provide a critical set of data points for use during and immediately after a disaster event. Information from Twitter, Facebook or crowdsourcing/ed sites can capture details submitted by individuals in real-time. This data can help fill in the many uncertainties confronted by first responders and officials during and after any disaster. These sites operate as critical emergency communication venues during crises, telling us where people are located during a disaster, where they are moving and in some cases where they are checking in as “safe.” In practice, this functionality leads people to use them as a supplement to overloaded 911 and other emergency systems. To adequately take advantage of this reality, traditional emergency systems would need to integrate this data. As response moves to recovery, the information collected from these sites can help assess damage at a finer grain by combining it with immediate damage estimates from the Federal Emergency Management Agency (FEMA) and other entities.

The ubiquity of mobile devices, the internet of things and social media platforms creates new opportunities to rethink our current disaster response system. These platforms allow individuals not traditionally part of a conversation to have an equal footing in disseminating information. Leveraging crisis informatics is critical to the response of an emergency disaster because they present a live view of disasters as they unfold. However, crises informatics are not a replacement for traditional damage and need assessments, but rather, as this report proposes, they can substantively supplement our current system to create a robust response system.

By developing new protocols built from live information into response and recovery systems, we can improve efforts by more accurately assessing damage and using that information to allocate vital recovery resources. While social media data has a variety of potential applications, this report uses data from Twitter and from CrowdSource Rescue, a crowdsourcing platform set up during Hurricane Harvey, to show how crisis informatics could supplement traditional damage estimates.

Twitter is chosen because it has been historically used as a method of sharing instant news, thoughts and ideas about critical events. Twitter data is supplemented with information from CrowdSource Rescue because it was a site heavily used during the Hurricane Harvey.
Data

Using the PowerTrack 2.0 API aggregation tool created by Gnip, Inc., over 1 million tweets are collected spanning August 25, 2017 through September 8, 2017. This period was the most relevant window for rescue requests and distress signals. This study does not rely on users with location services enabled. If users have location services enabled, tweets are geotagged to an approximate latitude and longitude, however, only one percent of users enable geotagging. This study relies on geographic information retrieval inferencing to approximate the location of rescue and evacuation requests from Twitter users. Damage estimates from all three datasets are buffered to present a more robust picture; buffering estimates increases robustness because FEMA estimates are not always accurate, not all individuals are active Twitter users or heard about CrowdSource Rescue.

The search criterion for tweets was conducted by selecting public tweets using any of a number of Harvey-related hashtags in their tweets. Hashtags are a mechanism to broaden the discussion of the tweet by tagging it to a specific topic. See Appendix A for the complete list of hashtags used to collect data for this report.

This information is supplemented with by using water rescue requests from CSR as a proxy for flooding damage. Using a crowdsourcing approach, CSR allows users to request a water rescue and share some basic information (health issues, number of adults and children, number of pets, cell phone numbers, etc.) while allowing spontaneous rescuers to share information if they have a boat or truck to provide a rescue. This method allows users to quickly connect resources during the disorder of immediate disaster response. During Hurricane Harvey, CSR was able to collect over 15 thousand records and 5,200 rescue requests.
After Hurricane Harvey, FEMA created flood damage estimate maps for areas impacted by the storm. These maps in turn, are often used as an essential resource for nonprofit and government bodies looking to respond and mitigate any damage to residents. However, the use of these maps is limited if their delivery is slow or if they fail to capture significantly damaged areas. Twitter and CSR data were available virtually as individuals requested rescue and evacuation. These estimates show how crisis informatics could be leveraged in the future to better understand the geography of damage and, by extension, needs.

FEMA

The study uses parcel-level data from the 2017 Homeland Infrastructure Foundation (HIFLD) and creates parcel-level damage estimates from the FEMA flood estimate data to represent the official estimates of potential damage. Addresses are then identified in the Twitter data and geocoded to the parcel-level data. The estimates are compared and buffering is used to improve the likelihood of real damage.

The parcel data from HIFLD was used to derive a structure type, inundation extents were created based on event data, the inundation data layer was used to assign a z-value (Raster value) and associated damage level. The data assumes that the parcel centroid is where the building was located, it did not include elevation of structure, vacant parcels were removed, and the methodology does not account for flooding as a result of storm water backups, irrigation ditch failures, flooding from dam or levee breaks, use of spillways and weirs or from other storm damage.

For FEMA damage level estimates, any level of damage is considered, whether it was major or minor, to be an indication of flooding. Figure 1 displays the familiar image many nonprofit and governmental entities considered when evaluating where to allocate recovery resources.

FEMA damage estimates are buffered by 75 meters; because the information is an estimate at the parcel level, this represents only an expected damage for an individual unit, however, when building units are located next to each other, similar damage among neighboring buildings is expected.
RESULTS

FIGURE 1

FEMA Parcel Level Flood Damage Estimates (75 Meter Buffer)

FIGURE 2

Twitter Parcel Level Flood Damage Estimates (250 Meter Buffer)
Lessons from Harvey: Crisis Informatics for Urban Resilience

Twitter

Twitter data was mined to find tweets with numbers and evidence of street endings (ln, rd, st, court, blvd, hwy, sq, plaza, cir, etc.). City names and zip codes were then identified for the subset of tweets and geocoded to parcel level data.

Results yielded over 1,300 useable addresses from Twitter, of these 72 percent matched an exact street address and 67 percent matched an exact parcel. Twitter estimates are buffered by 250 meters. Tweets are buffered at a greater level because of the sparsity of the number of tweets. Realistically, most individuals were not compelled to tweet their address information for a rescue unless they are active users of the platform. Figure 2 displays the tweets that were geolocated to a parcel, buffered to include parcels within 250 meters.

CrowdSource Rescue

CSR provided de-identified non-rescuer user data with latitude and longitude estimates of rescue requests. The data included over 5,200 rescue requests with approximate latitude and longitude estimates, the centroid of 34 percent of the rescue requests were geolocated to an exact parcel, which was then buffered by 50 meters. Figure 3 displays the point estimate of all rescue requests.

The Whole Picture

Each of the three datasets provides useful information for the recovery phase after a disaster. They provide an indication of damage, but only the Twitter and CSR information would have been available to nonprofit and governmental agencies. It is significant to say by themselves, Twitter and CSR are not a replacement of FEMA damage estimates but rather, corroborate the same information earlier and manage to capture areas missed by the FEMA estimates. Figure 4 displays a more complete image using all three data sources, with a zoomed in image of the Highway 90 corridor in the eastern part of Houston.

The zoomed image of Figure 4 displays the three datasets at work, corroborating each other in several instances. Purple shows validation between Twitter and FEMA; brown-yellow, validation between CSR and FEMA; blue-green, validation between CSR and Twitter. At first glance the image is evocative, but at a finer inspection it reveals a more complete story than the FEMA estimates alone.
FIGURE 4  
FEMA, Twitter, and CSR Damage Estimates  
(75M FEMA, 50M CSR & 250M Twitter Buffer)
**Case Study:**
**C.E. King Middle and High Schools**

As previously stated, damage estimates based on limited information are problematic if they fail to cover all impacted areas. As a case study, Figure 5 shows a zoomed in image of two local schools, C.E. King Middle School and C.E. King High School. The figure shows the school relatively safe from expected damage levels in the FEMA dataset.

However, the schools were, in fact, flooded. Figure 6 displays images of the CE King High School in the aftermath of Harvey. Of course, this information would have been disseminated through word of mouth to the relevant officials, but this would have traveled at a much slower rate than what could be mined from Twitter.

Figure 7 adds the Twitter damage estimates using a 250-meter buffer. The surrounding neighborhoods suffered similar flooding and residents asked for rescue and evacuation requests using the Twitter and CSR platform. The crisis informatics leveraged from these datasets aligned with the FEMA estimates—corroborating their efficacy—but provided additional information and provided this information immediately as it was disseminated.

---

**FIGURE 5**

CE King High School and FEMA Damage Estimates

---

**Tweet:**

please send help, my friend’s boyfriend is stuck with two cats and two dogs, water is rising

425 Shane St.
Houston TX 77036
United States
11:40 PM - 27 Aug 2017
115 Retweets 53 Likes
FIGURE 6
Images of C.E. King High School

FIGURE 7
CE King High School, FEMA Damage Estimates, and Twitter Damage Estimates
This report does not directly rely on location services data, i.e., it does not rely on geolocated information from service providers. However, the study shows how data from social media and crowdsourced sites can be leveraged to supplement official data sources.

Location services data has several usages, from advertising and commercial purposes, political campaign messaging or emergency services, the latter of which, is increasingly being explored and contextualized. Geolocation information lets us know when there is danger nearby or when weather events are imminent, but this information can be further leveraged to assist nonprofits and governmental entities in providing aid to those affected. By providing de-identified data, social media firms, such as Twitter and Facebook, can help nonprofits and governmental entities identify who is in need, where they are located, where they are moving and if their needs have been satisfied. To use this data effectively, public agencies will need to have plans in place before future crises hit.

De-identified data is essential to protect the identity of individuals. While this information could be provided directly by the social media platforms in a condensed and aggregated fashion, there could also be partnerships developed with research and nongovernmental institutions to act as the interpreters of data for nonprofit and governmental bodies. Such a relationship could be managed with the social media firm providing raw data to a research institute, an entity that can operate with the utmost data security standards, while the nonprofit and governmental entities provide the inquiries and a list of available resources to the research institute. The partner research institutes could then use these datasets to identify the most relevant information for humanitarian aid.

Of course, this data is only representative of the users who actually use the social media platforms and enable location services or voluntarily share geographical information. This, naturally, leads to doubts about social media information being generalizable to a larger disaster situation. However, while very specific individual data is not generalizable, this does not mean useful information cannot be drawn from social media users. Indeed, as this report shows, such a decision could hamper damage assessment and other recovery efforts.
The nature of Twitter and other social media platforms allows for immediate and informal participation in an online discussion. This creates a space of latent data and information during disasters and more generally, emergency situations. The study shows how Twitter and crowdsourced sites can be used to leverage this data into credible information. Significantly, these methods don’t replace traditional data sources but rather make response and recovery more robust.

People often do not know where to get assistance after a disaster or how the process works, similarly those helping do not know who needs assistance and wish to optimize their funds by giving to the most vulnerable and in need first. Twitter data can contribute to this discussion by offering a platform for information sharing which can be leveraged by individuals to supplement their knowledge and the providers of aid by assessing a platform for who actually needs aid.

While this information is not representative of all Twitter users, or the general population for that matter, the data, nonetheless, significantly increase awareness and unveil a more complete image than traditional data sources.
Tweets for this study were requested through Gnip, Inc., a social media API aggregation firm, using the PowerTrack 2.0 API. The PowerTrack leverages the full Twitter firehose, or the complete stream of public messages, to filter and deliver relevant data.

Tweets were filtered on conditional words, mentions or hashtags containing:
- “harvey,”
- “hurricane,”
- “hurricaneharvey,”
- “hurricaneharvey2017,”
- “hurricaneharvy,”
- “hurricaneharvery,”
- “harveystorm,”
- “houston,”
- “houstonstrong,” or
- “soshouston.”

These words represent some of the widely used mentions and hashtags throughout the storm. On top of these unique mentions the tweets needed to contain the word, mention or hashtag:
- “urgent,”
- “rescue,”
- “cajunnay2016,” or
- “uscg”.

Only tweets identified as English and original content (not a retweet) were collected. Further, tweets were filtered for germaneness; tweets with the word, mention or hashtag containing “steve,” “weinstein,” “pray,” or “praying” were excluded. Tweets with the words “thoughts” and “prayers” within three tokens and the words “thinking” and “victims” within five tokens were excluded to omit well wishes. Tokens are characters separated by a separator, such as a space. The dataset represents the closest approximation to tweets focused on rescue, emergency and distress calls during the storm.
References


Lessons from Harvey: Crisis Informatics for Urban Resilience

U.S. Army Corps of Engineers
The Kinder Institute thanks the following contributors for their transformational support of our mission to build better cities and improve people’s lives.

KINDER FOUNDATION

HOUSTON ENDOWMENT

bp

Reinnette and Stan Marek*

Chevron

*K denotes multi-year commitments to the Kinder Institute.

CenterPoint Energy

SIS and Hasty Johnson

United Way

Becky and Ralph O’Connor*

THE ROCKEFELLER FOUNDATION

Aliyya and Herman Stude‡

*K denotes multi-year commitments to the Kinder Institute.

‡ Includes gifts made through family foundations, donor-advised funds, or other organizations.

Current as of July 31, 2018

Additional support comes from the Friends of Kinder Institute, Kinder Institute Corporate Council and Kinder Institute Supporting Foundations and Funds.
Mission
The Kinder Institute for Urban Research builds better cities and improves people’s lives by bringing together data, research, engagement and action.