

## Integrated Natural Language Processing and Meta-network Analysis for Social Sensing of Location-Event-Actor Nexus in Disasters

Chao FAN<sup>1</sup>, Yucheng JIANG<sup>2</sup>, and Ali MOSTAFAVI<sup>3</sup>

<sup>1</sup> Urban Resilience, Networks and Informatics Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77840-3136; PH (530) 761-3633; email: chfan@tamu.edu

<sup>2</sup> Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77840-3136; PH (812) 349-8215; email: jiang50@tamu.edu

<sup>3</sup> Urban Resilience, Networks and Informatics Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77840-3136; PH (765) 543-4036; email: amostafavi@civil.tamu.edu

### ABSTRACT

The objective of this study is to propose and test a methodological framework that integrates natural language processing and meta-network analysis to leverage social media data for situation assessment and action prioritization in disasters. Situation assessment based on evaluation of location-event-actor nexus (i.e., what event is occurring where and who is working on what response actions) is essential for residents, volunteers, and relief organizations to effectively manage response actions to cope with natural disasters. This study proposed an integrated framework to parse social media data and analyze location-event-actor meta-networks for examining location-specific disaster situations. The proposed framework consists of three components: named entity recognition, convolutional neural networks, and meta-network analysis. First, a named entity recognition model is utilized to identify locations and organizations that are reported in posts on social media. Then, using a taxonomy for humanitarian event categories, a Convolutional Neural Network model is developed and trained to classify the tweets and identify disruptive events that occurred in specific locations. Finally, the relationship among locations, disruptive events, and actors/organizations are captured and modeled as a meta-network. The meta-network analysis provides indicators to assess the urgency of the events and the needs in the locations for prioritization of relief and rescue efforts. The application of the proposed framework is implemented on a case study of Hurricane Harvey in Houston. The results illustrate the capabilities of the proposed framework for enhanced social sensing of disaster situations and prioritizing relief actions.

### INTRODUCTION

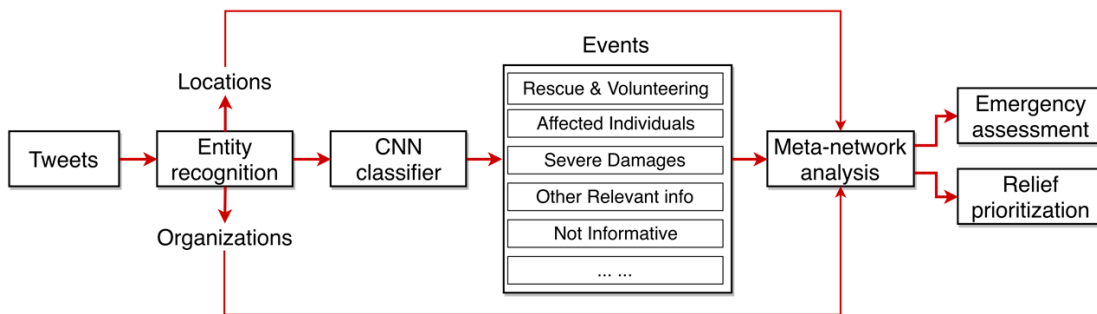
Being aware of the disaster situations and assessing the performance of response actions at a specific location is important for efficient and effective disaster management. To achieve this, one of the key components is to collect reliable information in disasters and extract insights to inform disaster response. Current data collection and analytics technologies such as remote sensing and satellite images have a number of shortcomings such as time lags, high costs, and limited spatial resolution. Complementing physical sensing, social sensing provides unique opportunities for timely gathering of inexpensive and high-resolution information related to disaster

situations. The rapidly increasing adoption and use of social media platforms in disasters has made social sensing an emerging element of situation information mapping in disasters and crises.

Although the potential of social media data for disaster management has been emphasized in existing studies (Zhang et al. 2019), important methodological limitations still persist. In particular, existing studies mainly focus on developing and implementing machine learning algorithms and techniques to classify images and texts (Fan and Mostafavi 2019b), map the scales of impacts (Jongman et al. 2015), and assess human sentiments (Kryvasheyeu et al. 2016) for better understanding of disaster situations. An important methodological gap exists to examine the location-event-actor nexus. The location-event-action nexus is the foundation for effective disaster management since it informs about what disruptive events (e.g., road inundation) has occurred at a specific location and what organization/actor is acting upon it. Hence, this study proposed a methodological framework to parse social media data and automatically capture and analyze location-event-actor meta-networks. The application of the proposed methodology is shown in a case study of disaster situations during Hurricane Harvey in Houston.

**THE INTEGRATED FRAMEWORK**

The proposed framework involves three components: Named Entity Recognition (NER), Convolutional Neural Networks (CNN), and Meta-network analysis (see Figure 1). The inputs for this framework are the tweets collected during the period of disasters or crises, and the outputs include the efficiency of relief execution in meta-networks for different locations. The NER component can identify the locations and organizations/actors mentioned in the tweets. The CNN classifier classifies the tweets into different event categories. The links among locations, actors, and events in meta-networks would be built when they appear in the same tweets. Through the analysis on the meta-networks, we can assess disaster situations and status of relief needs to better allocate the resources to affected areas.



**Figure 1. The proposed integrated methodological framework**

**Named entity recognition**

Locations, events and actors are three major types of entities in disaster management systems and processes. Precisely understanding what organizations/actors work at which locations in response to what events can improve situational awareness and relief operations. Named entity recognition is an entity classifier which classifies the entities that are present in a text into predefined categories such as “person”,

“location”, and “organization/actor” (see Figure 2). Meanwhile, the named entities are not always single words and do not remain the same across all disasters and crises. To achieve high accuracy and generalization of the method in different disaster contexts, a general pretrained model to recognize the locations and organizations in tweets is essential. Stanford NER has been demonstrated as a valid tool for a general implementation of entity recognition (Finkel et al. 2005). Natural Language Toolkit (NLTK), a Python package, provides a set of natural languages corpora and APIs of wide varieties of NLP algorithms including Stanford NER. Hence, we employed the NLTK package to identify and tag the locations and organizations/actors in our tweets (see Figure 2) and extract them for creating the meta-networks.

**Tweet 1:** Truckers for Texas is a Facebook *Organization* page set up by George Parker *Organization* and Bill Weaver *Organization* to raise money for victims of Harvey.

**Tweet 2:** Houston Patch: Hurricane Harvey: Flood Control District *Organization* Releases More Water Into Buffalo Bayou *Location*

**Figure 2. Examples of Named Entity Recognition in tweets**

### Convolutional neural networks

Convolutional Neural Networks (CNN) perform very well on text classification problems. Current CNN models to classify the texts have varying architecture with different layers. In this study, we adopt a single layer CNN which is shallow and wide but can achieve high accuracy for our classification tasks (see Figure 3). That is because, the meaning of the tweets lies in some keywords rather than the sequence of the words and their combination patterns, which is emphasized in Recurrent Neural Networks. In particular, the sequence of words is of less importance in classification purposes (compared with other applications). Based on this rationale, we do not consider the sequence of the words and use pre-trained word embedding approaches to convert the tweets into vectors.

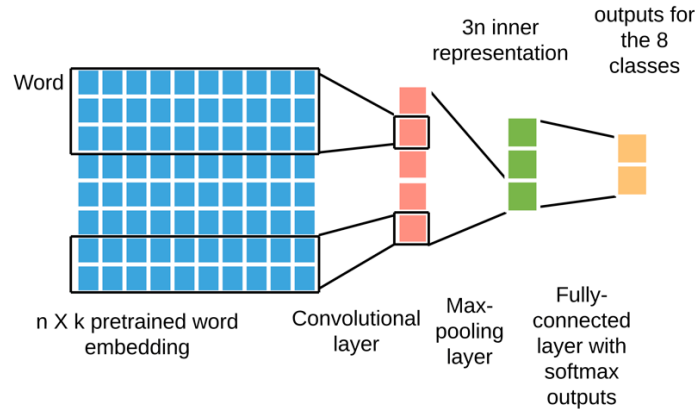
First, we employ a commonly used pre-trained embedding model, named “Glove”, which is trained based on 2 billion tweets (Pennington et al. 2014) to convert the words into vectors. The Glove model includes 27 billion tokens and 1.2 million vocabulary elements can convert each word into a 200-dimensional vector and has been shown to have high performance in capturing the meaning of the words and minimizing the loss of information in multiple existing studies (Levy et al. 2018; Levy and Goldberg 2014). The combination of the word vectors forms embedding matrices representing each tweet.

Then, the embedding matrices go through the convolutional layer which has three types of filters to reduce the dimensionality of the matrixes. The heights of the filters are 3, 4, and 5, while the width of the filter is 200 in accordance with the fine-tuning of parameters in existing studies (Zeppelzauer and Schopfhauser 2016). We perform the convolution operator between the embedding matrix and the internal filter to learn a feature  $c_i$ , which is generated from a window of words  $x_{i:i+j-1}$  using:

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+j-1} + b) \quad (1)$$

where  $b \in \mathbb{R}$  is a bias term and  $f$  is a non-linear function (Mouzannar et al. 2018). We use ReLU as the activation function in this model.

Finally, the outputs of the convolutional layer go into the max-pooling layer, a non-linear subsampling function enabling a summarization of the outputs of neighboring groups of the elements in matrixes. By feeding the outputs from the max-pooling layer into a fully-connected layer, we can apply dropout and Softmax functions to assign the humanitarian categories to the tweets.



**Figure 3. The architecture of the CNN classifier**

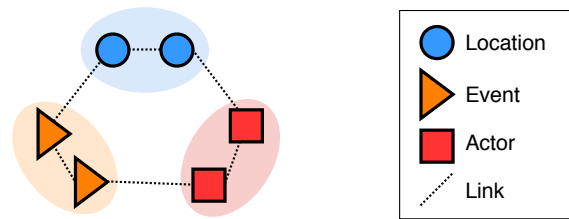
### Meta-network modeling

Meta-network modeling enables modeling systems with heterogeneous and interactive entities in complex emergencies (Fan and Mostafavi 2019a). Different from conventional network models with a single type of nodes, meta-networks include multi-type nodes and relationships that can capture the complex interactions among organizations/actors, locations, and events during disasters (Zhu and Mostafavi 2018). In addition, meta-network modeling allows us to develop measurements for node centrality adapting to specific contexts. In this study, meta-network models are adopted to examine location-event-actor nexus in disaster situations.

The first step in the meta-network analysis is to specify entities and their relationships. As shown in Table 1, there are 3 types of nodes and 7 types of links, presented in the meta-matrix. The elements of the meta-network model is shown in Figure 4. The links can be abstracted between single-type nodes as well as multitype nodes. Each set of nodes and their relationships form a distinct network (Fan et al. 2018). For example, the actors/organization nodes and their relationships form the cooperation network. A cooperation network in the context of disasters represents communication or collaboration among various relief organizations. In addition, the location and event nodes and the links connecting these types of node form the location-condition network, representing the disruptions occurring in these locations. For example, water release (event) can connect to downstream areas of Addicks and Barker reservoirs (location) in the location-condition network. Similarly, other types of networks shown in Table 1 can be specified and modeled in the meta-network. These 7 individual networks are interconnected via shared nodes. As such, the meta-network model can be established.

**Table 1. Meta-matrix for Conceptualization of the Disaster Situation**

<b>Networks</b>	<b>Location (L)</b>	<b>Event (E)</b>	<b>Actor (A)</b>
Location (L)	Spatial interdependency network	Location condition network	Location relief network
Event (E)		Event interdependency network	Relief execution network
Actor (A)			Actor cooperation network



**Figure 4. An illustration of meta-network modeling for disaster situation**

Second, based on the meta-network model, we develop measures to assess the performance of the entities involved in the meta-network. An important indicator for examining the performance of disaster response is the efficiency of relief execution. That is, each disruptive event (e.g., road closure, building damage, and utility outage) should be responded by at least one relief organization at a specific location. A disruptive event would be resolved more efficiently if more than one relief organizations would work on it. Otherwise, the disruptive event would continue due to lack of response actions and low efficiency of relief management. Accordingly, the efficiency of relief execution can be formulated as:

$$E = \frac{N_e^o}{N_e^t} \quad (2)$$

where  $E$  is the efficiency of relief execution,  $N_e^o$  is the number of events that relief organizations work on, and  $N_e^t$  is the total number of events in the meta-network. The value of the execution efficiency ranges from 0 to 1, which is independent of the number of events. The formula can be generalized to different locations and disaster context. The closer the value of  $E$  to 1, the more efficient the whole disaster response actions. Finally, based on the outcome of the analysis, one can evaluate the performance of disaster responses and develop response strategies. By comparing the efficiency of multiple locations, we can identify which locations need more urgent attention. Furthermore, different events may occur simultaneously in a location. The meta-network allows us to identify which events lead to the decrease of relief efficiency and prioritize the response actions to increase efficiency. For example, some relief organizations are working on search and rescue, while none of the organizations put efforts on repairing broken power cables. The proposed social sensing methodology can identify such missing response actions.

**CASE STUDY**

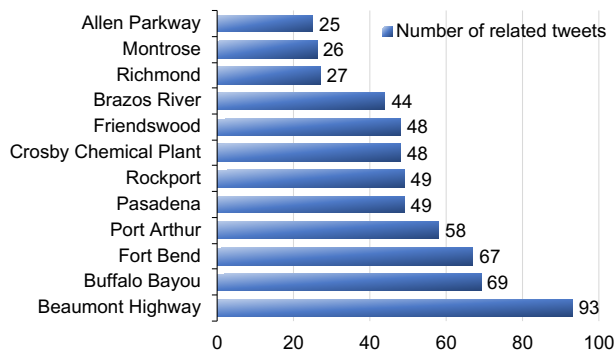
To demonstrate the application of the proposed integrated framework, a case study related to Hurricane Harvey in Houston was conducted.

**Disaster context and data collection**

Hurricane Harvey is a category four tropical storm which landed in Houston on August 27 and passed on August 30, 2017 (Amadeo 2019). The heavy rainfall and strong wind brought led to large-scale flooding, which inundated hundreds of thousands of homes. To implement and test the framework, we collected about 21 million tweets over the Houston area from August 22 to September 30 using user-profiles and tweet geotags. This dataset is composed of all the tweets posted by users whose profile locations are in Houston or the tweets tagged by geo-coordinates which are in the bounding box of Houston. Because this paper considers the situation during Hurricane Harvey to demonstrate the application of the proposed methodology, we used a smaller dataset which ranges from August 27 to Sept 2 (one week) and filtered the original tweets that are related to Hurricane Harvey using keywords filtering. Once the dataset was prepared, we conducted NER and CNN for constructing meta-network models to analyze the disaster situations.

**Results**

Using the Stanford NER model, we identified the locations and actors/organizations present in the tweets. Figure 5 shows the number of related tweets for different locations in our dataset. To demonstrate the capability of our framework, in this case study, we primarily investigated two cases of disruptive events: Allen Parkway and Crosby Chemical Plant. Allen Parkway was submerged during Hurricane Harvey. The road has spatial co-location with many other infrastructure including Buffalo Bayou and Memorial Drive. Crosby Chemical Plant is a location where an explosion was reported due to flooding during Hurricane Harvey. The explosion caused severe impacts, such as toxic air and polluted water. Multiple relief organizations were involved in dealing with these problems.



**Figure 5. Top-ranked locations mentioned in Harvey-related tweets**

To train the CNN model, we adopted a published dataset from (Alam et al. 2018), which contains 4,000 labeled tweets (3533 valid tweets) related to Hurricane Harvey, starting from Aug. 26 to Sep. 20. The humanitarian categories labeled for the tweets include:

- **Other relevant:** the text is related to the disaster, but it does not belong to any of the following humanitarian categories; for example, *“The weather is improving, and they are expecting 1-2” of rain on our end of town today and tonight so fingers crossed the water levels will remain or possibly start to go down slowly”*.
- **Rescue:** the text is related to the rescue efforts such as donations, search and rescue, and volunteering; for example, *“The Harris County Office of Homeland Security & Emergency Management invites residents affected by Hurricane Harvey to a recovery fair this weekend”*.
- **Infrastructure:** the text is related to infrastructure disruptions such as road closure, power outage, and utility damages; for example, *“Has anyone found a way to get to I10 from south of the bayou? I am at Eldridge and Enclave”*.
- **Non-relevant:** the text is not related to the disasters.
- **Affected individuals:** the text is related to the people who are in the disaster situation and need relief aid; for example, *“Does anyone have a canoe/boat/float of some sort to use tomorrow morning to get to our homes on Silvergate?”*.
- **Injured people:** the text is related to injured people.
- **Vehicle:** the text is related to vehicle damages such as bus, car, train, and boats.
- **Missing people:** the text is related to missing people.

Figure 6 shows the distribution of the humanitarian categories of the tweets in the training set. The number of related tweets in different categories is uneven. Most of the tweets are in the category of “other relevant” and “rescue”. This may lead to learning bias in our CNN model. To mitigate this issue, we constructed new training set by sampling examples with replacement when training our CNN model (bootstrap).

The training performance and validation accuracy of the CNN model are shown in Figure 7. We observe that the training loss of the model decreases gradually with the increase in the number of training steps. The final training loss is lower than 0.1. We further apply the model to the validation dataset. Figure 7(b) suggests that the validation performance increases very fast after several steps and then maintain a high level of accuracy, which is about 0.689. The measurements show the stability and robustness of the CNN model in learning tweets for humanitarian categories.

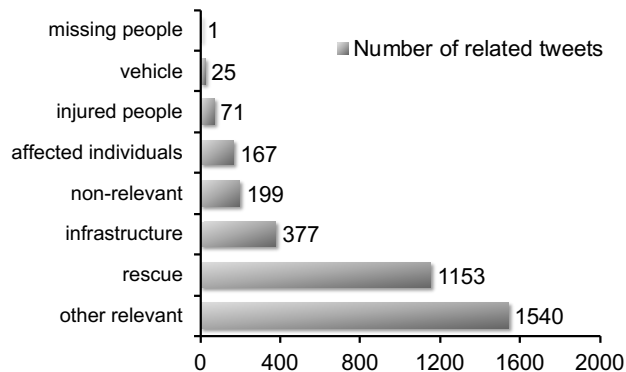
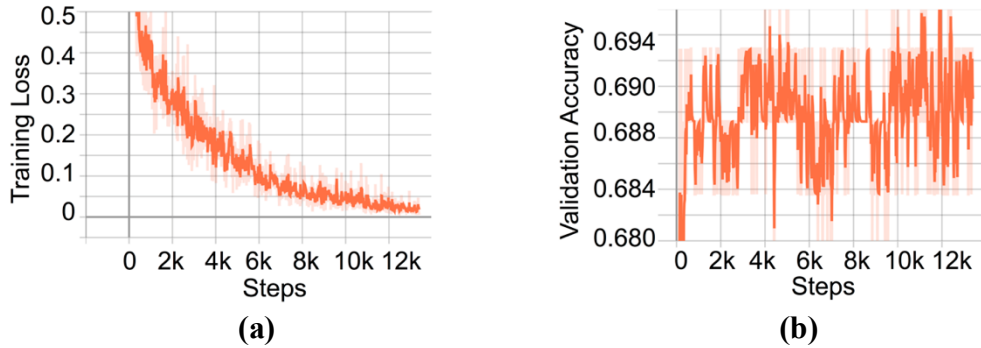


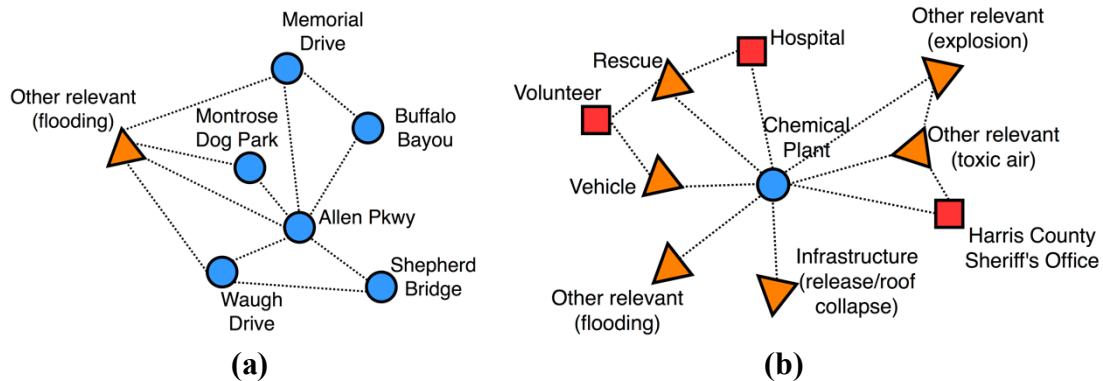
Figure 6. Distribution of the humanitarian categories in training set





**Figure 7. Training loss (a) and validation accuracy (b) of the trained CNN model**

Once the CNN model is trained, we applied it to our filtered datasets for learning relevant events in tweets and creating meta-networks. Figure 8 shows the meta-networks for both cases: Allen Parkway and Chemical Plant. The meta-network for Allen Parkway is the nexus of spatial interdependency network and location condition network, because it includes multiple interdependent infrastructures and one significant event, flooding. The efficiency of relief execution at Allen Parkway is 0 because no relief organization/actor was mentioned in tweets to deal with the flooding event. Scanning the information in tweets, some nearby connected infrastructure such as Montrose Dog Park and Waugh Drive were flooded together with the Allen Parkway, while other interdependent infrastructure such as Buffalo Bayou is a source of floodwaters. The meta-network for Chemical Plant includes multiple events and relief actors, but no interdependent location was mentioned. The relief efficiency at the Chemical Plant is 0.5 because 3 relief actors (i.e., hospital, volunteer, and Harris County Sheriff’s Office) involved in 3 events (i.e., rescue, vehicle, and toxic air). The explosion of the Chemical Plant did not affect other locations but led to significant impacts in the surrounding ecosystem (e.g., air).



**Figure 8. Meta-networks for Allen Parkway (a) and Arkema Chemical Plant (b) in Houston during Hurricane Harvey**

**Validation**

To validate the credibility of the results generated by our integrated framework in this case study, we examined the extracted information in both cases by comparing to recorded information from news articles published during and in the aftermath of



Hurricane Harvey. Houston Chronicle reported the flooding situation at Allen Parkway and other connected roads (Gordon 2017). Meanwhile, “*The flood sparked multiple fires and explosions, and caused authorities to evacuate about 200 people in the Chemical Plant in Houston*” (Osborne 2018). The news articles indicate that the information identified from the tweets and mapped in the meta-networks are credible.

## **CONCLUDING REMARKS**

This paper proposed an integrated framework combining natural language processing and meta-network analysis for social sensing of location-event-actor nexus in disasters. The application of the proposed framework was demonstrated in a case study of Hurricane Harvey in Houston. The proposed method contributes to increased visibility of the disaster situations and management process. As the first attempt in social sensing of location-event-actor nexus in disasters, this study has some limitations. First, Twitter data may contain noise such as non-sense symbols, punctuation, and misspelled words. Therefore, to mitigate the adverse impacts of the noise in tweets, preprocessing the texts in the tweets by removing punctuations and emoticons is important. However, some emoticons may be informative for the content of tweets. Thus, a fine-grained preprocessing is needed. Second, the representativeness of the social media data for mapping the meta-network is always an important consideration. For example, the absence of actors’ info in a tweet does not necessarily mean that an organization is not acting upon a disruption event. However, as we discussed earlier, social sensing is a complementary approach to provide evidence for physical sensing and other data sources. The proposed framework can also be generalized to other sources of data (e.g., Facebook and Instagram). By fusing multiple data sources and implementing our framework on it, we can have a more comprehensive list of nodes and links. Then, the analysis and results would enable an enhanced understanding of disaster situations to inform response and relief processes.

## **ACKNOWLEDGEMENT**

This material is based in part upon work supported by the National Science Foundation under Grant Number IIS-1759537, and the Amazon Web Services (AWS) Machine Learning Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and Amazon Web Services.

## **REFERENCES**

- Alam, F., Ofli, F., and Imran, M. 2018. “CrisisMMD: Multimodal Twitter Datasets from Natural Disasters.”
- Amadeo, K. 2019. “Hurricane Harvey Facts: Damage and Costs.” *The Balance*, <<https://www.thebalance.com/hurricane-harvey-facts-damage-costs-4150087>> (Jun. 3, 2019).
- Fan, C., and Mostafavi, A. 2019a. “Metanetwork Framework for Performance Analysis of Disaster Management System-of-Systems.” *IEEE Systems Journal*, 1–12.
- Fan, C., and Mostafavi, A. 2019b. “A graph-based method for social sensing of infrastructure disruptions in disasters.” *Computer-Aided Civil and Infrastructure Engineering*, John Wiley & Sons, Ltd (10.1111), mice.12457.

- Fan, C., Zhang, C., and Mostafavi, A. 2018. "Meta-Network Framework for Analyzing Disaster Management System-of-Systems." *2018 13th Annual Conference on System of Systems Engineering (SoSE)*, IEEE, Paris, France, 372–378.
- Finkel, J. R., Grenager, T., and Manning, C. 2005. "Incorporating non-local information into information extraction systems by Gibbs sampling." *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05*, Association for Computational Linguistics, Morristown, NJ, USA, 363–370.
- Gordon, M. 2017. "Then-and-now photos show how Hurricane Harvey flooded Houston neighborhoods - Houston Chronicle." *Houston Chronicle*, <<https://www.houstonchronicle.com/news/houston-weather/hurricaneharvey/article/Then-and-now-photos-show-how-Hurricane-Harvey-13155174.php>> (Aug. 5, 2019).
- Jongman, B., Wagemaker, J., Romero, B., and de Perez, E. 2015. "Early Flood Detection for Rapid Humanitarian Response: Harnessing Near Real-Time Satellite and Twitter Signals." *ISPRS International Journal of Geo-Information*, 4(4), 2246–2266.
- Kryvasheyev, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. 2016. "Rapid assessment of disaster damage using social media activity." *Science Advances*, 2(3).
- Levy, O., and Goldberg, Y. 2014. "Neural Word Embedding as Implicit Matrix Factorization." *Advances in Neural Information Processing Systems (NIPS)*.
- Levy, O., Goldberg, Y., and Dagan, I. 2018. "Improving Distributional Similarity with Lessons Learned from Word Embeddings." *Transactions of the Association for Computational Linguistics*.
- Mouzannar, H., Rizk, Y., and Awad, M. 2018. "Damage Identification in Social Media Posts using Multimodal Deep Learning." *15th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2018)*.
- Osborne, M. 2018. "Arkema chemical company indicted for plant fire after Hurricane Harvey - ABC News." *abcNews*, <<https://abcnews.go.com/US/arkema-chemical-company-indicted-plant-fire-hurricane-harvey/story?id=57029025>> (Aug. 5, 2019).
- Pennington, J., Socher, R., and Manning, C. D. 2014. "GloVe: Global Vectors for Word Representation." *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Sebastian, T., Lendering, K., Kothuis, B., Brand, N., Jonkman, B., van Gelder, P., Godfroij, M., Kolen, B., Comes, T., and Lhermitte, S. 2017. "Hurricane Harvey Report: A fact-finding effort in the direct aftermath of Hurricane Harvey in the Greater Houston Region." Netherlands.
- Zeppelzauer, M., and Schopfhauser, D. 2016. "Multimodal classification of events in social media." *Image and Vision Computing*, 53, 45–56.
- Zhang, C., Fan, C., Yao, W., Hu, X., and Mostafavi, A. 2019. "Social media for intelligent public information and warning in disasters: An interdisciplinary review." *International Journal of Information Management*.
- Zhu, J., and Mostafavi, A. 2018. "Enhancing Resilience in Disaster Response: A Meta-Network Analysis Approach." *Construction Research Congress 2018*, (1), 2250–2259.