

# A Twitter Data Credibility Framework —Hurricane Harvey as a Use Case

Jingchao Yang, Manzhu Yu, Han Qin, Chaowei Yang

George Mason University

NSF Spatiotemporal Innovation Center



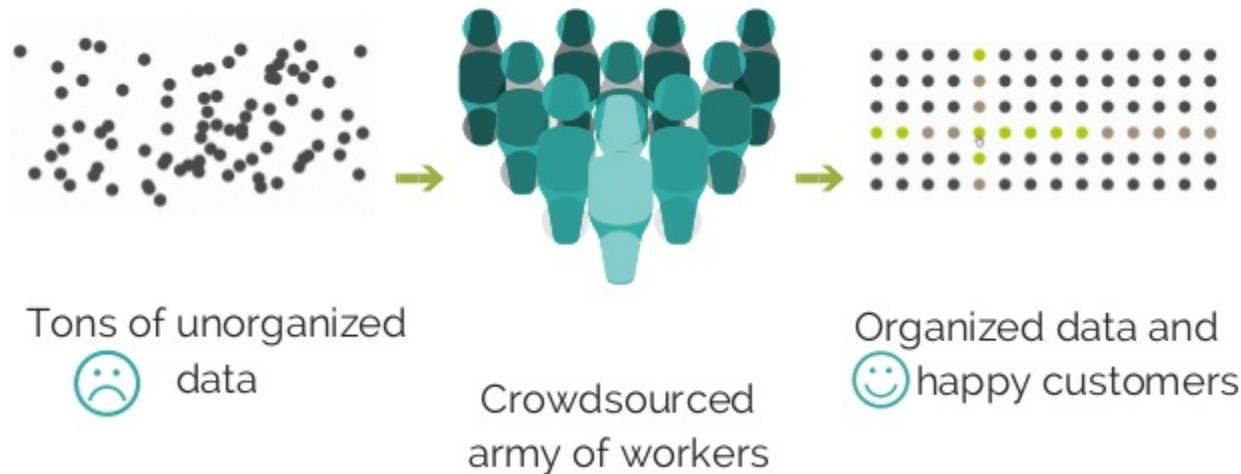


# Outline

- Crowdsourcing Data in Emergency Management
- Hurricane Harvey Background and Data
- Overall Structure of The Credibility Framework
- Methodology
- Credibility Analysis
- Discussion



# Crowdsourcing Data in Emergency Management



<https://www.slideshare.net/SquadRun/when-to-use-crowdsourcing-to-manage-your-data-45694706>

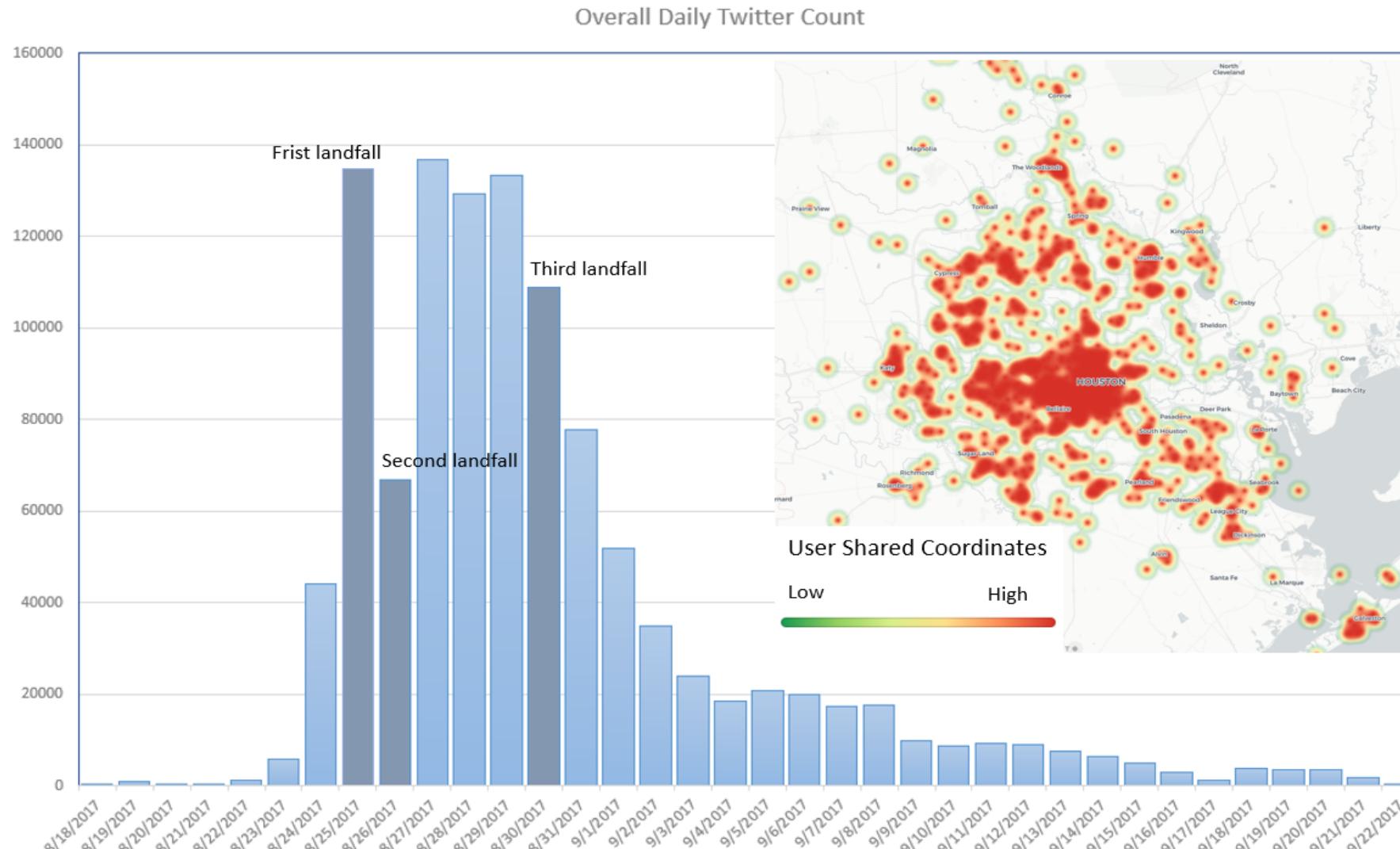
- During a natural disaster, collected social media data are commonly massive, and only a small portion of the data is related and contributing to situation awareness. Therefore, extracting essential and related information from social media data is crucial for situation awareness.

- Year 2017
- One of the costliest tropical cyclone on record
  - Inflicting \$125 billion in damage
- Most affected area
  - Houston metropolitan area and Southeast Texas

## Hurricane Harvey Background



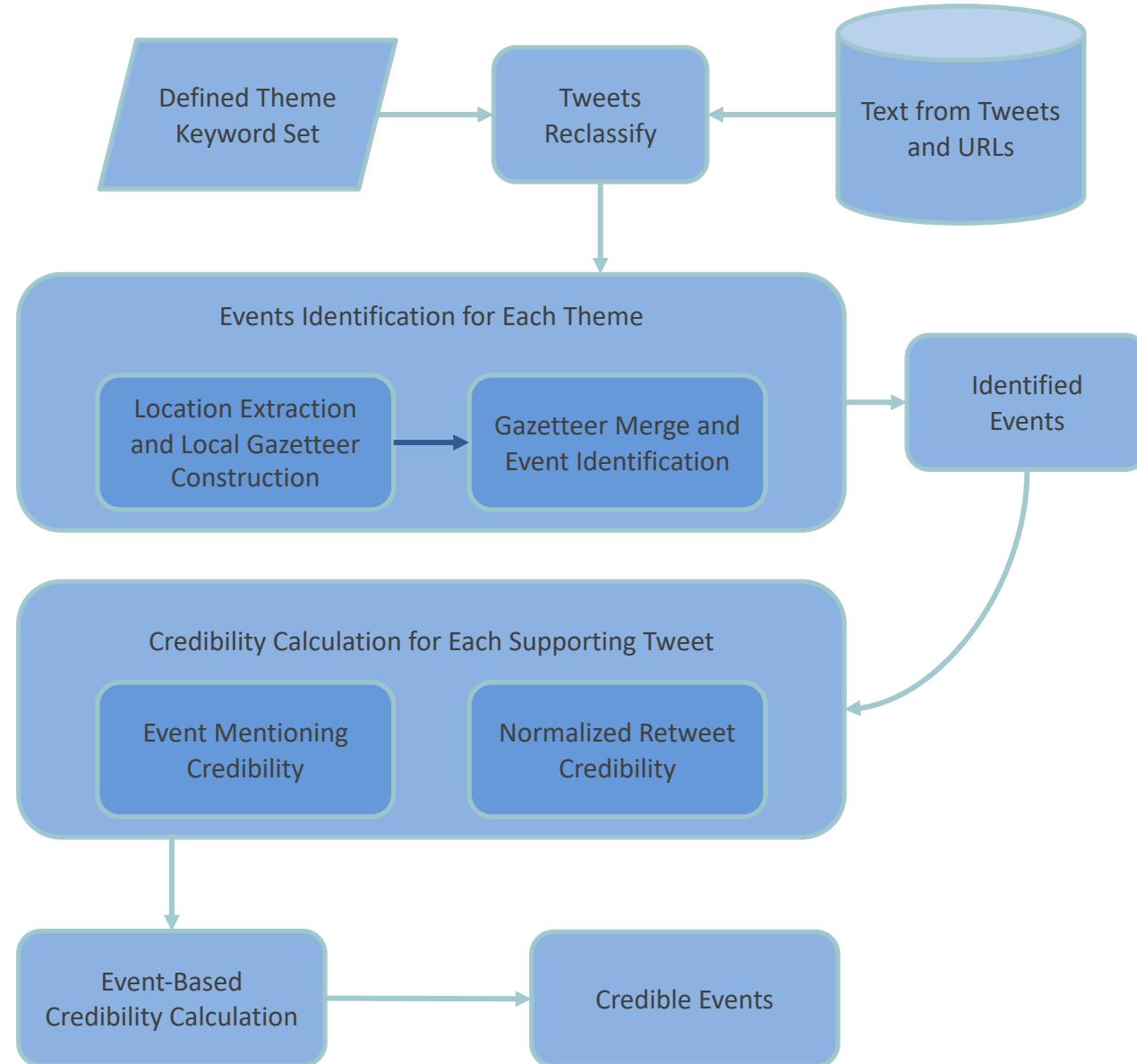
<https://www.youtube.com/watch?v=YzQGgyrxXil>



Phillips, M.E. Hurricane Harvey Twitter Dataset. Twitter, 2017. Available online:  
<https://digital.library.unt.edu/ark:/67531/metadc993940/> (accessed on 7 December 2018).



# Overall Structure of The Credibility Framework

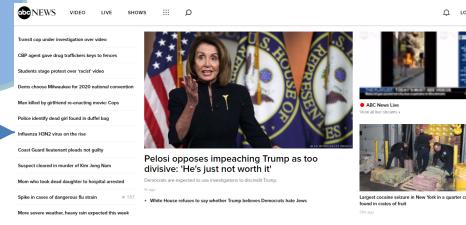
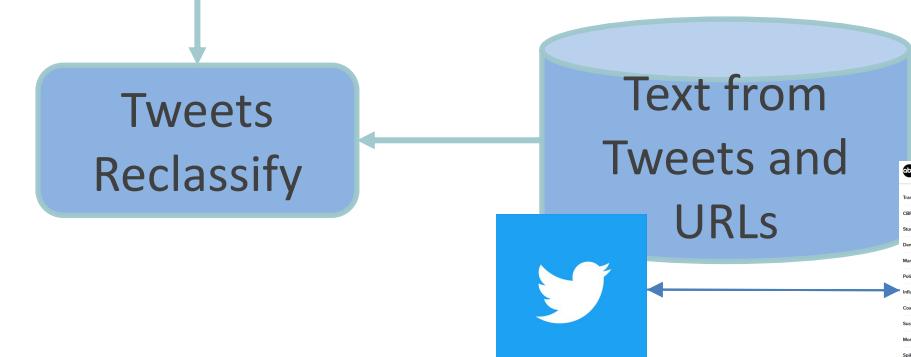


# Twitter Reclassify

## Defined Theme Keyword Set

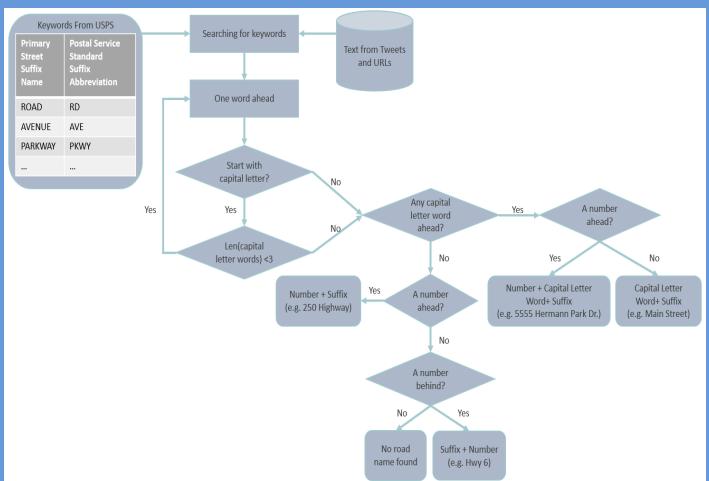
Theme	Keywords
<b>Sheltering</b>	hotel, housing, shelter
<b>Casualty</b>	dead, death, death toll, drowned, kill
<b>Damage</b>	catastrophe, collapse, damage, damaged, damaging, debris, destroy, destroyed, destruct, destructed, destructing, destruction, destroying, devastate, devastation, rip off, ruin, ruined, ruining, wreck, wrecking
<b>Flood</b>	flood, flooded, flooding, spill over, surge overflow, under water, underwater, wash away, washing away, water over the roof, water overflow, water rushing, drown
<b>Power/Electricity</b>	black out, blackout, coned, dark, darker, downed electrical wires, POWER down, POWER not expected, POWER off, POWER out, POWER outage, goodbye POWER, knock out POWER, lose POWER, losing POWER, lost POWER, no POWER, <u>noPOWER</u> , off the grid, powerless, shut off POWER, taken POWER, transformer exploding, transformer explosion, w/o POWER, wait POWER return, without POWER, without power

Huang, Q.; Xiao, Y. Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery. *ISPRS Int. J. Geo-Inf.* 2015, 4, 1549–1568.



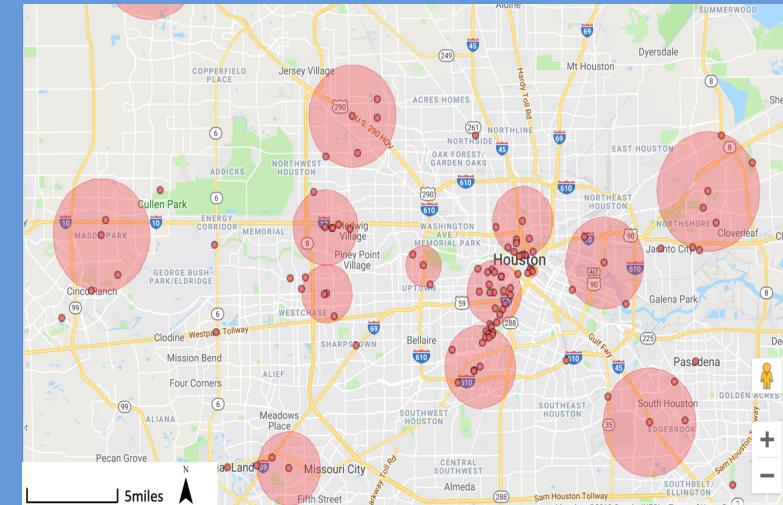
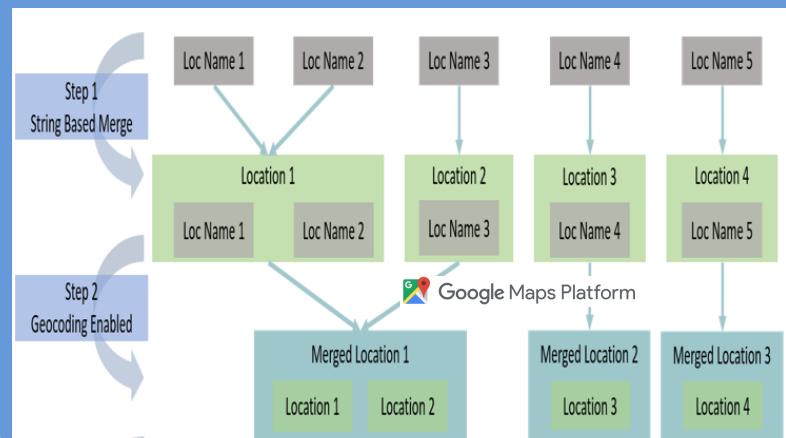
# Event Identification

## Location Extraction



## Events Identification for Each Theme

## Gazetteer Construction and Event Identification

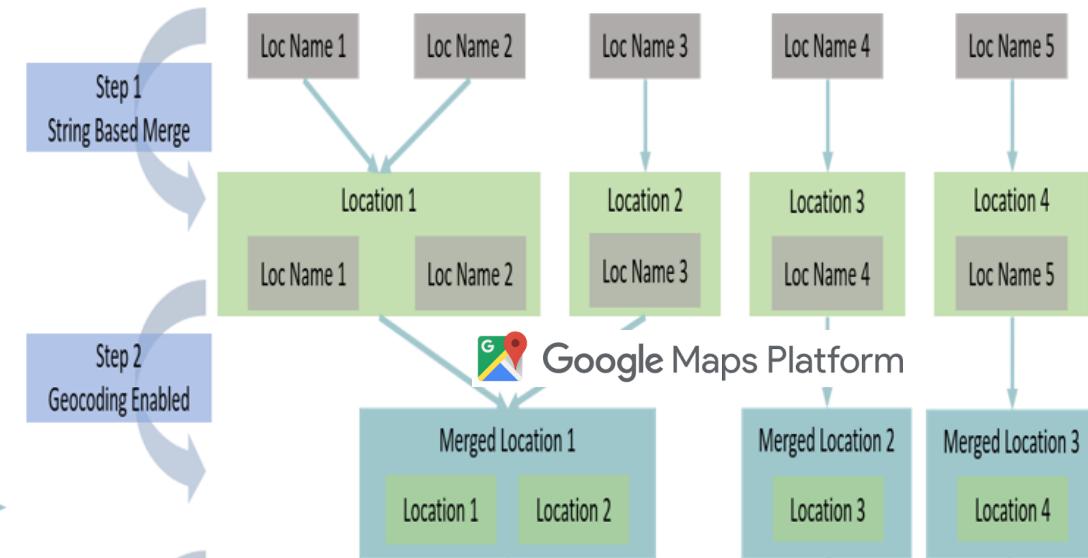
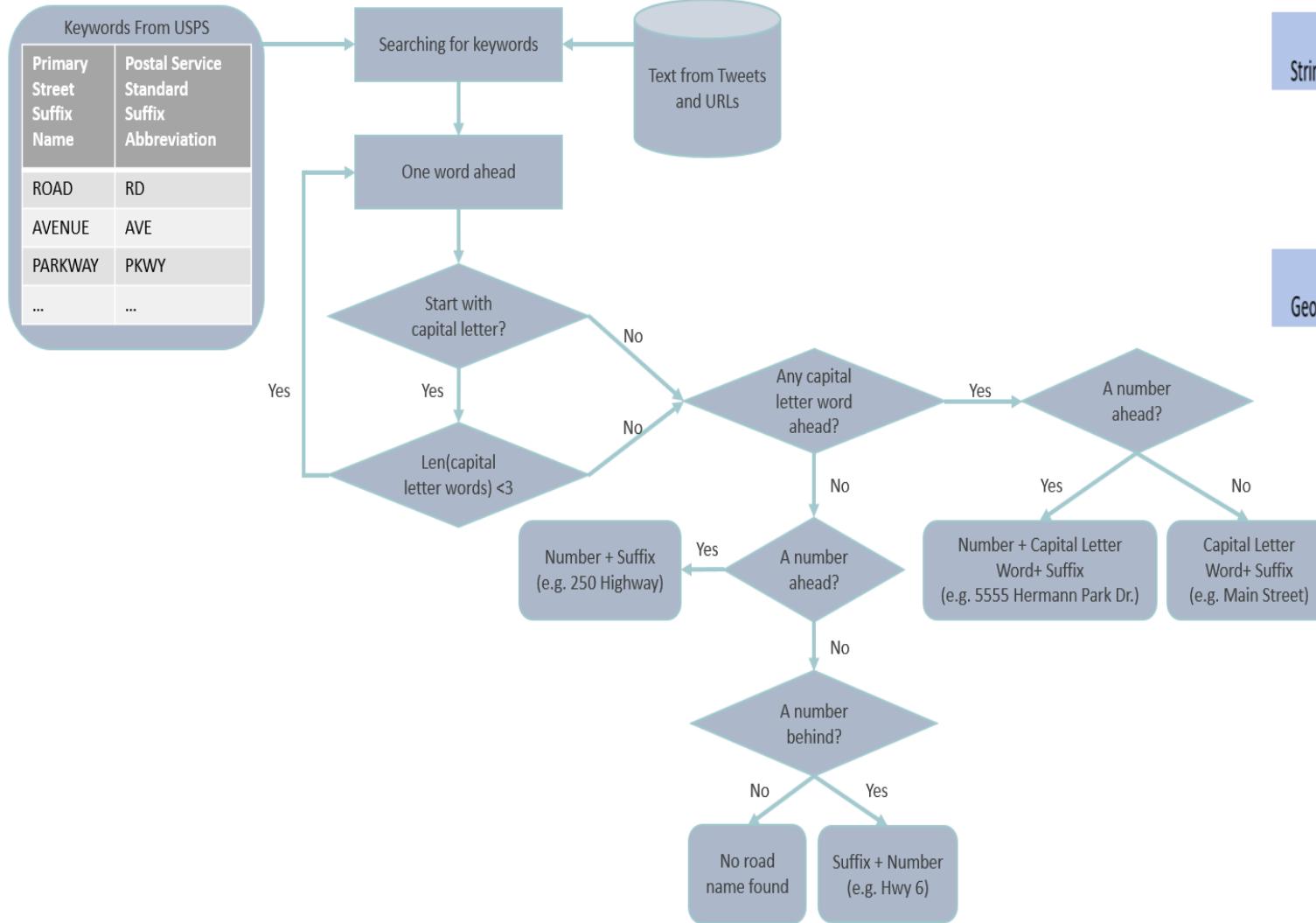


Identified  
Events



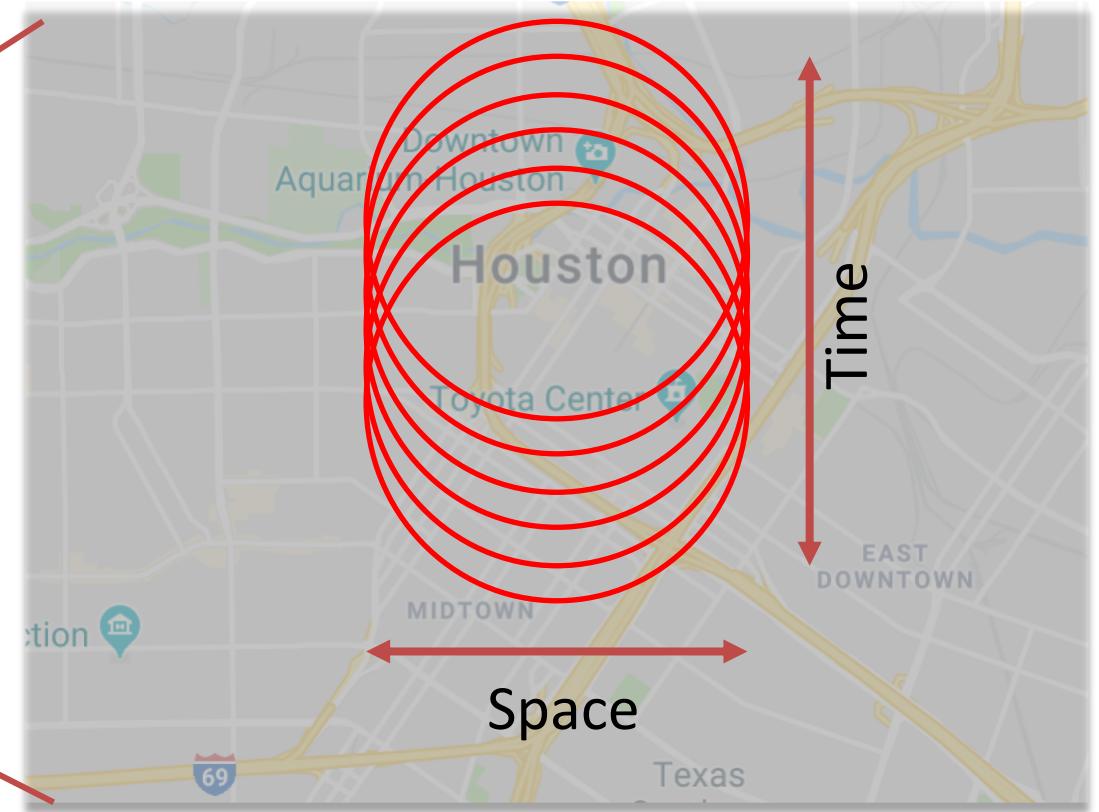
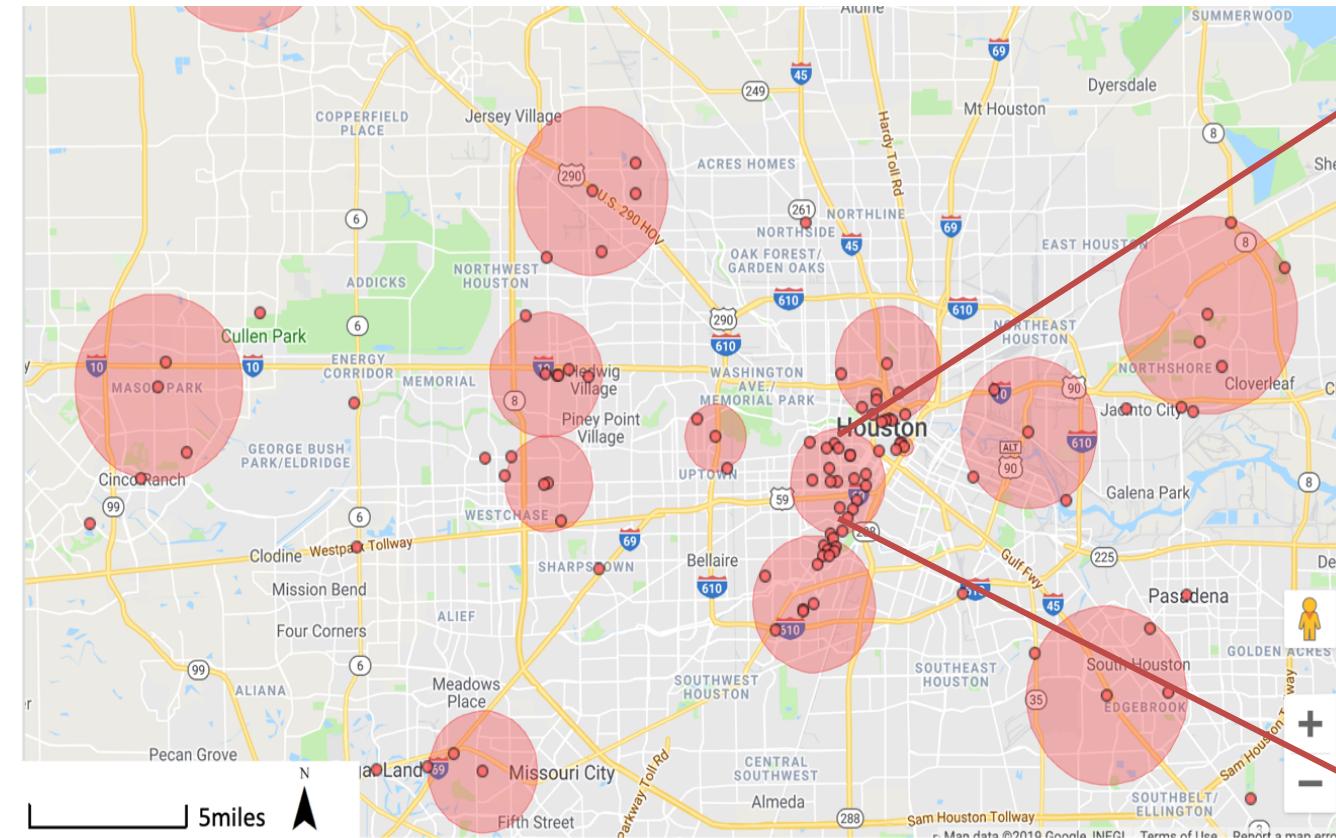


# Gazetteer Construction





# Space-Time Event Identification



**SatScan™**  
Software for the spatial, temporal, and space-time scan statistics



# The Credibility Scheme

Credibility Calculation for Each Supporting Tweet

Event Mentioning  
Credibility

Normalized  
Retweet Credibility

Event-Based Credibility Calculation

$$EBC_{NORM} = 0.5 \times \text{text} + 0.5 \times \text{URL} \quad (1)$$

$$\text{Retweet}_{\text{NORM}} = \frac{\text{Retweet}_{\text{tweet}}}{\text{Retweet}_{\text{max}}} \quad (2)$$

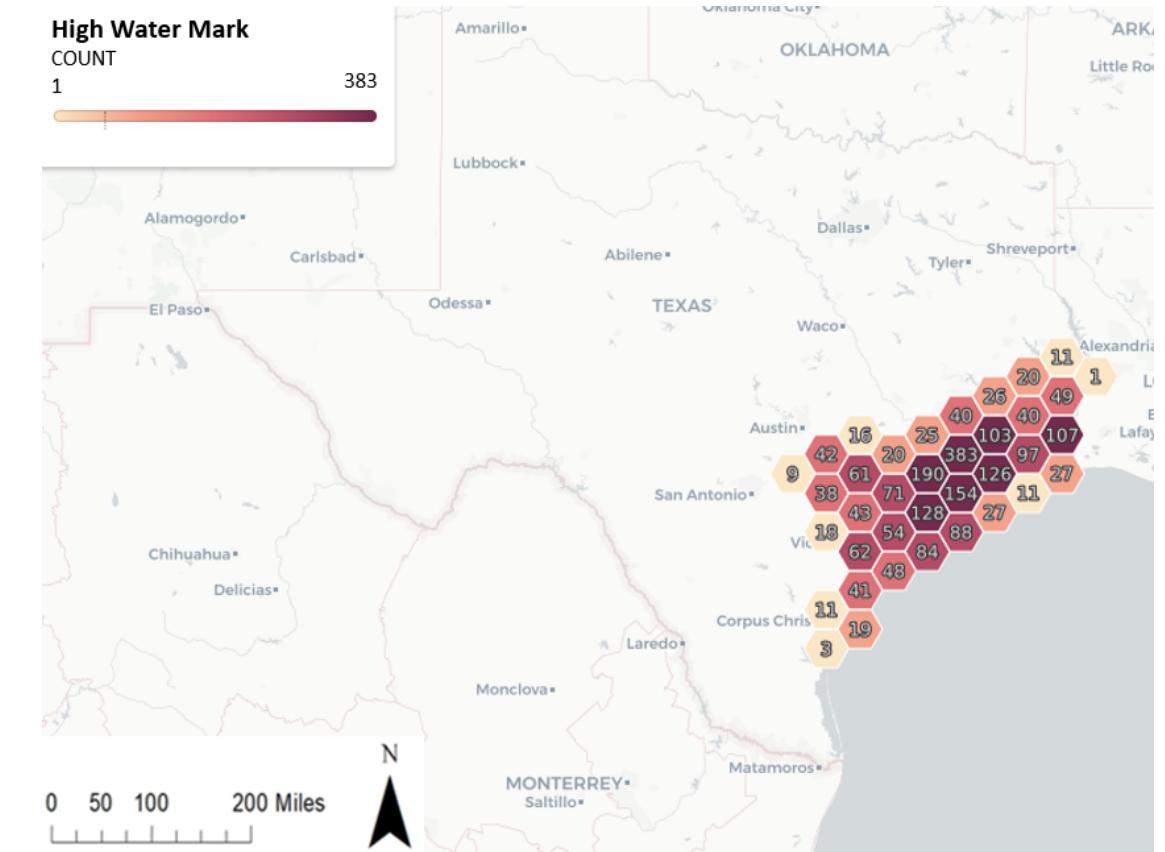
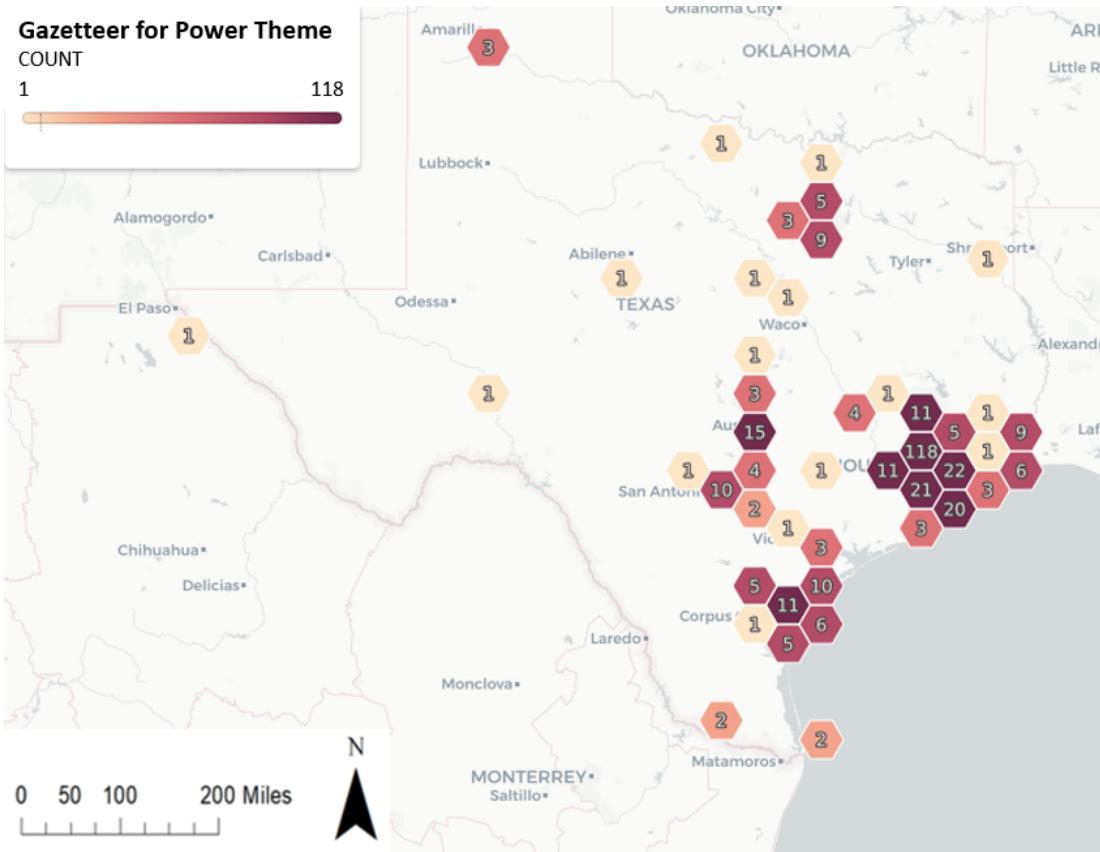
$$EBC_{\text{tweet}} = EBC_{\text{NORM}} + \text{Retweet}_{\text{NORM}} \quad (3)$$

$$EBC_{\text{total}} = \sum_{i=1}^{i=n} EBC_{\text{tweet}}^i \quad (4)$$

Credible  
Events

# Spatial Reliability

- Pearson correlation for location distribution →
  - The  $r = 0.93$  for location distribution of Power and HWM



# Temporal Trend



**Sylvester Turner**

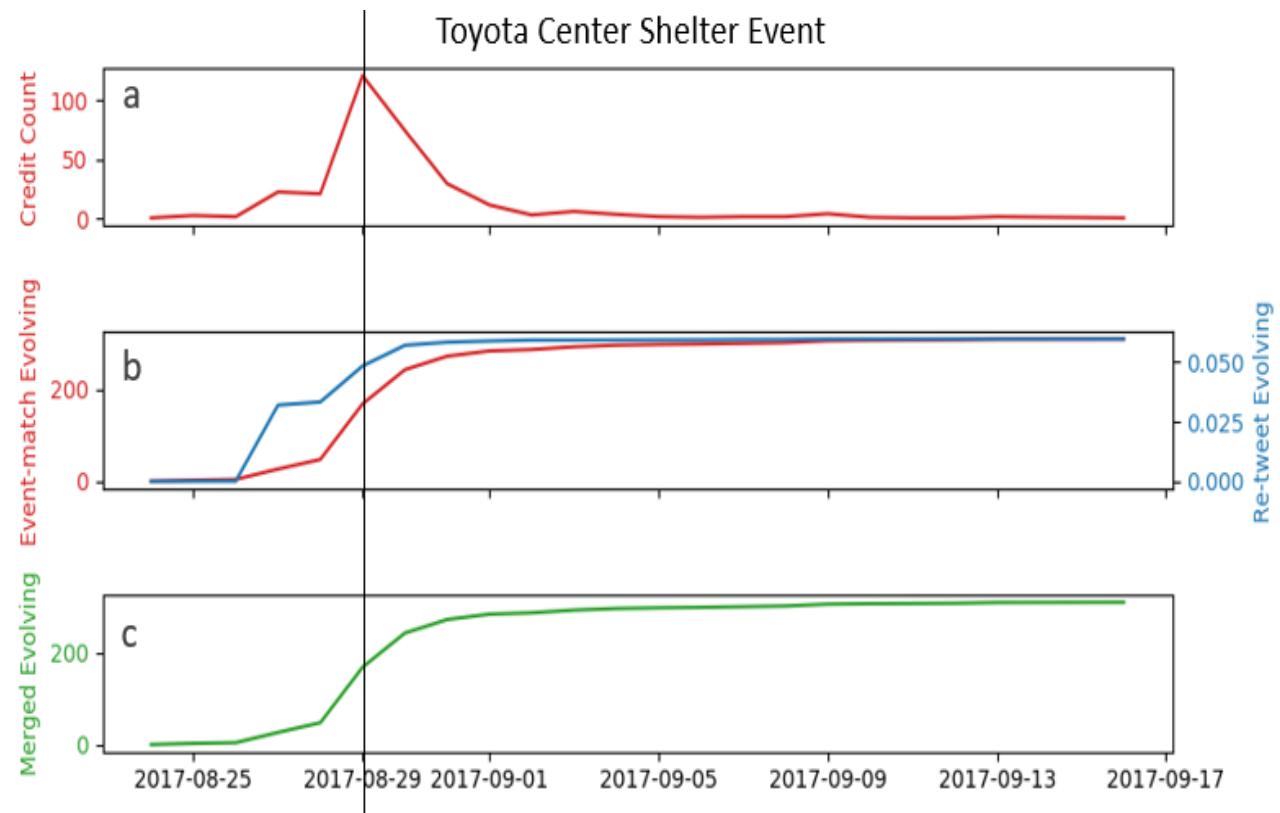
@SylvesterTurner



Opening Toyota Center to alleviate any crowding of evacuees at nearby Brown convention center downtown.

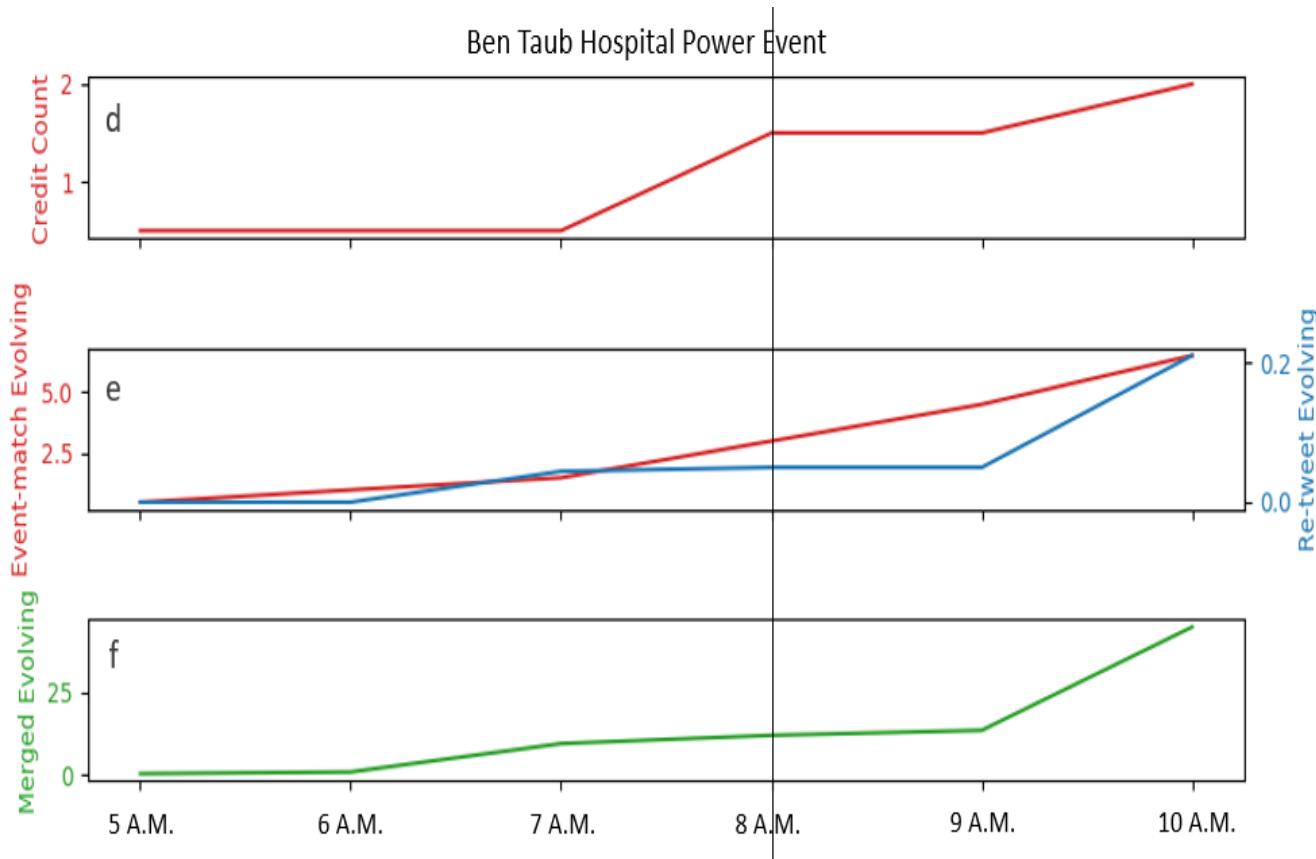
1,672 7:27 PM - Aug 29, 2017

1,156 people are talking about this



- Evacuation event detected by the framework due to the dramatic increase of reporting the flooding and power outages
- 4 hours early than the news reports at 12:23 P.M.

## Temporal Trend



- Applying this credibility framework with streaming data
- Utilizing the social impacts
  - Enhance the framework by making the reliability of the individual user as an extra indicator
- Advanced location extraction
  - Auto expandable keyword set

## Discussion





Thanks

- Q&A
- Contact Info:
  - Jingchao Yang, [jyang43@gmu.edu](mailto:jyang43@gmu.edu)
- Yang, Jingchao; Yu, Manzhu; Qin, Han; Lu, Mingyue; Yang, Chaowei. 2019. "A Twitter Data Credibility Framework—Hurricane Harvey as a Use Case." *ISPRS Int. J. Geo-Inf.* 8, no. 3: 111.

