中央财经大学
CENTRAL UNIVERSITY OF FINANCE AND ECONOMICS

# Application of Big Data in Insurance

------Early Warning and Damage Estimation of Catastrophe Risk

Sujin Zheng    Hairuo Guo
Shuning Song    Haitao Hu

CAS Annual Meeting
2019.11.11 Hawaii

# CONTENTS

# 01

Background

# Background



**insurance**

Ex post risk management financing vehicle

**question**

Damage estimation of Catastrophe Risk

Early Warning of Catastrophe Risk

# Social media platforms

- Gather an **unprecedented scale** of **data** and information

- Offer possibilities to observe social responses in a timely and continuously manner

- Speculate on the implications of expressed social attitude

- Distinguish interactions on social media relating to **significant events**

# Twitter  Microblog

- Its **information network properties** accelerate information diffusion

- Its **social attributes** allow access to personal information such as geo-location, social connections, personal emotions

- **Limitation of message** length promote the rapid exchange of information

- It became a research **hotspot** in the field of emergency management

# 02 Contents

# Case Study 1 : Hurricane Sandy

How cyber citizens react to a catastrophe using analyses of both tweet volume and content

**A**

**C**
Whether the combination of tweets and geo-information has an Advantage in early warning.

Hurricane Sandy

Identify general features of the public behavior

**B**

**D**
Analyze the correlations between tweets information and the damage

➤ Hurricane Sandy：the most destructive hurricane of 2012
                    the third-costliest hurricane in American history
                    (Harvey,2017; Katrina,2005)
                    damage: >$70 billion

**2012.11.1**
**23:00**

**2012.10.2**
**9**
**23:00**  landfall

departure

New Jersey

**Tweet data**
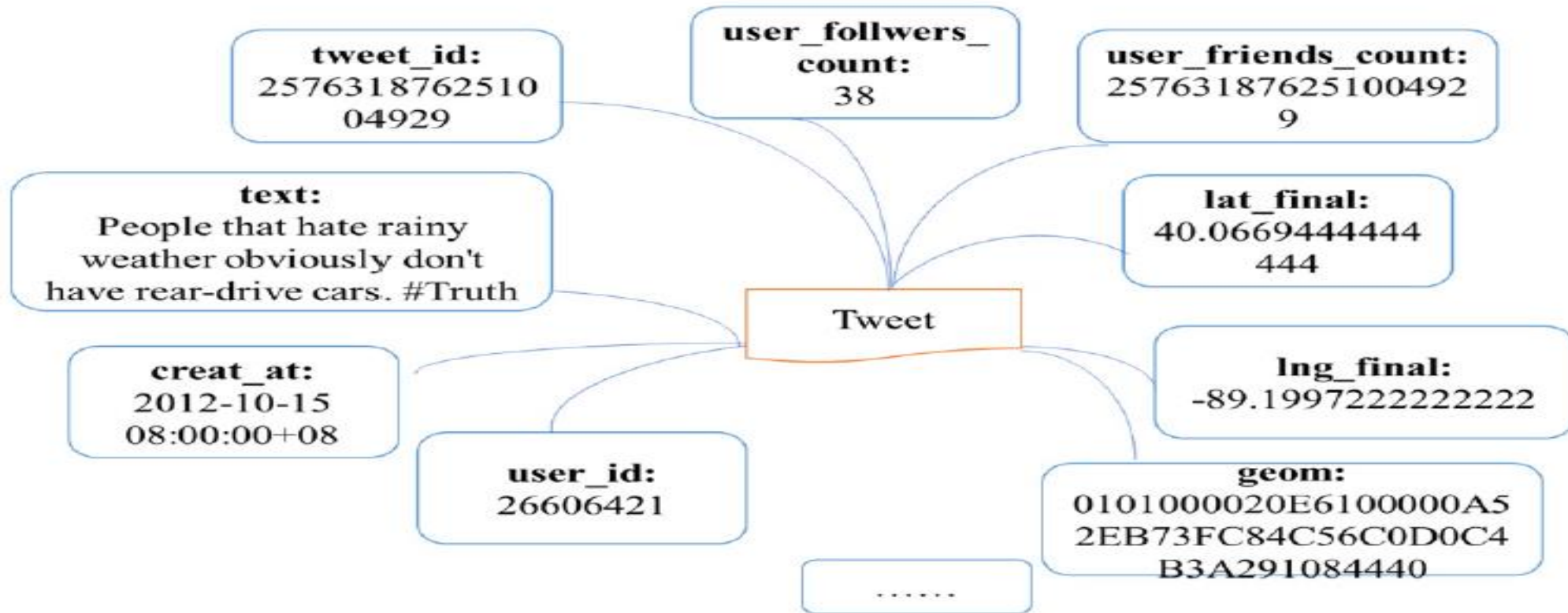New York: 500 thousands
New Jersey: 470
thousands

**Hurricane damage losses data**
1.FEMA's Enterprise Coordination& Information Management(ECIM) reporting team
2.https://www.fema.gov/what-disaster-assistance

➤ An example of the extracted tweets



tweet_id:
2576318762510
04929

user_follwers_
count:
38

user_friends_count:
25763187625100492
9

text:
People that hate rainy
weather obviously don't
have rear-drive cars. #Truth

lat_final:
40.0669444444
444

Tweet

creat_at:
2012-10-15
08:00:00+08

lng_final:
-89.1997222222222

user_id:
26606421

geom:
0101000020E6100000A5
2EB73FC84C56C0D0C4
B3A291084440

......

03 Methodology

# Sentiment analysis

**Measure** the emotion or mood expressed in each tweet and classify it as **negative, neutral, or positive**

*Sentiment lexicons*
*Machine learning*

Whether a correlation exists between postings on tweet and factual data on disaster damage.

**Reverse geocoding**

**Content analysis**

Reverse geo-information of tweets into identifiable geographic information

# frequency analysis

**Top hushtags and keywors**

**Import:** Tweets set $T = \{t_1, t_2, t_3, L, t_n\}$, sentiment lexicon $Dict$, emoticons set $Emotions$.

For each tweet $t_i$, the scoring formula $Scores(t_i)$ is:

$$Scores(t_i) = \sum_{w_{i,j} \in t_i} f(w_{i,j}, Dict) + \sum_{e_{i,j} \in t_i} \delta(e_{i,j}, Emotions)$$

Where $w_{i,j}$ and $e_{i,j}$ represent the word and emoticon in tweet $t_i$ respectively.

$$f(w_{i,j}, Dict) = Pos(w_{i,j}) - Neg(w_{i,j})$$

Each word $w_{i,j}$ in $Dict$ has two numerical scores $Pos(w_{i,j})$ and $Neg(w_{i,j})$, which indicate the degree of positivity or negativity.

$$\delta(e_{i,j}, Emotions) = \begin{cases} Pos(e_{i,j}) \text{ if } e_{i,j} \text{ represents positive in Emotions} \\ Neg(e_{i,j}) \text{ if } e_{i,j} \text{ represents negative in Emotions} \end{cases}$$

**Output:** For each tweet, $Scores(t_i)$ is calculated, where $Scores(t_i) > 0$ indicates positive emotional orientation and vice-versa.

Five topics of Twitter messages.

| Topics | Description | Related words |
|---|---|---|
| Disaster-related | The most important role of Twitter during a disaster is information sharing and dissemination. When people use some words directly related to the disaster, which could attract people's attention as an early warning. | "disaster," "Franken-storm," "hurricane," "flood," "destroy," "Sandy," "storm," "tornado." |
| Weather-related | Hurricanes are often accompanied by climate changes, being different from other kinds of natural disasters. | "cold," "climate change," "black out," "climate." |
| Emotion-related | Emotional support is important during and after a disaster. When undergoing a disaster, people naturally express their personal feelings and concerns, providing social and emotional support, comfort, and sympathy to others. | "panic," "hope," "terrifying," "insane," "scared," "keep calm," "scary," "hungry," "fuck," "hate," "horrible," "pray for us," "omg," "god." |
| Action-related | When a disaster happens, the government plays an irreplaceable role in disaster relief and support. This category includes the words that people are using to appeal to the government to tack action. | "government," "MTA," "FEMA," "governor," "red cross," "federal," "Cuomo," "NY," "NY," "wall st," "CNN." |
| Situation-related | Situation updates and influence descriptions are important for early warnings and disaster assistance. | "power," "no power," "food," "house," "energy," "hospital," "airport," "electric," "electricity," "cancelled," "gas," "blocked," "emergency," "life," "warning," "lost," "help," "need," "keep safe," "safe," "stay home," "stay safe," "recovery," "survived," "dead," "kill," "dying," "home," "problem," "move." |

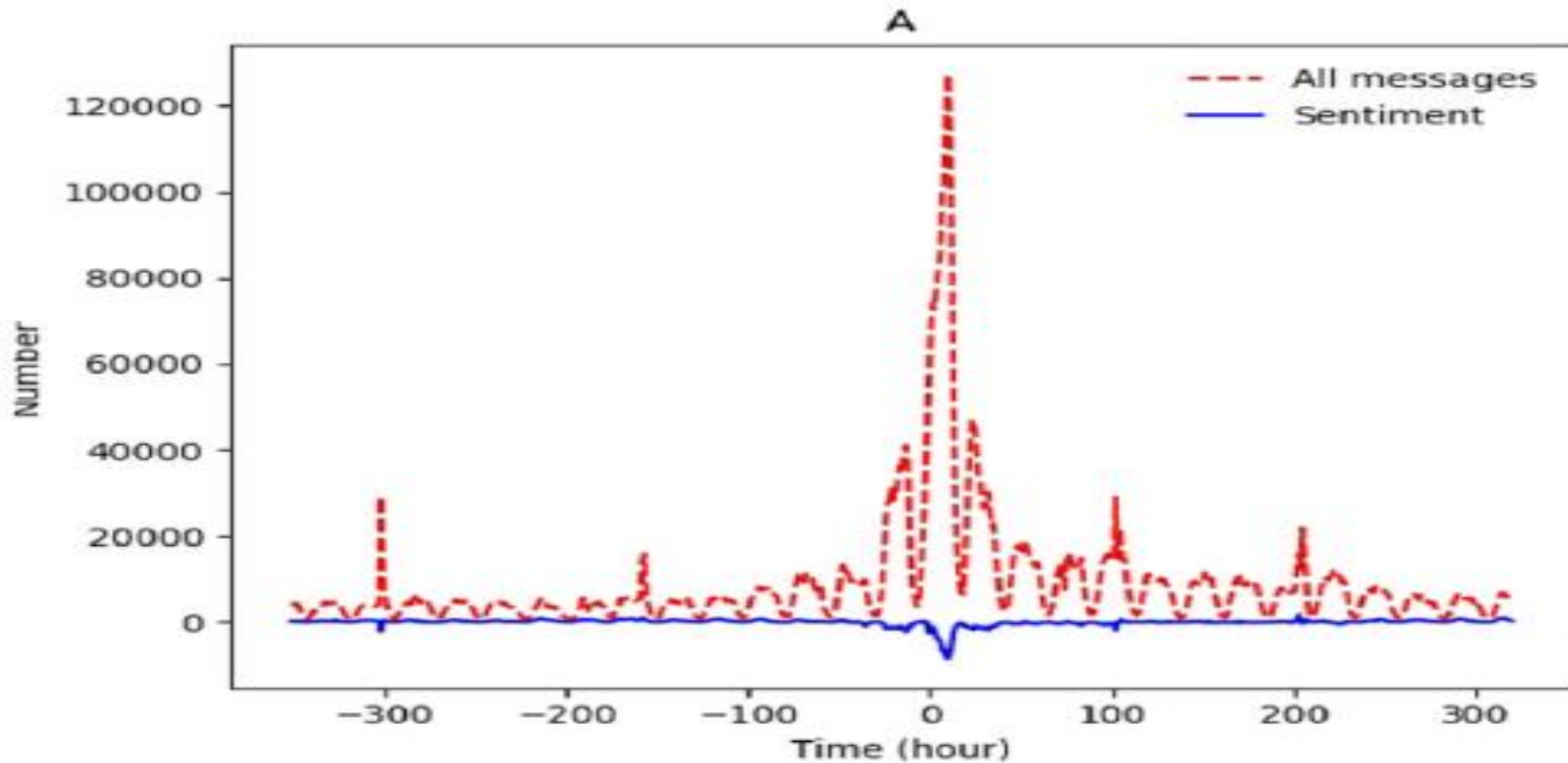# Data processing and analysis-1
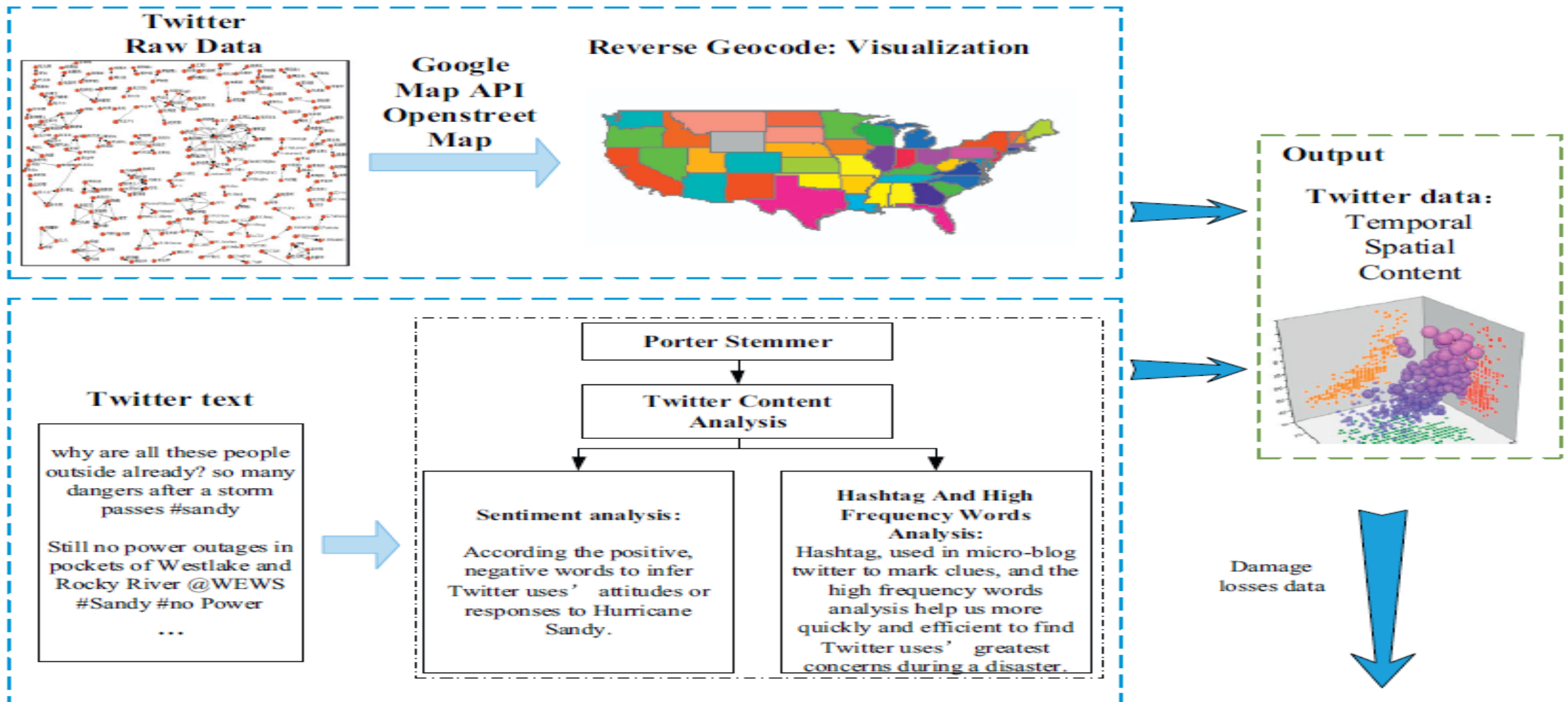
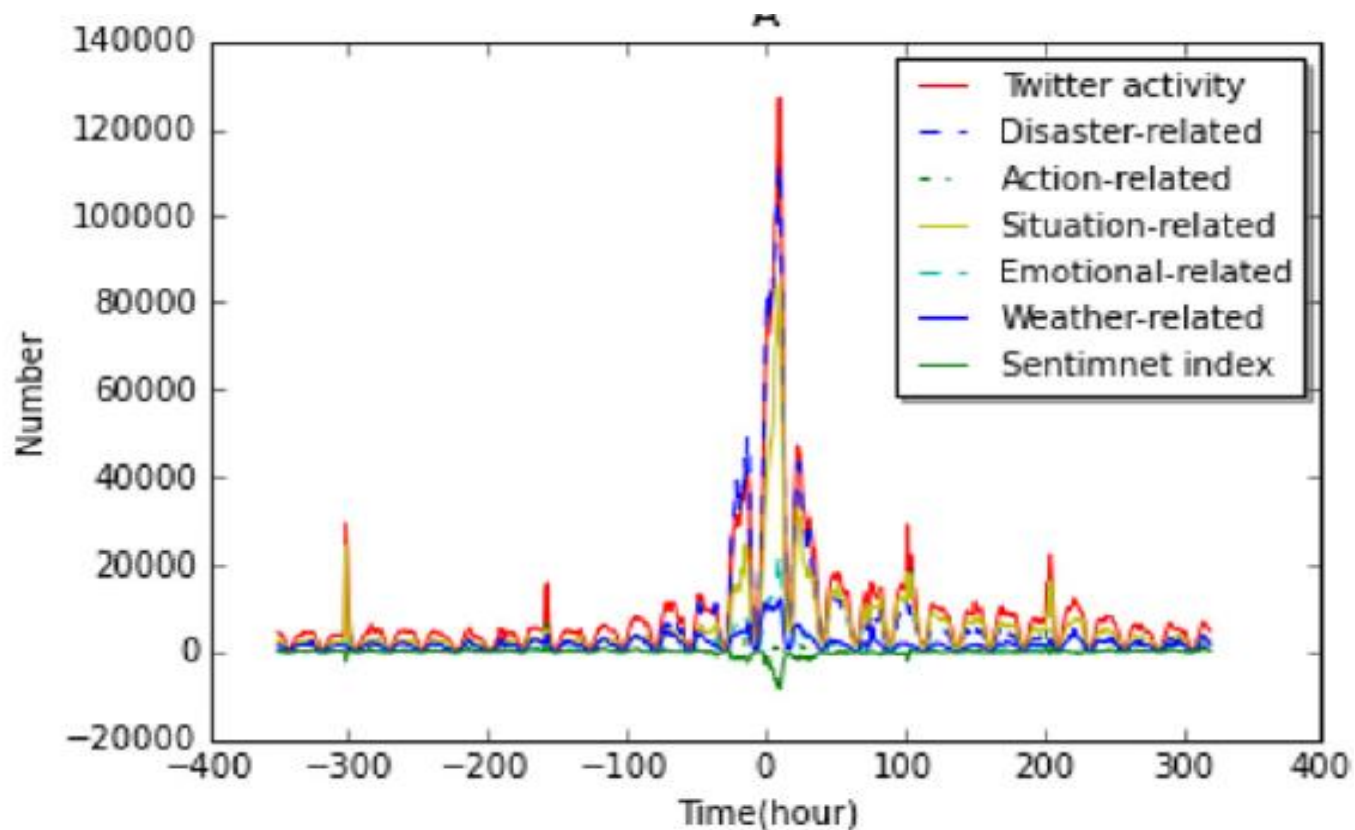➢ Data processing and analysis-2
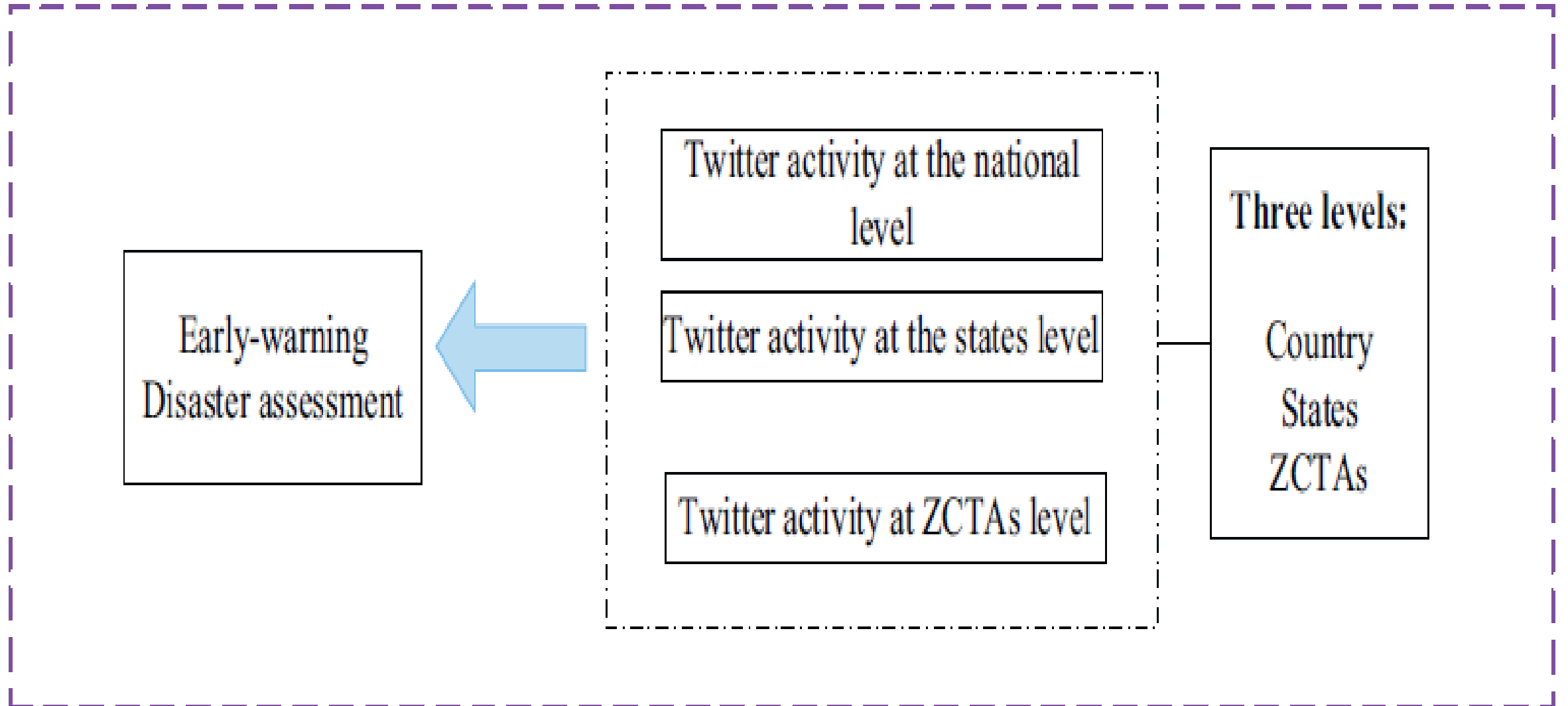
# 04 Conclusion

# Tweet **nationwide** analysis--- <u>volume of Twitter activity</u>


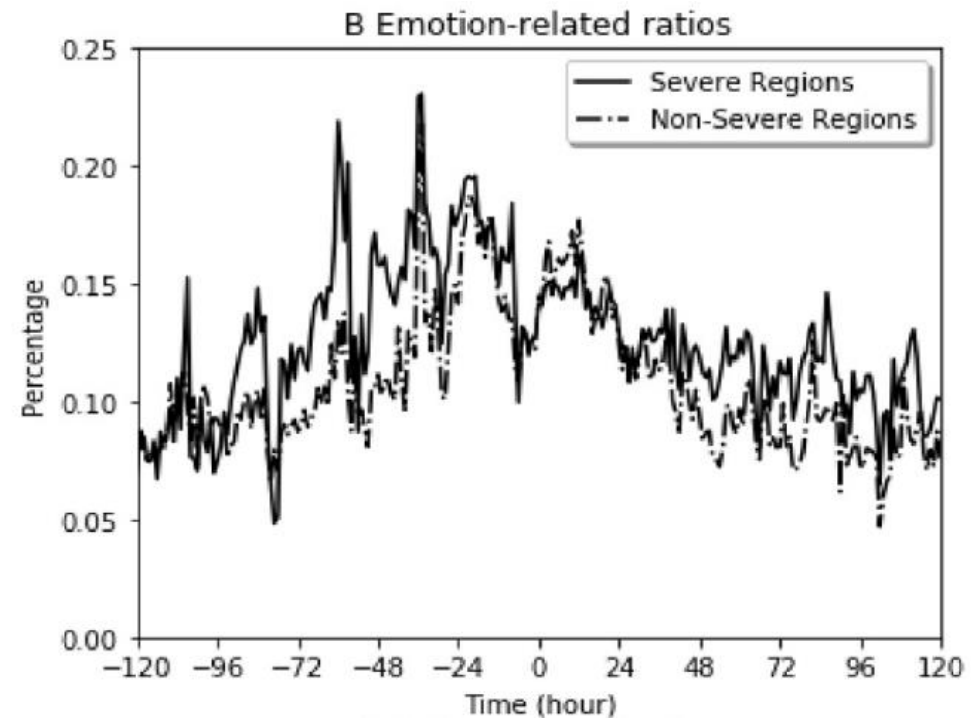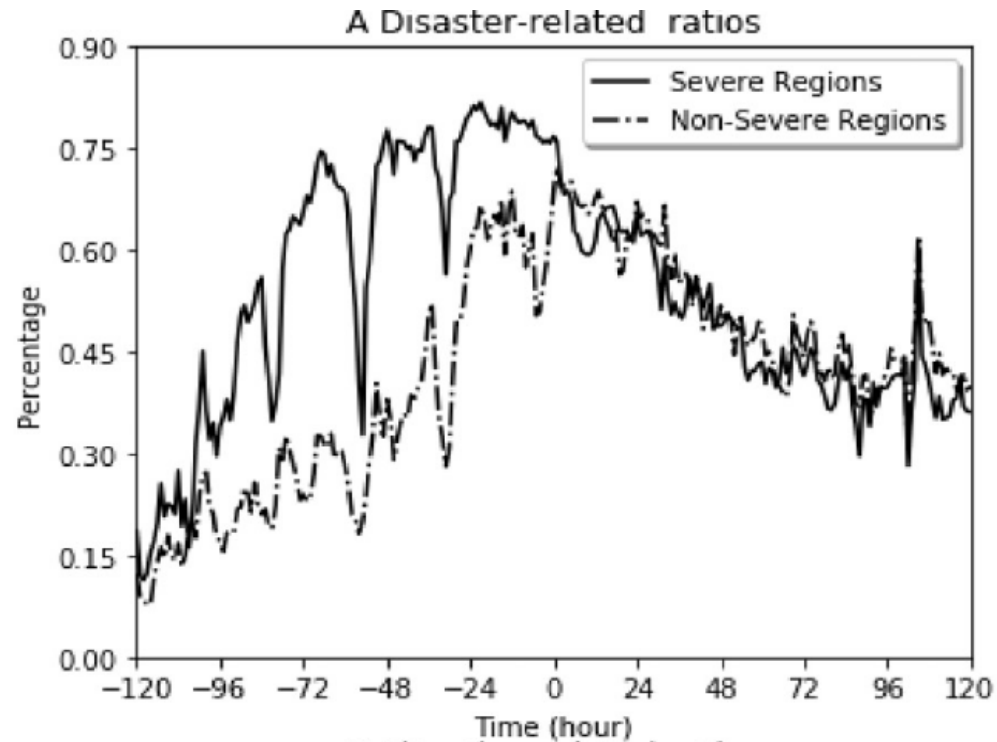
Horizontal axis is an offset representing the time of Sandy landfall, time 0 means 23:30 10.29

# Tweet nationwide analysis--- content features

# Tweet nationwide analysis——
## consistency between Twitter activity, five topics and sentiment index
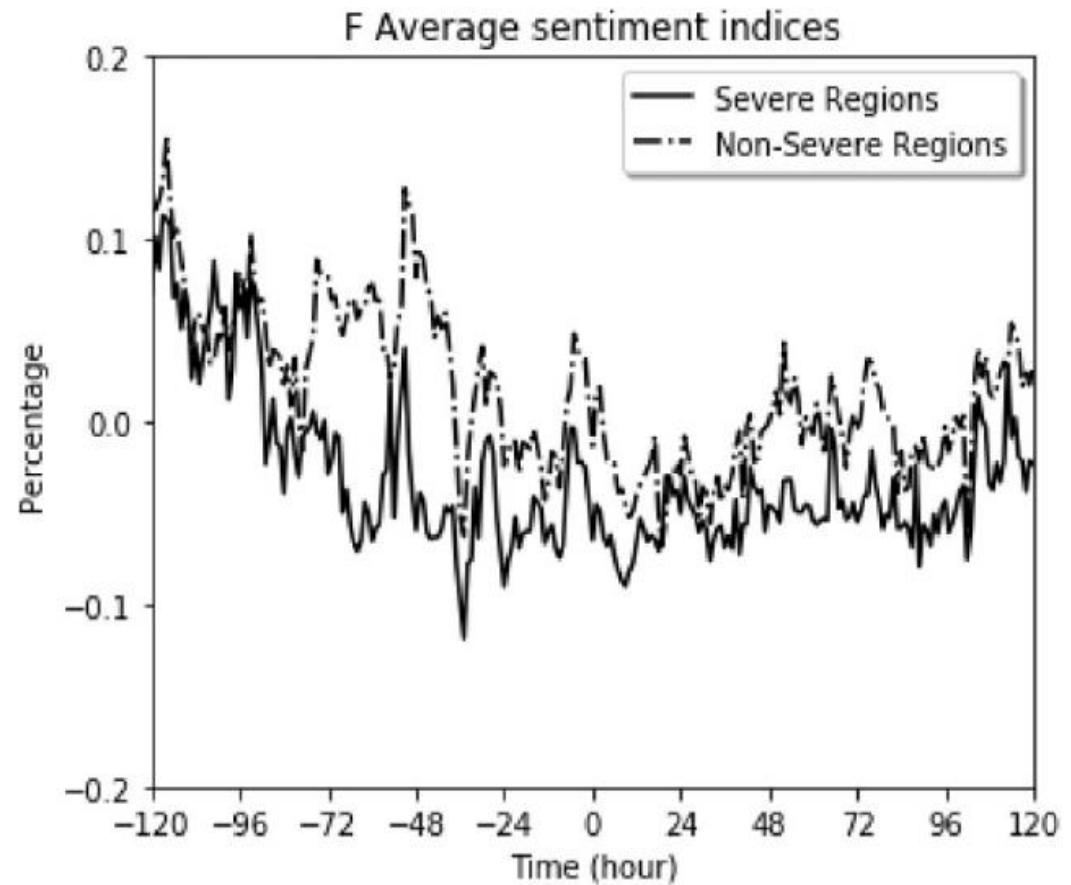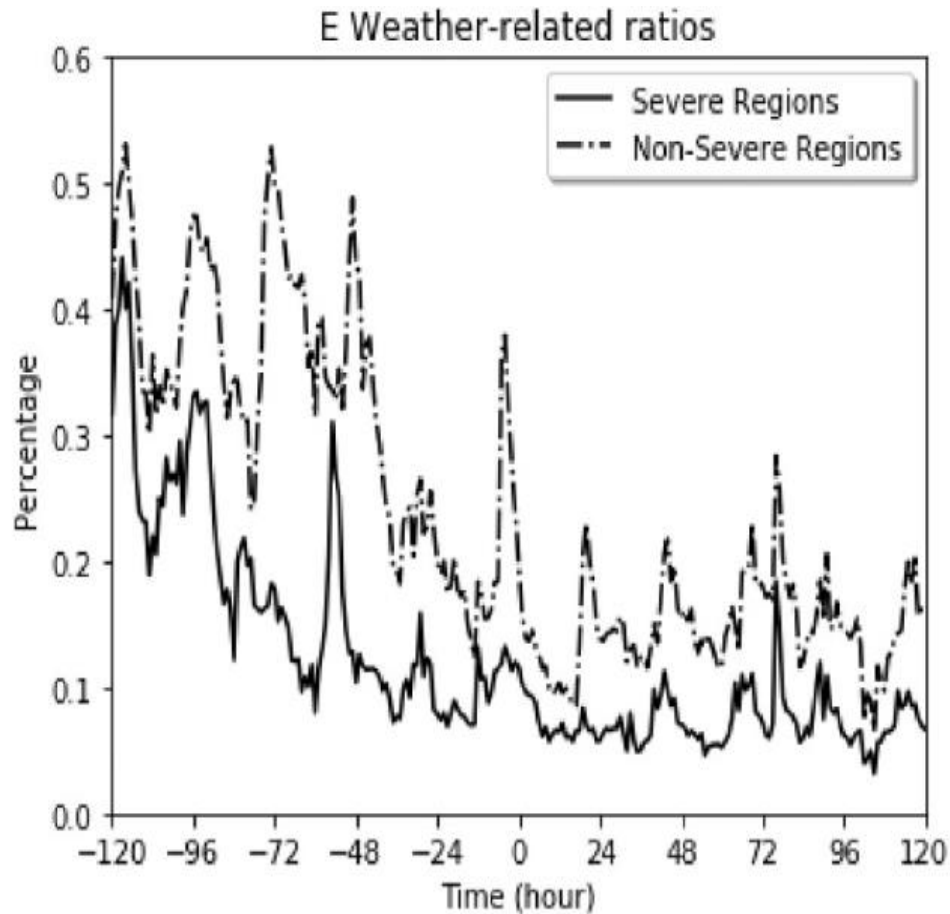
# Tweet **state** analysis--- <u>early warning</u>



Severe Region : ten states(Connecticut, Delaware, Maryland, New Jersy, New York etc.) and District of Columbia——declaration of a state of emergency

# Tweet **state** analysis--- <u>early warning</u>



Severe Region : ten states(Connecticut, Delaware, Maryland, New Jersy, New York etc.) and District of Columbia——declaration of a state of emergency

# Tweet state analysis--- early warning



E Weather-related ratios

F Average sentiment indices
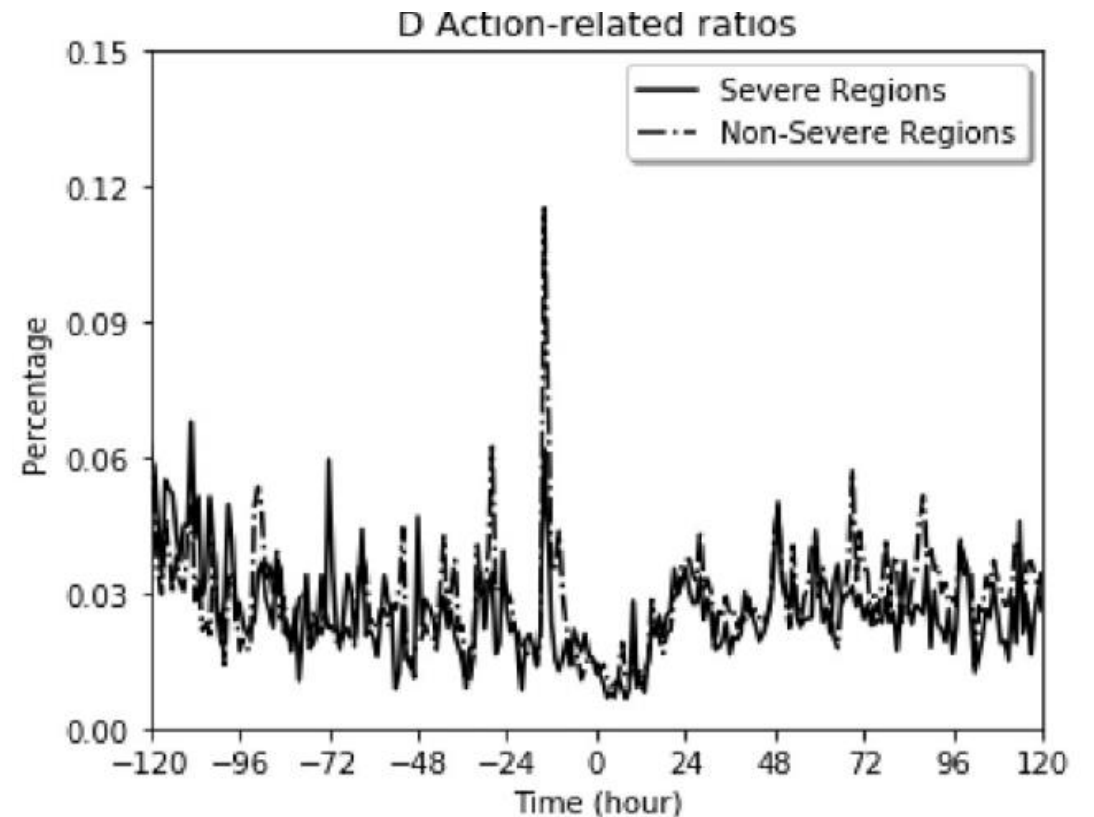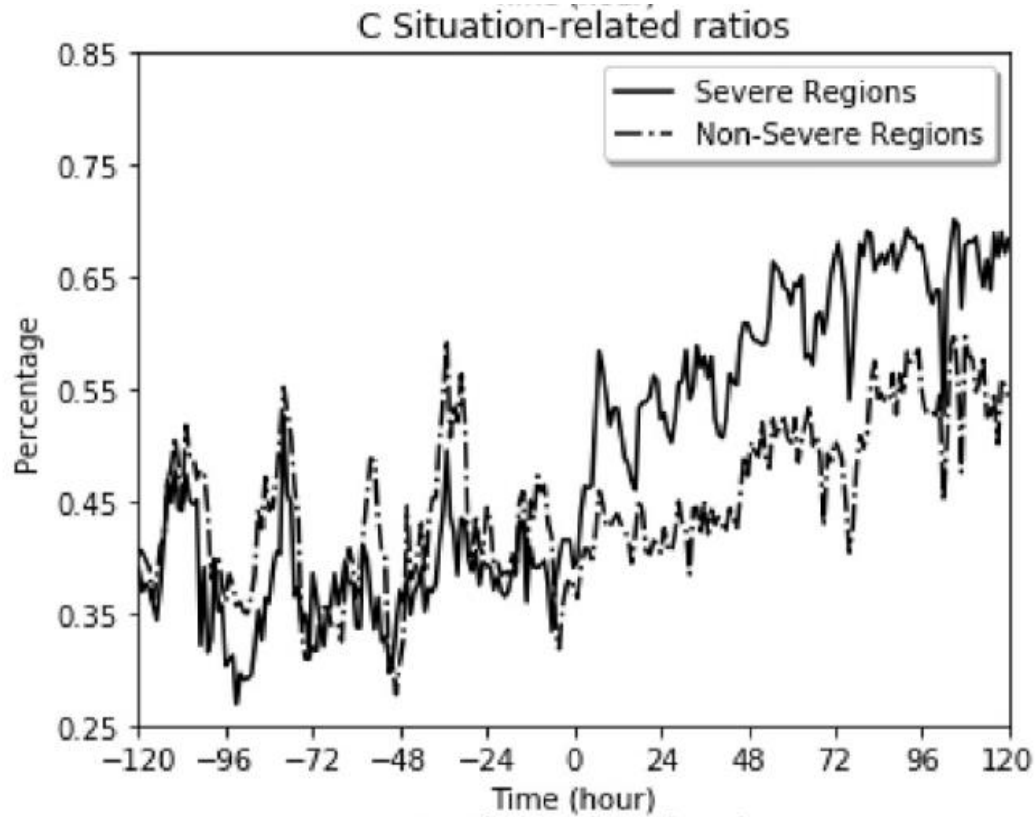
Severe Region : ten states(Connecticut, Delaware, Maryland, New Jersy, New York etc.) and District of Columbia——declaration of a state of emergency
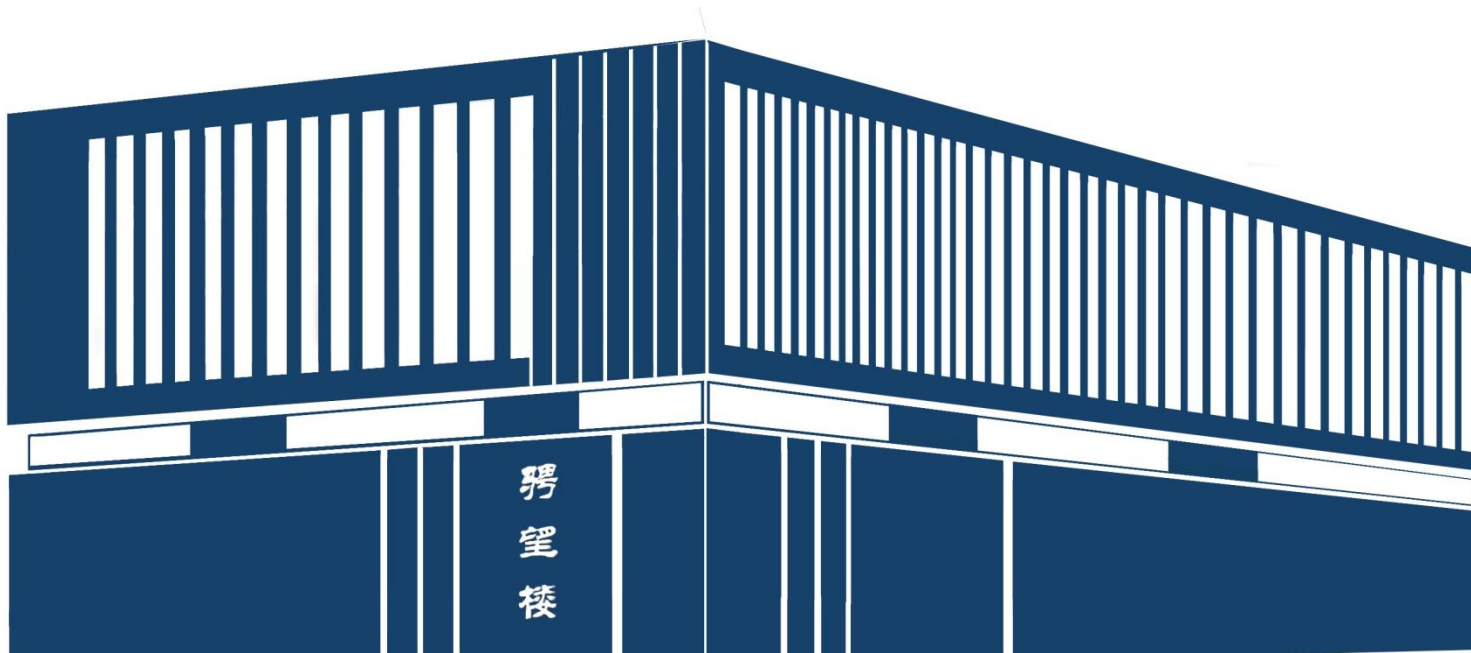
# Tweet ZCTA analysis--- disaster assessment
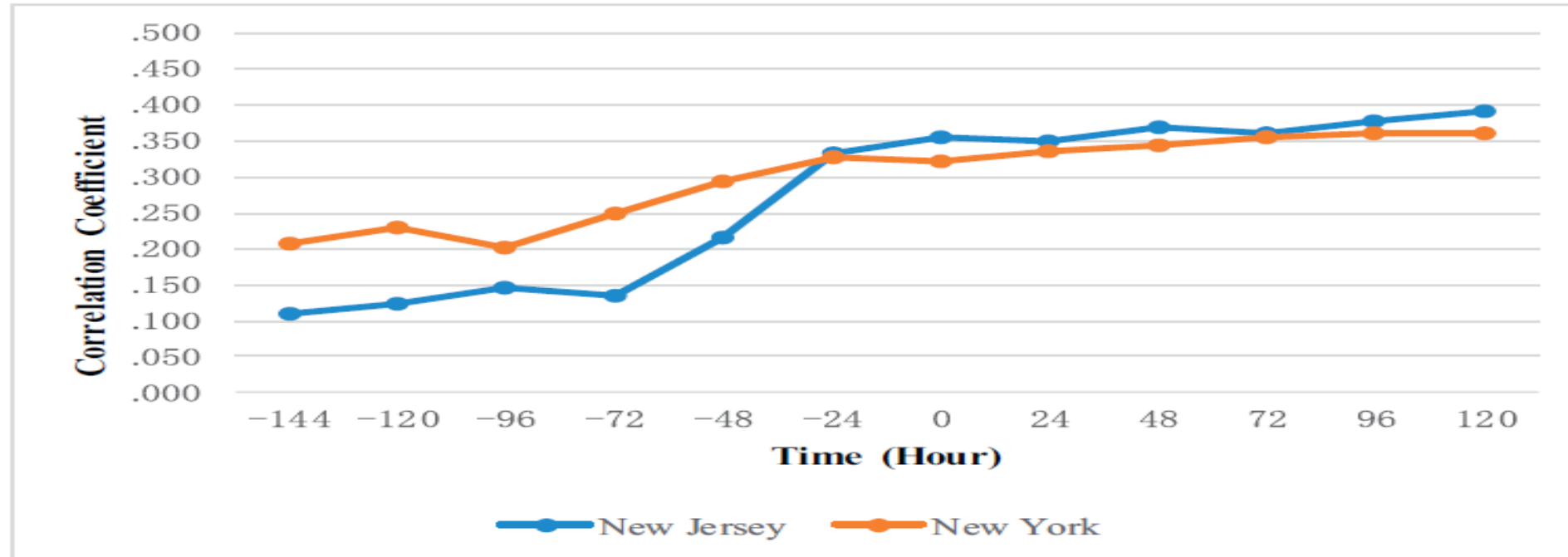


Trend of correlation coefficient between the total damage losses(FEMA) and Twitter activity in New Jersey and New York

Case Study 2 : **Typhoon Lekima**      8.9.8:00/2019-8.14.8:00/2019

# References

[1]Desheng Wu, Yiwen Cui, Disaster early warning and damage assessment analysis using social media data and geo-location information, Decision Support Systems.[J].111 (2018) 48-59.
[2]Sujin Zheng, Hairuo Guo, Shuning Song, Haitao Hu,

Thanks for listening