Journal of Operations Management

Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises

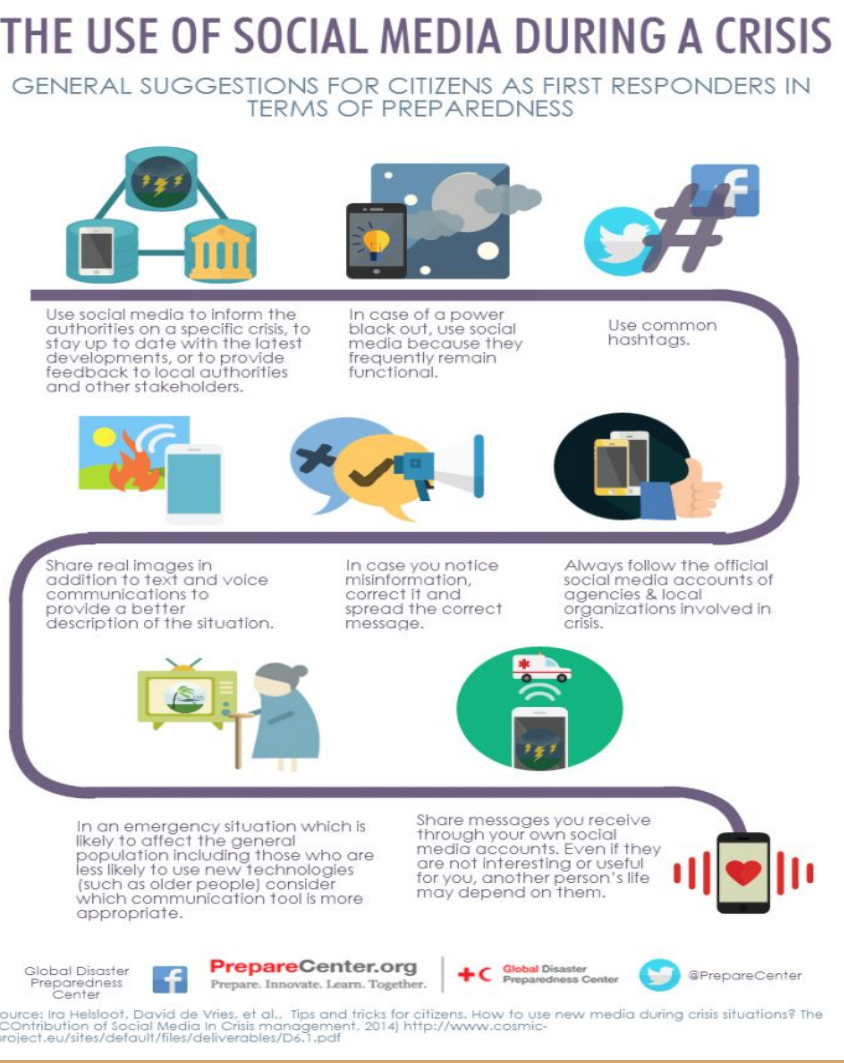
Özge Garip 21400216

Yusuf Duyar 21501768

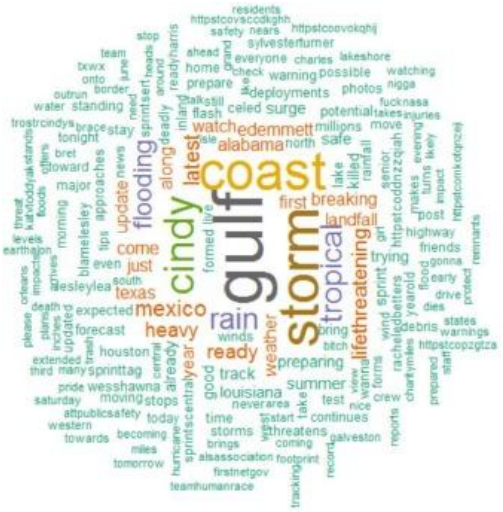
Ege Semih Kozoğlu 21602904

Introduction

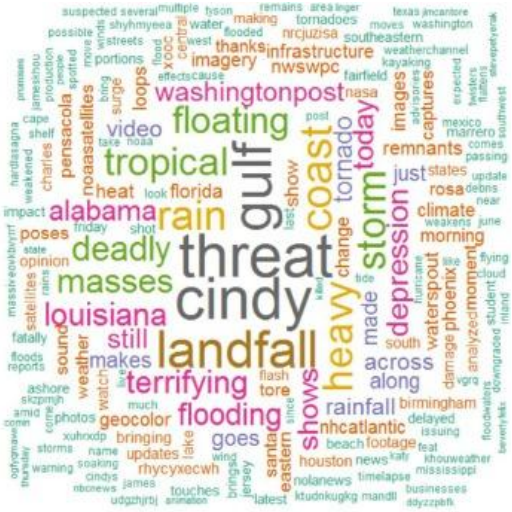
The rapid diffusion of information is critical to combat the extreme levels of uncertainty and complexity that surround disaster relief operations. As a means of gathering and sharing information, humanitarian organizations are becoming increasingly reliant on social media platforms based on the Internet



In this presentation, we show a field study that examines how effectively information diffuses through social media networks embedded in these platforms. Using a large dataset from Twitter during Hurricane Sandy, we first applied Information Diffusion Theory to characterize diffusion rates



(a): keywords before June 22 8am

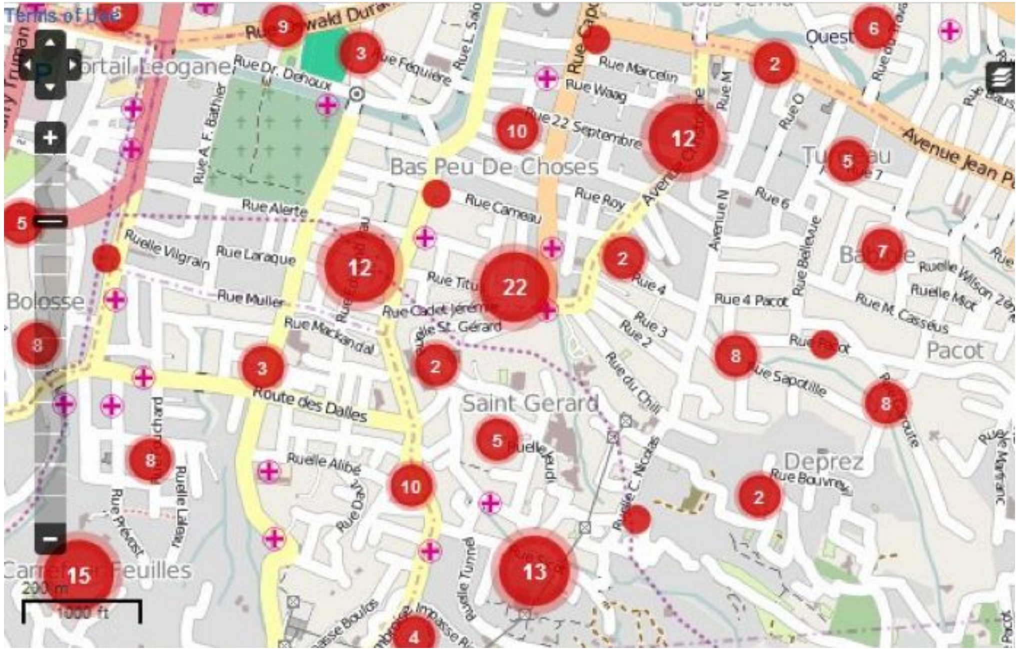


(b) : keywords after June 22 8am

Flood in Virginia in 22 June 2016

Current Situation

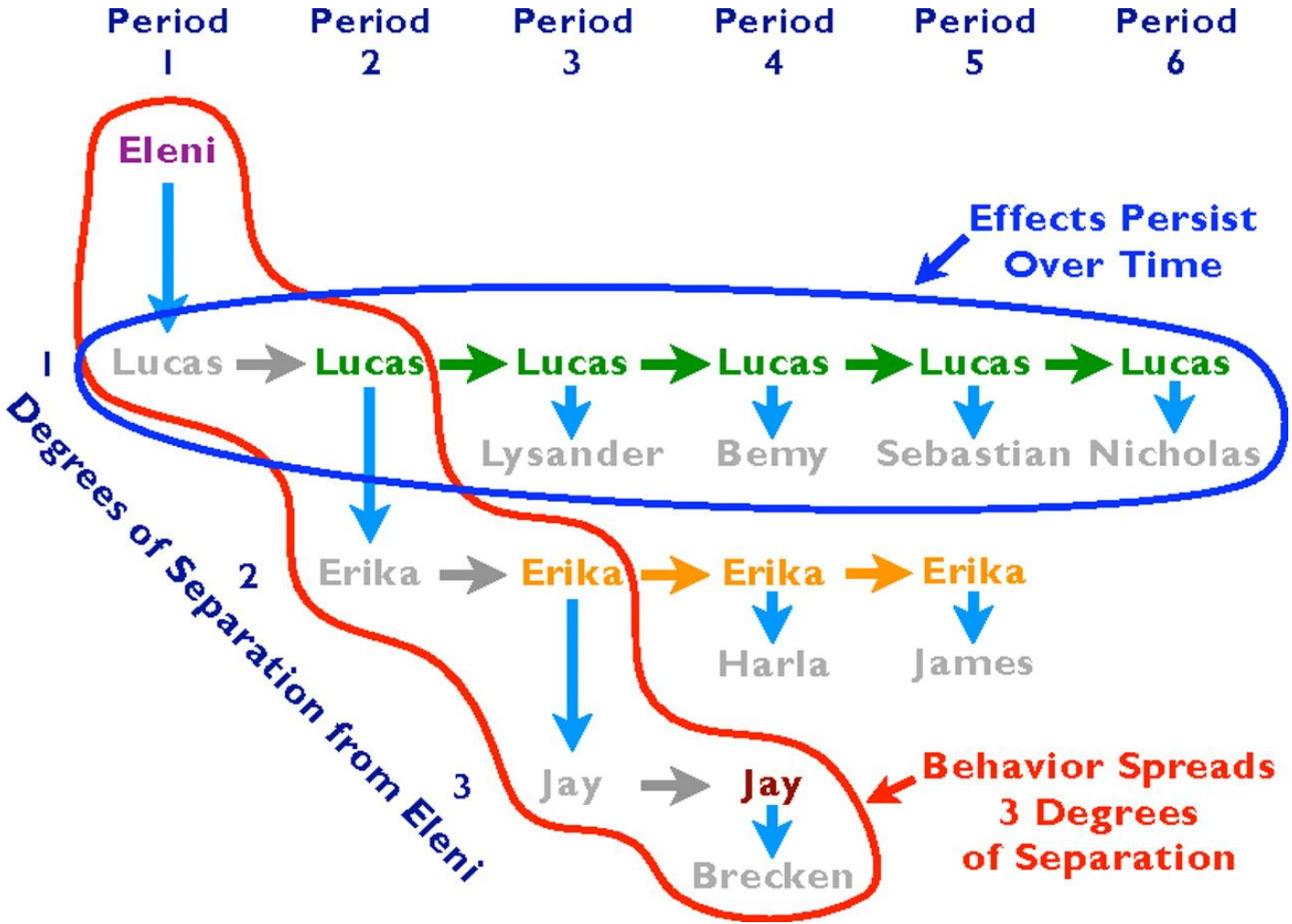
Having data from informants with first-hand knowledge of what is occurring in affected areas , and recently, HOs have aggregated these data to create crisis maps showing landmarks like damaged infrastructure and shelters



Current Situation

HOs have also used social media to share capacity levels and resource availability to enhance coordination among stakeholder.
These connections will form social media networks in which information produced by a user (an originator) will create cascades when those connected directly to her receive it and, in turn, share it with those with whom they are connected.

??? CASCADES ???


Period 1
Period 2
Period 3
Period 4
Period 5
Period 6
Eleni
Effects Persist Over Time
1
Lucas
Lucas
Lucas
Lucas
Lucas
Lucas
Lysander
Bemy
Sebastian
Nicholas
2
Erika
Erika
Erika
Erika
Harla
James
3
Jay
Jay
Brecken
Degrees of Separation from Eleni
Behavior Spreads 3 Degrees of Separation

Improvement

To address this objective, they develop and test a set of theoretical propositions regarding the role played by three key determinants of information diffusion dynamics in social media networks :

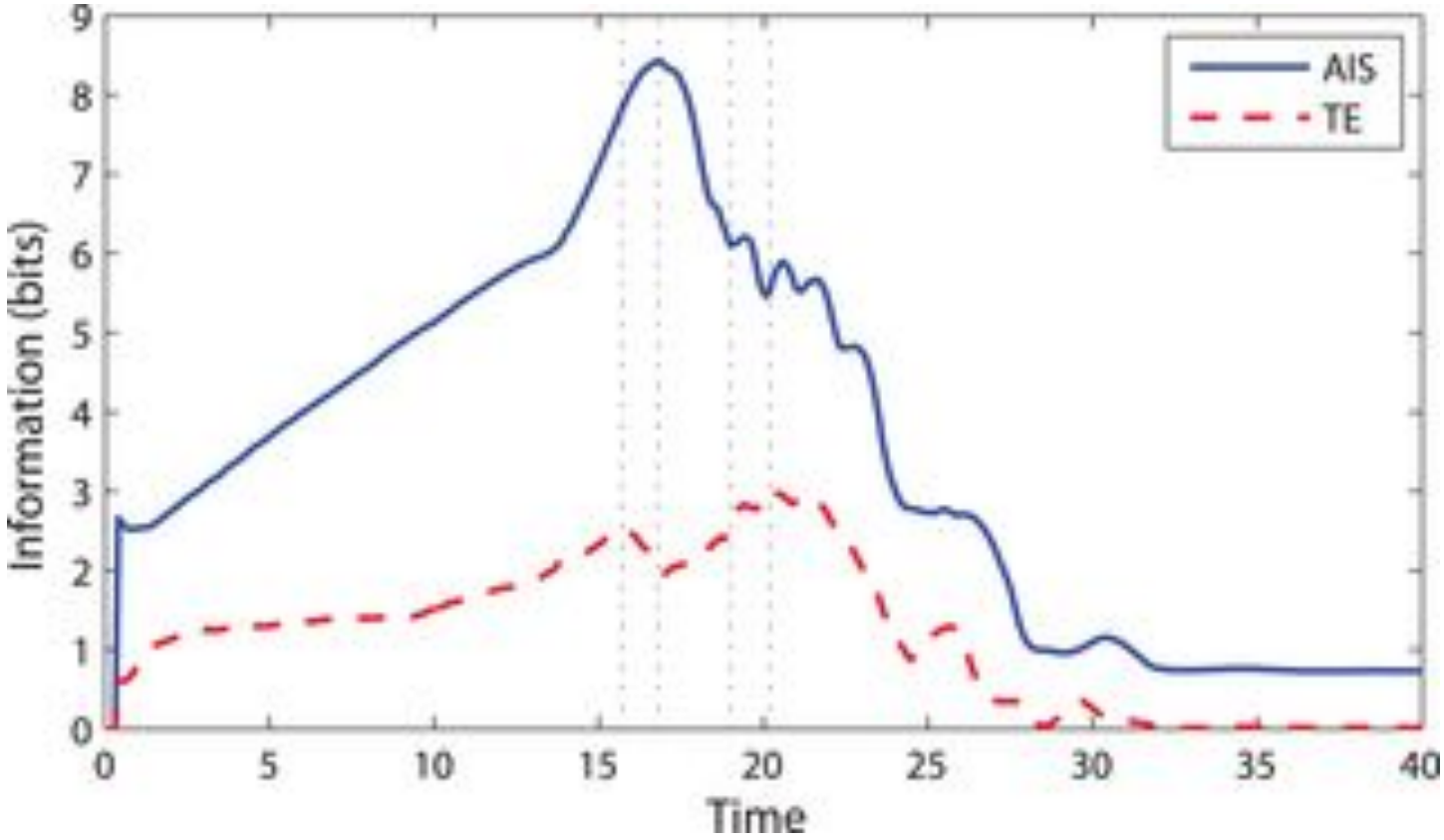
- Influence
- Type of Content
- Timing

Improvement

- **Influence** : In the context of cascades carrying disaster-related information throughout social media networks, the influence of a cascade's originator contributes positively to the cascade's speed of diffusion
- **Type of Content**: In the context of cascades carrying disaster-related information throughout social media networks, speed of diffusion will be higher for cascades carrying information that heightens situational awareness than for cascades carrying other types of information.

Improvement

- **Timing**: In the context of cascades carrying disaster-related information throughout social media networks, the speed of diffusion will be lower for cascades that are launched later than for cascades launched earlier during the progression of a disaster event.


AIS
TE
Information (bits)
Time
0
1
2
3
4
5
6
7
8
9
0
5
10
15
20
25
30
35
40

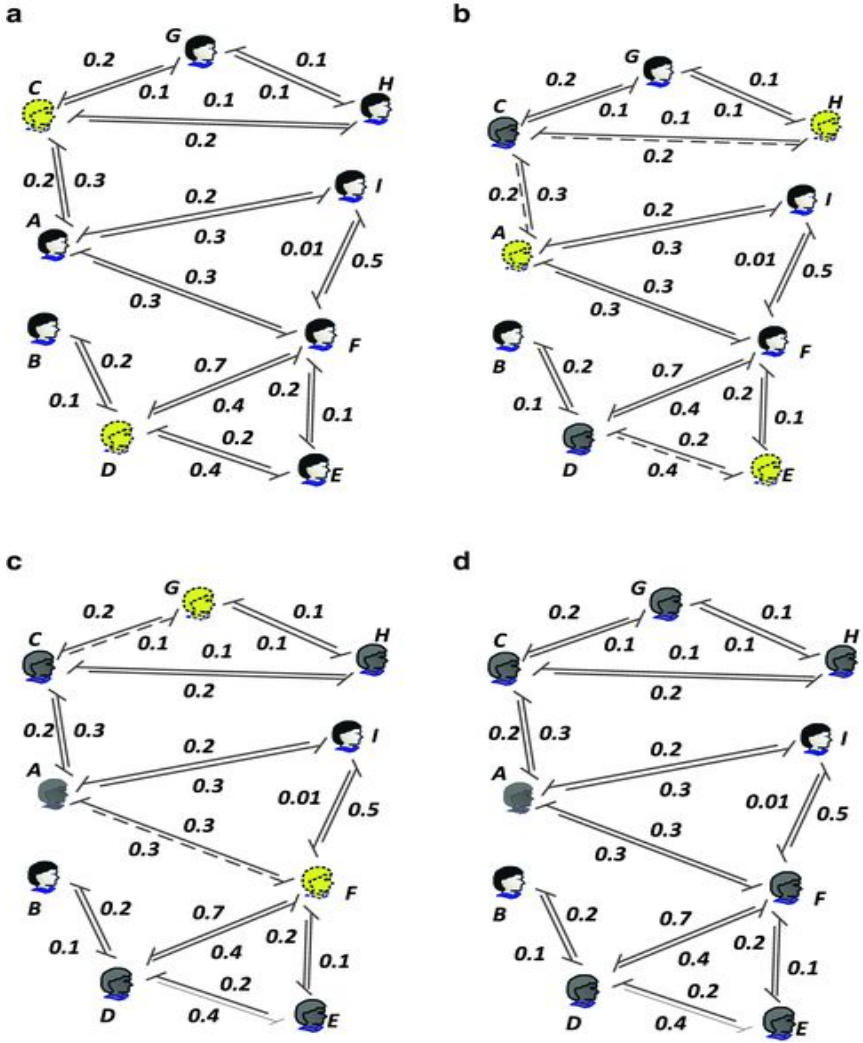
Improvement

Information Diffusion Theory: According to dynamics of information cascades' diffusion in network settings. Two of the seminal models are :

- Independent Cascade (IC) model
- Linear Threshold (LT) model

Independent Cascade Model

IC models assume that an informed member has one chance at a time of independently sharing information with one uninformed member adjacent to her in the network

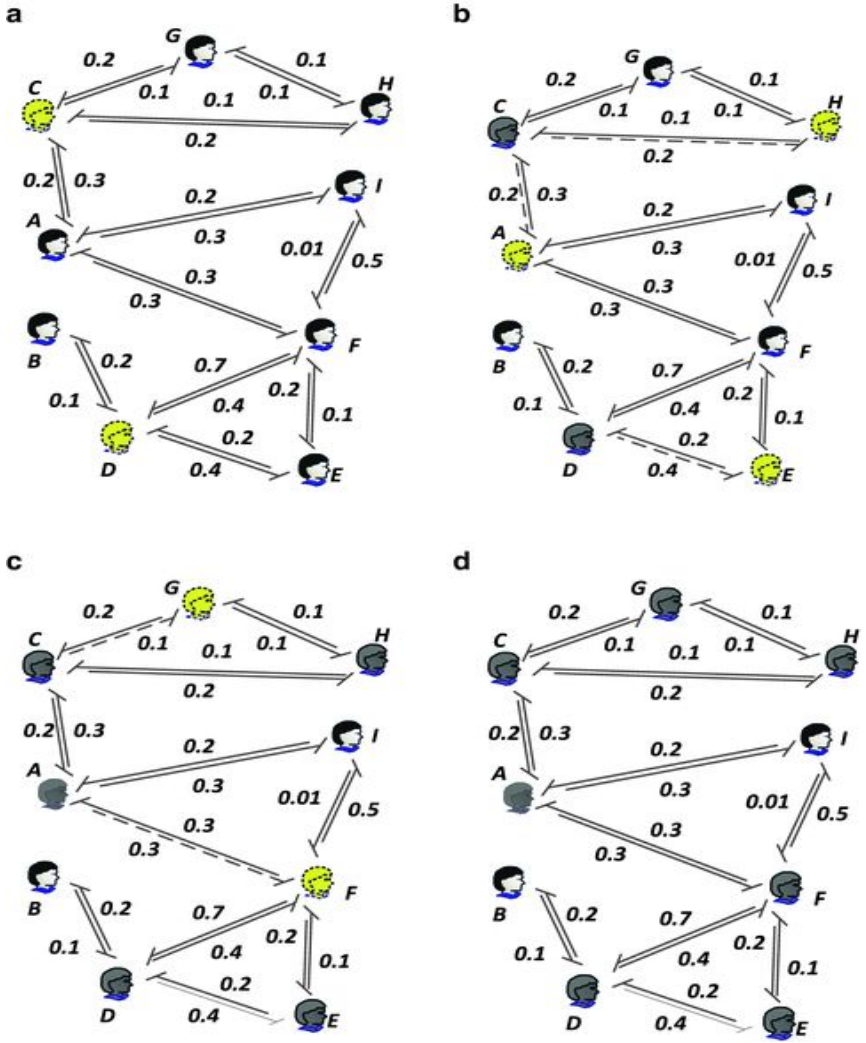


Independent Cascade Model

Parameters :

q: likelihood of uninformed member has a becoming aware of the information when at least one of her neighbors in the network has already become aware of the information

p: that the individual will become aware of this information from external sources

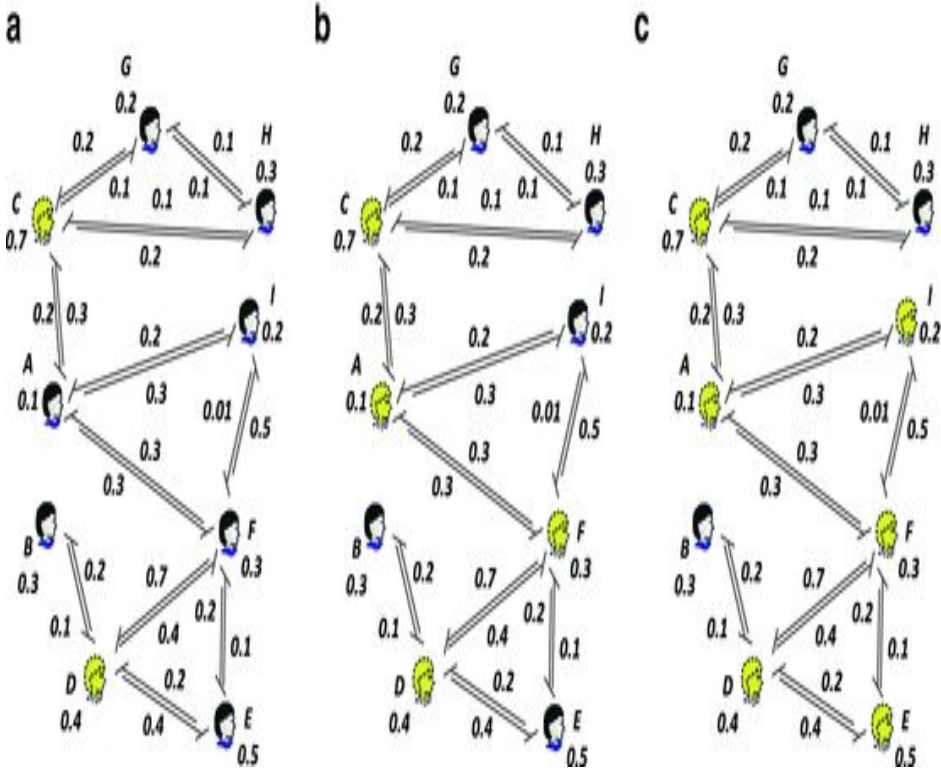


Linear Threshold Method

Participant will share information with her uninformed neighbors in the network. If the number of informed members adjacent to her in the network exceeds her own influence threshold.

Parameter:

ϕ : Threshold rate of informed neighbors. (ϕ=1-q)



Method- Research Methodology

Twitter and Hurricane Sandy :

- Twitter texts are **tweets**, reposting it is called **retweets**
- Largest Hurricane in Northern US
- Appropriate since affected region is developed/ language is English

Method- Research Methodology

Data Collection:

- Tweets and retweets between 26-30 October 2012
- Period includes preparation, during and post processes
- Data gathered from Twitter Search API, filtered by keywords: “Sandy”,”Hurricane”,”Storm” etc.
- Original tweets and retweets extracted to eliminate irrelevant tweets (Manually)

Method- Research Methodology

Data Collection:

Table 1
Irrelevant tweets.

Irrelevant category	Example tweet
Emotional response	"actually really scared of the hurricane coming:("
Joke	"Hurricane Sandy sounds like a delicious mixed drink."
Not related to Sandy	"Yay!!(: hanging out with my bestfriend @strong_sandy"
Opinion	"I get the feeling this hurricane in gonna be just like irene and barley [sic] hit us.."
Song lyric	"The voice that calmed the sea would call out through the rain and calm the storm in me .
Vague forecast	"Sandy is coming"

Exact Method- Research Methodology

Data Collection:

- After irrelevant tweets gone, 333.968 tweets left
- The unit of analysis is **Cascade**
- Original tweets are start of each cascade
- 5683 cascades determined by original tweets that got at least 10 RTs
- 249 cascades eliminated because of lack of information
- 5434 cascades and 311.429 RTs left as final sample

Method- Research Methodology

Operational Measures:

- Agent Based Model (ABM), to measure the cascades' diffusion speed on Twitter's network.
- To evaluate how well IC and LT models work from six aspects
- Two entities of ABM: Knowledge share among users, user connections
- Longest cascade with 4076 participants taken as pattern
- 1.322.814 links in total and 3315 of them is between cascade participants.

Method- Research Methodology

Adoption of New Info

External Influence

- Coefficient p
- Random coef between 0,1
- If less than p, info adopted
- Identical for LT and IC

Internal Influence

- Coefficient q
- **In IC model**:

Random coef between 0,1
If less than q, info adopted

- **In LT model:**

If # of adopted neighbours/ # total neighbours is greater than ϕ=1-q
Information adopted

Method- Empirical Analysis

Simulated Annealing Approach (SA):

- To approximate optimum of the given function
- Y(t): the number of agents in the network who had adopted the information at each time step, t
- Empirical(t): actual number of adopters per time step from our empirical data
- Mean Absolute Percentage Error (MAPE) used for comparison

$$MAPE = 100 \times \frac{1}{n} \sum_{t=0}^{n} \frac{|Empirical(t) - Y(t)|}{Empirical(t)} \qquad (1)$$

Method- Empirical Analysis

Simulated Annealing Approach (SA):
Table below provides distribution of the MAPE values:

Table 4
Descriptive statistics for *MAPE* values.

Model	Median	Mean	SD	Min	Max
IC	13.36	22.91	94.58	0.61	5611.07
LT	12.97	21.89	86.26	1.08	4821.76

- Both IC and LT performed well, IC is chosen to focus
- Cascades' diffusion speed DIFFUSION = q/p
- Easier to obtain q and p in IC model

Method- Empirical Analysis

Determinants:

1. The cascade originator's influence
2. Cascades' contribution to improve situational awareness
3. Timing of the launching of the cascade

Method- Empirical Analysis

Table 3
Variable operationalization.

Construct	Variable label	Operationalization
Information cascade's diffusion speed	*DIFFUSION*	Ratio of *q*/*p* values obtained from the *IC* model
Cascade originator's influence	*INFLUENCE*	Number of users following the cascade originator (at the time of cascade launch)
Cascade content's contribution to situational awareness	*AWARENESS*	Dummy variable coded 1 if the cascade content contributed to situational awareness; 0 otherwise
Lateness in the launch of the cascade during the disaster	*LATENESS*	Lag in the launch of the cascade relative to the start of the data collection (measured in hours)
Incidence of cascade boosts by originator	*BOOST*	Dummy variable coded 1 if originator boosted the cascade; 0 otherwise
Misleading cascade	*FALSE*	Dummy variable coded 1 if the cascade content was misleading; 0 otherwise

Method- Empirical Analysis

$$DIFFUSION_i = \beta_0 + \beta_1\ INFLUENCE_i + \beta_2\ AWARENESS_i + \beta_3\ LATENESS_i + \beta_4\ BOOST_i + \beta_5\ FALSE_i + u_i$$

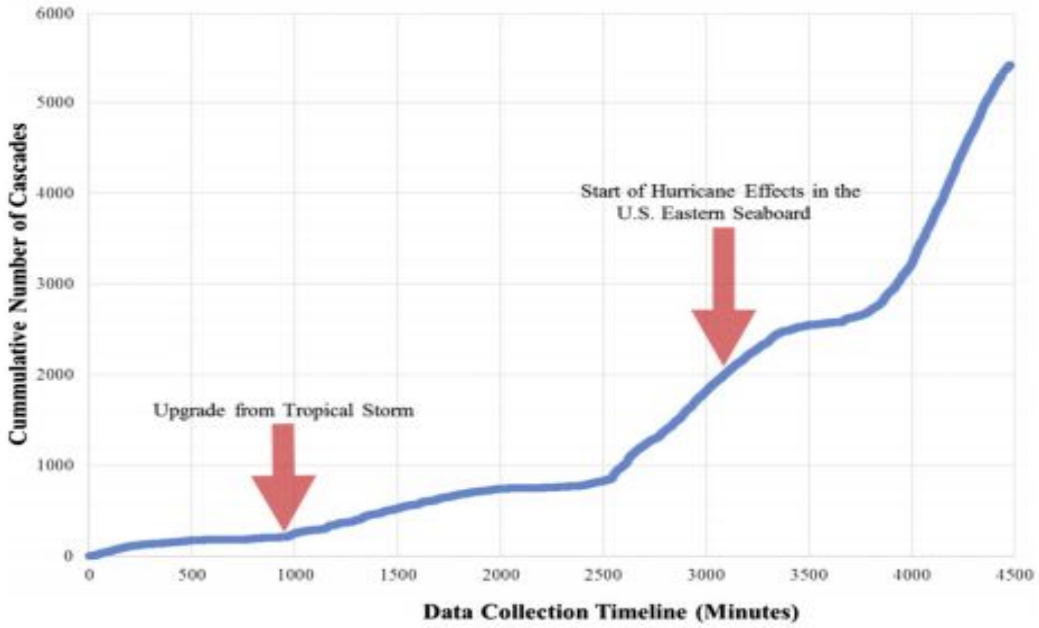


Fig. 1. Cumulative distribution of cascades over time.

Method- Empirical Analysis

Table 5
Correlations and descriptive statistics.

	1	2	3	4	5	6
1. DIFFUSION	1					
2. INFLUENCE	0.03*	1				
3. AWARENESS	−0.01	−0.04**	1			
4. LATENESS	−0.25**	0.01	0.16**	1		
5. BOOST	0.28**	0.01	0.01	−0.09**	1	
6. FALSE	−0.01	−0.04**	0.11**	0.16**	−0.03	1
Mean	37.28	234,447.66	0.78	55.34	0.02	0.04
Std. Deviation	55.63	848,551.55	0.42	18.04	0.13	0.19
Minimum	12.67	0	0	0.00	0	0
Maximum	737.80	9,133,950	1	74.65	1	1

*$p < 0.05$, **$p < 0.01$.

Method- Empirical Analysis

- Out of 5434 cascades, only 648 of the IC model fits had p values at their maximum value
- Of those 648, only 12 had the minimal q values
- Approximately 88% of our cascades, the best model fit was one where internal influence of network connections was much higher than external influence
- Removing the runs where p reached its maximum value changes the mean for DIFFUSION to 40.25

Method- Empirical Analysis

Generalized Linear Model (GLM)(Regression)

- DIFFUSION only took on positive values and displayed a right-skewed distribution of values

Table 6
GLM results.

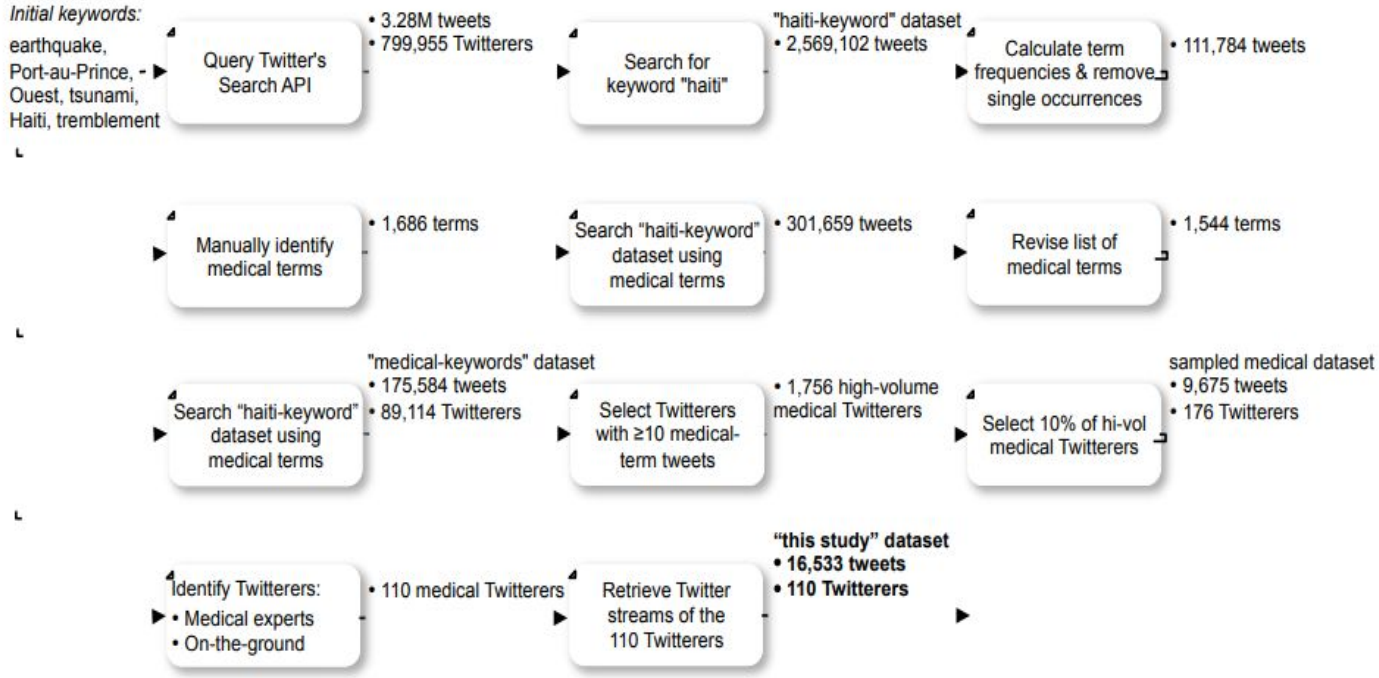
	GLM 1 Coefficients (Std. errors)	GLM 2 Coefficients (Std. errors)
INFLUENCE		1.05E-6 (4.92E-7)*
AWARENESS		−0.27 (1.25)
LATENESS		−0.75 (0.05)**
BOOST	124.31 (20.37)**	100.65 (18.12)**
FALSE	−1.76 (1.64)	9.13 (1.61)**
Intercept	35.36 (0.68)**	77.15 (3.36)**
Scale factor	1.84	1.45
Likelihood Ratio Chi-Square	181.51**	612.32**
AIC	49864.94	49320.01
BIC	49884.74	49359.61
Deviance	3170.49	2619.56
Obs.	5434	5434

*$p < 0.05$, **$p < 0.01$.

Method- Empirical Analysis

01	Proposition 1 (Influence of the Originator)	• Since the INFLUENCE coefficient was positive and statistically different from zero ($p < 0.05$). • **CONFIRMED**
02	Proposition 2 (Situational Awareness)	• Received no support since the coefficient for AWARENESS was not significantly different from zero • **DECLINED**
03	Proposition 3 (Late Launched tweets are more crucial)	• The effect by LATENESS on the dependent variable was negative and statistically different from zero ($p < 0.01$) • **CONFIRMED**

Additional Example-Haiti Earthquake



Alternative Actions

- Diffusion of wrong information about disaster might be easy in social media. Should we find better way to eliminate these wrong information?

  Determining the location of people owned relevant tweets can eliminate spreading wrong information.
  Developing a machine learning tool which measures reliability of the accounts from their previous behaviour.

- RT tool has changed in Twitter.

- Lack of parameters while determining diffusion speed rate.

- Hashtag indexing.

- Mark yourself as safe tool in Facebook and Twitter.

Comparison In Turkey

Van Earthquake:

How was social media used in Van Earthquake?

- Needs were met.
- Twitter was used for communication while Facebook and Blogs were used for organizing.
- Rescue teams, charity campaigns and organizations were implemented in 24 hours.

Comparison In Turkey

#EvimEvindirVan

2 yıl önce geldiler başımıza tac oldular, bugün Van'a dönüyorlar! Uğurlar olsun
Daima baştacısınız! #EvimEvindirVan

#vanicintekyurek

#VanDayanisma

Comparison In Turkey

- More than **500,000** tweet were posted.
- Number of people followed in Google Person Finder: **4900**
- In the first hour of the EvimEvindirVan campaign, more than **15,000** homes offered to share.
- The amount of aid collected for the **#kardeslikzamani** program was over **65,000,000** TL.
- The amount of aid collected for the **#vanicintekyurek** program was over **61,000,000** TL.

Additional References

- Bhatnagar, Abhinav, et al. “The Independent Cascade and Linear Threshold Models.” *SpringerLink*, Springer, Cham, 1 Jan. 1970, link.springer.com/chapter/10.1007/978-3-319-23105-1_4.
- Kıvılcım, Ayşenur. “Sosyal Medya Ve Afet Yonetimi.” *LinkedIn SlideShare*, 4 Nov. 2012, www.slideshare.net/AyenurKvlcm/sosyal-medya-afet-yonetimi.
- “Models, Solutions and Enabling Technologies in Humanitarian Logistics.” *European Journal of Operational Research*, North-Holland, 26 Nov. 2014, www.sciencedirect.com/science/article/pii/S0377221714009539.