Social Media Analytics for Natural Disaster Management: Framework and Implementation

Xinyue **Ye**, Ph.D.,
Founding Director, Computational Social Science Lab
Kent State University, Kent, OH

Kent State University
COMPUTATIONAL SOCIALSCIENCE LAB

Research on Sustainable Communities

- Urban and regional planner
- Economic geographer on public policy evaluation
- Geographic information scientist and spatial econometrician

- A collaborator teaming with social/environmental scientists and computational scientists on **sustainable development** and **smart cities** projects funded by NSF, DOC, and DOE

Torsten Hägerstrand

Farmers are more likely to adopt innovations if they are in close proximity to earlier adopters.

Hägerstrand, Torsten (1967) **Innovation diffusion as a spatial process**. Chicago: University of Chicago Press.

The study of just groups creates a homogenization of reality and hides the truth.

Hägerstrand, Torsten (1970). **What about people in regional science**?. Papers of the Regional Science Association. 24 (1): 6–21.

Science AAAS

Home | News | Journals | Topics | Careers

HHS Public Access
Author manuscript
Peer-reviewed and accepted for publication

About author manuscripts | Submit a manuscript

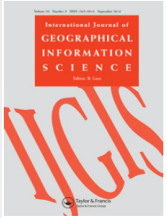
Science. Author manuscript; available in PMC 2009 Sep 16.
Published in final edited form as:
Science. 2009 Feb 6; 323(5915): 721–723.
doi: 10.1126/science.1167742

PMCID: PMC2745217
NIHMSID: NIHMS98137

Life in the network: the coming age of computational social science

David Lazer, Alex (Sandy) Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne

*(big data) offer increasingly comprehensive pictures of both **individuals and groups**, with the potential of **transforming our understanding of our lives, organizations, and societies** in a fashion that was barely conceivable **just a few years ago**.*



International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

Editorial: human dynamics in the mobile and big data era

Shih-Lung Shaw, Ming-Hsiang Tsou & Xinyue Ye

Advancements in location-aware technology, information and communication technology, and mobile technology have transformed the focus of urban science towards **spatial, temporal, and dynamic relationships** of human behaviors and the environment.

Detailed data of individual activities and interactions are being collected by communication service providers, online applications, private companies, and government agencies.

Activities and interactions taking place in **virtual space** are related to the activities and interactions in **physical geographic space**.

Sustainable and Smart Communities

- ***2014-2018, Spatiotemporal Modeling of Human Dynamics across Social Media and Social Networks, National Science Foundation, $999,887***

- 2015-2018, SI2-SSE: Collaborative Research: TrajAnalytics: A Cloud-based Visual Analytics Software System to Advance Transportation Studies Using Emerging Urban **Trajectory** Data, National Science Foundation, $300,000

- 2016-2018, S&CC: Support **Community-Scale Intervention** Initiatives by Visually Mining Social Media **Trajectory** Data, National Science Foundation, $100,190

Motivation

- What about people in human-environment interaction and emergency response?

- Social media in natural disaster management
  - Broadcast situational announcements
  - Solicit on-the-ground information

- Applications
  - Event detection
  - Rapid assessment of disaster damage
  - Situational awareness

Four Dimensions

- *Space*. Two types of spatial information in social media messages: exact coordinates (i.e., longitudes and latitudes) and toponyms (e.g., a city name).

- *Time. Every social media message comes with a high-resolution timestamp, recording the exact time when a message was posted.*

Four Dimensions

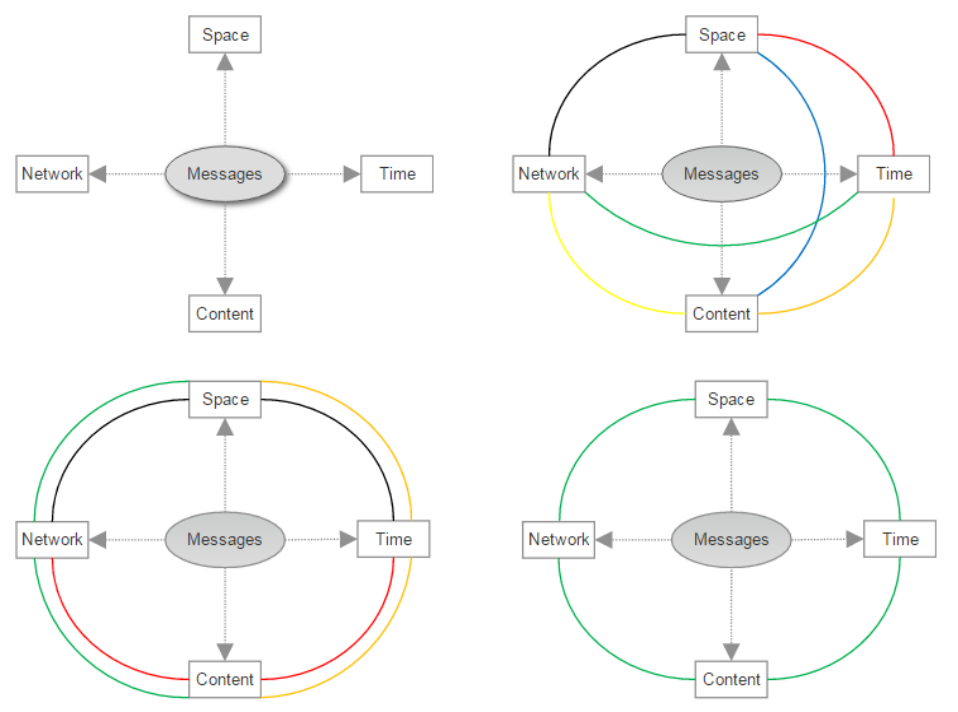
- *Content.* The content of social media messages varies from texts to images. Twitter mainly serves as a social networking site for users to exchange text messages, while a major function of Instagram is to allow users to share images and photos.

- *Network.* Various relationships (e.g., retweet, reply, mention, and friends/followers) recorded by social media sites could be employed to formulate networks.

Focusing on social media information

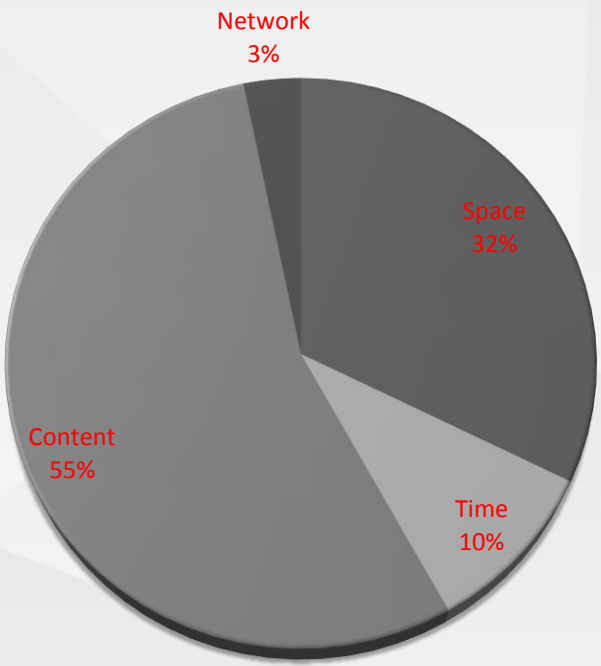
- A large part of studies involve multiple dimensions of social media data in their analyses.
- There are both separate analyses and simultaneous analyses for dimensions.
- There are few simultaneous analyses as dimensions increase.

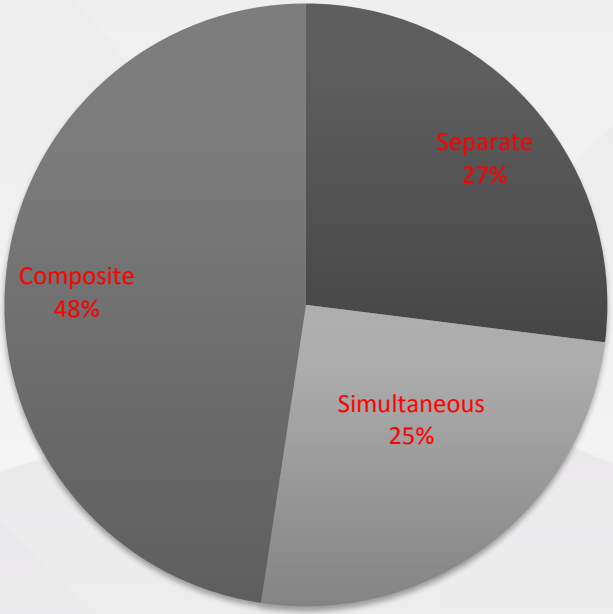
Focusing on social media information



Combinations of four dimensions in social media data

Focusing on social media information





Focusing on social media information

Combination of dimensions	Data analysis tasks
Space	Where is the hot spot of people's responses to a disaster? For example, are the impact areas the hot spots of disaster-related social media activities?
Time	How do people's responses change with the evolution of a disaster (before, during and after)? For example, when do disaster-related social media activities reach peak in the process of a disaster?
Content	How do people's responses vary according to their posted content? For example, how many social media feeds report power outage in a disaster?
Network	Who are the important players in spreading disaster-related information on social media in a disaster? For example, how many reposted messages are originally from emergency management agencies?

Focusing on social media information

Combination of dimensions	Data analysis tasks
Space∩Time	How do people's responses to a disaster vary across space and over time? For example, do people's social media activities from the impact area form a significant hot spot immediately after being struck by a disaster?
Space∩Content	How do people's conversational topics related to a disaster on social media vary across space? For example, do people proximate to the impact area have more on-topic messages than distant people do?
Space∩Network	What is the spatial manifestation of the network structure in a disaster? For example, who are the local opinion leaders in disseminating disaster-related information for a given place?
Time ∩Content	How do people's conversational topics vary with the evolution of a disaster? For example, do people change their topics from preparedness (e.g., survival kits and food stock) to impact (e.g., damage and casualty)?
Time∩Network	What is the temporal manifestation of the network structure in a disaster? For example, is the same set of opinion leaders dominant in all phases of a disaster?
Content∩Network	Which topic goes viral in a disaster situation? For example, how do rumor messages spread across the social network?

Focusing on social media information

Combination of dimensions	Data analysis tasks
Space∩Time∩Network	What is the space-time manifestation of the network structure in a disaster? For example, is the same set of local opinion leaders dominant in all phases of a disaster for a given place?
Space∩Content∩Network	How does geographical space characterize the diffusion of social media messages under a certain topic? For example, what is the spatial extent of the spreading of rumor messages in a disaster?
Time∩Content∩Network	What is the temporal dynamics of the diffusion of social media messages under a certain topic? For example, how long do rumor messages last for spreading?
Space∩Time∩Content∩Network	How do space and time jointly characterize the diffusion of social media messages under a certain topic? For example, what is the space-time extent of the diffusion of rumor messages in a disaster?

Fusing social media data with authoritative data

❖ *Fusing with remote sensing data*

	Remote sensing	Social media
Strengths	• Providing information for poorly accessible areas or areas with sparse ground measurements • Capturing physical features	• Real-time data • Freely accessible • Recording human activities
Limitations	• Not all data are freely accessible • Could be influenced by cloud and vegetation cover • Lengthy revisit time	• Quality and reliability problems • Unstructured data • Digital divide

Fusing social media data with authoritative data

- *Fusing with census data*
  - Socio-political ecology of disasters
  - Demographic and socioeconomic characteristics shape risk/disaster perceptions
  - Little information on demographic and socioeconomic characteristics from social media

Case Study of Twitter for wildfire hazards

Wang, Z., **Ye, X *,** & Tsou, M. (2016) Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Natural Hazards,* doi: 10.1007/s11069-016-2329-6

01 Introduction

As more and more fire-prone areas have been urbanized, people's livelihoods in the western USA have been severely influenced by the increasingly frequent wildfires.

1. Introduction

1

wildfire exposure modeling

(Ager et al. 2014a, b; Thompson et al. 2015; Youssouf et al.2014)

2

wildfire risk assessment

(Chuvieco et al. 2010, 2012; Martı´nez et al. 2009; Padillaand Vega-Garcı´a 2011; Rodrigues et al. 2014)

3

wildfire and wildland–urban interface(WUI)

(Herrero-Corral et al. 2012; Massada et al. 2009; Schulte and Miller 2010)

4

wildfire–climate interactions

(Gillett et al. 2004; Liu et al. 2014; Westerling et al. 2006)

In order to achieve a better understanding of the occurrences and patterns of spread of wildfires, efforts by domain scientists have been made from various perspectives

1. Introduction

5

wildfire management agencies have incorporated various wildfire detection systems, e.g., the general public, lookout towers, terrestrial mobile brigades, and aerial reconnaissance (Rego et al. 2013)

6

The Wildland Fire Decision Support System (WFDSS) has been developed (Calkin et al. 2011)

In order to achieve a better understanding of the occurrences and patterns of spread of wildfires, efforts by domain scientists have been made from various perspectives

1. Introduction

Space and time are strongly related to situational awareness in emergency events.

Case studies

- **De Albuquerque et al. (2015)**: carried out a spatial analysis and found a strong spatial relationship between locational proximity to floods and the usefulness of the messages for crisis management
- **Guan and Chen (2014)**: found that the ratio of tweets associated with Hurricane Sandy to general tweets increased gradually before this disaster, peaked when it landed, and then gradually decreased
- **Huang and Xiao (2015)**: indicated that messages posted by Twitter users varied with the temporal process of a disaster and thus could provide useful information for improving situational awareness

1. Introduction

Mining the actual content of social media messages to improve knowledge about disaster situations.

Qu et al. (2011)

Developed a platform for emergency situational awareness, which could detect emergent incidents and classify tweets as interesting or not

Imran et al. (2013)

Designed an Artificial Intelligence for Disaster platform

Divided the earthquake related microblog messages with valuable information for improving situational awareness into four categories

Cameron et al. (2012)

Utilized machine learning methods to extract informative Twitter messages

Imran et al. (2014)

1. Introduction

In disaster situations, people may also tend to obtain situational updates and gain situational awareness from the informative messages shared by opinion leaders.

Case studies

- Cheong and Cheong (2011) — found that local authorities, traditional media reporters, and, etc. are important players in spreading situational information during 2010–2011 Australian floods.
- Kogan et al. (2015) — indicated that local government authorities and the media are the most important nodes in the retweet network during the 2012 Hurricane Sandy.
- Starbird and Palen (2010) — a similar phenomenon was also observed.

Case study in San Diego

This case study presents the findings from **examining the spatial and temporal variations of wildfire-related tweets** and from our attempt to **characterize wildfire by the discussion topics in the collected tweets**, as well as from investigating the **role of opinion leaders** in people's acquisition of wildfire-related information.

Introduce data

Related methodology

Findings and implications

Next step

02 Data and methodology

2.1. Data

We used Twitter search API (https://search.twitter.com/) to collect wildfire-related Tweets. Our collection process included two phases.

- First, collect any tweet that contained either of the two keywords: fire and wildfire.
- Second, glean tweets associated with specific wildfires based on keywords which are places where wildfires occurred. The keywords were randomly selected from a list of places (see Table 2).
- Third, check whether a fire or wildfire also appeared in the collected tweets.

2.1. Data

Table 2 Overview of the major wildfires in May, 2014. *Source*: complied from http://www.fire.ca.gov/

Major wildfires	Time of outbreak (UTC)	Time of 100 % contained (UTC)	Location	Long/lat	Acres
Bernardo Fire	May 13, 11:00	May 17, 20:14	Off Nighthawk Lane, southwest of Rancho Bernardo	−117.133/33.003	1548
Tomahawk Fire	May 14, 9:45	May 19, 9:20	Traveled from Naval Weapons Station, Fallbrook to Camp Pendleton	−117.285/33.353	5367
Poinsettia Fire (Carlsbad fire)	May 14, 10:30	May 17, 12:00	Off Poinsettia Ln & Alicante Rd in Carlsbad	−117.278/33.112	600
Highway Fire	May 14, 13:00	May 15, 18:30	Off Old Hwy 395 and I-15 in the Deer Springs area	−117.162/33.312	380
River Fire	May 14, 12:12	May 19, 9:20	North River Road and College Blvd., Oceanside	−117.747/33.251	105
Cocos Fire (San Marcos Fire)	May 14, 16:00	May 22, 18:15	Village Drive and Twin Oaks Road, San Marcos	−117.160/33.114	1995
Freeway Fire	May 14, 17:43	May 20, 11:30	Naval Weapons Station, Fallbrook	−117.260/33.370	56
Pulgas Fire	May 15 14:45	May 21, 17:00	Off Interstate 5 at Las Pulgas Rd, north of Oceanside	−117.463/33.303	14,416
San Mateo Fire	May 16, 11:24	May 20, 23:30	In the Talega area of Marine Corps Base Camp Pendleton	−117.300/33.286	1457

2.1. Data

Tweets collected in the first phase could be used in analysis of all dimensions (i.e., space, time, content, and network).

Tweets gleaned in the second phase are of particular importance for spatial analysis.

Our study period spans from May 13, 2014, when the first wildfire occurred, to May 22, 2014, when most of the destructive wildfires were 100 % contained. A radius of 40 miles was set to specify a circular area (centered at downtown) to cover the majority of San Diego County.

2.2. Methodology

Several specific methods were used in our study:

Kernel density estimation (KDE)

analyze the spatial pattern of wildfire-related tweets

Text mining

identify conversational topics

Social network analysis

detect the opinion leaders in wildfire hazards

2.2. Methodology: KDE

- KDE imported the coordinates of tweets and exported a raster formatted map where each cell was assigned a value to represent the intensity level (Han et al. 2015).

- To deal with the impact of population, a dual kernel density estimation (Dual KDE) was employed.

Dual KDE = Each Cell Value of Tweets / Each Cell Value of Population

2.2. Methodology: Text Mining

Identifying important terms and term clusters in wildfire-related tweets using the“tm”package in R

FIRST

cleaned the raw tweets by removing URLs and stop words

SECOND

Obtained a term-document matrix, where a row stood for a term and a column for a tweet

THIRD

With k-means clustering method, terms which appeared frequently in the same document were grouped into one cluster.

03

Spatial and temporal analysis of wildfire Twitter activities

3. Spatial and temporal analysis of wildfire Twitter activities

First

Check the temporal evolution of wildfire tweets and compare it with the wildfire's temporal evolution.

Second

Examine whether the impact areas are clusters of wildfire tweets.

3. Spatial and temporal analysis of wildfire Twitter activities

Some basic spatiotemporal information of the major wildfires occurred in the study period

Table 2 Overview of the major wildfires in May, 2014. *Source*: complied from http://www.fire.ca.gov/

Major wildfires	Time of outbreak (UTC)	Time of 100 % contained (UTC)	Location	Long/lat	Acres
Bernardo Fire	May 13, 11:00	May 17, 20:14	Off Nighthawk Lane, southwest of Rancho Bernardo	−117.133/33.003	1548
Tomahawk Fire	May 14, 9:45	May 19, 9:20	Traveled from Naval Weapons Station, Fallbrook to Camp Pendleton	−117.285/33.353	5367
Poinsettia Fire (Carlsbad fire)	May 14, 10:30	May 17, 12:00	Off Poinsettia Ln & Alicante Rd in Carlsbad	−117.278/33.112	600
Highway Fire	May 14, 13:00	May 15, 18:30	Off Old Hwy 395 and I-15 in the Deer Springs area	−117.162/33.312	380
River Fire	May 14, 12:12	May 19, 9:20	North River Road and College Blvd., Oceanside	−117.747/33.251	105
Cocos Fire (San Marcos Fire)	May 14, 16:00	May 22, 18:15	Village Drive and Twin Oaks Road, San Marcos	−117.160/33.114	1995
Freeway Fire	May 14, 17:43	May 20, 11:30	Naval Weapons Station, Fallbrook	−117.260/33.370	56
Pulgas Fire	May 15 14:45	May 21, 17:00	Off Interstate 5 at Las Pulgas Rd, north of Oceanside	−117.463/33.303	14,416
San Mateo Fire	May 16, 11:24	May 20, 23:30	In the Talega area of Marine Corps Base Camp Pendleton	−117.300/33.286	1457

3. Spatial and temporal analysis of wildfire Twitter activities

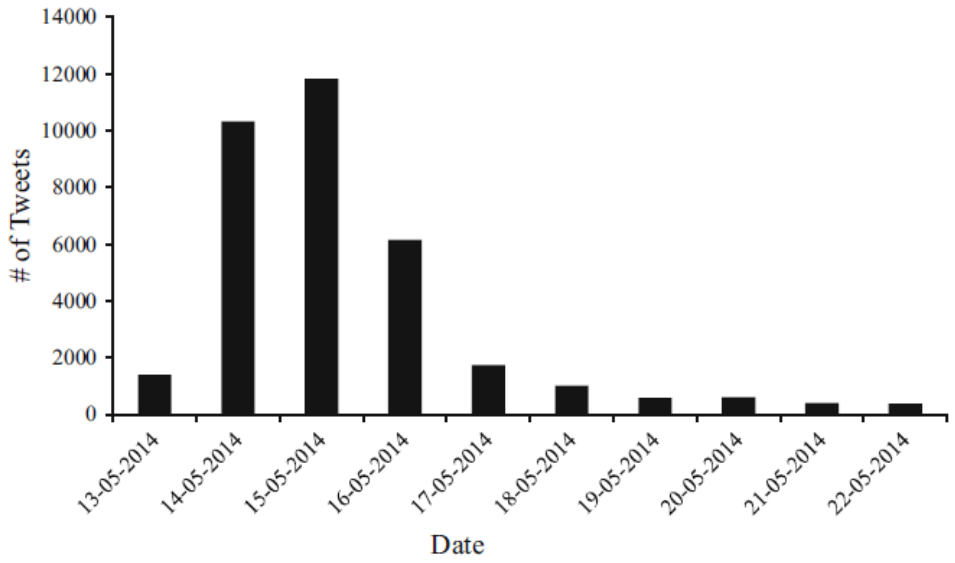
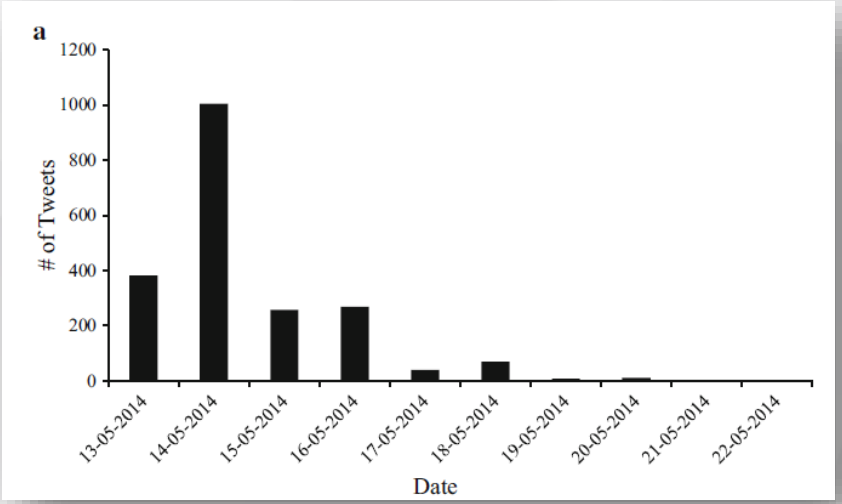


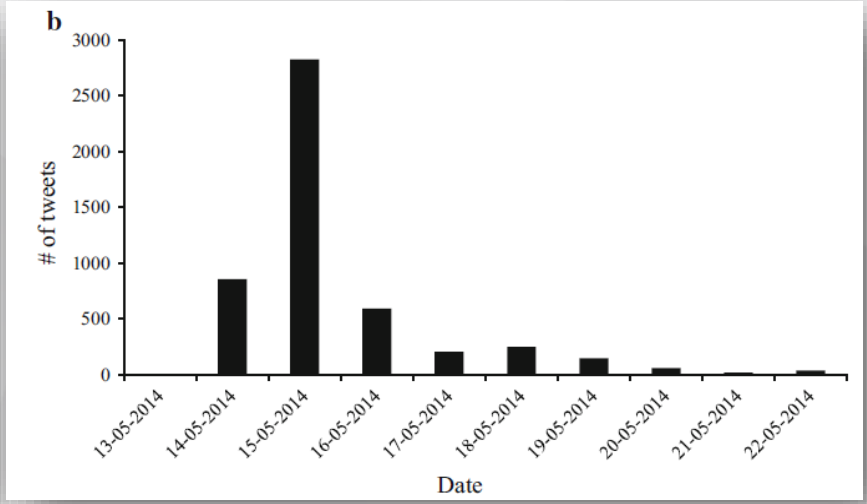
Fig. 1 Temporal evolution of wildfire-related tweets with keywords of "fire" and "wildfire"

Six of the nine wildfires occurred on May 14, explaining why May 14 experienced a sudden increase in wildfire tweets (Fig. 1).

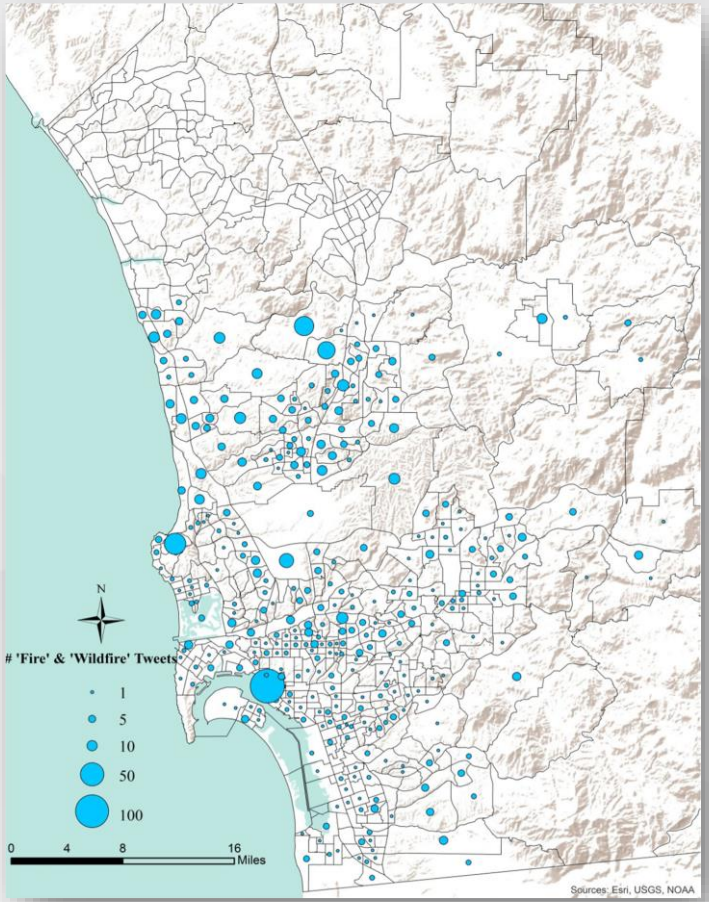
3. Spatial and temporal analysis of wildfire Twitter activities

- A temporally concurrent evolution of wildfire and its related tweets can be observed.
- Both Bernardo fire (a) and San Marcos fire (b) had their corresponding tweets peak on the day after the breakout day. This 1-day time lag is probably because it takes time to spread information.





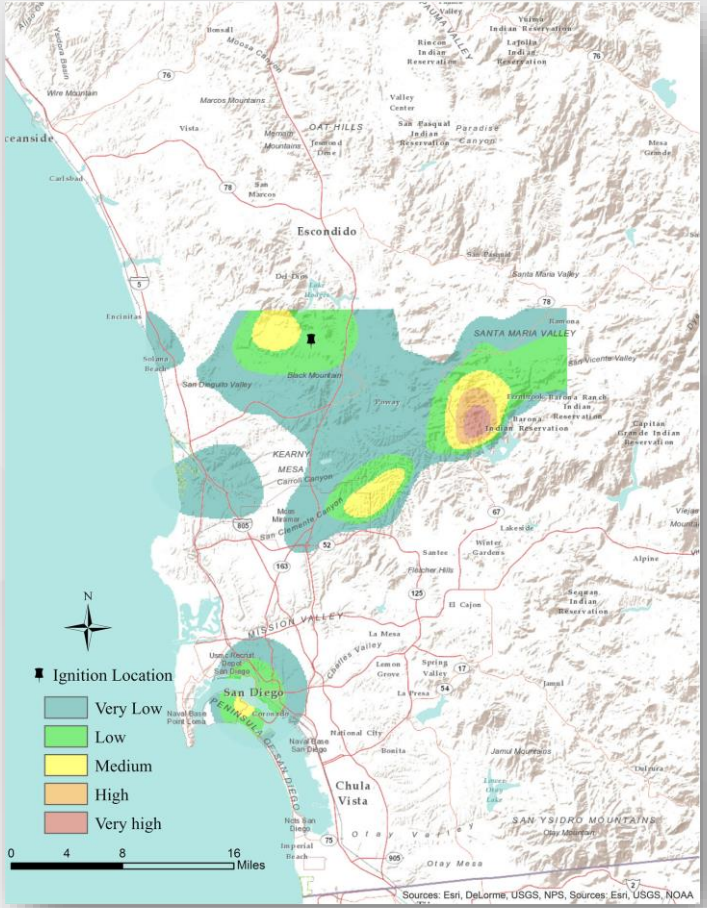
3. Spatial and temporal analysis of wildfire Twitter activities



Downtown area is the largest hot spot in terms of the number of fire and wildfire tweets.

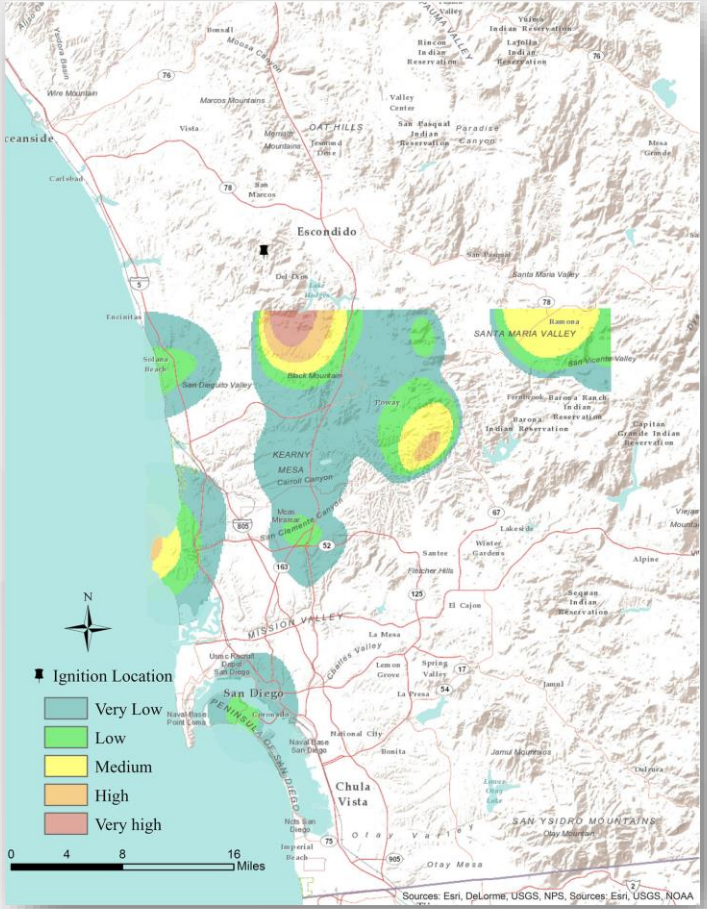
This may be due to the fact that a large population generate numerous Twitter activities.

3. Spatial and temporal analysis of wildfire Twitter activities



To filter out the influence of population, dual KDE is performed to detect the clusters of tweets related to Bernardo fire and Cocos fire

3. Spatial and temporal analysis of wildfire Twitter activities

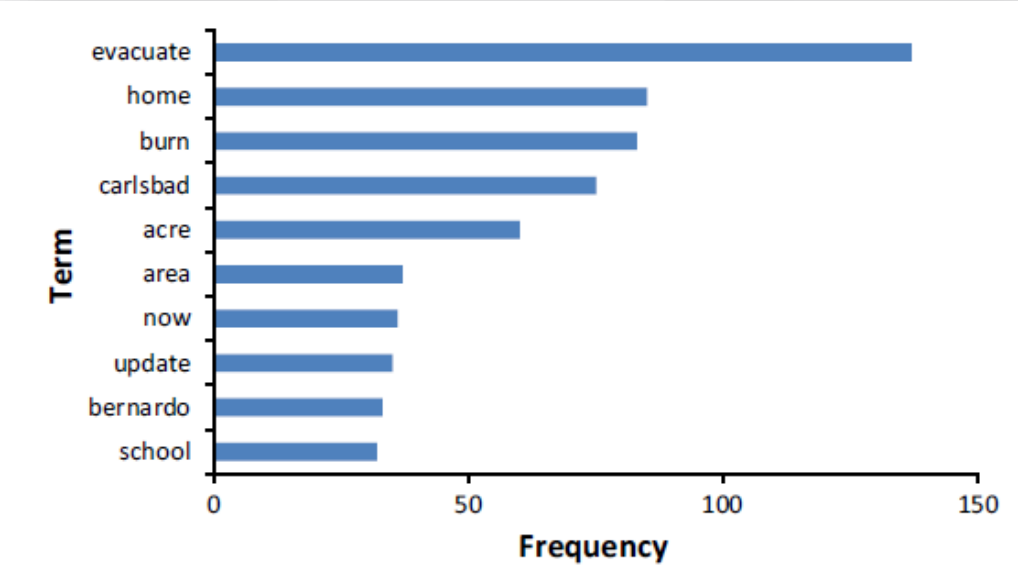


The downtown area has become a low-value cluster, whereas clusters with values higher than medium are close to the wildfires ignition locations

04 Topics and network

4. Topics and network

The importance of a term in tweets (the top 10 frequent words): if a term appears frequently in tweets, it is regarded as important.



the most important term is evacuate, because the most urgent thing in wildfire situations is to evacuate

a large part about the evacuation of homes results in a high frequency of home

4. Topics and network

The seven clusters and only top three terms within each cluster are shown. The number of clusters specified here is to ensure that we get the most but differentiated topics.

Number	Term clusters
Cluster 1	know; thank; firefight
Cluster 2	home; Carlsbad; burn
Cluster 3	wind; Carlsbad; area
Cluster 4	Carlsbad; contain; acre
Cluster 5	burn; evacuate; 4S Ranch
Cluster 6	acre; burn; contain
Cluster 7	evacuate; home; Bernardo

- cluster 1 stands for the topic related to thankfulness to firefighters
- cluster 2 is about the burned homes in Carlsbad
- cluster 3 is about the wind fanning the wildfire in Carlsbad area
- cluster 4 discloses the containment percentage and impacted acres of Carlsbad wildfire

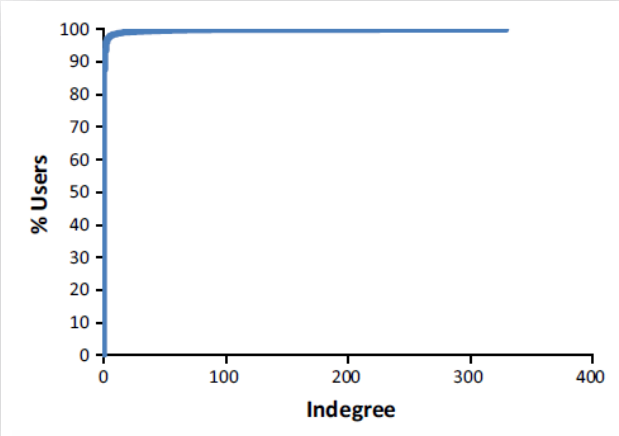
4. Topics and network

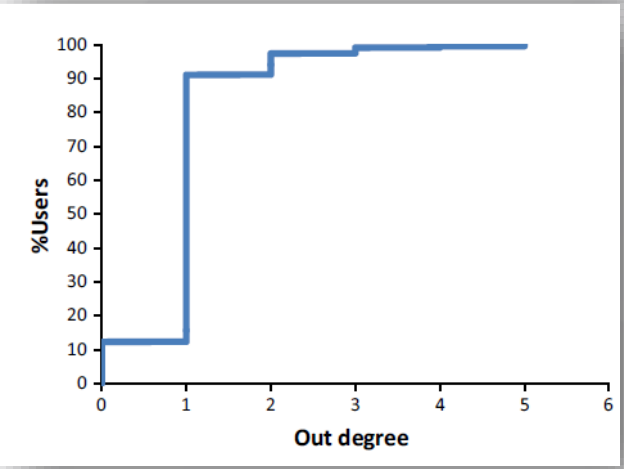
The seven clusters and only top three terms within each cluster are shown. The number of clusters specified here is to ensure that we get the most but differentiated topics.

Number	Term clusters
Cluster 1	know; thank; firefight
Cluster 2	home; Carlsbad; burn
Cluster 3	wind; Carlsbad; area
Cluster 4	Carlsbad; contain; acre
Cluster 5	burn; evacuate; 4S Ranch
Cluster 6	acre; burn; contain
Cluster 7	evacuate; home; Bernardo

- cluster 5 represents the evacuation caused by a burning wildfire in 4S Ranch
- cluster 6 is on damage report
- cluster 7 is on the evacuation in Bernardo.

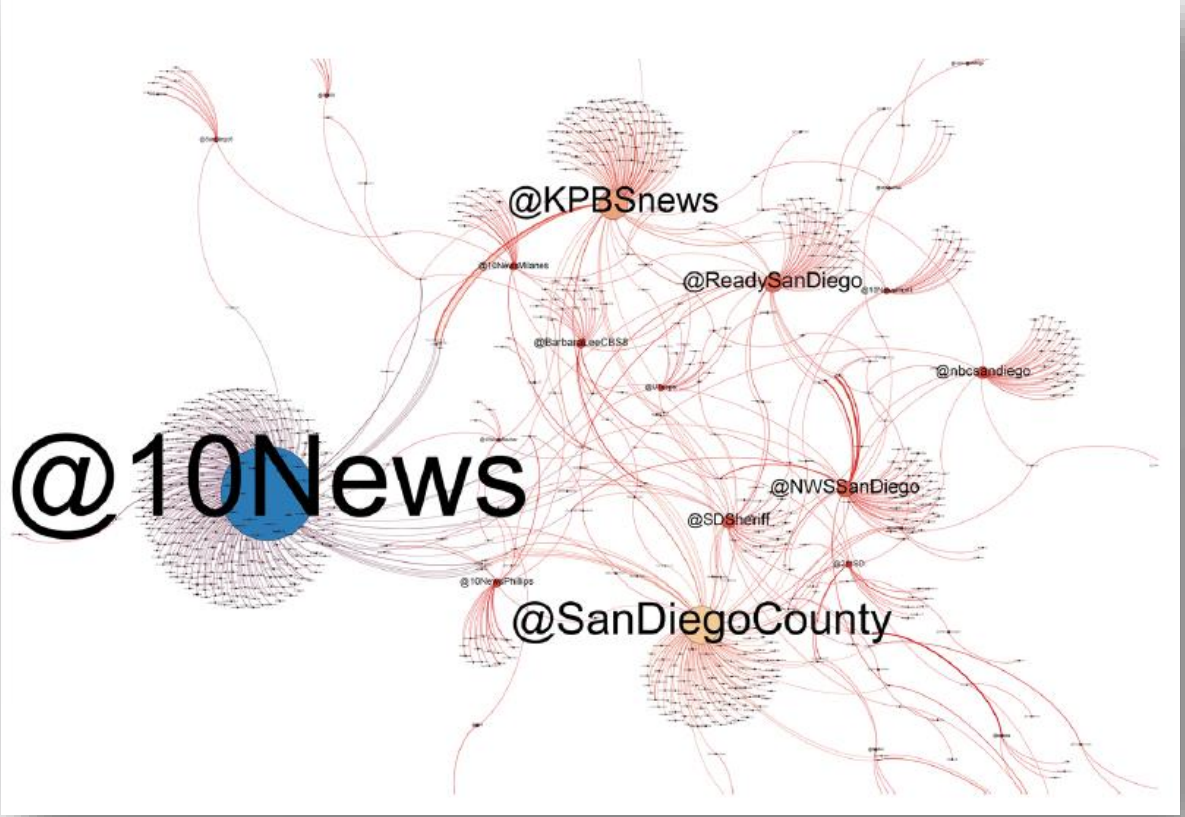
4. Topics and network





- The social network analysis is built based on the retweet relationship. We calculate the indegree and outdegree for each node.
- More than 85% nodes had no users retweet their messages.
- About 90% of users retweeted only one user or none.
- There are dominant users acting as hubs in the information exchange.

4. Topics and network



The nodes of @10news, @KPBSnews, and @nbcsandiego are Twitter accounts owned by three local news media in San Diego.

05 Conclusion and discussion

5. Conclusion and discussion

spatial and temporal patterns of wildfire-related tweets

a temporally concurrent evolution of wildfire and wildfire-related Twitter activities

Mining topics can extract useful information

strong situational awareness during emergency events

some elite users such as local authorities and traditional media reporters dominant in the retweet network

opinion leaders play an important role

simultaneous analysis of the four dimensions might be able to provide new insights

simultaneous analysis

Limitations and future work

First

although the searching range could cover the majority of San Diego County, some places where wildfire occurred were not contained.

Second

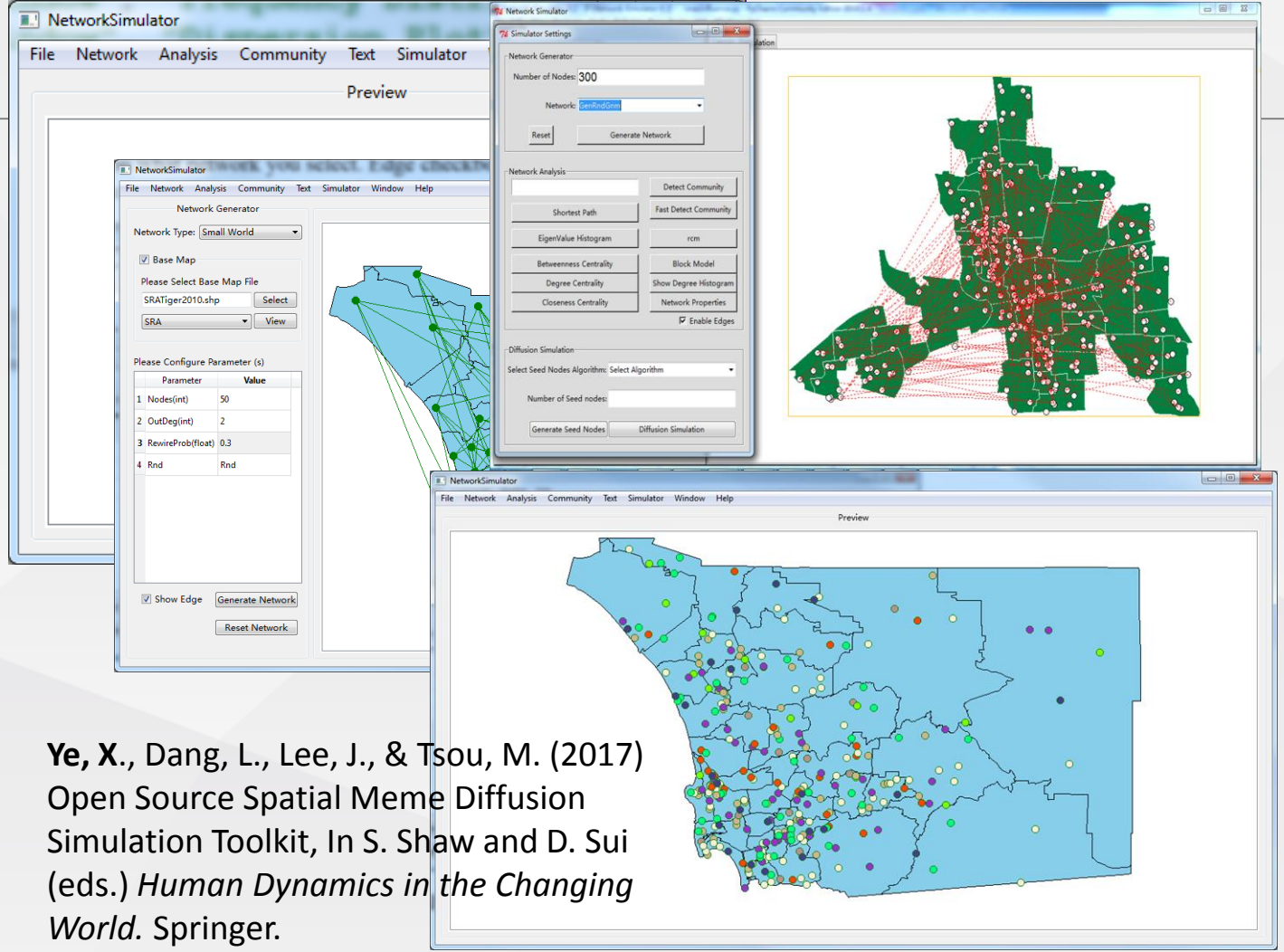
whether the 1% sampled data are a valid representation of the overall wildfire Twitter activities.

Third

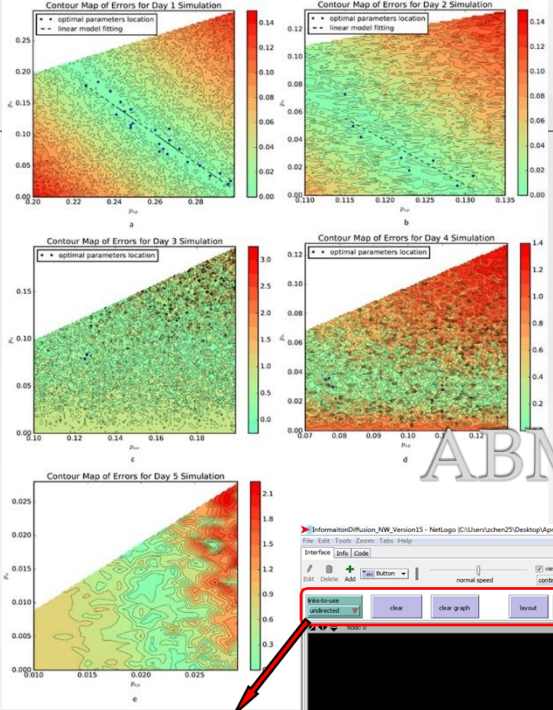
the social network in our research is only based on the retweet relationship, while other types can be used in the future study.

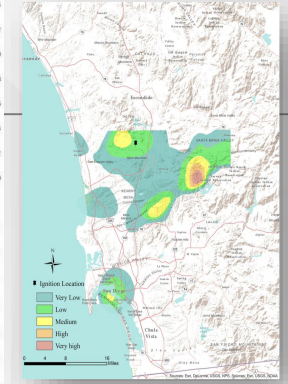
Fourth

overlooks the information diffusion process including its components, phases, and characteristics.

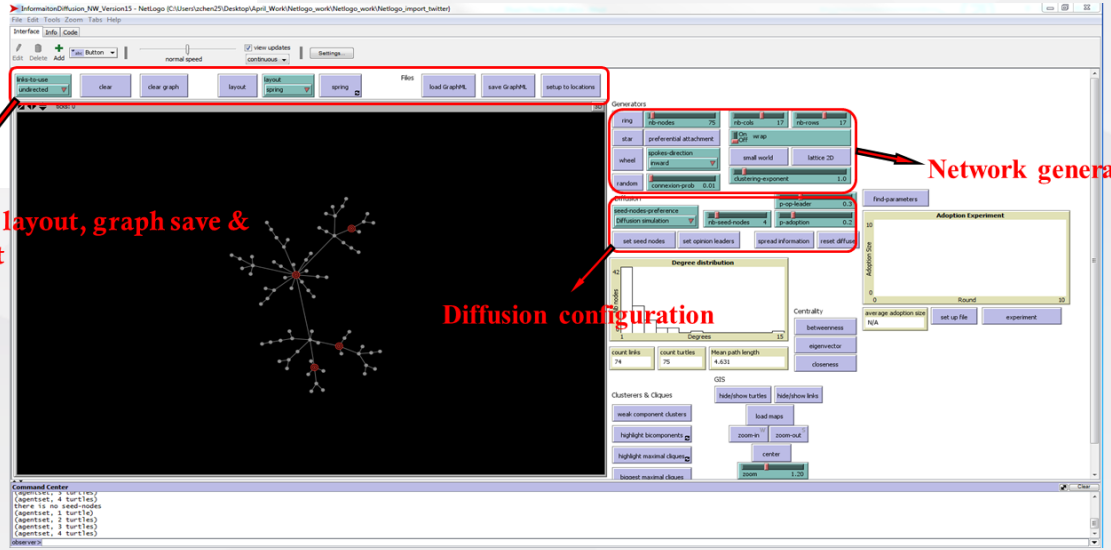


Ye, X., Dang, L., Lee, J., & Tsou, M. (2017) Open Source Spatial Meme Diffusion Simulation Toolkit, In S. Shaw and D. Sui (eds.) *Human Dynamics in the Changing World.* Springer.





ABM designed in NetLogo



Social Media Analytics for Natural Disaster Management: Framework and Implementation

Xinyue **Ye**, Ph.D.,
Founding Director, Computational Social Science Lab
Kent State University, Kent, OH

Kent State University
COMPUTATIONAL SOCIALSCIENCE LAB