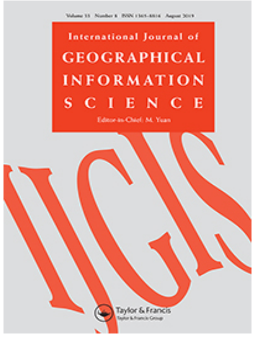

International Journal of Geographical Information Science

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

Simulating the spatial diffusion of memes on social media networks

Lanxue Dang, Zhuo Chen, Jay Lee, Ming-Hsiang Tsou & Xinyue Ye

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2019, VOL. 33, NO. 8, 1545–1568
https://doi.org/10.1080/13658816.2019.1591414

RESEARCH ARTICLE

Simulating the spatial diffusion of memes on social media networks

Lanxue Dang[a], Zhuo Chen[b], Jay Lee[b,c], Ming-Hsiang Tsou[d] and Xinyue Ye[e]

[a]Henan Key Laboratory of Big Data Analysis and Processing, School of Computer and Information Engineering, Henan University, Kaifeng, Henan, China; [b]Department of Geography, Kent State University, Kent, OH, USA; [c]College of Environment and Planning, Henan University, Kaifeng, Henan, China; [d]Department of Geography, San Diego State University, San Diego, CA, USA; [e]Urban Informatics & Spatial Computing Lab, Department of Informatics, New Jersey Institute of Technology, Newark, NJ, USA

ABSTRACT

This article reports the findings from simulating the spatial diffusion processes of memes over social media networks by using the approach of agent-based modeling. Different from other studies, this article examines how space and distance affect the diffusion of memes. Simulations were carried out to emulate and to allow assessment of the different levels of efficiency that memes spread spatially and temporally. Analyzed network structures include random networks and preferential attachment networks. Simulated spatial processes for meme diffusion include independent cascade models and linear threshold models. Both simulated and real-world social networks were used in the analysis. Findings indicate that the numbers of information sources and opinion leaders affect the processes of meme diffusion. In addition, geography is still important in the processes of spatial diffusion of memes over social media networks.

ARTICLE HISTORY
Received 5 September 2018
Accepted 4 March 2019

KEYWORDS
Spatial diffusion; social media networks; agent-based modeling; memes

1. Introduction

Location-Based Service has been increasingly and widely adopted by many industries, with an essential component of human-created online information reflected in a spatial context (Jiang and Yao 2006). Many users are willing to share their geographic positions with others through wireless devices by switching on the Global Positioning System trace. Benefiting from this, a great amount of real-time spatial data source can now be accessed by researchers (Ratti *et al*. 2006). Furthermore, this revolution in tracking human activities, behaviors, and communications in the electronic format has been facilitating the collection and analytics of multiple attributes and status updates at the finest scales of footprints of human dynamics (Shaw *et al*. 2016, Sharag-Eldin *et al*. 2018).

This research is inspired by the availability of abundant geo-tagged information in social media and georeferenced communication content which form many location-based social networks (Ye and Liu 2018) and spatial social networks (Tsou and Yang,

CONTACT Jay Lee jlee@kent.edu; Xinyue Ye xye@njit.edu

This article was originally published with errors, which have now been corrected in the online version. Please see Correction (http://dx.doi.org/10.1080/13658816.2019.1626098)

2016). For instance, users can share their opinions on specific topics using a special sign symbol such as hashtag (#) on Twitter, which is a widely used micro-blogging online service established in 2006. On Twitter, specifically, the information diffusion can be easily detected through retweet activities. Retweets allow users to spread new social media messages (memes) from other users by simply restating original information while keeping the source of the information in their posts. Thus, as the new meme becomes popular, a network of retweets can be formed showing how a meme is propagated through the social network.

This article realizes a detailed implementation of agent-based modeling (ABM) to simulate the processes of message diffusion in social networks that disseminate memes in the open source environment (for the conceptual framework, please check Ye *et al.* 2018, Lee and Ye 2018). The model is designed to investigate and analyze different configurations for meme networks with user-driven footprints. Especially, it analyzes various structures of meme networks and simulates the diffusion processes of memes based on different theories of information diffusion. Unlike other simulation models, it supports the investigation of the effect of the event's *Public Attention Levels* and the spatial weights based on common diffusion theories. By simulating how social media information diffuses through social networks of different structures and spatial configurations, outcomes from studying such processes would provide valuable insights to understanding what factors, including spatial factors, may affect how the social media messages spread and what impacts do these factors have during the processes.

2. Literature review

2.1. *Social network and GIS*

Combining spatial data with connections between individual users in a social network, researchers would be able to explore the spatial properties of the social network (Boessen *et al.* 2018). Spatial properties are critical in social networks because they help to develop practical applications that would have great potential for benefiting our societies. For example, as natural disasters occur, optimal evacuation routes can be developed based on the information being passed on via social networks. Using such routes, parents may then pick up their children on the way of leaving the town if the degree of urgency of the disaster is learned from social media early enough and if all family members are able to communicate with each other via social media.

Spatial social network is an emerging research topic focusing on spatial patterns and geo-visualization of social networks with either real-world coordinates or simulated map coordinates. Spatial social networks can be considered as linking people's communications in the forms of phone calls, e-mails, telephones, and other social network services. These networks can also link real-world events, including (but not limited to) such activities as social gatherings for political events or music concerts (Tsou and Yang 2016).

The study of the interactions between locations and social relationships becomes possible only in recent years due to the growth of location-based services. In the meantime, these services generate large volumes of geospatial data that can facilitate the analysis of social relationships (Wang and Ye 2018). A large body of literature generated from and based on research on such characteristic is closely related to geography. For instance, the online activities in social networks, such as those commenting on a particular event, are in fact closely reflecting activities in the physical world. This may include, but not limited to, a Superbowl football game or an unexpected wildfire. More specifically, Wang *et al*. (2018) conducted a large-scale study that reveals social properties and spatial properties of location-based online social networks, providing evidence of how these social connections were shaped over space. Matsuda *et al*. (2014) extract various patterns of the spatiotemporal distribution from Foursquare, a location-based social networking system. In Yin *et al*. (2016), space-time information of messages posted by social media users was used to improve the accuracy of communities in social media networks so to better understand behavioral patterns of social media users.

In analyzing the communication networks of social movements, Conover *et al*. (2013) found that 'Occupy Wall Street' movement, an American anti-capitalist movement, though spread widely spatially, nevertheless exhibited high levels of locality. Their findings echoed the finding of geographic constrains on social networks by Onnela *et al*. (2011). Tsou *et al*. (2013) used kernels to estimate densities and map algebra methods for the visualization of the spatial patterns of the 2012 U.S. Presidential Election and the dynamics in them. In effect, the location-based function in online social networking services is a perfect location-sensor when it is considered in the field of GIS or volunteered GIS (VGIS) in particular. As Sui and Goodchild (2001), Sui and Goodchild (2011) suggested, geographic information systems (GISystems) were increasingly becoming social media. In the meantime, social media were rapidly transforming into part of GISystems.

Despite many researchers using spatial properties of social media for the convenience in life such as tourist trajectory systems (Liu *et al*. 2018), very few of them focused on the spatiotemporal trends of this geotagged information while a meme spread. To the best of our knowledge, only Wang *et al*. (2012), (2013) addressed the influence of different spatial patterns on how information on social media spread. This is important because it helps us to understand the spatial trend of online information while it is spreading. This understanding often reveals the trend of social behaviors and events in the context of geography. However, the diffusive logistic model proposed by Wang *et al*. (2012) is non-graphic and it does not consider network structures, which, as we discussed before, is an essential part of understanding the social networks. By employing agent-based modeling (ABM) approach, we anticipate a statistical and visual discovery of spatiotemporal patterns along with various forms of meme diffusion processes under different social network structures.

2.2. *Social network and meme diffusion*

With the advent of the Internet, an enormous amount of information is being generated online by every second, every minute, every hour, and every day. Although it seems that the information on the Internet is disorderly and unsystematic, it is still an invaluable resource if handled properly. Social media allows the creation and exchange of User Generated Content (Bakshy *et al*. 2012, Luca 2015). With User Generated Content, Watts (2007) and Lazer *et al*. (2009) indicated that a revolution of our understanding of collective human behaviors has arrived.

One of the essential parts of understanding the influence on social behaviors by social media is to understanding social networks themselves. Indeed, a social network matters in many ways. It helps us understand the macro patterns of social behaviors by aggregating individual relations. For example, Christakis and Fowler (2007) quantitatively analyzed a social network that is densely inter-connected to reveal that links were likely between obesity epidemic and people's behavioral and biological traits spread through social ties.

Social networks can also play an essential role in effective management of emergencies (Wang and Ye 2017). On the one hand, information that is generated and is disseminated through social networks is incredibly valuable for important matters in life, such as disaster responses. On the other hand, the study of the relationships, the behaviors, and the interactions that may exist in social networks could offer important hints for rescue efforts during a disaster, such as collecting information and planning of evacuations and sheltering. Social network perspectives provide us a keen method for analyzing different structures of social entities and help to examine different theories embedded in sociology (Wasserman and Faust 1994). Benefiting from the versatility of social networks, we can then utilize them to study information diffusion such as diffusion of innovations where structures of networks may determine how quickly an innovation is diffused and the timing of each individual's adoption of the innovation (Valente 1996). In doing so, the structure of social networks is of main importance in studying the diffusion efficiency (Ye and Lee 2016). For example, in the field of ecology, Bodin and Crona (2009) testified that the structure of a social network did make a difference in affecting natural resource governance. In addition, online social networks have evolved as a new type of social networks on the Web 2.0 environment. In such networks, people may not know each other in person even though they are connected to each other directly. However, they can exchange or share their opinions and interests from their participation in online social networks.

It has been confirmed that online social networks exhibit small-world properties and are scale-free (Mislove *et al*. 2007). Many models have been introduced to emulate these structures. The first and most famous model is Barab´asi-Albert model, also known as *preferential attachment* model. It generates a scale-free network that has two simple mechanisms. One mechanism is continuously growing by adding new nodes into the network. The second mechanism is by connecting the newly added nodes to other existing nodes with a preference to well-connected nodes or nodes of high-degrees (Barab´asi and Albert, 1999). Other models include *copying* model (Kleinberg *et al*. 1999) and *ranking* model (Fortunato *et al*. 2006). These structural properties play essential parts when we consider the information diffusion into online social networks.

As online social networks are gaining popularity, networks scientists have begun to investigate how online content spreads. Meme diffusion has received increasing attention in recent years (Vespignani 2009). Meme refers to an idea, behavior, topic, or style that transfers from a person to another through social media networks. It was first introduced by biologist Richard Dawkins (Dawkins 1976). Traditionally, a meme can spread via writing, speech, or gestures. In modern social media platforms, a meme could also be a URL, an image, a video, or a 'hashtag' on Twitter on the Internet. All of these are transmissible online. A large amount of research was conducted by using twitter memes to help business advertisement and analyze consumer opinions (Jansen *et al*. 2009). Research was also carried out to analyze the outbreaks of infectious diseases in public health (Paul and Dredze 2011). Results from such analyses contributed to situational awareness during natural hazards events (Wang and Ye 2018).

Indeed, with its characteristics of the high frequency and enormously large volumes, memes offer researchers an ideal data source over a variety of domains in social sciences. To analyze such large amount of irregular data sources is not an easy task, let along predicting how the content spreads. In early years, one of the common models for meme diffusion was to mimic memes as infectious diseases (Rapoport 1953, Goffman and Newill 1964, Bailey 1975). These studies were based on the concept that a meme can be as viral as an epidemic and can be transmitted among people. The 'infected' individuals can then spread the 'virus' to others, thus forming a widespread contagion. To that end, Granovetter (1978), Centola (2010), and Romero *et al*. (2011) presented an extended model to demonstrate that meme spread in more complex ways than diseases do.

2.3. *Agent-based model and diffusion simulation*

Agent-based model (ABM) is a model that can simulate how autonomous agents act and interact in computational environment. ABM integrates concepts in game theory, mechanisms in complex systems, concepts in computational sociology, and the operational processes in evolutionary programming. In recent years, ABM has been widely used in biology (An *et al*. 2009), ecology (Karsai *et al*. 2016), and social sciences (Livet *et al*. 2010). It attempts to reproduce and predict complex phenomena by establishing simple behavioral rules for each agent and simulating simultaneous operations and interactions among multiple agents. Once an agent-based model is developed, the user can carry out suites of simulations using the model by adjusting values of the model parameters to explore the relationships among different model parameters and the effect of each parameter on the whole system.

Information diffusion is an activity in which individuals, media, and organizations exchange information and transmit information, ideas, attitudes or feelings to other individuals or groups. Information diffusion includes communicators, the communication channels, and the receivers. Currently, online social networks have been argued to play a major role in how information is spread (Sharag-Eldin *et al*. 2018). In this regard, a social network is a channel through which information is disseminated. Many studies have been carried out in attempts to understand how information diffuses over a social network. Some of which are based on simulations. Neri (2004) described an agent-based tool for analyzing market behavior under several varying rates of information diffusion. He and Hu (2010) analyzed the structural characteristics of the worldwide web and users' social network, providing an agent-based model of information diffusion on the web. The model was used to establish rules for information diffusion under normal situations. Gatti *et al*. (2013) proposed a stochastic approach to analyzing information diffusion via online social networks that use multi-agents. It aimed at predicting human behaviors such as posting a message regarding certain topics and examining how such emerging behavior acts.

In this study, we developed an agent-based model to simulate information diffusion on social networks across space and over time. Comparing to other ABM, we not only examine the different levels of influences by different types of nodes, but we also examine the spatiotemporal processes of how memes diffuse over social media networks. This study aims at exploring the extent to which the structure of a network may impact on the diffusion of social media messages. Such an understanding of the diffusion process facilitates the mechanical process of the spreading of the information and the relationships between network parameters and diffusion parameters for information diffusion over social networks.

3. Design of an agent-based model

3.1. *Models and algorithms*

Information diffusion on a social network starts with one or more source locations where information is generated and starts diffusing. These source locations are known as *seed nodes*. Information spreads outwards from these locations, gradually through the network. There are two most common models in the propagation of information through social networks (Kempe and Kleinberg 2003): the *independent cascade model* (Goldenberg *et al*. 2001) and the *linear threshold model* (Granovetter 1978).

The independent cascade model (IC) is a probabilistic model. When a node u is active, meaning it is a source node or it has received the information being passed through it. It attempts only once to activate its inactive neighbor node v with a probability p_{uv}. These attempts are independent of each other. That is, the activation of u to v is not affected by the actions of u attempting to activate other nodes.

The information propagation process of the independent cascade model is described as follows:

- Assuming that a node u is activated at the moment t, it has an opportunity to activate its neighbors that are still inactive at time $t + 1$. Whether a neighbor v is activated by node u or not depends on the probability p_{uv}, which can be randomly assigned, based on a pre-defined probability model, or specified with a user-defined probability for the node u by the system.
- Higher p_{uv}, means more likely the neighbor node v would be activated by node u. If node v has multiple adjacent nodes w that have been activated, these points would attempt to activate node v in any order. If node v is successfully activated by node u, it will be active at the subsequent moment, $t + 1$.
- At the moment $t + 1$, node v will affect the other nodes and repeats the above process. Node u would no longer be influential at this moment. That is, the node has been activated at the moment t and has tried to activate its own neighbor nodes. It may still be active at moment $t + 1$. However, it cannot activate any other nodes by itself. When there are no influential active nodes in the network, the propagation process ends.

The IC model of the processes of diffusing social media messages is described in Figure 1. When a social media network is represented as a network (G) of nodes and links (V, E), each link e_{ij} is associated with a probability p_{ij} to indicate how likely a message is passed from v_i to v_j. The probability p_{ij} can be randomly generated in simulations. Alternatively, p_{ij} can be calculated as suggested by Kempe and Kleinberg (2003): Let $p_{ij} = 1/\deg(v_j)$, where $\deg(v_j)$ is

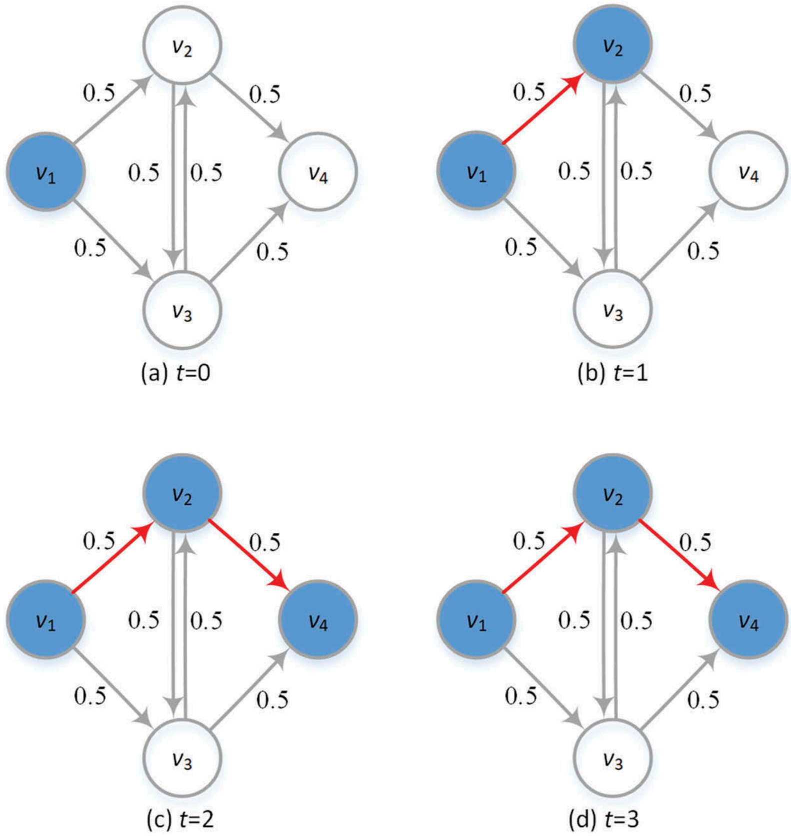

Figure 1. Independent cascade model.

the degree of node v_j. As shown in Figure 1(a), when $t = 0$, v_1 can be a seed node, which is marked as activated so it is colored in blue. Nodes in other colors represent inactive nodes. At $t = 1$, v_1 activates nodes that are connected to according to calculated probability values.

In Figure 1(b), v_2 was activated using a probability level of 0.5. Links in red show the paths that memes passed through. In Figure 1(c), when $t = 2$, v_2 activated v_4. At $t = 3$, v_4 has no more connected nodes to activate so the processes of passing messages through nodes are terminated. With varying probability levels, the orders and the paths of passing messages are not always the same.

Alternatively, the linear threshold model is different from the independent cascade model in determining whether a node is activated or not. It assigns a threshold for each node u, which represents the difficulty that the node is affected by other connected nodes. If a node w, that is connected to v, influences node v with a nonnegative weight and the sum of all neighbors like w of v is less than or equal to 1. For an inactive node v, node v is activated only if the sum of the influence of its active neighbor node is greater than its threshold. That is, the decision for activating an individual node in the network depends on the decision of all its neighbor nodes. And the active neighbor node of node v can participate in the activation of v multiple times. The process of information diffusion in the linear threshold model is as the following:

- The initial set of active nodes S_{active}.
- At the moment t for node v, if the sum of the influence of all neighboring active nodes exceeds the activation threshold, node v becomes active at the moment $t + 1$.
- The process is repeated until the sum of the influence of any active node that already exists in the network cannot activate any remaining inactive neighbor nodes. At which moment, the propagation process ends.

LT model proceeds as shown in Figure 2. In Figure 2(a), at $t = 0$, v_1, as a seed node, is active and colored in blue. Nodes not colored are inactive nodes. Figure 2(b) shows at $t = 1$, the influence levels of v_1 on v_2 and v_3 are both 0.5. Given that the threshold levels for v_2 and v_3 to be activated are 0.6 and 0.5, only v_3 is activated by v_1, with v_2 remained inactivated. Figure 2(c) shows at $t = 2$ when both v_1 and v_3 have the potential for activating v_2, the sum of the two influence levels exceeds the threshold level of v_2. Consequently, v_2 is activated. At this point in time, the influence level of v_3 on v_4 is 0.5, which is less than the threshold level of v_4 (i.e., 0.8). So v_4 remains inactive. At $t = 4$, as shown in Figure 2(d), the combined influence from v_2 and v_3 on v_4 is 1.0, exceeding the threshold level of v_4, v_4 is activated. Since all nodes are activated, the diffusion process is terminated.

The process of information dissemination on a network is greatly affected by the number and locations of seed nodes (information sources). To maximize the speed or

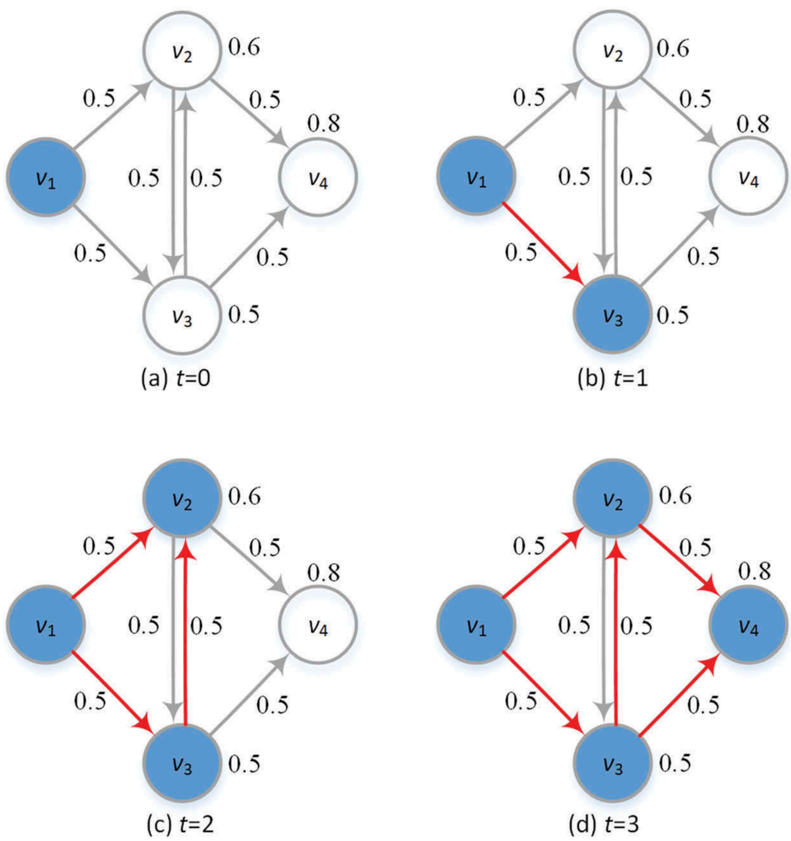

Figure 2. Linear threshold model.

efficiency of dissemination of information in a network, the most suitable person or media organization must be the publisher or originator of the information. How to choose seed nodes is a key issue for maximizing their influence on social networks (Ye *et al*. 2018). A variety of measures of *centrality* can be used to evaluate the importance of nodes in a network. For this, there exists literature on different developed algorithms that can be used for choosing seed nodes in a network according to different levels of centrality that these seed nodes need to have. In this study, we implemented five algorithms for selecting seed nodes. In addition to a random algorithm, the other four are centrality-based algorithms, including *degree centrality, betweenness centrality, closeness centrality*, and *eigenvector centrality*.

In a network, some nodes that are connected to more follower nodes or to 'friends' nodes can have strong influence on how a message is passed through the network. These nodes are referred to as opinion leader nodes in this article. In simulating the diffusion of memes, a simple way to select (from all nodes in a network) opinion leader nodes can be: (1) first sorting all nodes by the number of links that each node has, (2) determine a cut-off level of the number of links that a node should have in order for it to be selected as an opinion leader node, and (3) select nodes as opinion leader nodes according to a predetermined number of such nodes needed. In simulating a network with more than one community, different communities can have different number of opinion leader nodes.

3.2. *Agent-based model structure*

An agent-based simulation model is made up of *agents* and the corresponding rules that govern how the agents behave in the model. Here we defined three types of agents in our model. They are nodes, links, and observers. Each type of agent has its own set of attributes and rules.

3.2.1. *Nodes*

In a social network, a node can be a user or a location such as a twitter user or a city. In order to simulate information diffusion over a network, a node agent is defined as the following:

Node = (*ID, Coordinate, Degree, Links, Type, Status, Color*).

ID is a unique identification index which is used to identify a node: one ID per node.

Coordinate is a position that shows where a node is: defined by a pair of (x,y) coordinates.

Degree of a node is the number of links connected to the node.

Links is a set of objects that include all the links connecting the nodes. Each link has a source node and a target node.

Type of a node identifies the type of the node. Specifically, 0: Normal node; 1: Seed node; 2: Opinion leader node.

Status of a node records whether a node is active or not. Specifically, 0 means the node is inactive and 1 means active.

Color of a node displays the type (among different types of nodes) of the node for visualization purposes.

A node can be initialized with its attribute values randomly assigned to avoid a simulation falling into any specific prescribed process. Alternatively, a node can also be initialized based on its role in an imported network (e.g., real-world Tweet data) of nodes and links. Upon initiation, all nodes are set to be normal nodes and inactive. A node may be a seed node, an opinion-leader node, or a normal node, depending on the attributes it is assigned to have. The numbers of different types of nodes in the simulation are based on user-defined values of model parameters and the user-selected algorithm for generating a network.

In our implementation, the number of each type of nodes, defined by a model parameter, can be the actual number of nodes or as a proportion (percentage) of the total number of nodes in the network. Depending on the selected algorithm and the parametric values, *Seed Nodes* can be randomly selected or manually selected. *Opinion-leader Nodes* are designated based on calculating the number of links they are connected to. Once designated, both seed nodes and opinion-leader nodes are first set to their corresponding *Types* that are visualized with their corresponding *Colors*. To simulate the process of information diffusion on a social network over time, each node has a chance to be activated at any moment (time step) based on the value of its parameter, *Threshold*, its active neighborhood, and the status of nodes that it is connected to.

3.2.2. *Links*

A link connects two nodes. It is associated with a probability that determines the likelihood a meme may be passed on from one of its end nodes to the other end node. In social media networks, a link can be connecting nodes that are 'friends' or that are a message disseminator and its follower. Here we define a link agent as follows:

Link = (*ID, Source-node, Target node, Visible, Direction, Probability*)

ID is the identification of a link.

Source-node is the first node of a link. If the link is directional, then the source node of the link is the node where the link starts.

Target-node is the second node of a link. If the link is directional, then the target node is the node where the link ends.

Visible is an attribute of a link for visualization. 0 means the link is not visible when user visualizes the network.

Direction is a Boolean variable. If it is TURE that means two-way passage is allowed via this link. Otherwise, False means only one-way passage is allowed.

Probability is the probability of a link allowing the passage of a meme through the link.

3.2.3. *Observers*

In addition to node agents, link agents, our model also considers Observers as a type of agent. Each observer is an agent that monitors the simulated process and outputs the results under specified values of model parameters. During the process of simulation, an observer counts the number of iterations of simulation processes as the simulation proceeds. An observer is also instituted for assessing and remembering the percentage of nodes in the network that have adopted a meme. Finally, an observer is added in the model for reporting in tabular or graphic forms the results maintained by the other observers.

The following is the structure of an observer agent.

Observer = (*ID, Enable, Function, Format*)

ID is the index to access a unique observer.

Enable is used to signify whether the observer is enabled.

Function is a flag for marking the function of the observer.

I: an observer that counts and remembers the number of iterations in the ongoing simulation.

P: an observer that assesses and remembers the percentage of nodes who have adopted a meme or, the *active* nodes.

R: an observer that reports the observed *I* and *P* in *REPORT* form.

Format is the report category for an observer. T: text file, G: Graph.

Each observer is initialized by setting *Enable* = *TRUE*. Terminating an observer's activity is done by setting its *Enable* = *FALSE*. Upon setting an observer to have *Function* = *I*, it maintains the number of iterations carried out in the ongoing simulation. If an observer is *Function* = *P*, it assesses the percentage of all nodes adopting a meme at each iteration. When set to *Function* = *R*, the observer reports the values of *I* and *P* in the form of *T* (text file) or *G* (Graph).

3.2.4. *Community*

Community, as a concept, is introduced in our model to emulate the real-world condition that social media outlets are used by users who read and sent messages in the languages that they are the most familiar with or use the most frequently. A *community* may be explained as a sub-structure of a network in which nodes and links in a community form a subnetwork. Consequently, a network may be made up of a number of subnetworks and some links may be included in multiple communities.

Examples for communities may be tweeter users who speak or read/tweet messages in a particular language. For example, Spanish-speaking tweeter users or Chines-speaking tweeter users may form their own subnetworks in an overall social network of tweets. It should be noted also that nodes within a community are linked more tightly than the nodes between communities. Mechanically, there would be opinion leader nodes in each community who play the critical role of disseminating social media messages in their perspective communities. In each community, there would also be some nodes, though not many, who are linked to multiple communities because they read and sent messages in different languages. These nodes function as bridges between communities. While they are not large in numbers, they are key nodes that connect communities to a greater social network.

3.3. *Parameters and behavioral rules for information diffusion simulation*

In our agent-based model, we defined a group of parameters that are listed in Table 1 for simulating information diffusion. Some of them are designed for conducting experiments under different settings of model parameters. The others are used in the algorithms for simulations running in background.

Table 1. Simulation parameters.

Parameters	Description	Type
N_s	The number of seed nodes (count or percentage)	User & System
N_o	The number of opinion leader nodes (count or percentage)	User & System
$Prob_{opinion}$	The probability of propagating a meme of an opinion leader node	User & System
$Prob_{normal}$	The probability of propagating a meme of a normal node	User & System
$Adoption_N$	The number of the active nodes (count or percentage)	System
Set_{nodes}	The set of nodes	System
Set_{seed}	The set of seed nodes	System
$Set_{opinion}$	The set of opinion leader nodes	System
Set_{normal}	The set of normal nodes	System
Set_{active}	The set of active nodes	System
$Set_{inactive}$	The set of inactive nodes	System
$Set_{lastactive}$	The set of nodes that was set to be active at latest time step	System

Formally, the simulation algorithm is as below:

Initialization

(1) Set up N_s, N_o, $Prob_{opinion}$, $Prob_{normal}$, $Stop_N$
Select N_s seed nodes and put them into Set_{seed},
Select N_o opinion leader nodes and put them into $Set_{opinion}$,
Set $Set_{normal} = Set_{nodes} - Set_{seed} - Set_{opinion}$.
Set $Set_{active} = \varnothing$, $Set_{lastactive} = \varnothing$, put all nodes into $Set_{inactive}$
Conduct community detection. Initialize observer agent.

(2) Node Classification
Seed nodes, opinion leader nodes, bridge nodes.

(3) Community Detection
$Community_C$, Number of communities
$Community_{CN}$, number of nodes in each community
$Community_{COE}$, number of opinion leader nodes in each community
$Community_B$, number of bridge nodes between communities, a bridge node is a node that connects two or more communities.

Behavioral Rules for information diffusion:

(1) At the first-time step, for each $v \in Set_{seed}$:
$Set_{active} = Set_{active} \cup \{v\}$
$Set_{inactive} = Set_{inactive} - \{v\}$
$Set_{lastactive} = Set_{lastactive} \cup \{v\}$.

(2) At any regular time step, for each node $u \in Set_{lastactive}$, check all the inactive nodes that node v connected to. Supposed the link is (u,v), generate a random number r between 0 and 1.
if $u \in Set_{opinion}$ and $Prob_{opinion} < r$,
$Set_{active} = Set_{active} \cup \{v\}$
$Set_{inactive} = Set_{inactive} - \{v\}$
$Set_{lastactive} = Set_{lastactive} \cup \{v\}$
if $u \in Set_{normal}$ and $Prob_{normal} < r$,
$Set_{active} = Set_{active} \cup \{v\}$
$Set_{inactive} = Set_{inactive} - \{v\}$
$Set_{lastactive} = Set_{lastactive} \cup \{v\}$

```
if v ∈ Set_inactive and Prob_out < r,
  Set_active = Set_active ∪ { v}
  Set_inactive = Set_inactive − { v}
  Set_lastactive = Set_lastactive ∪ { v}
if v ∈ Set_active, Adoption_N += 1,
  Set_lastactive = Set_lastactive − { U}
```

(3) Update observer agents:
Diffusion Process (Visualization process)
Propagate proportion (graph)
if $Adoption_{N}$= $Adoption_{N-1}$, stop the process of diffusion.
Otherwise execute the step (2) iteratively.

The level of influence of an event is related not only to the structure of a network where information propagates, where and how many seed nodes and opinion leaders there are in the network, but also the level of public attention to the event. For example, a fire in a small town can be seen as a local event because the fire may affect only local communities. Most people who are concerned with it may only be those living in the vicinity close to the fire. While a wildfire in a national forest can draw nationwide attention. Thus, such an event is at a higher level of public attention. We define and use a *public attention value* to emulate the different levels of public attention (or, significance) of any given event.

We assume that events have different geographical influences at different localities. With that, we assume, for a particular event that has a local theme, the interests in that event by people in different locations would be different, depending on how far the locations are from the event. Furthermore, we assume such variation of interests at different locations can be expressed as inversely proportional to the distances that locations have to the location of the event.

To simulate such effects, we added in our simulation model a function that users can specify different spatial weights to the rate of spreading memes at different locations (or regions). Here we defined a distance-decaying spatial trend that is with different weights assigned to regions that are of different distances to the event being modeled. For simulating meme diffusion at the city level, we also defined a decaying ratio based on the distance between two cities to represent different levels of the decaying trend of the interests in the event. By allowing both spatial weights to and ratios of decaying interests of different cities, we were able to model the different patterns of the spatial diffusion of memes related to events of local interests spread over social media networks.

Public attention value and spatial weight are two model parameters that we added to the simulation model to emulate how real-world events are concerned by people at different locations. In the simulation model, nodes are defined with $Prob_{\text{opinion}}$ or $Prob_{\text{normal}}$, depending on the functional roles of the nodes in the network. With the introduction of public attention value, both $Prob_{\text{opinion}}$ and $Prob_{normal}$ would be affected. Nodes assigned with higher levels of public attention values would have higher $Prob_{opinion}$ and $Prob_{normal}$ than those with lower levels of public attention value. Similarly, after the model is structured with spatial weights assigned to different locations, nodes further away from a concerned event would have lower $Prob_{opinion}$ or $Prob_{normal}$ than those closer to the event. Such changes in $Prob_{opinion}$ and $Prob_{normal}$ are determined by the ratio of distance decay as specified by users.

4. A simulator for information diffusion

To explore the spatial process of information diffusion on a social network over time, we need to prepare a simulated network and analyze the characteristics of the network so that we can observe how information diffusion functions in different types of networks. To do this, we designed a toolkit that includes four components: Network Generator, Network Analysis, Community Detection, and Information Diffusion.

4.1. *Network generator*

Social network is the platform where information diffuses. Each type of networks has a unique structure and corresponding characteristics. For example, nodes in a network can be different degrees; a network can be controlled to have different average degrees of all nodes or to have different modularities or different average shortest paths in the network, and so on. This module generates different network models to allow investigations of the differences among different network structures. This module also allows users to explore how different structures of a network may affect the extent and the speed of spreading information. Additional network structures may be added to this module if needs arise in the future because the module is highly customizable using Python language.

4.2. *Network analysis*

In preparation for processing and analyzing real-world social media networks, this module provides a group of analytic methods for exploring the structure of a network. On one hand, it has some metrics to examine a network in its entirety, such as the number of nodes, number of edges, average the degree, modularity and the diameter in or of a network (Scott 2011). Using measures described in Newman (2008), the toolkit we developed can also be used to assess nodes in networks for different centrality measures, including betweenness, closeness, and eigenvector. In addition, nodes can also be evaluated for their degrees, out-degrees and in-degrees to identify their relative importance in a network.

4.3 *Information diffusion*

This module is the core function in the toolkit that simulates how information is disseminated through nodes (e.g., twitter users) in a social network among nodes over time. It implements the agents, rules, information diffusion models, and algorithms for selecting seed nodes or opinion leader nodes. It supports users to do the simulations under different conditions through defining different values for the set of model parameters. In the process of a simulation, the output agents would output the result in the form of graphs or txt files. This helps us to find out how the network structure, seed nodes, and the chosen model for the diffusion of information may affect how the dissemination of information proceeds in a social media network.

5. Simulations and results

To explore the impacts of different network structures on information diffusion and to analyze the influence of spatial distances on a network's ability in propagating messages on social network, we use our toolkit to simulate three types of networks: small world networks, random networks, and preferential attachment networks. In order to facilitate comparisons among these network types, we use the same number of nodes and average degrees when generating the simulated networks. This is so that the simulation analysis is carried out in the same scale and in the same network environment.

Seed nodes are where meme diffusion starts. Selecting suitable seed nodes not only better reflects how memes diffuse in real world but also affects the speed and extent of meme diffusion. In the simulation model, we offer five different algorithms to simulate meme diffusion. With the exception of using randomly generated seed nodes, the other four algorithms are based on the centrality of nodes in the network, including Degree, Betweenness, Closeness, and Eigenvector. These algorithms first order nodes by their centrality values. Next, seed nodes are then selected in the order of high-to-low centrality values. In sections 5.1 to 5.3 where we analyze how different model parameters affect meme diffusion, we compare the results of using five algorithms for selecting seed nodes.

5.1. *The relationship between influence size and the number of seed nodes*

In the agent-based simulator, an automatic program is set up to display the relationship between the levels of influence and the number of seed nodes. The result is shown in Figure 3.

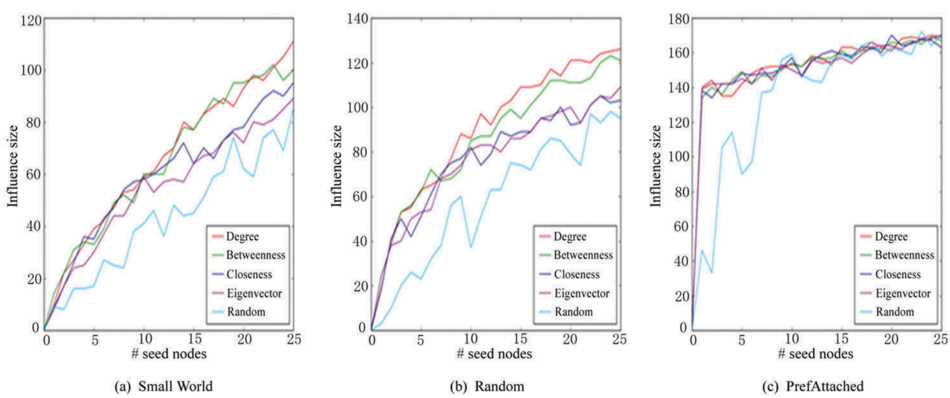

Figure 3. The relationships between levels of influence (number of activated nodes in the network) and different numbers of seed nodes on three types of networks with four centralities. (The total number of nodes is N = 400; Propagation probabilities for opinion leader nodes and normal nodes are p_{op} = 0.3, p_n = 0.2; the proportion of opinion leader nodes is 10%).

Table 2. The characteristics of the three networks.

Network Model	#Nodes	# Edges	Average Degree
Small World	400	798	4
Random	400	800	4
PrefAttached	400	797	4

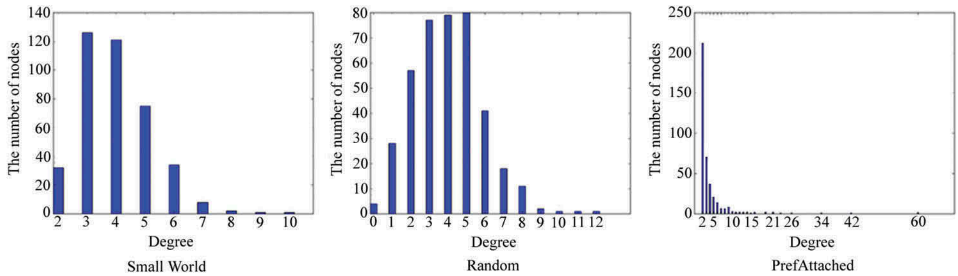

Figure 4. The distribution of degree of node for the three networks.

As can be seen in Figure 3, the level of influence increases gradually with the increase of the number of seed nodes. To facilitate direct comparisons, three types of networks were generated with the same number of nodes and links, and the same average degree as shown in Table 2.

Figure 4 shows the distribution of different degrees of nodes in these networks. It is obvious that the distribution of the degrees of a preferential attachment network was extremely uneven. In fact, the distribution of this type of network is consistent with a power law distribution. That is the reason why on a preferential attachment network, the diffusion scale increases rapidly when the number of seed nodes increases in the early stages and the growth curve becomes very flat in later stages.

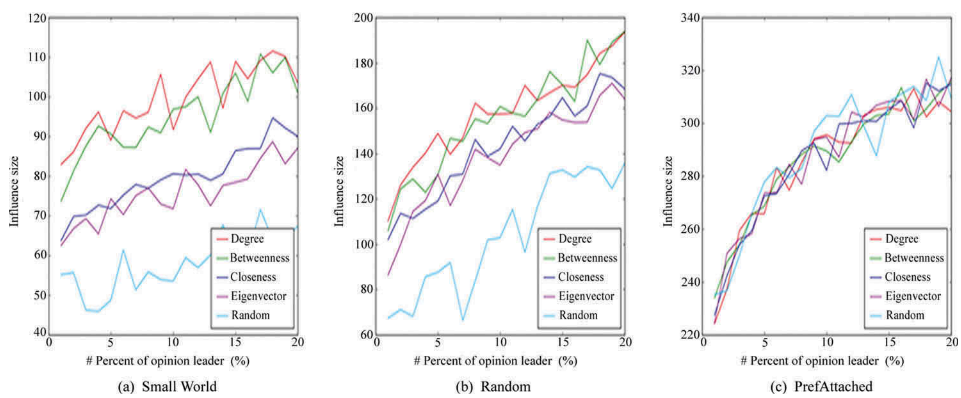

Figure 5. The relationship between levels of influence and the proportion of opinion leader nodes on three types of networks (The total number of nodes is N = 800; Propagation probabilities for opinion leader nodes and normal nodes are p_{op} = 0.3, p_n = 0.2; the number of seed nodes is 15).

5.2. *The relationship between influence size and the proportion of opinion leaders*

Figure 5 demonstrates the relationship between levels of influence (i.e., influence size) and the percentages of opinion leader nodes. Because opinion leader nodes have a higher probability of propagating memes as compared to normal nodes, the level of influence increases gradually with the increase of the percentage of opinion leader nodes in the network. Note that there may exist a situation where the percentage of opinion leaders were too high, causing the level of influence also increases abnormally. In extreme cases, if every individual becomes an opinion leader, then the effects of opinion leader nodes would not exist. In real-world, we found that opinion leaders are usually around 10% of total population.

5.3. *Different probability pairs for opinion leaders and normal users*

Other important parameters that determine the efficiency of information diffusion are the probabilities of opinion leaders and normal users. Figure 6 shows the relationships between levels of influence and the numbers of seed nodes under different propagating probabilities for opinion leader nodes and normal nodes. It can be seen from the figures that, with the same numbers of seed nodes, when the propagation probability increases, the level of influence ascends accordingly. A higher propagation probability for both opinion leader nodes and normal nodes makes a quicker diffusion to the maximum diffusion size, which can be seen in the Random network in Figure 6.

5.4. *Effect of the levels of public attention on information diffusion*

The Public Attention Values as we defined are meant to reflect the intensity of the different levels of concerns by the general public. To a great extent, such varying attention does influence how memes related to the event spread. For example, any news or other media reports of the event may increase public's attention to the event thereby speed up the diffusion of related memes. It should be noted also that such public attention values may change over time and may change with the intensity of reports by news or other media.

In the model, we allow a fixed public attention value to be set for an event at each time step or different public attention values for different time steps. Such function allows users to find the effects of different public attention values on the efficiency of information diffusion. By introducing varying public attention values, we are able to simulate different diffusion processes of different topics that have different levels of local versus nationwide concerns.

To explore the effect of public attention values of an event and also their spatial effects (to be discussed in the next section), we used city networks instead of individual networks. Figure 7 shows the results that the higher an urgency value is, the more quickly the information spreads. In addition, varying urgency values would change the processes of meme propagation in different ways.

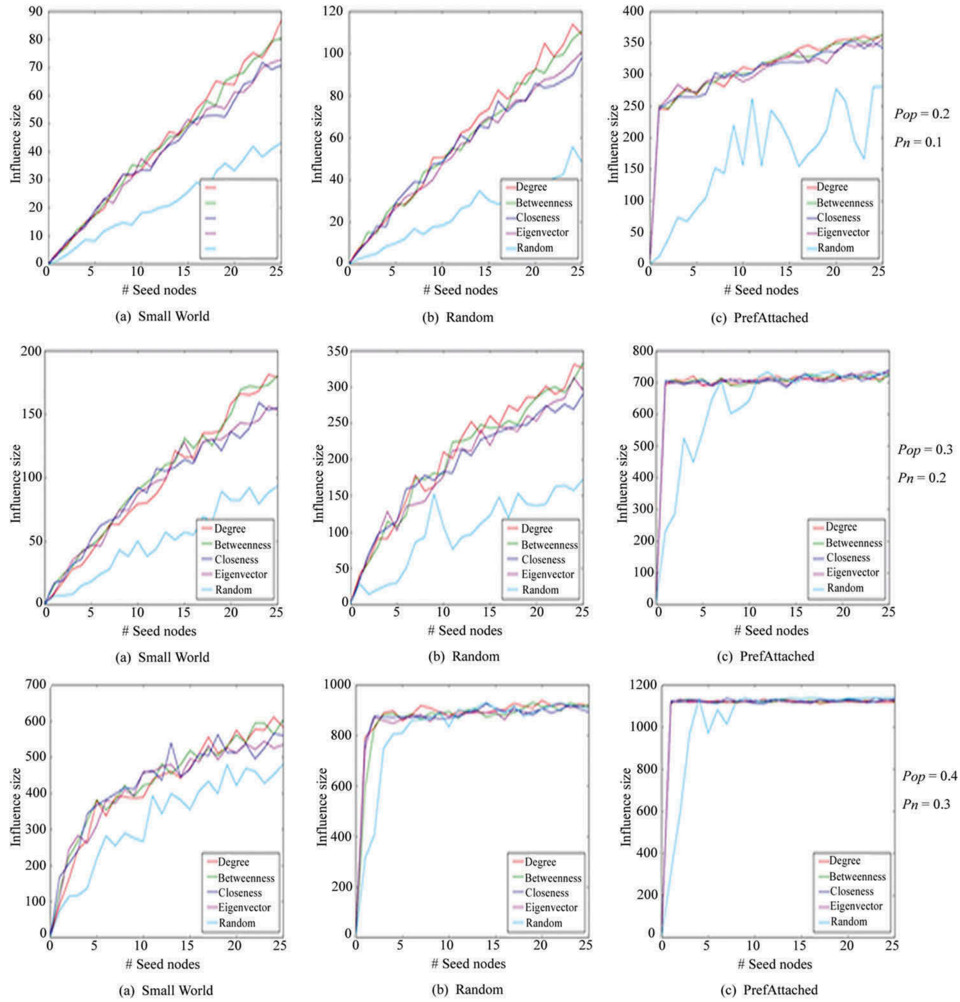

Figure 6. Information diffusion on three types of networks with four centralities under different probability pairs for opinion leader nodes and normal nodes (The total number of nodes is N = 2000; P_{op}: Probability of opinion leaders disseminating messages to connected nodes; P_n: Probability of normal nodes disseminating messages to connected nodes.).

5.5 *Effect of spatial weights on information diffusion*

We structured the simulation model such that spatial weights can be fixed (i.e., the same for all locations) or decaying (i.e., inversely proportional by distance to the concerned event). Note that, when using the decaying spatial weights, two parameters are needed for specifying the decaying distances and decaying ratios. These two parameters determine the intensity of meme diffusion at different locations based on their distances to the concerned event.

Figures 8 and 9 show the different diffusion processes of memes over major cities in the US. They demonstrate the effects of varying spatial weights during information propagation. In time, information started to diffuse from source city or cities that had already adopted the message to other cities. The red nodes denote the cities that have received

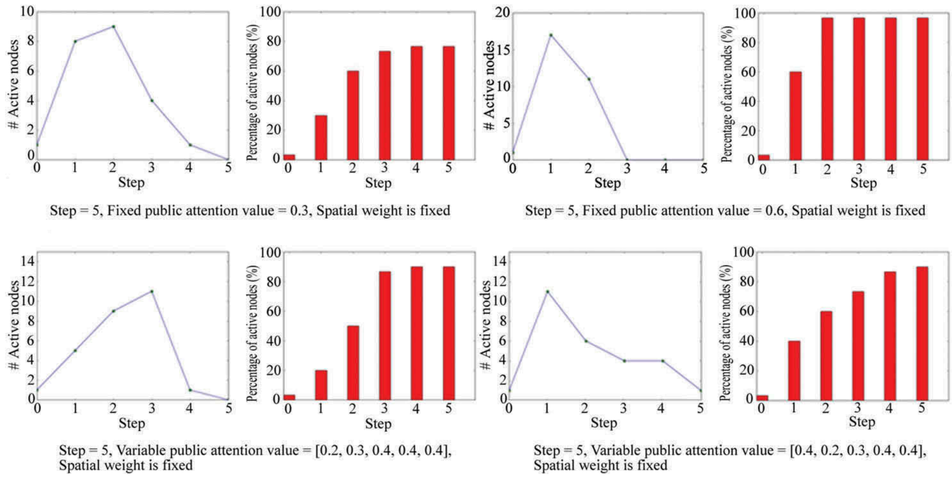

Figure 7. Effects of the public attention values on information diffusion; blue lines in each chart show the relationship between the number of influenced cities (y axis) and time steps of diffusion process (x axis); red charts on the right show the accumulated percentage of influenced cities and time steps (y axis).

Figure 8. Effects of spatial weights on information diffusion (public attention value: 0.3; spatial weight: fixed).

the message. The red lines represent current diffusion process and the blue lines represent diffusion edges in previous time-steps. Without the spatial effect (Figure 8), it is easy to notice that the diffusion links are easy to be formed between cities even if they are distant to each other geographically. However, as the spatial effect comes into play (Figure 9), diffusion links are preferably formed among nearby cities, as it becomes more difficult for messages to propagate to distant cities, which is when the decaying distance increased.

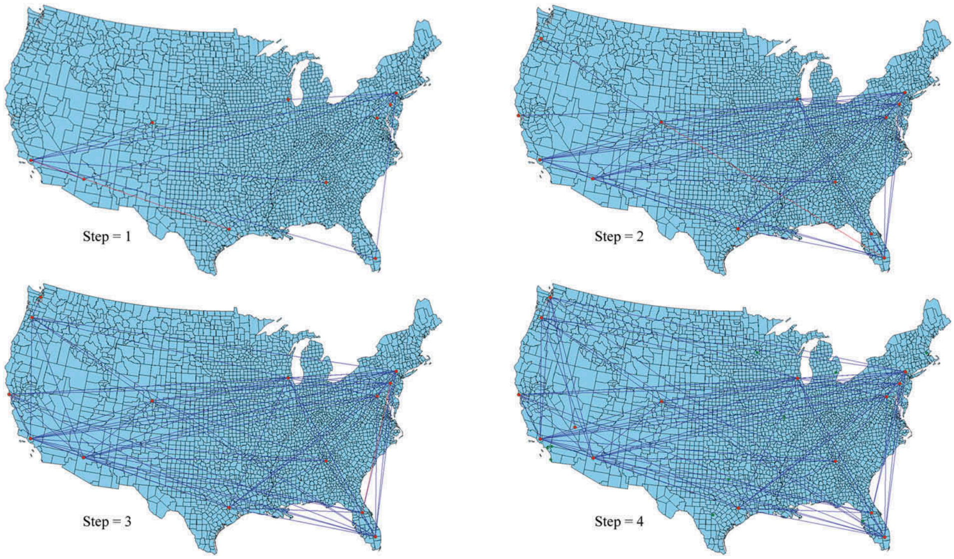

Figure 9. Effects of spatial weights on information diffusion (Fixed public attention value = 0.3, Spatial weight: decaying, decaying radius = 200 miles, Decaying ratio = 5%).

Table 3. Average diffusion distances by decaying rates and time steps.

		Time step			
Decaying radius/mile	Decaying ratio	1	2	3	4
100	5%	1092	771	757	764
100	10%	854	845	462	764
200	10%	1092	760	897	492
200	15%	854	790	637	507

To better emulate the effects that different spatial weights have on information diffusion, we experimented with different parametric values of decaying ratio at each time step. Table 3 shows the average diffusion distances at different time steps by different decaying rates.

As can be seen in Table 3, the information diffusion started with one seed node to represent a single information source. The resulting information diffusion was slower when the decaying ratio was set to be greater and the spatial extents of the information diffusion were smaller, too. Please note that, after information diffusion continued over some time steps and even with more nodes to further diffuse the information, the extents of the diffusion were not greatly expanded.

6. Conclusions

In this article, we developed a toolkit for establishing ABMs for spatial diffusion of memes. By using the toolkit, we investigated the factors that would impact the influence of diffusion over social networks. The structure of a network is confirmed to be the factor affecting the efficiency and influence size of a diffusion process. The same was found for factors such as seed nodes and opinion leaders in the networks. The locations of seed nodes and opinion leaders are critical to the influence size of meme diffusion. It is also important to notice that the propagation probabilities for normal nodes and opinion leaders, respectively, would also affect the diffusion process.

Based on classic diffusion models such as linear threshold (LT) and independent cascade (IC) model, we used ABM to investigate spatial meme diffusion by introducing two more factors, urgency values and spatial weights. These factors allow us to explore spatial meme diffusion by considering the effects of event's own hierarchy and spatial decaying. We believe such ABM is more aligned with real spatial meme diffusion over real-world social networks, where the hierarchy of an event and physical geography also play the roles in online social networks.

In our ABM, we use discrete time-steps to simulate the diffusion process as it is more realistic for us to track the influence order during the diffusion process. In reality, diffusion should be continuous in time, not discrete. However, in contiguous time, the diffusion process can occur in more complex order and modeling such process becomes more difficult. To this end, one of our future works would be to integrate contiguous time into our ABM so we can simulate the diffusion process in a way that approximates reality more closely than using the discrete time steps.

Outcomes from simulated nature normally can be best assessed by comparing them to what happened in reality to validate if the simulated processes are indeed describing how the simulated phenomena progress. However, in our case, the focus was on finding out the relationships between values of model parameters and the efficiency, both in the context of spatial and temporal spread of memes, of the diffusion processes over social networks. In this regard, we acknowledge that the ABM developed and presented here cannot be validated but for the purpose of understanding how different model parameters affect diffusion of memes, we believe the findings help us better understand the roles of network structures and the characteristics/functions of nodes in a network when disseminating memes.

Acknowledgments

This research was supported partially by the National Science Foundation (NSF) through the award NSF #1416509 'IBSS: Spatiotemporal Modeling of Human Dynamics across Social Media and Social Networks'. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the US National Science Foundation [1416509].

Notes on contributors

Dr. Lanxue Dang is an Associate Professor of Computer Science at Henan University. He received a M.S. in Computer Science and a Ph.D. in Geography from Henan University in 2006 and 2014, respectively. In the past years, he published more than 20 articles in the field of Computer

Application or Geographic Information System. His research interests include spatio-temporal analysis, data mining technology and operational research.

Zhuo Chen is a Ph.D. student of Geography at Kent State University. His research interest lies on social network analysis and geospatial analytics. More recently, he is particularly interested in GeoAI techniques and its integration with built environment and environmental health.

Dr. Jay Lee is professor of Geography at Kent State University. His research works stem from a broad interest in bridging operations research and GIS. Over the last several years, he has focused on extending spatial analytics to spatiotemporal analytics.

Dr. Ming-Hsiang Tsou is a Professor in the Department of Geography, San Diego State University (SDSU) and the Founding Director of the Center for Human Dynamics in the Mobile Age (HDMA). He received a Ph.D. (2001) in Geography from the University of Colorado at Boulder. His research interests are in Human Dynamics, Social Media, Big Data, Visualization and Cartography, and Web GIS. He is co-author of Internet GIS, a scholarly book published in 2003 by Wiley and served on the editorial boards of the Annals of GIS, Cartography and GIScience and the Professional Geographers. Tsou was the Principal Investigator (PI) of, "Mapping ideas from Cyberspace to Realspace" research project (2010-2014, $1.3 millions) funded by National Science Foundation. This NSF-CDI project integrates GIS, computational linguistics, web search engines, and social media APIs to track and analyze public-accessible websites and social media (tweets) for visualizing and analyzing the diffusion of information and ideas in cyberspace. In Spring 2014, Tsou established a new research center, Human Dynamics in the Mobile Age (HDMA), a transdisciplinary research area of excellence at San Diego State University to integrate research works from GIScience, Public Health, Social Science, Sociology, and Communication.

Dr. Xinyue Ye is Associate Professor of Spatial Data Science, College of Computing at New Jersey Institute of Technology, where Dr. Ye directs Urban Informatics and Spatial Computing Lab. He integrates social science and computational science towards information visualization, urban informatics and spatial social network analysis – the mapping of relationships among individuals in networks, integrated with spatial and environmental factors. His works has been funded by National Science Foundation, National Institute of Justice, Department of Commerce, and Department of Energy.

References

An, G., *et al*., 2009. Agent-based models in translational systems biology. *Wiley Interdisciplinary Reviews Systems Biology & Medicine*, 1 (2), 159–171. doi:10.1002/wsbm.45

Bailey, N.T., 1975. *The mathematical theory of infectious diseases and its applications*. 5a Crendon Street, High Wycombe, Bucks HP13 6LE: Charles Griffin & Company Ltd.

Bakshy, E., *et al*., 2012. The role of social networks in information diffusion. *In*: *Proceedings of the 21st international conference on World Wide Web (WWW '12)*. New York, NY, USA: ACM, Vol. 8, 519–528.

Barabási, A.L. and Albert, R., 1999. Emergence of scaling in random networks. *Science*, 286 (5439), 509–512.

Bodin, Ö. and Crona, B.I., 2009. The role of social networks in natural resource governance: what relational patterns make a difference? *Global Environmental Change*, 19 (3), 366–374. doi:10.1016/j.gloenvcha.2009.05.002

Boessen, A., *et al*., 2018. The built environment, spatial scale, and social networks: do land uses matter for personal network structure?. *Environment and Planning B: Urban Analytics and City Science*, 45 (3), 400–416.

Centola, D., 2010. The spread of behavior in an online social network experiment. *Science*, 329 (5996), 1194–1197. doi:10.1126/science.1185231

Christakis, N.A. and Fowler, J.H., 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357 (4), 370–379. doi:10.1056/NEJMsa066082

Conover, M.D., *et al.*, 2013. The geospatial characteristics of a social movement communication network. *PloS one*, 8 (3), e55957. doi:10.1371/journal.pone.0055957

Dawkins, R., 1976. *The selfish gene*. New York: Oxford university press.

Fortunato, S., Flammini, A., and Menczer, F., 2006. Scale-free network growth by ranking. *Physical Review Letters*, 96 (21), 218701. doi:10.1103/PhysRevLett.96.218701

Gatti, M.A.D.C., *et al.*, 2013. A simulation-based approach to analyze the information diffusion in micro-blogging online social network. *In*: *Winter Simulations Conference*. Washington, DC, USA: IEEE, 1685–1696.

Goffman, W. and Newill, V.A., 1964. Generalization of epidemic theory. *Nature*, 204 (4955), 225–228. doi:10.1038/204225a0

Goldenberg, J., Libai, B., and Muller, E., 2001. Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 9 (3), 1–18.

Granovetter, M., 1978. Threshold models of collective behavior. *American Journal of Sociology*, 83 (6), 1420–1443. doi:10.1086/226707

He, X.Y. and Hu, X.F., 2010. Modeling and simulation for agent-based information diffusion on worldwide web. *Journal of System Simulation*, 10, 2426–2431.

Jansen, B.J., *et al.*, 2009. Twitter power: tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60 (11), 2169–2188. doi:10.1002/asi.v60:11

Jiang, B. and Yao, X., 2006. Location-based services and GIS in perspective. *Computers, Environment and Urban Systems*, 30 (6), 712–725. doi:10.1016/j.compenvurbsys.2006.02.003

Karsai, I., Montano, E., and Schmickl, T., 2016. Bottom-up ecology: an agent-based model on the interactions between competition and predation. *Letters in Biomathematics*, 3 (1), 161–180.

Kempe, D. and Kleinberg, J., 2003. Maximizing the spread of influence through a social network. *In*: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Washington, D.C., USA: ACM. August, 137–146.

Kleinberg, J.M., *et al.*, 1999. July. The web as a graph: measurements, models, and methods. *In*: International Computing and Combinatorics Conference. Berlin, Heidelberg: Springer, 1–17.

Lazer, D., *et al.*, 2009. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323 (5915), 721. doi:10.1126/science.1167742

Lee, J. and Ye, X., 2018. An open source spatiotemporal model for simulating obesity prevalence. *In*: Thill, J., and Dragicevic S, eds. *GeoComputational analysis and modeling of regional systems*. Cham: Springer, 395–410.

Liu, Q., Wang, Z., and Ye, X., 2018. Comparing mobility patterns between residents and visitors using geo-tagged social media data. *Transactions in GIS*, 22, 1372–1389. doi:10.1111/tgis.12478

Livet, P., *et al.*, 2010. Ontology, a mediator for agent-based modeling in social science. *Journal of Artificial Societies & Social Simulation*, 13 (1), 3. doi:10.18564/jasss.1538

Luca, M., 2015. User-generated content and social media. *In*: Anderson, S. P., Waldfogel, J. and Stromberg, D., eds. *Handbook of media economics*. North-Holland, Vol. 1, 563–592.

Matsuda, Y., Yamaguchi, K., and Nishioka, K., 2014. Discovery of spatio-temporal patterns from foursquare by diffusion-type estimation and ICA. *In:* Wermter, S. eds. *Artificial Neural Networks and Machine Learning - ICANN 2014*. Lecture Notes in Computer Science, Vol 8681. Springer, Cham.

Mislove, A., *et al.*, 2007. Measurement and analysis of online social networks. *In*: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. San Diego, USA: ACM. October, 29–42.

Neri, F., 2004. Agent-based simulation of information diffusion in a virtual market place. *In*: *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*. Beijing, China: IEEE. September, Vol.23, 333–336.

Newman, M. E., 2008. The mathematics of networks. In: Palgrave Macmillan, eds., *The new palgrave dictionary of economics*. London: Palgrave Macmillan.

Onnela, J.P., *et al.*, 2011. Geographic constraints on social network groups. *PLoS one*, 6 (4), e16939. doi:10.1371/journal.pone.0016939

Paul, M., and Dredze, M., 2011. You are what you Tweet: Analyzing twitter for public health. In: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain. pp. 265–272.

Rapoport, A., 1953. Spread of information through a population with socio-structural bias: I. *Assumption of Transitivity. The Bulletin of Mathematical Biophysics*, 15 (4), 523–533. doi:10.1007/BF02476440

Ratti, C., *et al*., 2006. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 33 (5), 727–748. doi:10.1068/b32047

Romero, D.M., Meeder, B., and Kleinberg, J., 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In Proceedings of the 20th international conference on World wide web, Hyderabad, India: ACM, 695–704.

Scott, J., 2011. Social network analysis: Developments, advances, and prospects. *Social Network Analysis & Mining*, 1 (1), 21–26.

Sharag-Eldin, A., Ye, X., and Spitzberg, B., 2018. Multilevel model of meme diffusion of fracking through twitter. *Chinese Sociological Dialogue*, 3, 17–43. doi:10.1177/2397200917752646

Shaw, S., Tsou, M., and Ye, X., 2016. Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30 (9), 1687–1693. doi:10.1080/13658816.2016.1164317

Sui, D. and Goodchild, M., 2011. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, 25 (11), 1737–1748. doi:10.1080/13658816.2011.604636

Sui, D.Z. and Goodchild, M.F., 2001. GIS as media? *International Journal of Geographical Information Science*, 15 (5), 387–390. doi:10.1080/13658810110038924

Tsou, M.H. and Yang, J.A., 2016. Spatial social networks. *In*: D. Richardson, *et al*., eds.. *the international encyclopedia of geography*. Oxford, UK: John Wiley & Sons, Ltd. doi:10.1002/9781118786352.wbieg0904

Tsou, M.-H., *et al*., 2013. Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 US presidential election. *Cartography and Geographic Information Science*, 40 (4), 337–348. doi:10.1080/15230406.2013.799738

Valente, T.W., 1996. Network models of the diffusion of innovations. *Computational & Mathematical Organization Theory*, 2 (2), 163–164. doi:10.1007/BF00240425

Vespignani, A., 2009. Predicting the behavior of techno-social systems. *Science*, 325 (5939), 425–428. doi:10.1126/science.1171990

Wang, F., *et al*., 2013, July. Characterizing information diffusion in online social networks with linear diffusive model. *In*: *2013 IEEE 33rd International Conference On Distributed Computing Systems (ICDCS)*. Philadelphia, PA, USA: IEEE, 307–316. doi:10.1016/j.cbpa.2012.10.019

Wang, F., Wang, H., and Xu, K., 2012. Diffusive logistic model towards predicting information diffusion in online social networks. *In*: *2012 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*, Macau, China: IEEE. June, 133–139. doi:10.1177/1753193412451383

Wang, Z., *et al*., 2018. A spatial econometric modeling of online social interactions using microblogs. *Computers, Environment and Urban Systems*, 70, 53–58. doi:10.1016/j.compenvurbsys.2018.02.001

Wang, Z. and Ye, X., 2017. Social media analytics for natural disaster management. *International Journal of Geographical Information Science*. doi:10.1080/13658816.2017.1367003

Wang, Z. and Ye, X., 2018. Space, time, and situational awareness in natural hazards: a case study of Hurricane Sandy with social media data. *Cartography and Geographic Information Science*, 1–13. doi:10.1080/15230406.2018.1483740

Wasserman, S. and Faust, K., 1994. *Social network analysis: methods and applications*. Vol. 8. Cambridge: Cambridge university press.

Watts, D.J., 2007. A twenty-first century science. *Nature*, 445 (7127), 489. doi:10.1038/445489a

Ye, X., *et al*., 2018. Open source social network simulator focusing on spatial meme diffusion. *In*: S.-L. Shaw and D. Sui, eds. *human dynamics research in smart and connected communities*. Cham: Springer, 203–222.

Ye, X. and Lee, J., 2016. Integrating geographic activity space and social network space to promote healthy lifestyles. *ACM SIGSPATIAL Health GIS*, 8 (1), 24–33.

Ye, X. and Liu, X., 2018. Integrating social network and spatial analyses of the built environment. *Environment and Planning B*. doi:10.1177/2399808318772381

Yin, H., *et al*., 2016. Discovering interpretable geo-social communities for user behavior prediction. *In*: *IEEE, International Conference on Data Engineering*. IEEE, 942–953.