Ekoloji 28(107): 453-462 (2019)

Community Structure-based Re-ranking Information Influence Maximization Algorithm for Typhoon Disasters

Zufeng Zhong [1], Hongyan Yang [2*]

[1] Business School, Lingnan Normal University, Zhanjiang 524048, CHINA

[2] School of Chemistry and Chemical Engineering, Lingnan Normal University, Zhanjiang 524048, CHINA

[*] Corresponding author: hongyan7910@126.com

Abstract

This paper proposes a new community structure-based influence maximization algorithm (CSRR) based on the topological structure of diffusion networks for typhoon disasters. We developed the proposed algorithm by using pre-existing community detection algorithms to detect the community structures hidden in the networks and identify the TOP-*M* nodes which span a relatively large number of communities; We found the TOP-*K* nodes at large distances and used them as the initial diffusers of influence through re-ranking of the *M* nodes. A typhoon disaster event in the Leizhou Peninsula, China, is used as an example to validate the effectiveness and feasibility of the proposed Algorithm. Our test results show that this algorithm provides higher diffusion speed and larger diffusion range than traditional-influence maximization algorithms in the typhoon disaster.

Keywords: typhoon disasters, online social network, influence, community structure, re-ranking, China

Zhong Z, Yang H (2019) Community Structure-based Re-ranking Information Influence Maximization Algorithm for Typhoon Disasters. Ekoloji 28(107): 453-462.

INTRODUCTION

The most frequent and destructive type of natural disaster currently threatening global society is the typhoon (China, & Marine Bureau 2015, 2016, Schmidt et al. 2010). Typhoons characterized by cyclonic vortexes occurring in tropic or subtropic oceans accompanied by strong winds, tidal waves, rainstorms, and storm tides. Typhoon disasters are destructive. Statistics from the Swiss Re-insurance Company show that eight of the ten disasters that caused the largest insurance losses worldwide between 1970 and 2007 were associated with typhoon events (Ou et al. 2002). In addition to the destruction directly caused by a typhoon event, information diffusion problems such as information interruption, false information, information fragmentation, and information distortion are highly problematic.

Social networks as media now play a fundamental role in the interactions between individuals and the diffusion of information and views in the typhoon disaster. Social network services are computer applications that have quickly spread across a vast array of demographic groups since their relatively recent emergence. They have also been under continual, swift, and extensive development. Social network platforms have become the most important information diffusion media by their high user coverage, self-produced content, and timeliness of information diffusion, and accordingly are a very popular research object. They can be utilized to form various virtual interpersonal relationship networks which exhibit diverse social relations and interpersonal interactions. A particular opinion or piece of information can be spread among several users or rapidly disappear depending on this diffusion. Understanding the dynamic, complex manner in which information diffuses will also help elucidate how information is accepted by the public.

The processes and characteristics of information diffusion in social networks can be modelled in various ways. Existing information diffusion models are based on two principles: 1) Each information diffusion process involves several starting points, such as the initiator of the information; 2) Each piece of information received by a user comes from other users he or she follows (e.g., friends on Sina Weibo). Popular information diffusion models include the independent cascade model, the linear threshold model, and the statistical model. The independent cascade model

 Received: 29 Feb 2018 / Accepted: 3 Nov 2018

(Goldenberg et al. 2001) is based on the theory of probability and describes the information diffusion process as a dynamic "domino effect". The linear threshold model (Bozorgi et al. 2016, Granovetter 1978) is based on the theory of that each node has an information diffusion threshold; when a node receives information from neighbours with influence greater than said threshold, the information is propagated. The statistical model (Aggarwal et al. 2011) is based on the hypothesis of independence probability, where an initial set A_0 isassumed for information diffusion – any node in A_0 will propagate a piece of information at a probability of 100%, while a node outside A_0 will diffuse the information at a probability determined by the probabilities of its neighbours.

In the typhoon disaster, the question of how a limited amount of capital can be utilized to find an initial audience, to let these initial audience influence other individuals on a network, and to maximize the number of affected individuals is referred to as "influence maximization". This question, as first proposed by Domingos and Richardson (2002), is: Given a social network *G*, a specific network influence propagation model, and a number of initial nodes, what is the best method to determine the set of initial nodes, to utilize these initial nodes for influence propagation, and ultimately to maximize the number of affected nodes?

The key to establishing an effective influence model is to identify the root node in influence maximization. This part of the algorithm can be roughly divided into two categories: Greedy algorithms and heuristic algorithms. The basic idea of greedy algorithms is to select the most influential nodes (*A* nodes) as a set of initial seed nodes and to determine all the affected nodes according to the propagation mechanism of the corresponding model. The basic greedy algorithm (Kempe et al. 2003) is superior in regards to influence maximization research owing to its simple structure, ease of understanding, and ability to yield stable solutions; the result of this algorithm is at least 63% of the optimal solution. The heuristic algorithm with degree conversion (Chen et al. 2009) in the independent cascade model achieves very similar influence diffusion efficiency, but its time complexity is much lower. Jiang et al. proposed a simulated annealing evolution method, used two heuristic algorithms to accelerate its convergence, and designed an evolutionary algorithm of influence maximization; their simulated annealing algorithm runs faster than the greedy algorithm by 2-3 orders of magnitude and with improved precision (Jiang et al. 2011). Kimura et al. proposed a shortest path heuristic strategy and designed a series of high-efficiency influence maximization algorithms (Kimura and Saito 2006). The feasibility of these algorithms is largely restricted due to the ever-increasing scale of the social network, however.

There is a unique relationship from user to user within the social network because the information being shared is interactive. Users who share interests, attributes, and close relationships gather into special groups called "community structures". Algorithms which detect community structures can be divided into three types: Statistical, segmentation, and modularity-based. Some famous algorithms include the module maximization algorithm (Clauset et al. 2004, Wakita and Tsurumi 2007), the Given-Newman algorithm (Newman and Girvan 2004), the Louvain algorithm (Blondel et al. 2008), the clique percolation method (Palla et al. 2005), and the link community (Ahn et al. 2010). These algorithms identify communities by the manner in which they are structured. Alongside the increasing development of social networks in recent years, researchers have begun to add network node attribute information into pre-existing community detection algorithms. Steinhaeuser et al. proposed an edge-weighted node attribute similarity (NAS) algorithm and integrated it with the conventional random walk algorithm (Steinhaeuser and Chawla 2010). Dang et al. estimated the weighted summarization of a modularity function and a node attribute similarity function (Dang and Viennet 2012) and succeeded in module maximization via Louvain's algorithm (Blondel et al. 2008), thereby identifying community structures effectively. In a study on community detection on Twitter, Naresh et al. attempted to combine conventional clustering with the similarity of multiple attributes (Kewalramani 2011). Deitrick et al. investigated tweets within a given period in the effort to improve detection performance (Deitrick and Hu 2013).

In the typhoon disaster, existing information diffusion models and influence maximization algorithms are not necessarily applicable to social networks. This is because information diffusion is more rapid in social networks and there are various levels of connection among users. Existing influence maximization models rarely consider community structures in social networks.

During information diffusion, the nodes in a community are strongly connected; to be more specific, the diffusion within the community is rapid and accessible, but the diffusion is slow between weakly-connected communities. A user who has been recently very active is certainly more influential than an inactive user. This creates a few important problems to be solved: 1) Individuals are ranked by influence, and when a community has an individual with large influence, the information will be rapidly diffused within this community but less outside the community. 2) The most influential individuals are not necessarily those connecting the largest number of communities. 3) The activity in a recent period should be taken into account when selecting initial nodes; users more recently active are more influential under the same conditions.

In this study, we built a new diffusion model based on the community as the basic unit of measurement in the typhoon disaster. We added the sequence activeness index to the model to identify the initial diffusion points and maximize the influence of information. Our CSRR algorithm is based on our analysis of pre-existing community detection algorithms and, again, has the community as the basic unit of information diffusion. The algorithm can be utilized to accelerate information diffusion and maximize the diffusion scope in the typhoon disaster.

RELATED WORKS

Research on Information Diffusion in Disaster

When a natural disaster events occurs, the public utilizes social networks for the generation, interaction with, and diffusion of information, effectually providing dynamic suggestions for emergency management. This has been proven by social media's effects on disasters such as the Wenchuan earthquake and Japanese tsunami (Acar and Muraki 2011, Plotnick et al. 2011). When an emergency event occurs, more than 10% of related information is attributed to individuals present at the scene; the information is reposted by individuals who are closely (e.g., regionally) correlated with the original posters, and thus the scene (Starbird and Palen 2010). The key factors influencing the quantity of information transmitted are uniform resource locators (URLs) and labels (O'Connor et al. 2010). The emotional reaction generated by content is significantly correlated with the amount of times the content is reposted (Hansen et al. 2011).

Network Propagation Model

Independent cascade (IC) model

The IC model (Bozorgi et al. 2016) was abstracted from previous research findings on social networks from a sociological perspective. This model has been well recognized in the sociology field. The diffusion mechanism of this model works through a five-step process (Li 2010, Mashayekhi et al. 2018).

1) In a network graph $G(V, E)$, each node is at either in the active or inactive state. The transition direction is only from the inactive to the active state, never the reverse.

2) For any active node v, all neighbor nodes are represented by a set $N(v)$. For any node w in $N(v)$, node v can activate node w at the probability $p(v, w)$. When $p(v\ w)$ is larger, node w can be more easily activated by node v.

3) For any inactive node v, when several neighbor nodes of node v are activated, then node v can be activated by any of these neighbor nodes.

4) Any node v may be successfully or unsuccessfully activated (from the inactive to active state) by its active neighbor nodes. If node v is successfully activated, it can influence the nodes to which it points.

5) Repeat steps 2, 3, and 4, until there is no node in $G(V, E)$ that can be activated.

Regarding the diffusion mechanism, we mentioned that the probability of node v activating the nodes it points to is decided by $p(v, w)$; when $p(v, w)$ is larger, node v requires more changes to activate node w. As of now, $p(v, w)$ can be estimated through the data mining technique (Lu et al. 2012), or selected randomly from [0, 1].

Linear threshold model

The linear threshold model is another highly influential maximization model. Its diffusion mechanism works through a four-step process (Granovetter 1978).

1) In a network graph, each node is in the active or inactive state. The transition direction is only from the inactive to the active state, never the reverse, which is consistent with the IC model.

2) For any node v in the network, the possibility of activating node v is expressed by a threshold $\theta_v(threshold)$. A larger threshold θ_v indicates the activation of node v is more difficult, and vice versa. The nodes that can influence node v are included in a set $AN_v(active\ neighbors)$. For any node w in this set, the probability of node w activating node v is expressed by

b_{vw}. If the threshold θ_v of any node is within [0, 1], then in set AN_v the sum of influences on node v does not exceed 1.

3) If the sum of all influences on node v in AN_v is $>=$ its threshold θ_v, then node v is activated; otherwise, node v is not activated. When node v is activated, all nodes it points to become influential.

4) Repeat steps 2 and 3 until there is no node in the graph that can be activated.

Two parameters in the linear threshold model should be determined: The activation threshold θ_v of node v, and the mutual influence between nodes. The setting of these two parameters also depends on the network structure, which can be determined through the same strategy as the IC model. If the mutual influence between nodes can be expressed by the weight of lines, then the linear threshold model can be expressed as a weighted graph.

Influence Maximization Algorithm

Hill climbing greedy algorithm

The selection route of k seed nodes is divided into k steps (Bozorgi et al. 2016) in the hill climbing greedy algorithm. At each step, the current most influential node (i.e., the node with the largest marginal effect $m(u|S_i)$ is selected. From S_0 to step i, node u is selected on basis of the local optimization strategy and u satisfies $u = argmax_v m(v|S_{i-1})$. Then let $S_i = S_{i-1} \cup \{u\}$. The flow of the hill climbing greedy algorithm is shown below.

Input: The initial diffusion set of the network graph G(V, E) containing k nodes.
Output: The initial diffusion set S_k.
Initialize the set $S_0 = \emptyset$.
For i=1 to k do
Compute the marginal effect $m(u|S_i)$ of each inactivated node u.
Select the node u with the largest marginal effect.
Let $S_i = S_{i-1} \cup \{u\}$.
End For

The implementation of the hill climbing greedy algorithm has many strengths. Its properties are very simple and understandable, and it can yield stable and approximate solutions; we found that it guarantees at least 63% optimal solution. It also has some notable limitations, however, such as high time complexity. Many nodes change from the inactive state to the active state after one seed node in the network is determined. Thus, at each step, the marginal effects $m(u|S_i)$ of all inactive nodes u must be re-computed. This makes the hill climbing greedy algorithm very time-consuming and thus unsuitable for large-scale online social networks.

We used the hill climbing greedy algorithm proposed by Kempe and Kleinberg to solve an influence maximization problem. To find the initial diffusion node set S_k at each step, one standard was used to determine one node in this set until k nodes were identified. We defined $S_0 = \emptyset$, where $I(S_i)$ is the set of activated nodes after diffusion of the set S_i; $m(u|S_i) = |I(S_i \cup \{u\})| - |I(S_i)|$ is the marginal sphere of influence of node u relative to set S_i.

Heuristic algorithm

Influence maximization can be solved via some heuristic algorithms such as Random, MaxDegree, and Degree Centrality. These heuristic algorithms are superior regarding time complexity, but since the static selection of nodes does not consider the diffusion process, they cannot provide any stable solution or ensure optimal influence effect. The Random algorithm is the oldest and simplest heuristic algorithms. Specifically, k nodes are randomly selected from the network as the initial set, then used to activate other nodes; this makes the optimal solution of influence maximization rather obscure.

Many simple heuristic algorithms have been to identify k-node sets in social networks for the influence maximization after the influence maximization was proposed. For instance, the HighDegree algorithm (Sun and Tang 2011), which aims to find the k nodes with the largest degrees as the initial node set. The main problem inherent in these algorithms is that their solutions are not sufficiently stable and cannot adapt to the dynamic changes in the network. Owing to a large number of spam users and inactive users in Weibo, in particular, there is no algorithm that yields a favourable influence effect.

The High Degree algorithm can be used to maximize the information influence in large-scale networks by its low time complexity, so we compared it against the new algorithm proposed in this study. The pseudo-codes of the High Degree algorithm are as follows.

Input: Graph G and initial set k.
Output: Initial set U with a size of k.
Initialize the set $U = \emptyset$
for i=1 to n
Degree(i)
end for
U←High Degree (k)
Return U

Label Propagation Algorithm (LPA)

Many effective algorithms have been proposed to detect the community structures of social networks. The main objectives are to detect as many accurate community structures as possible by reducing prior information and focusing on accurately characterizing the structures of real-world social networks.

Raghavan et al. proposed the rapid community detection label propagation algorithm (LPA) (Raghavan et al. 2007), which is significant by ts ability to perform dynamic and real-time community detection in large-scale online social networks. The LPA concept originated from that of information propagation and has a similar theoretical basis to the "random walk intermediation" of GN algorithms. LPA first assigns each node a specific label which is propagated along the between-node edges and thereby plays the role of the "random walker". In each run, the labels of neighbour nodes, are summarized and the most often-appearing node is selected as the new label of the current node. After several iterations, the labels on the network gradually stabilize. Usually, the terminal condition holds that no label change occurs after several computations, indicating the algorithmic convergence. Finally, the network can be divided into several communities according to the names of the labels on different nodes.

LPA is superior to other algorithms in many aspects. It is simple, intuitive, understandable, fast, and with high-resolution accuracy. For the label propagation in each run, the time complexity of label renewal at each node is O(*d*), where *d* is the degree of the node. In each image, the sum of the degree of all nodes is twice the number of edges as follows:

$$\sum_{v_i \in V} d_{v_i} = 2m \tag{1}$$

Where m is the number of edges, v is the set of all nodes and d_{v_i} is the degree of the node v_i. Thus, the time complexity of label renewal of the whole network is O(m). Nearly all experiments proved that LPA converges only after a few runs of computation. It can be assumed that the number of computations is a constant, so the total time complexity of LPA is O(m). In other words, the number of edges is a linear complexity algorithm which quickly divides communities for most sparse images. LPA does not require any parameter input due to its simplicity and likewise does not necessitate parameter adjustment. The algorithm has been modified in many ways (Gregory 2010, Peng et al. 2016, Sun et al. 2015, Xie and Szymanski 2012, 2013, Xie et al. 2011, 2013).

RE-RANKING INFLUENCE MAXIMIZATION ALGORITHM BASED ON COMMUNITY STRUCTURE

As discussed above, the pre-existing influence maximization models and algorithms focus on individual influence. However, in social networks, people with common interest tend to gather together and form virtual communities as facilitated by rapid information diffusion. In other words, birds of a feather flock together. The existence of virtual communities affects information diffusion in community networks, However, the connection within a virtual community is strong while the connection between virtual communities is weak. The information diffusion between individuals from the same community is different from that between individuals from different communities. The initial diffusion points can be selected from the more influential points with the traditional method; these points are located in one community or several neighbouring communities, which inhibits the speed and breadth of information diffusion in the typhoon disaster.

We established the proposed influence maximization algorithm as a response to the problems described above in the typhoon disaster. Information is guided to the individuals spanning several communities and initial nodes that areas recently active as possible. First, we identify the communities in a social network, then compute the number of communities for each point and identify TOP-M nodes. Then the activeness in a recent period of all nodes in M is computed and top-k is selected through ranking the initial diffusion points.

The community structure-based algorithm proceeds as follows.

Step 1: Identify communities.

LPA is utilized for community segmentation.

Input: adjacent matrix $A_{n \times n}$.

Output: individual sets of k communities $group = \{group_1 - group_k\}$.

1) Initialize nodes with unique labels.

2) Set each node's label to the label shared by most of its neighbours.

3) If not converged, continue to step 2.

Step 2: Find out M nodes as a candidate set M that connects the largest numbers of communities.

Input: adjacent matrix $A_{n\times n}$; individual sets of k communities $group = \{group_1 - group_k\}$; L (preset node size);

individual set V', m

Output: L relatively better nodes.

1) Scan the adjacent matrix $A_{n\times n}$, $N(v_1) - N(v_n)$.

2) For all individuals in V, compute $Com\ N$ ($Com\ G$ is the number of communities connected to the node) for each node.

3) Select the node v_i with the largest $Com\ G$; if two or more nodes with largest $Com\ G$ are available, select the node with the largest value.

4) $m = m + 1$.

5) Delete all individual sets connected with v_i and update group.

6) Delete nodes v_i and update V'.

7) $t \leftarrow |V'| - 1$.

8) If $m = L$, then end.

9) If $m < L$: if $t \neq 0$, then the number of remaining groups is not 0 and return to Step 2; otherwise, from the remaining nodes in V, select $L - m$ nodes with the largest $Com\ G$. End.

Step 3: From the candidate set L, find most recently active K.

In Weibo, the main user behaviours of option are the release of new micro blogs, reposting of micro blogs, commenting, sending private emails, sign-ins, following other users or cancelling a following, browsing micro blogs of all followed users, and favouring microblogs. Each of these behaviours is considered an "activity" in Weibo. The sum of daily activities is a_t. Since the time closest to the present moment is the most important, we used an attenuation window to compute activeness. The latest element is assigned a weight that gradually attenuates at a slightly smaller interval ratio as time progresses. The equation is expressed as $\sum_{i=0}^{t-1} a_{t-1}(1-b)^{1-i}$, where a_t is the number of activeness on day 7. Let b be the attenuation constant and t the number of days to be tested; t=5is the activeness index in over 5 days. The activeness computations are ranked from largest to smallest and the top K nodes are extracted.

EMPIRICAL ANALYSIS

We used an event in the Leizhou Peninsula, located at the southmost reaches of Mainland China, as a case study. Leizhou has a coastline of 1,180 km and its major city, Zhanjiang, has a population of 9 million. It is located in a typhoon zone at the west of the tropical oceans of the southwest North Pacific. It is directly attacked by 2-3 typhoons per year (in summer months, i.e., typhoon season) from the westward path and northwestern path.

The Leizhou Peninsula was attacked by typhoon Caihong from October 4-7, 2015. We extracted the open data from the data access interfaces provided by Sina Weibo and used it as for our experiment. With one of the largest user bases among all microblogs in China, Sina Weibo harbours more than 0.5 billion users; by the end of September 2015, the daily number of active users of Sina Weibo was 212 million. Since the release of the mobile Weibo client, users have been able to browse and post microblogs anytime, anywhere. Accordingly, Weibo has become an indispensable part of daily life for many – this also makes the use of Sina Weibo data for the analysis of influence diffusion very meaningful and valuable. Sina Weibo Open Platform provides an API through which we were able to acquire a substantial amount of microblog information, relationships among fans, user information, and information fissile diffusion channels that occur across a wide array of times and locations.

The period of data extraction spanned from October 4, 2015, to October 7, 2015. The extracted data were pretreated, and spam users and corpse users were excluded. We extracted relevant data from the corresponding IDs through the Sina-provided API. The specific data included the following: 1) User information corresponding to the user ID, including user name, number of followers, number of others followed, and number of microblogs; 2) Following relations among users including the IDs of followed users within a given micro-group; and 3) Information forwarded and commented on by users, including IDs of forwarded information and IDs of users doing the forwarding/commenting. The dataset contains 13,553 user nodes and 388,624 lines formed by the following relations.

We compared our proposed algorithm against three common influence maximization and Weibo influence

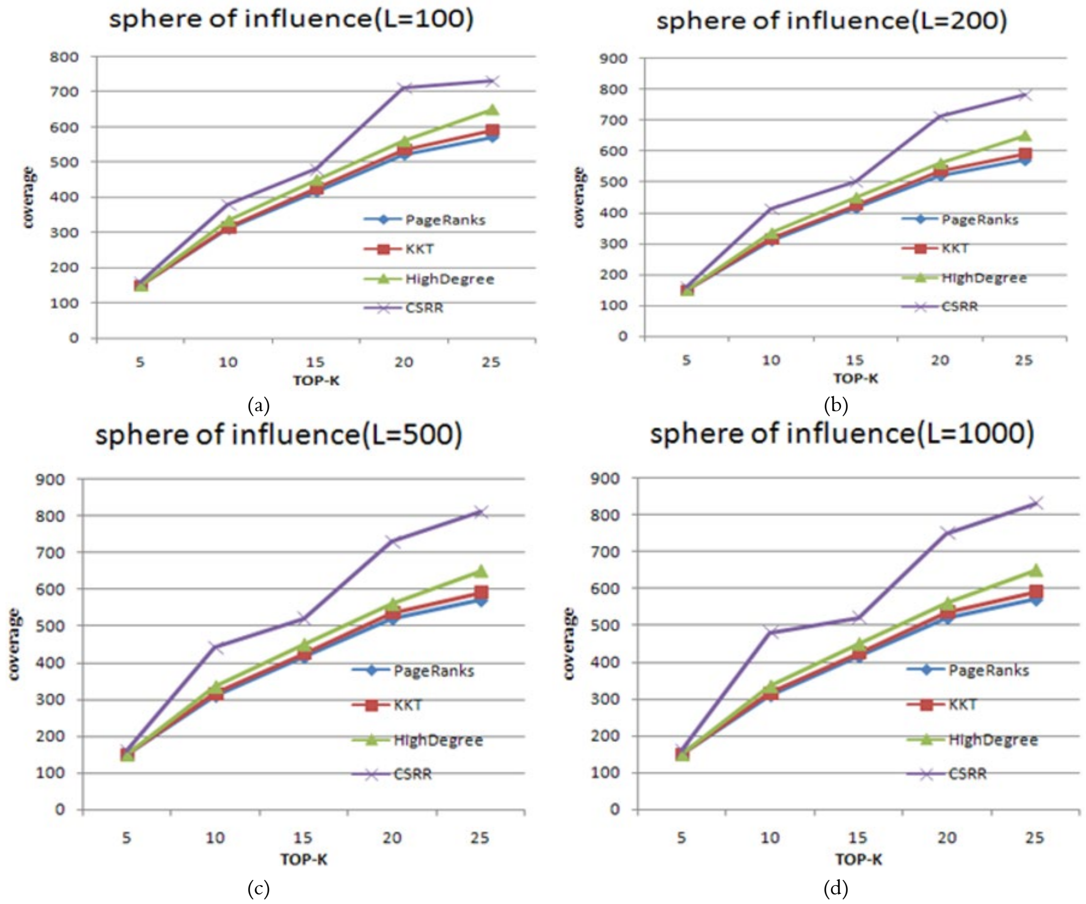

Fig. 1. Coverage of different algorithms and different L values

measurement algorithms to validate the effectiveness of CSRR in the influence maximization of Weibo networks.

1) KKT algorithm: A natural hill is climbing greedy algorithm proposed by Kemple and Kleinberg. At each step, the currently most influential node is selected and put into the seed set. The KKT algorithm commonly used in social network maximization is formed when the algorithm is applied to the linear threshold model.

2) PageRank algorithm: This is a common influence measurement algorithm that assigns influence depending on the degree of a node.

3) HighDegree algorithm: The top k nodes with the highest degrees are selected as the initial node set.

In previous research on influence maximization algorithms, the extracted Top-K nodes were only applied to some influence diffusion models to simulate their covering effects. In this study, the Top-K nodes were extracted via HGAE algorithm, and the comparative algorithms were applied into real networks to validate its diffusion and coverage effects.

The real number of influence-covered users, M, is defined as an indicator assessing the node diffusion ability as follows:

$$M_i = \frac{\sum_{j=1}^{N} RT_j}{N} \tag{2}$$

where RT_j is the number of diffusion-covered users by the j-th microblog from the i-th user and N is the statistical number of microblogs posted by the i-th user within a certain period.

The test results from the several algorithms show that the CSRR algorithm outperformed the others as far as information diffusion speed (**Fig. 1**) in the typhoon disaster. The sphere of influence from the CSRR-selected initial nodes was larger than that of the KKT algorithm; the number of nodes increased by 9%. In effect, the introduction of the community structure-based model allowed the CSRR algorithm to efficiently improve the information diffusion in Weibo networks and find nodes with high diffusion influence in Weibo communities. We also found that as L increased, the speed of information diffusion improved. This is because, with a larger L, the K nodes after the second

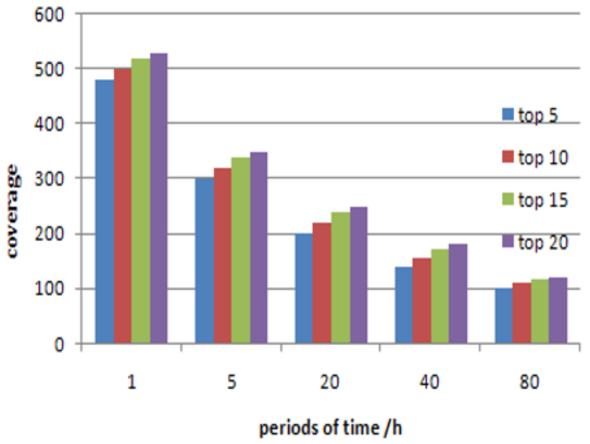

Fig. 2. Coverage of different time and different TOP-k values

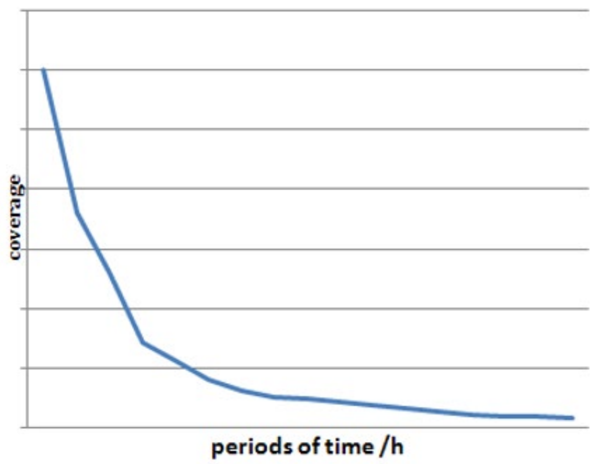

Fig. 3. Time variation of coverage

selection are more active – in other words, the selected initial nodes are more active and distributed more widely, so the diffusion breadth and speed are larger.

As shown in **Fig. 2**, varying *t* also affected the experimental results. A smaller *t*, i.e., more recently active nodes, led to faster propagation while a larger *t* slowed down the propagation. As *t* prolonged, the change trend decreased in intensity (**Fig. 3**).

CONCLUSIONS

In this study, we examined influence maximization during information diffusion for typhoon disasters. The diffusion range and speed are reduced if the initial nodes are located in the same or a nearby community. To avoid this problem, before selecting the initial nodes in information diffusion, it is necessary to consider not only the influence of nodes but also the communities containing the nodes, as well as the number of communities a node spans. We developed a re-ranking algorithm based on community and distance accordingly. The number of communities and the ratio of long-to-short between-node distance can be adjusted by setting the value *L* in TOP-*L*. Through simulation tests, we found that the proposed community-based information diffusion re-ranking algorithm outperforms similar, traditional algorithms in the breadth and speed of information diffusion for typhoon disasters.

ACKNOWLEDGEMENT

This research is supported by the Ministry of education of Humanities and Social Science project (18YJCZH267) and Zhanjiang Social Science Association (ZJ15QN03).

REFERENCES

Acar A, Muraki Y (2011) Twitter for crisis communication: lessons learned from Japan's tsunami disaster. International Journal of Web Based Communities, 7(3): 392-402.

Aggarwal CC, Khan A, Yan X (2011) On Flow Authority Discovery in Social Networks. SDM: 522-533.

Ahn Y Y, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature, 466(7307): 761-764.

Blondel VD, Guillaume JL, Lambiotte R, et al. (2008) Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, (10): P10008.

Bozorgi A, Haghighi H, Zahedi MS, et al. (2016) INCIM: A community-based algorithm for influence maximization problem under the linear threshold model. Information Processing & Management, 52(6): 1188-1199.

Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM: 199-208.

China, & Marine Bureau (2015) China marine disaster Bulletin.

China, & Marine Bureau (2016) China marine disaster Bulletin.

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Physical review E, 70(6): 066111.
Dang TA, Viennet E (2012) Community detection based on structural and attribute similarities. International Conference on Digital Society (ICDS): 7-12.
Deitrick W, Hu W (2013) Mutually enhancing community detection and sentiment analysis on twitter networks.
Goldenberg J, Libai B, Muller E (2001) Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing letters, 12(3): 211-223.
Granovetter M (1978) Threshold models of collective behavior. American journal of sociology: 1420-1443.
Gregory S (2010) Finding overlapping communities in networks by label propagation. New Journal of Physics, 12(10): 103018.
Hansen LK, Arvidsson A, Nielsen FÅ, et al. (2011) Good friends, bad news-affect and virality in twitter. Future information technology. Springer Berlin Heidelberg: 34-43.
Jiang Q, Song G, Cong G, et al. (2011) Simulated Annealing Based Influence Maximization in Social Networks. AAAI, 11: 127-132.
Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM: 137-146.
Kewalramani MN (2011) Community detection in Twitter. University of Maryland, Baltimore County.
Kimura M, Saito K (2006) Tractable models for information diffusion in social networks. European Conference on Principles of Data Mining and Knowledge Discovery. Springer Berlin Heidelberg: 259-271.
Li L (2010) Research on influence models and algorithms of social networks. Beijing Jiaotong University.
Lu Z, Zhang W, Wu W, et al. (2012) The complexity of influence maximization problem in the deterministic linear threshold model. Journal of combinatorial optimization, 24(3): 374-378.
Mashayekhi Y, Meybodi MR, Rezvanian A (2018) Weighted estimation of information diffusion probabilities for independent cascade model. 2018 4th International Conference on Web Research (ICWR). IEEE: 63-69.
Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Physical review E, 69(2): 026113.
O'Connor B, Balasubramanyan R, Routledge BR, et al. (2010) From tweets to polls: Linking text sentiment to public opinion time series. ICWSM, 11(122-129): 1-2.
Ou J, Duan Z, Chang L (2002) Typhoon risk analysis for key coastal cities in southeast China. Journal of Natural Disasters, 11(4): 9-17.
Palla G, Derényi I, Farkas I, et al. (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435(7043): 814-818.
Peng H, Zhao D, Li L, et al. (2016) An improved label propagation algorithm using average node energy in complex networks. Physica A: Statistical Mechanics and its Applications, 460: 98-104.
Plotnick L, Turoff M, White C (2011) Partially Distributed Emergency Teams: Considerations of Decision Support for Virtual Communities of Practice. Supporting Real Time Decision-Making. Springer US: 203-220.
Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. Physical review E, 76(3): 036106.
Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM: 61-70.
Schmidt S, Kemfert C, Höppe P (2010) The impact of socio-economics and climate change on tropical cyclone losses in the USA. Regional Environmental Change, 10(1): 13-26.
Starbird K, Palen L (2010) Pass it on? Retweeting in mass emergency. International Community on Information Systems for Crisis Response and Management.
Steinhaeuser K, Chawla NV (2010) Identifying and evaluating community structure in complex networks. Pattern Recognition Letters, 31(5): 413-421.
Sun H, Liu J, Huang J, et al. (2015) CenLP: A centrality-based label propagation algorithm for community detection in networks. Physica A: Statistical Mechanics and its Applications, 436: 767-780.
Sun J, Tang J (2011) A survey of models and algorithms for social influence analysis. Social Network Data Analytics. Springer US: 177-214.

Wakita K, Tsurumi T (2007) Finding community structure in mega-scale social networks: [extended abstract] Proceedings of the 16th international conference on World Wide Web. ACM: 1275-1276.

Xie J, Chen M, Szymanski BK (2013) LabelrankT: Incremental community detection in dynamic networks via label propagation. Proceedings of the Workshop on Dynamic Networks Management and Mining. ACM: 25-32.

Xie J, Szymanski BK (2012) Towards linear time overlapping community detection in social networks. Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer Berlin Heidelberg: 25-36.

Xie J, Szymanski BK (2013) Labelrank: A stabilized label propagation algorithm for community detection in networks. Network Science Workshop (NSW), 2013 IEEE 2nd. IEEE: 138-143.

Xie J, Szymanski BK, Liu X (2011) Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE: 344-349.