Network hot topic prediction of infectious disease model

Zhang Dandan
Hainan College of Economics and Business, Hainan Haikou China, 571127

Keywords: hot topic; propagation model; epidemic model; topic propagation model

Abstract: It is of great value to build an epidemic model for network hot topic, However, the existing epidemic models do not distinguish between the topic type and the probability of different users, To this end, a content-based network hot topic propagation model is introduced in this paper, The probability of a single user is defined based on the user topic sensitivity, Content classification of fusion topics, User propagation probability, Factors such as user re-entry probability, Learn from the basic idea of SIRS model, The topic propagation model (CSIRS) is constructed in the scale-free network, Small world network, Random network and Experiments on real social networks, The experimental results show that the CSIRS model can not only show the propagation model of general epidemic dynamics model, Also able to show a number of fluctuations, a small range of long spread, The rapid rise and decline of social network hot topics such as the spread of the model. The model provides a new research idea for the construction of the network structure and the topic content attributes.

1. Introduction

The topic of network hotspot has great influence on virtual society and real society. It is of great value to understand and model the spread of hot topics on social networks[1]. The dissemination of topical topics on the Internet is highly consistent with the spread of disease in human society. The use of infectious disease transmission models to understand and analyze the dissemination of messages in networks has a broad base. For example[2], the infectious disease model is used to understand the message propagation in the sensor network, and the dynamic network equations are constructed by combining complex networks and infectious disease dynamics theory in social networks. The modeling of topic development trends is difficult, not only because the topic development trend is related to the human behavior behind the topic[3], but human behavior is difficult to detect and quantify, in addition to the media's promotion of the attraction factor of the topic itself. Even related to the time when the topic started. Therefore, this paper develops a network hot topic prediction research by establishing an infectious disease model.

2. Overview of infectious disease models

The SI model is called one of the most basic popular virology models. It assumes that there are only two types of nodes in the virus's propagation environment, uninfected and infected. Every infected person tries to infect uninfected people with the probability of β, and β indicates the intensity of infection of the virus. Once a node is infected, it is always in an infected state and there is no possibility of "rehabilitation." Suppose there are N nodes in the propagation environment, and each node will not die during the propagation process. At time t, if there are I(t) infected nodes, then there are N-I(t) uninfected nodes. Based on the infectious disease transmission model, a large number of scholars have proposed new theories[4]. A representative literature proposes a study on the transmission mechanism of infectious diseases in a growth-free scale network with local structure. The model embodies two important characteristics of social networks: the heterogeneity of the degree distribution of nodes and the correlation of the degree of node degrees. It is pointed out that the node degree of the network obeys the power law distribution, and the existence of the local structure also leads to the positive correlation of the network node degree. The study of online social networks was originally limited to qualitative or semi-quantitative statistical analysis. The earliest research focused on analyzing the behavior of network users.

  DOI: 10.25236/mmmce.2019.131

3. Basic definitions and assumptions

3.1. Establish a network-based hot topic propagation model

In order to overcome the shortcomings of the infectious disease infection model, the factors that influence the topic spread on the social network include: (1) topic attraction index: the content characteristics used to express the topic[5]. Different types of topics attract different users to participate in different strengths, and the values are different. . (2) Network structure: A social network composed of users. (3) Length of topic retention: The length of time a node can perceive a topic, as indicated by the carrying period of the disease. The infected user has been infected for the duration of the topic. (4) Sensitivity of nodes to topics: Each user is sensitive to topics.Let β_i user i be sensitive to topic T. (5) User β_i reentry probability, with the user, the user can participate in the topic multiple times, that is, it can be infected multiple times. For the sake of simplicity, the probability that the user re-enters the susceptible group is a fixed value. The topical content-based propagation model CSIRS model is proposed using the SIRS model (see Figure 1)[6].

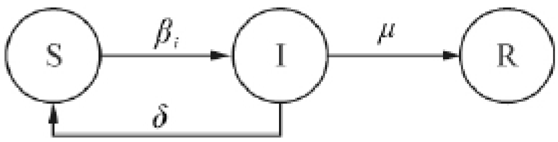

Figure 1 CSIRS model population migration map

As shown in the CSIRI model, there are three categories of hot topic user populations: one is S (susceptible) indicating potential users; the other is I (infected) indicates participating users; and the third is R (recoverer) indicating immune users. . Assuming that each user has a different probability of propagating a topic, the propagation probability of the CSIRS is given below.

3.2. Assumptions

Assuming that the total number of users in a social network is N, and the number of neighboring users N_i of a user i is, the user generates sensitivity to the topic, then the user becomes the probability of being infected at the time of infection, and the probability of the user topic propagation probability is as follows[7]:

$$p_i = 1-(1-\beta_i)\sum_{j=1}^{N_i}\delta_j \quad (1)$$

Where: δ_i represents the symbolic function of the infected user in the user i adjacent user, ie

$$\delta_i = \begin{cases} 0, \text{User j is not infected} \\ 1, \text{User j is infected} \end{cases}$$

According to the assumptions and propagation probabilities of the CSIRI model, the propagation process algorithm is designed as follows[8]: (1) initialization; (2) initialization parameters; (3) random generation of initial topic initiators; (4) While: loop; (5) For in Networks: for each node of the network; (6) If i == "1": if the node status is 1; (7) If i is within the hold time range: (8) keep; (9) Else: δ with probability (10) If i == "S": if the node state is S; (11) acquire neighboring nodes; (12) calculate participation probability p_i; (13) p_i infection with probability. The iterative loop reaches the upper limit of the number of iterations, or the algorithm ends when all the nodes are in the state of R. In the loop body, each node in the network is judged by state, and corresponding operations are performed.

4. Analysis of model experiment results

4.1. Simulation Design

In the simulation environment, the sensitivity of users to different topics is difficult to define. In order to truly simulate the degree of user attraction by the topic, the analog tag is used to simulate the sensitivity of users to different topics. Similarly, 3~8 tags are randomly set for the topic. Similarly, the user also randomly sets 3~8 tags, and then calculates the similarity between the user tag and the topic tag as the sensitivity of the user to the topic.Set the topic tag to $T_{tags} = \{tag(1), tag(2),..., tag(n)\}$, the user tag is, then the sensitivity of user I to $U_{tags} = \{tag(1), tag(2),..., tag(n)\}$ the topic is defined as[9]:

$$\beta_i = \frac{2 \times |T_{tags} \cap U_{tags}|}{|T_{tags}| + |U_{tags}|} \quad (2)$$

According to the above formula, the higher the similarity between the user tag and the topic tag, the higher the sensitivity of the user to the topic, and the easier it is to spread the topic, which is consistent with reality.

4.2. Analysis of simulation results

The experiment uses three kinds of network forms to simulate the propagation effect of the CSIRS model. The three network forms are:

(1)FS scale-free network

Scale-free networks have strong heterogeneity, and a small number of nodes occupy a large number of edges, so these few nodes have a great impact on the nature of the network.

(2)WS Small World Network

Small world networks have small world characteristics (smaller average shortest path) and clustering characteristics (larger clustering coefficients)[10]. Use the WS model to generate a small world network with a reconnection probability set to 0.2.

(3)ER random network

The structure of the random network has a great change, and a random network is generated by the ER model, in which the connection probability is set to 0.02.

(4)Compare the different network propagation effects of CSIRS

In order to compare the different network propagation effects of CSIRS, the same parameter settings are adopted for all three networks (see Table 1). The length of the topic retention time is a random variable, assuming that each user's duration of the topic is a normal distribution. For the sake of simplicity, let the mean of the distribution be 7.

Table 2 parameter settings

parameter	value
Number of nodesN	200
Topic attraction index θ	0.5
Topic duration π	N（7,1）
User reentry probability δ	0.5
User sensitivity to the topic β_i	$\beta_i = \frac{2 \times \|T_{tags} \cap U_{tags}\|}{\|T_{tags}\| + \|U_{tags}\|}$

Three different networks perform three simulation experiments under the same parameter configuration, and the three results are comprehensively compared. They are experimental results on scale-free networks, small world networks, and random networks. The conclusion of the comprehensive analysis of the communication model is as follows:

(1) CSIRS model On different networks, topic communication presents different modes. The CSIRS model infects the largest number of users on the small world network, and the speed is fast,

but the message fades at the fastest speed; on the scale-free network, the slowest speed is propagated, and the speed of the fade is also the slowest.

(2)The propagation of the CSIRS model on a scale-free network can exhibit multiple fluctuations. Since there are few nodes with very high degree in the scale-free network, the degree of most nodes is very small, so if the message cannot be transmitted to the node with high degree, the spread of the topic will be small. If the topic spreads to some high-degree nodes, it can cause multiple fluctuations. In addition, due to the structural complexity of the scale-free network, the topic has a long lifetime on a scale-free network and may re-transmit after a long period of silence. These three modes of communication have large differences. The reason is that the sensitivity of the node to the topic and the degree of the node are not necessarily related. Therefore, the determinants of the range and speed of the topic are not only the physical properties of the network, but also the node pairs. The sensitivity of the topic, etc., is consistent with previous literature studies[11].

(3) The three subgraphs show a basically consistent propagation pattern, and the topic rises rapidly and then falls in the form of an approximate power law. The topic has a short survival period, and the survival time of the three transmissions is less than 60h, while the survival period in the scale-free network and the random network exceeds 60h. This result and the previous literature found that the disease spread faster in the small world network than the rule network! The spread range is more consistent.

(4) The range and propagation speed of the CSIRS model on the random network are between the scale-free network and the small-world network. The long tail characteristics of topic propagation on random networks are more obvious than those on small world networks. This shows that due to the different sensitivity between content and users, topics can continue to be small in a long time after the peak period of communication. propagation. In addition, as the connection probability of random networks increases, the number of edges of random networks will continue to increase, and the average path between nodes will continue to decrease; therefore, the scope of model propagation will continue to expand and the speed will be significantly accelerated.

The results of different simulation experiments obtained by different networks under the same parameter configuration show that the CSIRS model can not only present the propagation mode of the general infectious disease dynamic model, but also exhibit multiple fluctuations, small-scale long-term propagation, and rapid rise. Slow decline and other social network topic communication models. This fully demonstrates that factors such as the characteristics of the topic, the sensitivity of the user to the topic, etc., can better describe the hot topic communication mode on the social network.

5. Real network communication experiment

In order to evaluate the effect of the CSIRS model on the real network, a real social network of a user was collected from the Renren network, and the number of nodes in the network was 445. Real social networks have a mixture of scale-free and small worlds, using the same parameter settings as the simulation experiments.

The test results (see Figure 2) confirm that the CSIRS model on the real social network exhibits a basically consistent propagation pattern, the difference being that the propagation process is rising faster, and that local fluctuations and more pronounced long tail effects occur during the propagation decline. The CSIRS model spreads very fast on real social networks, surpassing the speed of propagation under scale-free networks and small-world networks. The reason is that the experiment collects a real social network of users, and the average path is short, so it is easy. Form a rapid spread of topics. The decline mode of the CSIRS model is characterized by fast and slow first, which is consistent with the spread of the real topic. In addition, the long tail effect of the CSIRS model on real social networks is greater than that of small world networks, which is also consistent with the fact that the topic may be locally spread by a small number of users on a real network for a long time.

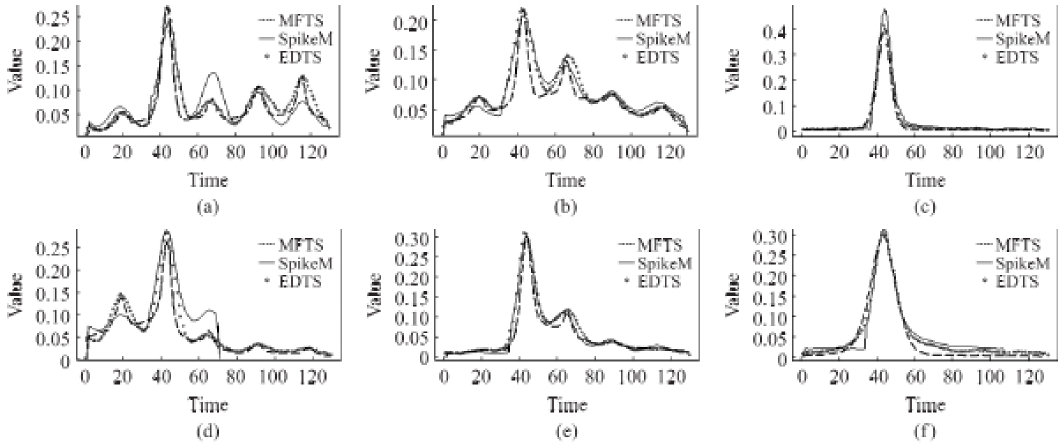

Figure 2 Comparison of empirical results

(1) The CSIRS model is difficult to predict the upward trend of the topic. It is found that after the prediction point, the CSIRS model fluctuates slightly and the overall decline trend. Because the CSIRS model has only one peak, the predicted point must fluctuate slightly after the peak, and the overall decline trend. If the topic propagation mode is after the first peak It is not a simple downward trend of small fluctuations, but the fluctuations continue to rise, so the CSIRS model cannot be predicted, and the model of this paper can better predict the pattern that can rise after the peak.

(2) The model can predict the propagation mode of multiple peaks. After the first peak, the topic will reach the second peak. As can be seen from the figure, although the model predicts the time point of the second peak and There are some differences in the actual sequence, but it is predicted that the topic will have a second peak.

(3) For the propagation mode with better morphology, for example, the two models can better predict the descending process of the topic.

When the topic communication is affected by fewer factors, both models can better fit and predict the trend of the topic.

(4) For the large amplitude fluctuations, the prediction effect of the CSIRS model is poor. The CSIRS model uses the time period as the factor of induced fluctuation, which is caused by the CSIRS model having a large fluctuation or peak and not a simple downward trend. The main reason why the topic is unpredictable. The model in this paper does not use time-simulating fluctuations, but adopts the mechanism of user revitalization, which is to simulate the development mode of user interaction to promote the topic, which has achieved good results.

6. Conclusions and future work

This paper analyzes the characteristics of topic-participating users and puts forward the reasonable assumptions for constructing the model. According to the topic development model, the CSIRS model is constructed. The experimental results show that the proposed model has better development trend in fitting and predicting topics. The effect, especially on the complex topic development model, the model proposed in this paper can be well fitted and predicted. These results indicate:

(1) When modeling hot topics, it is necessary to consider factors such as multiple re-entry of users and activity of different nodes, which affect the development trend of the topic. At the same time, it is also necessary to consider the attraction, time and external environment of the topic itself.

(2) Modeling single peaks and multiple peaks in line with the development model of real topics, this paper constructs a hot topic development model in real network environment. Under the condition of a lot of noise, the model can better fit the topic trend. .

(3) The CSIRS model has basically the same communication mode on the small world network and the real social network, which shows that the topic propagation on the complex network with

small world features has a similar pattern; but the form of topic communication in the actual social network is still There are many, and further factors need to be added to the model.

References

[1]Vinarti R A , Hederman L . Personalization of Infectious Disease Risk Prediction: Towards Automatic Generation of a Bayesian Network[C]// 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2017.

[2]Wenhua H . Hot topic prediction model based on improved echo state network[J]. Computer Engineering & Applications, 2014.

[3]Tong H , Liu Y , Peng H , et al. Internet Users' Psychosocial Attention Prediction: Web Hot Topic Prediction Based on Adaptive AR Model[C]// International Conference on Computer Science & Information Technology. IEEE, 2008.

[4]Meyers L . Contact network epidemiology: Bond percolation applied to infectious disease prediction and control[J]. Bulletin of the American Mathematical Society, 2007, 44.

[5]Hirose H , Wang L . Prediction of Infectious Disease Spread Using Twitter: A Case of Influenza[C]// International Symposium on Parallel Architectures. IEEE, 2013.

[6]Li J , Gu J Z , Mao S H , et al. [Preliminary application of Back-Propagation artificial neural network model on the prediction of infectious diarrhea incidence in Shanghai][J]. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi, 2013, 34(12):1198.

[7]Chen J , Yu J , Shen Y . Towards Topic Trend Prediction on a Topic Evolution Model with Social Connection[C]// IEEE/WIC/ACM International Conferences on Web Intelligence & Intelligent Agent Technology. IEEE, 2012.

[8]Zhou Q , Wang G , Chen H . A Topic Evolution Model Based on Microblog Network[J]. Lecture Notes in Electrical Engineering, 2014, 260:791-798.

[9]Eichinger, Ludwig. Editorial [Hot Topic: Model Organisms to Study Host - Pathogen Interaction: Prerequisites for the Identification of Novel Drug Targets (Guest Editor: Ludwig Eichinger)][J]. Current Drug Targets, 2011, 12(7):934-935.

[10]Mingsheng T , Xinjun M , Zahia G . Research on an Infectious Disease Transmission by Flocking Birds[J]. The Scientific World Journal, 2013, 2013:1-7.

[11]Burnett S M , Mbonye M K , Naikoba S , et al. Effect of Educational Outreach Timing and Duration on Facility Performance for Infectious Disease Care in Uganda: A Trial with Pre-Post and Cluster Randomized Controlled Components[J]. PLOS ONE, 2015, 10.