ISSN: 2831-6789

Research Article

Journal of Robotics and Automation Research

An Online Topic Diffusion Prediction Approach on Heterogeneous Social Networks

Beibei Zhang[1*], Linfeng Li[1], Xiaowan Li[1], Yichuan Wang[1], Robertas Damasevicius[2]

[1]*School of Computer Science and Engineering, Xi'an University of Technology, Shaanxi Key Laboratory for Network Computing and Security Technology Xi'an 710048, China*

[2]*Department of Multimedia Engineering, Kaunas University of Technology, Lithuania*

[*]Corresponding Author
Beibei Zhang, School of Computer Science and Engineering, Xi'an University of Technology, Shaanxi Key Laboratory for Network Computing and Security Technology Xi'an 710048, China.

Submitted: 13 Mar 2023; **Accepted**: 20 Mar 2023: **Published**: 10 April 2023

Citation: Zhang, B., Li, L., Li, X., Wang, Y., Damasevicius, R. (2023). An Online Topic Diffusion Prediction Approach on Heterogeneous Social Networks. *J Robot Auto Res, 4*(1), 343-358.

Abstract
Online topic indicates the diffusing process of various attitudes and emotions towards some events expressing and propagating over the Internet. It is an essential for monitoring all online information and ensuring internet information safety to search for the predicting approach of the trend of online topic. Previous methods focus on special social network and lack of versatility because of intensely relying on manual experience, that we are devoted to conquering these shortages. In the manuscript, we divide online topic into unimodal topic and multimodal topic which is the combination of a series of unimodal topics. We present a generally nonlinear dynamic model to describe unimodal topic diffusion mechanism on heterogeneous social network. The model can capture the long-term development tendency of unimodal topic, So, we present a long-term tendency prediction method of unimodal online topic with help of the dynamic model. But the model is not for multimodal online topic in that their long-term development tendency is made up of a series of unimodal topics' long-term development tendency, which leads to the unpredictability of their long-term future directions, we propose a short-term population prediction method of unimodal and multimodal online topic diffusion on heterogeneous social network based on the essential relationship between unimodal and multimodal topic. Experimental results are our two methods have good performance on predicting the long-term development tendency of unimodal topic and all topic's short-term population.

Keywords: Nonlinear Dynamic Model, Unimodal Topic Diffusion, Multimodal Topic Diffusion, Long-Term Trend, Short-Term Trend.

Introduction

Tracking online topic diffusion is a key component of developing successful and universal applications on different heterogeneous online social network. In most cases, online topic is possibly triggered by a series of related events. If topic is only one event-triggered, it is called unimodal topic, but otherwise it is called multimodal topic triggered by a series of related events, and a multimodal topic has a multi-stage topic development process. For example, while a film is being publicized, the film takes a series of commercial speculations before the premiere; each spike of the multimodal topic about the film exhibits the influence of each commercial speculation on the film's publicity. Furthermore, it is commonly seen that the multimodal and unimodal topics coexist widely in the internet. Because multimodal topic diffusion is compounded by a series of unimodal sub-topic diffusion, modeling unimodal topic diffusion is the kernel of our concern.

Although previous works focused on describing the patterns of information diffusion, they did not formulate a universal model to explain the mechanism of unimodal topic diffusion, and did not seek out a way to predict the rise and fallen of unimodal topic diffusion in the future, because they keep the key differences between the main social media platforms and how they are likely to influence information spreading, it is another fundamental problem with more cares to design universal approaches to eliminate those differences for predicting the short-term evolution of both the unimodal and multimodal topic diffusion. We generalize the major contributions of the manuscript as below. Firstly, we propose a novel time-dependent dynamic unimodal topic diffusion model to describe state transition of the population of users who participate in topic discussion. Secondly, we propose the long-term topic diffusion trend prediction method to predict the rise and fallen of topic diffusion based on our model. Thirdly, we propose the short-term multimodal topic diffusion prediction method to predict the short-term evolution of users' population who participates in both multimodal or unimodal topic discussion.

The residue of this manuscript is formed as below. In section 2, we provide an overview on previously online topic modeling; Section 3 reveals background knowledge on the epidemic-like dynamic model with vast applications in social networks. Section 4 shows our time-dependent unimodal topic diffusion model in detail, furthermore, theoretically and systematically we analyze the stability of time-dependent dynamic unimodal online topic diffusion model. Section 5 specifically describes two novel and systematic approaches for predicting long-term trend of unimodal topic, and short-term evolution of multimodal topic diffusion. Section 6 focuses on the evaluation of our two methods as well as their applications on the set of real data, we prove the validity of our model and our methods by comparing with previous information diffusion methods in literature. In section 7, we briefly sum up the performance of the manuscript and elaborate future research points.

Literature Review

In recent years, researchers on social networks have applied kinds of methods (such as dynamic modeling, stochastic approach and data mining method) to describe topic diffusion on social networks for aiding in understanding the dynamic diffusion mechanism of topic [1].

Epidemic-Like Modeling Approach

In the last several years, several models and approaches have been proposed to model and analyze topic diffusion process. Previous works focus on modeling rumor diffusion mainly by using epidemic modeling because rumor diffusion have similar diffusion features and schematics as to those of epidemics [2-4]. Then integrating complex networks and dynamics, proved that diffusion velocity will asymptotically drop according to a power law with a Susceptible-Infection (SI) model for incidental topics on blog networks based on individual fitness [5]. For investigating the role of different types of users in the diffusion of situational information through online social networks in disasters, investigated the influence of two types of users: crowd (regular users) and hubs (users with a large number of followers) on the speed and magnitude of information diffusion [6]. Then a novel intelligent information diffusion phenomenon is modeled, which introduces a new "uncertain" psychological state into classic susceptible-infected-recovered model [7]. In the diffusion process, the new proposed model characterizes a practical reinfection-reemergence scenario caused by the change of social attributes for individuals.

Stochastic Modeling Approach

Many works used a stochastic approach to model topic diffusion in weblogs [8, 9]. Used factor graph to indicate collapsed LDA, which succeed to describe online topic diffusion process in light of loopy belief diffusion (BP) algorithm estimating approximate inference and parameter. And Adoption-based Participation Ranking (APR) model focused on rank the actual participants in reality at higher positions to solve the Diffusion Participation Forecasting (DPF) problem by using influence maximization and boosting the accuracy of popularity prediction. Constructed a heterogeneous information network including two kinds of nodes (users and contagions) and discussed the representations of heterogeneous nodes with help of meta-path-based proximity random walk algorithm [10-12].

Data Mining-Based Approach

Data mining-based methods have been long and efficiently used to describe online topic diffusion [13, 14, 26]. Understanding how memes of a topic are propagated on web platforms, especially weblogs, is useful for knowing how information transfers through other social networks and how these structures either dampen or amplify this spread. Designed an approach for tracking short, distinctive phrases transformation of online text on blogs for discovering a broad class of memes with wide spread and rich variations daily with clustering the textual variants of such phrases [15]. And the K-Spectral Centroid (K-SC) community detection algorithm found community centroids by measuring the similitude of the appeared time series related with observed topics and discovering that there are six different participants in the online media space shape the dynamics of the attention. proposed an approach for calculating the long-term propagation trend of online post with the aid of early measuring netizen access on Digg and YouTube [16, 17]. The authors captured the propagation growth features in YouTube. Designed an approach to describe the mechanism of influence, authoritativeness and relevance under a topic-aware model. Proposed a general and flexible analytic model (termed as SPIKEM) to indicate the rise and fall patterns of topic influence diffusion [18-20].

Then a semi-supervised representation learning model was proposed using a graph attention network and a convolutional neural network (CNN) for HINs, called RANCH [21]. In the part of the graph attention network, we construct a heterogeneous graph attention network using heterogeneous edges to preserve the features of nodes and the structure of network. RNe2Vec (repost network to vector), which is a repost network embedding-based diffusion popularity prediction algorithm, predicted the information diffusion popularity only based on early repost information, given that the underlying user relation network precisely [22]. Then a recurrent neural network model with graph attention mechanism, which constructs a seq2seq framework to learn the spatial-temporal cascade features, was built to learn their structural context in several different time intervals based on timestamp with a time-decay attention and predict the next user with the latest cascade representation [23]. A formal model was developed to characterize the dynamic process of knowledge diffusion in the autonomous learning under multiple networks to expand the scope of knowledge through hybrid online learning for allocating educational resources [24]. Neural Diffusion Model (NDM) made relaxed assumptions and employs deep learning techniques including attention mechanism and convolutional network for better fitting the diffusion data and generalize to unseen cascades [25].

Unimodal Topic Diffusion Model on Online Social Network

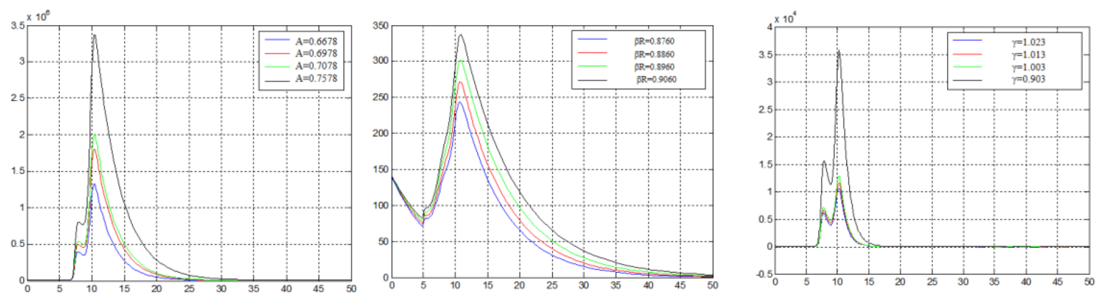
Netizens always diffuse their attitudes and information through Online social networks. Recently, have witnessed a flourish of researching online topic spread dynamics [26]. Furthermore, they presented some related macro- and micro-level social models for describing information diffusion modules in large-scale social platforms. But for heterogeneous online social network websites providing diverse services for users their spreading patterns have common features in the means of information diffusion [16, 27]. So, we consider group dynamics of information based on epidemic dynamic theory.

Model Definitions

Availably seeking a method to track the mechanism of topic diffusion, we will classify topics into two classes - unimodal topic and multimodal topic according to the number of topic spikes in the process of topic diffusion. However, a multimodal topic consists of a series of inter-correlated unimodal sub-topics, so the unimodal topic diffusion model is the basis for understanding the multimodal topic diffusion. Therefore, we firstly describe the dynamic mechanism of unimodal topic diffusion based on the discrete time-dependent SIRS (Susceptible–Infected-Removed-Susceptible) model in epidemic dynamics as shown in (3.1):

$$\begin{cases} \dfrac{dR}{dt} = B(t)R + \xi(t)E - \beta(t)RD - w(t)R \\ \dfrac{dD}{dt} = A(t)D + \beta(t)RD - \gamma(t)D \\ \dfrac{dE}{dt} = \gamma(t)D - \xi(t)E \end{cases} \qquad (3.1)$$

where $A(t)$ (the input rate of $D(t)$), $\beta(t)$ (the coefficient of spreading rate), (the input rate of), $B(t)$ (the output rate of $R(t)$), $w(t)$ (the output rate of $D(t)$) and $\xi(t)$ (the output rate of $E(t)$)are all time-varying, continuous and bounded (with the constraints of $\beta(t) > 0$, $w(t) > 0$, $B(t) > 0$, $A(t) > 0$, $\xi(t) > 0$, $w(t) > 0$, and $\gamma(t) > 0$) functions. $R(t)$, $D(t)$ and $E(t)$ are all state variables and bounded on time range ($t > 0$) with $D(t) > 0$, $E(t) > 0$, and $R(t) > g > 0 (\exists g > 0, \forall t \geq 0)$ as constrains; we define the Discussion Group as $D=\{d(t),t>0\}$,



(a) The population variation of the discussion group D with the change of it's input rate A

(b) The population variation of the discussion group D with the change of it's spreading rate βR

(c) The population variation of the discussion group D with the change of it's output rate γ

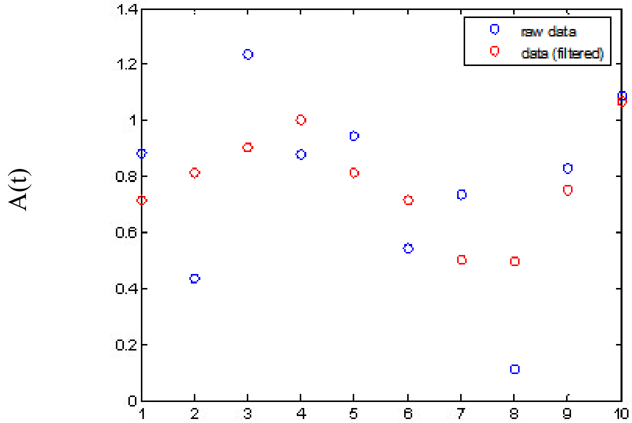
Figure 1: Influence relation graph between A(t), β(t) or γ(t) and the population of the Discussion Group D(t)

where $d(t)$ are users who post or comment on the unimodal topic T_d at time t; Secondly, a set of users $R=\{r(t), t>0\}$ are named as the Related Group, where $r(t)$ are users who post or comment on topic T_r at time t; we define the set of the Exited Group as $E=\{e(t), t>0\}$, where $e(t)$ are users who published or commented on unimodal topic T_d at t-1 epoch but withdraw from $D(t)$. $D(t)$, $R(t)$, and $E(t)$ usually represent the sizes of three groups at time t respectively. We establish hot online unimodal topic diffusion model with the help of time-dependent epidemic dynamics to describe the state dynamic transitions of $D(t)$, $R(t)$, and $E(t)$. The system (3.1) reflects the popularity level of the observed unimodal topic by the growth of discussion group $D(t)$.

Specially, we will describe dynamic topic diffusion by analyzing the complex interactions among three groups. Users should be eager for browsing or participating in following similar historical online information to topic if they are interest in the observed topic, so one of the input rates of the Discussion Group come from the Related Group by βRD but it is possible to ignore other input participants source, then we take advantage of the outside input rate $A(t)D$ to cover the resulted deficit. As far as the system (3.1) is concerned, we propose that some of the Exited participants lose interest in writing or commenting on webpage about any topic (T_r and T_d) which belongs to the same category as the topic and they may join in the Related Group in the future.

Analysis of Time-varying Dynamic Unimodal Hot-topic Diffusion Model

Because $A(t)$ (the input rate of $D(t)$), $\beta(t)$ (the coefficient of spreading rate), and $\gamma(t)$ (the output rate of $D(t)$) decide the population evolution of the Discussion Group $D(t)$, we will discuss quantity influence relationship between $A(t)$, $\beta(t)$ or $\gamma(t)$ and the population of the Disussion Group $D(t)$. According to $dD / dt = A(t)D + \beta(t) RD - \gamma(t)D$,when $\beta(t)$ and $\gamma(t)$ keep the value unchanged, the population of the Disussion Group $D(t)$ is directly proportional to $A(t)$, which is shown in Fig.1-a; when when $A(t)$ and γ(t) keep the value unchanged, the population of the Disussion Group $D(t)$ is directly proportional to $\beta(t)$, which is shown in Fig.1-b; when $\beta(t)$ and $A(t)$ keep the value unchanged, the population of the Discussion Group $D(t)$ is inverse proportion to γ(t), which is shown in Fig.1-c. So, from the equation and former analysis the population



Sampling time

Figure 2: The coefficient of input rate A (t) of the Discussion Group etween real data and recursive filtering data

change of the Discussion Group $D(t)$ is affected by the change of $A(t)$ (the input rate of) $D(t)$, $\beta(t)$ (the coefficient of spreading rate), and $\gamma(t)$ (the output rate of $D(t)$).

Furthermore, for analyzing the state of all parameters' change, we firstly complete to eliminate the major noise of data before data fitting of all parameters with help of recursive filtering algorithm, we take example as the $A(t)$ (the input rate of $D(t)$), which is shown in Fig. 2. We can find that all parameters, including $A(t)$ (the input rate of $D(t)$), $\beta(t)$ (the coefficient of spreading rate), and $\gamma(t)$ (the output rate of $D(t)$) are not constants over time, which is shown in Fig.3. So, it is reseanable that we consider all parameters to be time-varying.

Until now, we have discussed the some features of proposed time-varying dynamic unimodal hot-topic diffusion model based on time-dependent SIRS model; however, we need to catch the criterion to forecast the persistent or dying out of unimodal topic diffusion in the future by stability analysis of time-varying dynamic unimodal hot-topic diffusion model-the system (3.1).

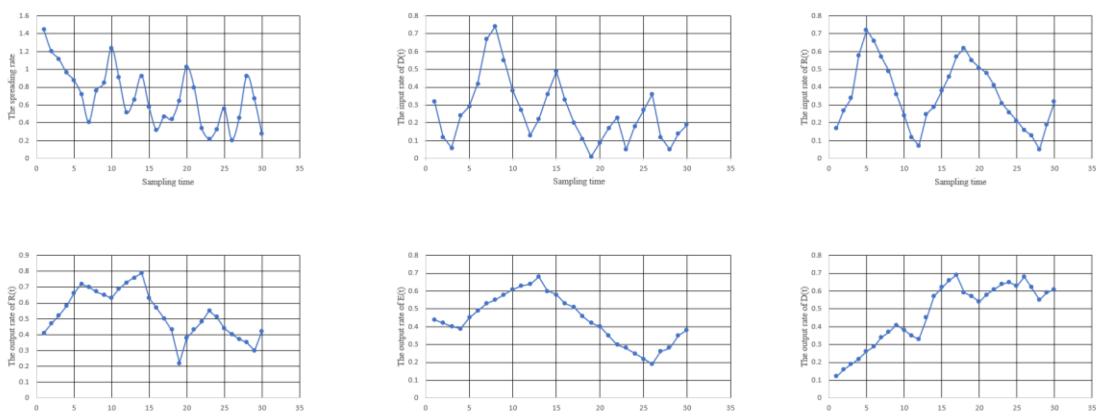


Figure 3: The coefficient of all parameters after recursive filtering

Before the above analyzing work, it is helpful to analyze the constant SIRS model where all coefficient parameters of system (3.1) are constant values. For the second equation of the system (3.1), the size of the Discussion Group is up to parameter D with $D = \int (A + \beta R - \gamma)\, dt$ hold. Actually, when $\beta R / (\gamma\text{-}A) > 1$, the size of D will tend to be a positive equilibrium of the constant dynamic topic diffusion model as $t \to \infty$. Furthermore, while $\beta R / (\gamma\text{-}A) > 1$, the scale of D will make trend towards die out as $t \to \infty$. Analyzing the stability of equilibrium of the constant system is huge inspiring for those analyses of the corresponding time-dependent system, i.e., analyzing the equilibrium stability of the system (3.1) which is more complicated than that of the corresponding constant SIRS system. We formulate it as the time-dependent system due to the fact that all most all the topic diffusion processes are affected by many potential stochastic factors in the internet environment, some of which may be important for topic development and others may be noise. It is difficult and unrealizable for us to distinguish between noise and real topic data. So it is better and more exact that we consider the parameters of system (3.1) to be time-dependent.

Therefore, with a view to coming up with a proper method and algorithm to predict the long-term trend of a hot online topic diffusion which represents two status of unimodal topic diffusion including the persistent or dying out in the future, it is necessary for us to catch the threshold of the equilibrium stability of system (3.1). By analyzing time-varying dynamic unimodal hot-topic diffusion model (3.1), we define the following thresholds expression:

$$Q^* = \frac{(\sup(\beta N) + w(t) - B)^*}{(\gamma - A)_*}$$

Where

$$(\gamma - A)_* = \lim_{t\to\infty} \inf(\frac{1}{t}\int_0^t (\gamma - A)ds)$$

$$(\sup(\beta N) + w(t) - B)^* = \lim_{t\to\infty} \sup(\frac{1}{t}\int_0^t (\sup(\beta N) + w(s) - B)ds)$$

By those proposed thresholds, we can determine a sufficient proposition as shown in Proposition 3.1 by which we can prejudge the long-term trend of unimodal topic diffusion.

Proposition 3.1 When $Q^* < 1$, the long-term trend of unimodal hot online topic diffusion will die out, because for every solution of system (3.1) we have $\lim_{t\to\infty} D(t) = 0$, as $t \to \infty$.

Proof:

$$\begin{cases} x = \dfrac{R(t)}{N(t)} \\ y = \dfrac{D(t)}{N(t)} \\ z = \dfrac{E(t)}{N(t)} \end{cases} \quad (3.2)$$

,

system (3.1) is rewritten by variable transformation equation (3.2) as:

$$\begin{cases} \dfrac{dx}{dt} = (B(t) - w(t))x + (w(t) - B(t))x^2 - (\beta N + A(t))xy + \xi(t)z \\ \dfrac{dy}{dt} = ((\beta N + w(t) - B(t)))xy - (\gamma - A(t))y - A(t)y^2 \\ \dfrac{dz}{dt} = \gamma y - \xi(t)z + (w(t) - B(t))xz - A(t)yz \end{cases} \quad (3.3)$$

By $Q^* < 1$, we have $(\sup(\beta N) + w(t) - B(t))^* < (\gamma - A(t))_*$.

For $\forall \varepsilon > 0, \exists r > 0, \ni$

$$\frac{1}{t}\int_0^t (\sup(\beta(s)N)ds + w(s) - B(s))ds - \frac{1}{t}\int_0^t (\gamma - A)ds \le -\varepsilon \quad (3.4)$$

We achieve the inequality (3.5) by calculating inequality (3.4)

$$\begin{aligned} &\frac{1}{t}\int_0^t \beta(s)N(s)xds + \frac{1}{t}\int_0^t (w(s) - B(s))xds - \frac{1}{t}\int_0^t (\gamma - A(s))ds - \frac{1}{t}\int_0^t B(s)yds \le \\ &\frac{1}{t}\int_0^t \sup\beta(s)N(s)xds + \frac{1}{t}\int_0^t (w(s) - B(s))xds - \frac{1}{t}\int_0^t (\gamma - A(s))ds \\ &\frac{1}{t}(\int_0^t \sup(\beta(s)N(s))xds + \int_0^t (w(s) - B(s))xds - \int_0^t (\gamma - A(s))ds \le -\varepsilon, \forall t \ge r \end{aligned} \quad (3.5)$$

Combining the second expression of variable transformation model (3.2) with expression (3.5), we infer to

$$\begin{aligned} &\frac{1}{t}\ln\frac{y(t)}{y(0)} \le \\ &\frac{1}{t}(\int_0^t \sup(\beta(s)N(s))xds + \int_0^t (w(s) - B(s))xds - \int_0^t (\gamma - A(s))ds \le -\varepsilon, \end{aligned} \quad (3.6)$$

From the inequality (3.6), $y(t) \le y(0)e^{\varepsilon t}$, $\forall t \ge r > 0$ So we have as $\lim_{t\to\infty} y(t) = 0$ as $t \to \infty$. That is to say, the trend of the hot online topic diffusion will die out in the future.

For example, Real on-line topics such as "Melamine-contaminated milk", "Policeman beating somebody", "Tencent", and "Shenzhou spacecraft" on the BMY and LQQM BBS and "Roh Moo-Hyun's death" on the Sina blog die out because all their threshold values satisfy $Q^* < 1$.

Meanwhile for catching the criterion by which we can judge whether the topic diffusion trend keeps persistent, we should introduce the concept of uniformly but weakly persistent as follows: [27]

$$\exists\, r \ge 0, \exists \varepsilon > 0,\ D^\infty = \lim_{t\to\infty} D(t) > \varepsilon. \text{ and } D(r) > 0.$$

When the size of the discussion group satisfies the above condition, we can affirm that the topic will be followed by some users of online social network. For example, the topic "Obama's inauguration" on the Sina blog is uniformly but weakly persistent.

By careful analysis that unimodal hot online topic diffusion keeps uniformly weak persistence, the threshold should satisfy the following condition:

$$Q_* = \frac{(\beta N)_*}{(\gamma - A(t))^*}$$

where

$$(\beta N)_* = \lim_{t\to\infty}\inf\left(\frac{1}{t}\int_0^t \beta N ds\right)$$

$$(\gamma - A(t))^* = \lim_{t\to\infty}\sup\left(\frac{1}{t}\int_0^t (\gamma - A(s))ds\right)$$

Based on this threshold, we obtain a sufficient condition to judge whether the long-term trend of unimodal topic diffusion is uniformly but weakly persistent or not.

Proposition 3.2 When $Q_* > 1$ and $(\gamma - A(t))^* > 0$, the long-term trend of unimodal hot online topic diffusion will stay uniformly but weakly persistent.

Proof: We suppose that there are some solutions satisfying $D^\infty = \lim_{t\to\infty} D(t) < \varepsilon$, by the third equation of system (3.1) and lemma[28], the inequality (3.7) is achieved as follows:

$$E^\infty \le hD^\infty \qquad (3.7)$$

where $h = \sup(\frac{\gamma}{\xi(t)}), (\xi(t) > 0)$ and $0 \le y^\infty < \varepsilon$

We deduce to the following expression (3.8) by transforming expression (3.7):

$$E^\infty + D^\infty \le (1+h)D^\infty \qquad (3.8)$$

Because of $N = R + D + E$, expression (3.8), and the second expression of the model (3.1), for $\forall \varepsilon > 0$

$$d(\ln D)/dt \ge \beta[N(t) - h\varepsilon] - (\gamma - A) \qquad (3.9)$$

And for $t >> 1$, we achieve inequality (3.10) by direct numerical integration of inequality (3.9).

$$\frac{1}{t}\ln\frac{D(t)}{D(0)} \ge \frac{1}{t}\int_0^t ((\beta N(s)) - \beta h\varepsilon - (\gamma - A))ds \ge (\beta N(t))_* - (\gamma - A)^* - \beta^* h\varepsilon \qquad (3.10)$$

where $\beta^* = \max(\int_0^t \beta ds)$.

For sufficiently large $t>0$, by $Q_*>1$ and the former inequality (3.10), there should exist $\theta(\varepsilon)>0$ to establish the inequality (3.11) as follows:

$$\frac{1}{t}\ln\frac{D(t)}{D(0)} \ge \theta(\varepsilon). \qquad (3.11)$$

So $D(t) \ge D(0)e^{\theta(\varepsilon)t}$.

However, it is inconsistent with the boundedness of D [28].

For better understanding, we illustrate the trend of the topic diffusion using Fig.4, the x-axis shows the time dimension of the observed topic, and the y-axis shows the time cumulative size of the discussion group related to the observed topic. The black line with asterisk as the symbol represents that the topic diffusion trend tends to be uniformly but weakly persistent, the dash red line represents the time cumulative size of the Discussion Group increasing ratio tends to some positive value, the black line with circle as the symbol represents that the topic diffusion trend tends to be dying out, the red line represents the time cumulative size of the discussion group increasing ratio tends to zero.

After we catch the thresholds for prejudging long-term trend of unimodal topic diffusion, we will systematically introduce the long-term unimodal topic diffusion trend prediction method and its corresponding specific algorithm for heterogeneous social network in Subsection 4.

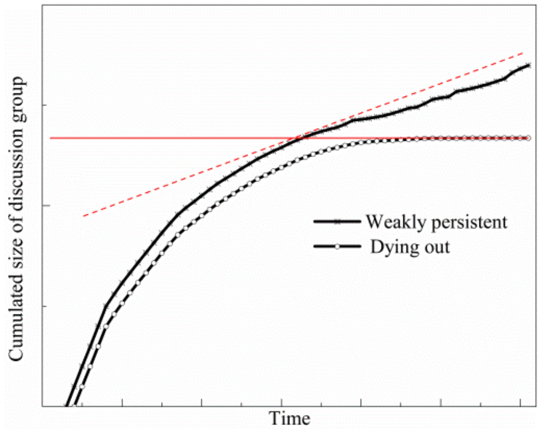


Figure 4: Illustrations for long-term the topic diffusion trends

Methodology

One natural application of analysis of the system (3.1) is to predict the trend of the topic diffusion on heterogeneous social network. Although unimodal and multimodal topics possess inter-related relationship, whole multimodal topics diffusion process owns the same unique long-term trend as those of unimodal topic. Firstly, a unimodal topic is caused by one event that happened; however, a multimodal topic is induced by a series of events correlated with one another. Secondly, a unimodal topic goes through integrated evolution and is almost immune to other events, but the evolution of a multimodal topic is overlapped and stimulated by a sequential process of unimodal subtopics. However, unimodal topic diffusion consisted of multimodal topic cannot determine the trend of the multimodal topic completely.

Thus, in the following discussion, we develop two methods to predict the topic diffusion trend based on different motivations above. One is long-term unimodal topic diffusion trend prediction method to predict the long-term trend of unimodal topic diffusion, and we have put forth an algorithm to show how to prejudge the long-term popularity of unimodal hot online topic diffusion. At the same time, we further propose the short-term multimodal topic diffusion prediction method to be a universal method for predicting the number of users who post or comment on webpage with respect to multimodal topics in following several days on heterogeneous social network. The short-term multimodal topic diffusion trend prediction method internally depends on the long-term unimodal topic diffusion trend prediction method; it seems that almost all steps of the long-term unimodal topic diffusion trend prediction algorithm are embedded and evolved in the short-term multimodal topic diffusion prediction algorithm with the time window properly chosen. Algorithms stop when either the raw data includes the topic trend information or a user-specified number of iterations are reached. The symbols definition of the algorithms are descripted in Table 1.

The parameters W and U are usually user-specified, and are basses to obtain desirable

Table 1: Parameters of the algorithm

Symbol	*Description*
L	Maximum number of days for prediction
W	Fit time
U	Time-update cycle
Q^*	Threshold of a unimodal topic “dying out” trend
Q^*	Threshold of a unimodal topic trend for uniformly weak persistence

Determined coefficient, precision (including spike precision), and $F1$ values later used to evaluate our approach. Specifically, we have found that determined coefficient and precision values are insensitive to the settings of W or U when W or U is above a certain value, as well as we describe the selective process of fit time in Subsection 6.2.

Long-term Unimodal Topic Diffusion Trend Prediction

We will describe the method predicting the long-term trend of unimodal topic diffusion. Topic information, extracted by the algorithm day by day, are stored for fitting distributions of main parameters including the coefficient of overflowing rate, the spreading rate, and the input rate of the Discussed Group, the output rate of the Related Group, the input rate of the Related Group, and the output rate of the Exited Group. Then the thresholds for the unimodal topic are calculated based on their definitions and compared against the long-term trend condition in propositions 3.1 and 4.2. If the criterion in proposition 3.1 is competent, the trend of topic diffusion is asserted to die out. If the criterion in proposition 4.2 is competent, the trend of topic diffusion is asserted to stay uniformly but weakly persistent. Observing the size of the Discussion Group on the topic, the algorithm terminates when the trend of topic diffusion is almost exhibited or the user-defined time iteration number is reached. The major idea for the unimodal topic diffusion prediction method is described in

Algorithm 1.
Algorithm 1: long-term unimodal topic diffusion trend prediction algorithm
Ensures: the long-term topic diffusion trend
1. *Capturing appropriate fit time W, it got activated at the long-term trend of unimodal topics, help system to describe the prospective trend of corresponding popularity level of the unimodal topics;*
2. *We summarize parameters of system (3.1), and catch the fitting distribution of system (3.1), the meaning of the summation is the available stimuli at time t; the effective goal is the long-term diffusion trend of the unimodal topic in the future, and the outcome gives the size of the discussion group day by day in the future;*
3. *The thresholds Q* and Q* handle cases such as long-term trend orientation of unimodal topic as discovered by previous works on real data on blogs or microblogs[9], and response to criteria on how to judge the long-term trend of unimodal topic diffusion.*

The output of the long-term unimodal topic diffusion trend prediction algorithm consists of the long-term trend of unimodal topic diffusion and the prediction size of the Discussion Group within the user-specified time range. It can vividly reflect the future popularity level of the unimodal topic diffusion. It is worth noting that the above unimodal topic prediction method is not suitable for predicting multimodal topic diffusion.

Short-term Multimodal Topic Diffusion Prediction

The short-term multimodal topic diffusion prediction method establishes a universal framework firstly proposed for topics based on the long-term unimodal topic diffusion trend prediction method. The key difference between the prediction methods for unimodal topic and for multimodal topic diffusion lies in that the method for unimodal topic can prejudge its long-term trend but for multimodal one can only prejudge its short-term orientation trend. Actually, the short-term multimodal topic diffusion prediction method also can predict the short-term trend of the unimodal topic diffusion.

After measuring numerous topics, we discovered that topic diffusion is stable when the time-update cycle is above ten days, as shown later in Subsection 6.2 and 6.3. Therefore, we choose the time-update cycle to be a ten-day period. The pseudocode for the short-term multimodal topic diffusion prediction method is shown as follows:

Algorithm 2: short-term multimodal topic diffusion prediction algorithm
Ensures: the prediction of the size of Discussion Group during the user-specified time interval
1. *It is crucial to capture appropriate time-window U in subsection 5.2, it got activated at the short-term trend of multimodal topics diffusion, and helps system (3.1) to perceive the continuous variations of the population size of multimodal topics;*
2. *We summate parameters of system (3.1) based on the raw data during the period of time-window U, and predict the size of the Discussion Group D(t) in the forecasting two days advanced availably.*

Although we cannot predict the long-term trend of multimodal topic in the future because of characteristics of multimodal topic, we can grasp the multistage orientation of unimodal subtopics constituting multimodal topics. In algorithm 2, the topic diffusion trend prediction process will terminate until the algorithm stopping conditions hold.

In our real experiments, the results show that the short-term unimodal as well as multimodal topic diffusion trend prediction method is much more stable and effective for predicting short-term population of the discussion group related to the multimodal topic without being supervised. However, the unimodal topic diffusion trend prediction method is more valid to predict the long-term popularity level of unimodal topic in the future.

Method Evaluations

In the following sections, we proposed three widely known parameters like the determined coefficient, precision, and F1 measure to evaluate the performance of our topic diffusion trend prediction methods. Then we report the evaluation results, and analyze the sensitivity effect of the fit time with predicting results of the size of the Discussion Group. Moreover, it is convict-able to compare our prediction methods with others in literature [25], [37] for validating the effectiveness and usefulness of our methods. Note that LUTDTPM is abbreviated as the unimodal topic diffusion trend prediction method; SMTDPM is for short-term multimodal topic diffusion prediction method respectively.

Method Evaluations

First, we propose the determined coefficient, precision, and $F1$ measures as follows and denoted as A, $P(K)$, and $F1$, respectively. These three quantities' metrics are subjective and neural in assessing the performance of our topic diffusion trend prediction methods. Adhoc measures such as precision, which are widely related to the confidence interval (CI), are the estimation of population parameters. furthermore, Adhoc measures can indicate the authority of prediction.

$$A = 1 - \frac{\sum_{i=1}^{n}(|N(i)^* - \overline{N}|)(|N(i) - N(i)^*|)}{\sum_{i=1}^{n}(N(i) - \overline{N})^2}$$

The determined coefficient (A) is defined as follows:
where $N(i)$ is the actual size of the Discussion Group D at time i; $N(i)^*$ is the prediction result of D at time i; $\overline{N}$ is the average size of the Discussion Group during the entire user-specified time interval for the observed topic; n is the total number of the time intervals.

In terms of confidence levels taken as 95 percent, we use the curve of the solution to the second equation in system (3.1) to estimate the size of the Discussion Group as to the unimodal topic based on method of the least residual within user-specified time interval. We use nonlinear retrieval method to catch both the upper confidence curve and the lower confidence curve. Then we calculate the proportion of the prediction size of the Discussion Group for a unimodal topic falling in the confidence interval compared to the whole prediction result of D. In addition, for evaluating the performance of the multimodal topic diffusion trend prediction method, we firstly determine the upper confidence curve and the corresponding lower one, the estimation terminates when all the spikes of the topic have been checked within the specified time interval. And we secondly calculate the proportion of the prediction size of the Discussion Group falling within the confidence intervals compared to the entire prediction size of the Discussion Group. We design the precision (P) of the estimation:

$$P = 1 - \frac{(N_{total} - N_{in})^2}{(N_{total} - \bar{N}_{in})^2}$$

where N_{in} denotes the prediction result of D in the confidence interval; N_{total} and is the prediction size of the Discussion Group during the entire user-specified time interval of the specified topic, $\bar{N}_{in}$ is the average size of the Discussion Group during the

$$K = \begin{cases} 1 - \frac{|M_{PSV} - M_{TSV}|}{M_{TSV}}, if \ |M_{PSV} - M_{TSV}| \le M_{TSV}; \\ 0, otherwise. \end{cases}$$

entire user-specified time interval. The prediction precision of the time epoch of the Discussion Group is defined as following expression:

where M_{TSV} is the actual maximum result size of D; M_{PSV} is the maximum prediction result of D. As we know that precision measurement alone is not sufficient enough to assess the performance of the hot online topic diffusion trend prediction algorithms. A high method prediction precision signifies that we can predict the size of the Discussion Group more accurately with an acceptable confidence level. The measure ($F1$) is designed as the harmonic value of A and P as below:

$$F1 = \frac{2PA}{P + A}$$

Effect of Fit Time on Proposed Methods

For explaining how to choose appropriate fit time of training system (3.1), we should analyze the effectiveness of fit time on the accuracy of our prediction methods by measuring evaluating parameters including the determined coefficient (A), precision (P), and $F1$ because fit time directly determines whether the data during the period of fit time contain the information which can reflect the almost all trend of topic diffusion. Either the long-term unimodal topic diffusion trend prediction method or short-term multimodal topic diffusion trend prediction method is based on system (3.1) which is implemented using different fit times (or time-update cycles (e.g.., 5, 10, 15, 20, 25, and 30). Table 2 shows results of determined confidence, table 3 and table 4 show the corresponding precisions values and $F1$ measures under different data sets with respect to the algorithm configuration settings, respectively. In the following, we use the topic on the Sina blog as an example to reveal and validate our results.

Firstly, by analyzing the variation of determined coefficient with fit time we can understand how the fit time influences the value of determined coefficient from one side. The highest determined coefficient (over different fit time settings) for the long-term unimodal topic diffusion trend prediction method is 92.6 % when the fit time is set as 35 days. The highest determined coefficient of the experiment based on the short-term multimodal topic diffusion trend prediction method is 91.5% when 35-day was used as the time-update cycle for training system (3.1). In Table 2, for the unimodal topic, determined coefficient of the long-term

Table 2: the Prediction result of D Based on the LUTDTPM and the SMTDPM

Items	5	10	15	20	25	30	35
LUTDTPM	0.254	0.317	0.521	0.857	0.897	0.895	0.901
SMTDPM	0.751	0.883	0.885	0.889	0.891	0.894	0.913

Table 3: Precision Measures and Spike Precisions of Prediction Based on the LUTDTPM and the SMTDPM

Items		5	10	15	20	25	30	35
LUTDTPM	*P*	0.241	0.270	0.432	0.810	0.873	0.876	0.879
	K	0.382	0.464	0.516	0.953	0.981	0.983	0.986
SMTDPM	*P*	0.691	0.864	0.867	0.870	0.872	0.876	0.912
	K	0.579	0.815	0.827	0.874	0.962	0.968	0.972

unimodal topic diffusion trend prediction method and the short-term multimodal topic diffusion prediction method slightly oscillate above twenty days and ten days with increasing fitting time, respectively. Furthermore, before the size of the discussion group of the unimodal topic reaches its peak number, determined coefficient of the unimodal topic diffusion prediction method is caught to be very sensitive to fit time. This strong sensitivity tells us how long the fit time should be taken as data preparation interval for predicting the long-term trend of unimodal topic diffusion. For other unimodal topics, this feature works still successful and fully held.

Based on the short-term multimodal topic diffusion trend prediction method, the determined coefficient stays relatively stable from about the 10th day for almost all topics. It is helpful to determine the suitable time-update cycle. The best precision for the long-term unimodal topic diffusion trend prediction method is 87.9% when fit time is set as 35 days. The precision is above81%, and the topic spike precision above 95% starting from the 20th day. Before the topic spike time (the time when the size of the Discussion Group reaches its maximum), the precision had oscillated somewhat strongly. Secondly, by analyzing the variation of precision with fit time we can understand how the fit time influences the value of precision from other side. The best precision (P) for the short-term multimodal topic diffusion trend prediction method is 91.2% within 35 days; however, starting from the 10th day, the precision stayed above 86.4% and the spike precision above 81.5%. For almost all topics taken here, this feature is still successful and helpful to ensure the time-update cycle.

The predicting values in Table 3 are designed as the ratio of the prediction size of D entering the confidence interval compared to the whole prediction size of D. Among all the experimental scenarios, it is not necessary for the long-term unimodal topic diffusion trend prediction method to choose the fit time matching with the best precision measure to prepare and calibrate raw data for predicting topic diffusion trend. By considering time efficiency, better precision with a shorter fit time also can determine the trend of unimodal topic diffusion. In Table 4, we randomly choose ten unimodal topics from the Sina blog, the LQQM BBS, the

Table 4: Topic Diffusion Long-term Trend Prediction Based on the LUTDTPM

Items		T^1	T^2	T^3	T^4	T^5	T^6	T^7	T^8	T^9	T^{10}
LUTDTPM	*UB*	0	0	0	0	0	0	0	1	0	1
	P	0	0	0	0	0	0	0	1	0	1
	LB	0	0	0	0	0	0	0	1	0	1

Note: Ti denotes topic: T^1-T^4 on BMY BBS, T^6-T^7 on LQQM BBS, and T^5 and T^{10} on the Sina blog; T^8 andT^9 on Sina microblog; "0" as dying out; "1" as uniformly but weakly persistent.

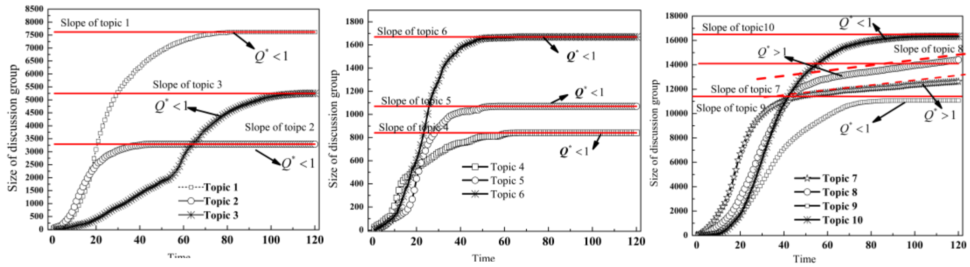
Table 5: the Prediction result of $F1$ Measures Based on the LUTDTPM and SMTDPM

Items	5	10	15	20	25	30	35
LUTDTPM	0.269	0.385	0.434	0.911	0.923	0.916	0.926
SMTDPM	0.822	0.902	0.904	0.908	0.910	0.912	0.915

BMY BBS and Sina microblog as test datum to show prediction results of the long-term trend of the unimodal hot topic diffusion. Different unimodal topics have different fit times to train system (3.1). Although it is important to obtain precise prediction values, it is counterproductive that too much time is spent on preparing raw data. If the P value at some time is close to the best P, we should sacrifice some precision for the sake of the entire with fit time we can understand how the fit time influences the value of $F1$ measure, which remedies the deficiency of both the determined coefficient (A) and precision (P). Raw data within the fit time interval can determine the long-term trend of unimodal topic diffusion. As shown in requirement including reducing complexity of computation and avoiding over-fitting. Firstly, by analyzing the variation of $F1$ measure.

Table 5, the best $F1$ measure using the long-term unimodal topic diffusion trend prediction method is 90.1% and the best $F1$ value from the short-term multimodal topic diffusion prediction method, but if the difference between the best $F1$ value and other $F1$ values are slight for the long-term unimodal topic diffusion trend prediction method, we choose an appropriate fit time for training system (3.1). As observed from most topics, the performance of precision based on the multimodal topic diffusion trend prediction is 91.3% with 35-days as the update time-cycle. diffusion trend prediction method is more stable than that of the unimodal topic diffusion trend prediction method when the fit time interval changes from ten days to the user-specified topic lifetime to train system (3.1). In summary, the performance of the multimodal topic diffusion prediction method for short-term prediction is better than that of the unimodal topic diffusion prediction method and most topics assign ten days as time-update cycle, see in Fig.5 shows the topic diffusion trends of ten topics as unimodal topic in real examples as before using the visualizable method as shown in Fig.4 in Subsection 4, the x-axis shows the total counted days numbered here as 120 days, while the y-axis shows the time cumulative size of the discussion group of target topics.

To better show and understand, we divide the ten topics into three groups with three topics



(a) (b) (c)

Figure 5: Diffusion trends of ten unimodal topics

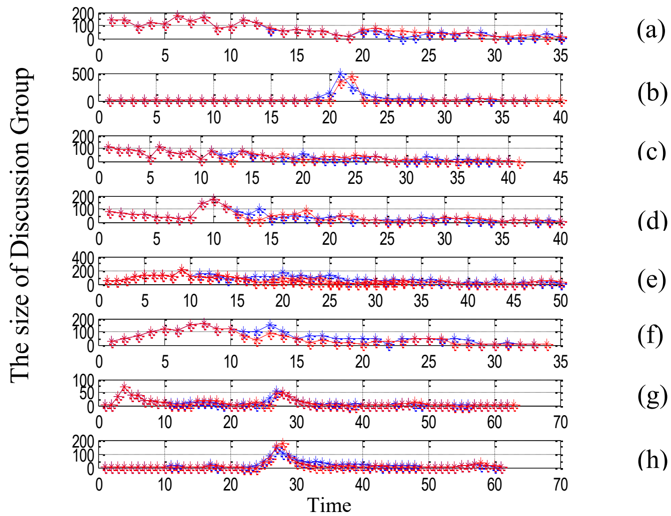


Figure 6: Size of the Discussion Group between the real and prediction values——* real data ; —— * predicting data; topic (a), (b), (d), (c), (e) ,(f) , (g) and (h) from blogs.

As one group, in the Fig.5 (a), we show that the increment ratios of time cumulative size of the discussion group of topic 1, topic 2, and topic 3 tend to zero from which we know the diffusion trends of topic 1, topic 2, and topic 3 tend to be dying out; in the Fig.5 (b), we show the increment ratios of time cumulative size of the discussion group of topic 4, topic 5, and topic 6 tend to zero from which we know that the diffusion trends of topic 4, topic 5, and topic 6 tend to be dying out; in the Fig.5 (c), we show the increment ratios of time cumulative size of the discussion group of topic 7 and topic 8 tend to some positive certain value, and that of topic 9 and 10 tend to zero, and thus logically we know that the diffusion trends of topic 7 and topic 8 tend to be uniformly but weakly persistent, while the diffusion trends of topic 9 and 10 tend to be dying out.

We took 63 topics either unimodal or multimodal as corresponding test scenarios, 89% topics' A, $P(K)$ and $F1$ are not sensitive to the time-update cycle after the 10th day based on the multimodal topic diffusion prediction method. Moreover, 96% of unimodal topics' parameters referring as A, $P(K)$, and $F1$ are sensitive to the fit time before the peak of the Discussion Group size based on the unimodal topic diffusion prediction method. This confirms the above conclusions about the overall performance of the unimodal topic diffusion trend prediction method for predicting the long-term diffusion trend of unimodal topic, moreover, about the overall performance of the multimodal topic diffusion prediction method for predicting the short-term population size of all topic diffusion. Although they provide significant approaches toward understanding topic diffusion patterns and the superiority of predicting the short-term and long-term trend of hot online topic diffusion, it is insufficient for providing a quick prediction in time at the level of the new relative event. This limitation leads to three problems: first problem is that there possibly exists a time discrepancy between the actual and the predicted spike time for some topics; another problem, we can remove this delay, if we understand how the new event affected the previous topic diffusion in

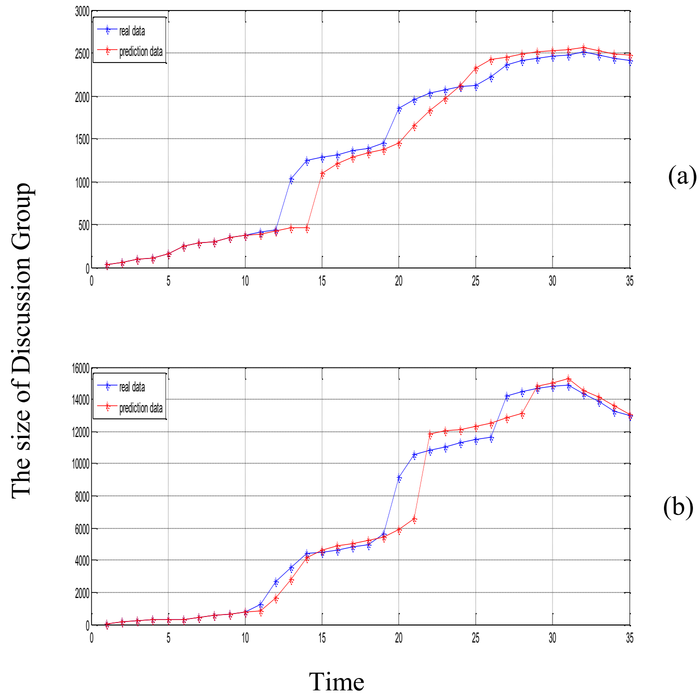


Figure 7: The size of D between real values and prediction: topic (a) and (b) from the Sina micro blogs

time. The solution to the second problem requires the improvement of Natural Language Understanding and distinguishing the appearance of a new spike in the whole data. For the third problem, we combine individual behavior with group behavior to establish reasonable topic diffusion model.

In summary, to conduct a comparative full-scale landscape to predict the long-term trend of unimodal topic diffusion, we take fully advantage of the features of both precision and the spike precision measure to explain how to choose an appropriate fit time. In the beginning, we randomly take 41 unimodal topics as test scenarios, and for most of these topics, the precision value oscillates only slightly from the topic spike time to the end of the time interval, with a spike precision of more than 82%. However, before the time epoch corresponding to the peak

Table 6: Multi-stage trend predictions of the eight topics based on SUMTPPM

Topic	***Websites***	**Sign{[The number of participants from *i*th day to (*i*+9)th day]−[The number of participants from (i+10)th day to (*i*+19)th day]}**					
		Real			***Predicting***		
		11−20	**21−30**	**31−40**	**11−20**	**21−30**	**31−40**
T^1	*BBS 2*	−	−	−	−	−	−
T^2	*BBS 2*	−	−	+	−	−	+
T^3	*BBS 1*	−	−	−	−	−	−
T^4	*Blog*	+	+	−	+	+	−
T^5	*BBS 1*	−	+	+	−	+	+
T^6	*Blog*	−	−	−	−	−	−
T^7	*Blogs*	−	−	−	−	−	−
T^8	*Microblog*	−	−	−	−	−	−

Note: Sign "−" denotes the number of participants from the (i+10)th day to the (i+19)th day is less than the number of participants from the ith day to the (i+9)th day and sign "+" denotes the number of participants from the (i+10)th day to the (i+19)th day is more than the size of D from the ith day to the (i+9)th. BBS 1 as BMY BBS, BBS 2 as LQQM BBS; Blog as the Sina blog; Microblog as Sina microblog.

Table 7: Methods comparisons of LUTDTPM, SMTDPM, IEPPA, IDCPM, RNEPPPA and RNNM

Items	*LUTDTPM*	*SMTDPM*	*IEPPA*	*IDCPM*	*RNEPPPA*	*RNNM*
Trend pattern	*Yes*	*No*	*No*	*No*	*No*	*No*
extinction	*Yes*	*No*	*No*	*No*	*No*	*No*
persistent existence	*Yes*	*No*	*No*	*No*	*No*	*No*
non-supervision	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*Yes*
UT3PM	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
MT3PM	*No*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Based on Cascade model	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*

UT3PM: unimodal topic diffusion population prediction method, MT3PM: multimodal topic diffusion population prediction method.

Value, the spike precision oscillates somewhat strongly. And next if the prediction maximum size of the Discussion Group continues to oscillate slightly around some value for at least ten days with an increasing variation of the fit time, thus we choose any of the last five days with previous days as the fit time interval. For example, 27-days are chosen as fit time for the unimodal topic in Table 2 because the long-term diffusion trend information of topic was embedded in raw data obtained within the previous the 27th days. In the final, we exhibit the corresponding results for predicting the long-term trend and the size of the Discussion Group in the future about topic in Table 6, and show the results of predicting short-term diffusion populations' size of topic based on the short-term multimodal topic diffusion prediction method in both Fig. 6 and Fig. 7.

Validation of Methods' Effectiveness and Efficiency

To validate our methods' effectiveness and efficiency, we aim to compare the application scope of the topic diffusion trend prediction, the prediction precision including the determined coefficient and precision among the unimodal topic diffusion trend prediction method (abbreviated as LUTDTPM), short-term multimodal topic diffusion prediction method (abbreviated as SMTDPM), information-dependent embedding-based propagation prediction algorithm (abbreviated as IEPPA), repost network embedding-based propagation popularity prediction algorithm (abbreviated as RNEPPPA), recurrent neural network model (abbreviated as RNNM) and conditional probability model of information diffusion (abbreviated as IDCPM) [29, 30]. IEPPA was used to deduce the information propagation in the latent embedding space according to propagation prediction model depending on information. RNEPPPA was applied for precisely predicting the information diffusion popularity only

Table 8: The proportion of short-term prediction precision comparison with LUTDTPM, SMTDPM, IEPPA, IDCPM, RNEPPPA and RNNM

Items	*LUTDTPM*	*SMTDPM*	*IEPPA*	*IDCPM*	*RNEPPPA*	*RNNM*
Proportion (UPA)/E	≈66%	≈88%	≈21%	≈78%	≈67%	≈79%
Proportion (UPP)/E	≈64%	≈85%	≈19%	≈74%	≈70%	≈80%
Proportion (MPA)/E	≈45%	≈89%	≈64%	≈73%	≈75%	≈77%
Proportion (MPP)/E	≈43%	≈90%	≈61%	≈76%	≈78%	≈74%

PA: performance of A, PP: performance of P: unimodal, U: unimodal, M: multimodal, E: the total number of topics

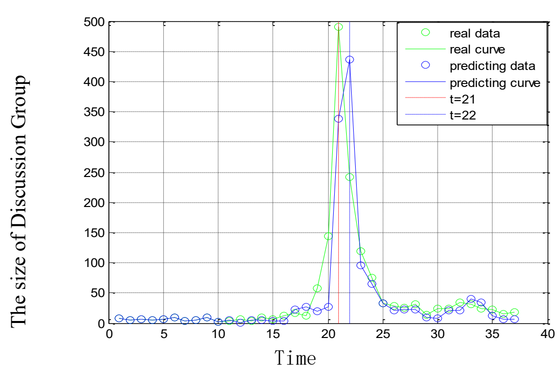


Figure 8: Topic spike times between actual and prediction value

based on early repost information and proposed repost network to vector. RNNM was designed to predict the next user with the latest cascade representation according to potential relationship among users based on social network through graph attention network. IDCPM was proposed to describe information propagation in heterogeneous information networks depending on path among nodes passing through various layers of a heterogeneous information network.

From TABLE.7, IEPPA only presented the population size prediction of the multimodal topic diffusion and population size but no prediction of unimodal topic diffusion trend pattern in an unsupervised way. Although IDCPM can be used to calculate the active degree of all nodes respectively, it cannot predict the topic population propagation trend. Though RNEPPPA captured different repost behaviors and predicted repost network embedding-based diffusion popularity, it failed to judge whether the topic's long-term diffusion trend is extinction or persistent existence yet. RNNM constructed a seq2seq framework to learn the spatial-temporal cascade features and predicted the next user with the latest cascade representation except forecasted the long-termed trend of topic diffusion. LUTDTPM presented detailed work about almost each item but no multimodal topic diffusion population prediction is mentioned, however, SMTDPM overcome this deficiency of LUTDTPM which fully shows the effectiveness of our methods.

From TABLE 8, we show that SMTDPM have better performance including determined coefficient and precision than other four methods for predicting the short-term topic diffusion process of no matter the unimodal topic or multimodal topic, however, the performance of LUTDTPM is not unsatisfactory because its function is to predict the long-term diffusion trend of unimodal topic. IEPPA, IDCPM, RNEPPPA and RNNM can predict the short-dated population size of topic diffusion with a certain accuracy even though their performance for predicting the short-term topic diffusion process is not as good as that of SMTDPM. Other sides, we compare the real and predicting spike times when the size of D reaches its maximum in order to measure the error of our methods. For example, for our target topic, there exists a one-day time delay between the actual and predicted spikes time epoch when using the short-term multimodal topic diffusion prediction method, as shown in Fig. 8 because the event happens quickly, there is not enough time for the raw data to catch the corresponding information of the near enough event. We took 63 topics as test scenarios based on the short-term multimodal topic diffusion prediction method, and for most topics there is only one day time delay at most between the actual and prediction spike times. However, if we are wild about quickly predicting the spike future. In all, it is definitely better bet for the long-term trend of unimodal topic diffusion prediction method when we are eager for understanding the future rise or fall long-term trend of unimodal topic diffusion. If we look forward for predicting the short-term topic diffusion process, the short-term multimodal topic diffusion prediction method can be more sufficiently qualified for the job than other methods or algorithms.

Conclusions and Future Work

With the popularity and overwhelming increase of smart mobile sensor device, people share information more easily across different online social network. So it is urgent and important to search for a universal online topic diffusion models for both needs of theoretical research and the demanding and realistic applications. Furthermore, the model has wide and high potentiality to be used in finance and marketing for effective assessment and analysis of commercial activity with the development and maturity of big data theory. Although previous approaches in topic diffusion modeling have resulted in some promising methods for the topic diffusion description and mathematical formulation in single online social network, they still failed to predict the diffusion characteristics all types of topics, especially multimodal topics, in heterogeneous online social network.

The outstanding contribution of our work is the development of novel universal method for predicting topic diffusion based on a new topic diffusion model in heterogeneous online social network and also its wide applications in prediction the topic diffusion trend. More specifically speaking, one obvious aspect of our contribution is the broad and successful use of a time-dependent epidemic-like propagation model to describe topic spreading process in heterogeneous online social network, which is a joint research area of epidemic dynamics and social networks and hot focus nowadays. Our method seamlessly expands the time-dependent diffusion model to the hot online topic diffusion modeling. The usefulness and validation of our methods are assessed through experimental evaluations and methods' results comparison with historical works in literature. Another worth-pointing aspect of our contribution is the proposal of a universal method based on system (3.1) which later divided into two separate topic diffusion trend prediction approaches in heterogeneous online social network due to topics' properties not to be neglected. Firstly, we design the approach to predict the long-term tendency of unimodal topic diffusion.

The long-term unimodal topic diffusion trend prediction method is underpinned by trend threshold in two propositions. Secondly, based on the long-term unimodal topic diffusion trend prediction method, the short-term multimodal topic diffusion prediction method accomplishes the goal of predicting short-term population diffusion for most topics. Using the experimental results of determined coefficient (A), precision (P), and F1 measures, we validate that the long-term unimodal topic diffusion trend prediction method is useful, and the short-term multimodal topic diffusion prediction method is successful for predicting the long-term trend of topic diffusion and predicting the short-term discussion group size of online topics in heterogeneous online social network. In our future work, we will endeavor to investigate the potential usability of our algorithms on other online forums. Moreover, we will develop model to describe the mutual interactive relationship between the individual behavior of the leader and group behavior of other participants. And we will study the effecting performance of new events during the topic diffusion.

Acknowledgments
This job is Supported by Natural Science Foundation of Shaanxi Province of China(No.2021JM-344). This work is supported by the independent research project of Shaanxi Provincial Key Laboratory of Network Computing and Security Technology (No.NCST2021YB-05), by the Key Research and Development Program of Shaanxi Province (No.2018ZDXM-GY-036).

References

1. Shiekh, M. A., Sharma, K., & Ganai, A. H. (2020). Information Diffusion: Survey to Models and Approaches, a Way to Capture Online Social Networks. In Intelligent Data Communication Technologies and Internet of Things: ICICI 2019 (pp. 25-32). Springer International Publishing.
2. Daley, D. J., & Kendall, D. G. (1964). Epidemics and rumours. Nature, 204, 1118-1118.
3. Moreno, Y., Nekovee, M., & Pacheco, A. F. (2004). Dynamics of rumor spreading in complex networks. Physical review E, 69(6), 066130.
4. McDonald, M., Suleman, O., Williams, S., Howison, S., & Johnson, N. F. (2008). Impact of unexpected events, shocking news, and rumors on foreign exchange market dynamics. Physical Review E, 77(4), 046110.
5. Zhao, L., Yuan, R., Guan, X., & Li, M. (2009). Propagation modeling and analysis of incidental topics in blogosphere. In Online Communities and Social Computing: Third International Conference, OCSC 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Proceedings 3 (pp. 401-410). Springer Berlin Heidelberg.
6. Fan, C., Jiang, Y., Yang, Y., Zhang, C., & Mostafavi, A. (2020). Crowd or hubs: information diffusion patterns in online social networks in disasters. International journal of disaster risk reduction, 46, 101498.
7. Yi, Y., Zhang, Z., Yang, L. T., Gan, C., Deng, X., & Yi, L. (2020). Reemergence modeling of intelligent information diffusion in heterogeneous social networks: the dynamics perspective. IEEE Transactions on Network Science and Engineering, 8(2), 828-840.
8. Zanette, D. H. (2002). Dynamics of rumor propagation on small-world networks. Physical review E, 65(4), 041908.
9. Zhou, Y., Guan, X., Zhang, Z., & Zhang, B. (2008, June). Predicting the tendency of topic discussion on the online social networks using a dynamic probability model. In Proceedings of the hypertext 2008 workshop on Collaboration and collective intelligence (pp. 7-11).
10. Zeng, J., Cheung, W. K., & Liu, J. (2012). Learning topic models by belief propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(5), 1121-1134.
11. Li, C. T., Lin, Y. J., & Yeh, M. Y. (2018). Forecasting participants of information diffusion on social networks with its applications. Information Sciences, 422, 432-446.
12. Su, Y., Zhang, X., Wang, S., Fang, B., Zhang, T., & Yu, P. S. (2019). Understanding information diffusion via heterogeneous information network embeddings. In Database Systems for Advanced Applications: 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22–25, 2019, Proceedings, Part I 24 (pp. 501-516). Springer International Publishing.
13. Li, J., Peng, W., Li, T., Sun, T., Li, Q., & Xu, J. (2014). Social network user influence sense-making and dynamics prediction. Expert Systems with Applications, 41(11), 5115-5124.
14. Gao, S., Pang, H., Gallinari, P., Guo, J., & Kato, N. (2017). A novel embedding method for information diffusion prediction in social network big data. IEEE Transactions on Industrial Informatics, 13(4), 2097-2105.
15. Leskovec, J., Backstrom, L., & Kleinberg, J. (2009, June). Meme-tracking and the dynamics of the news cycle. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 497-506).
16. Yang, J., & Leskovec, J. (2011, February). Patterns of temporal variation in online media. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 177-186).
17. Szabo, G., & Huberman, B. A. (2010). Predicting the popularity of online content. Communications of the ACM, 53(8), 80-88.
18. Figueiredo, F., Benevenuto, F., & Almeida, J. M. (2011, February). The tube over time: characterizing popularity growth of youtube videos. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 745-754).
19. Barbieri, N., Bonchi, F., & Manco, G. (2013). Topic-aware social influence propagation models. Knowledge and information systems, 37, 555-584.
20. Matsubara, Y., Sakurai, Y., Prakash, B. A., Li, L., & Faloutsos, C. (2012, August). Rise and fall patterns of information diffusion: model and implications. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 6-14).
21. Tong, N., Tang, Y., Chen, B., & Xiong, L. (2021). Representation learning using Attention Network and CNN for Heterogeneous networks. Expert Systems with Applications, 185, 115628.
22. Shang, J., Huang, S., Zhang, D., Peng, Z., Liu, D., Li, Y., & Xu, L. (2021). RNe2Vec: information diffusion popularity prediction based on repost network embedding. Computing, 103, 271-289.
23. Chen, Z., Wei, J., Liang, S., Cai, T., & Liao, X. (2021). Information cascades prediction with graph attention. Frontiers in Physics, 9, 739202.
24. Wan, P., Wang, X., Lin, Y., & Pang, G. (2021). A knowledge diffusion model in autonomous learning under multiple networks for personalized educational resource allocation. IEEE Transactions on Learning Technologies, 14(4), 430-444.
25. Yang, C., Sun, M., Liu, H., Han, S., Liu, Z., & Luan, H. (2019). Neural diffusion model for microscopic cascade study. IEEE Transactions on Knowledge and Data Engineering, 33(3), 1128-1139.
26. Wang, X., Wang, X., Min, G., Hao, F., & Chen, C. P. (2020). An efficient feedback control mechanism for positive/nega-

tive information spread in online social networks. IEEE transactions on cybernetics, 52(1), 87-100.
27. Sun, E., Rosenn, I., Marlow, C., & Lento, T. (2009, March). Gesundheit! modeling contagion through facebook news feed. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 3, No. 1, pp. 146-153).
28. Thieme, H. R. (1993). Persistence under relaxed point-dissipativity (with application to an endemic model). SIAM Journal on Mathematical Analysis, 24(2), 407-435.
29. Molaei, S., Babaei, S., Salehi, M., & Jalili, M. (2018). Information spread and topic diffusion in heterogeneous information networks. Scientific reports, 8(1), 9549.
30. Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2019). The dynamic embedded topic model. arXiv preprint arXiv:1907.05545.