Physica A 419 (2015) 55–63

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

An information propagation model considering incomplete reading behavior in microblog

Qiang Su [a], Jiajia Huang [a,*], Xiande Zhao [b]

[a] Economics & Management School, Tongji University, Shanghai, 200092, PR China

[b] Institute of Supply Chain Integration and Service Innovation, South China University of Technology, Wushan Road, Guangzhou, 510641, PR China

HIGHLIGHTS

- The microblog propagation is studied considering incomplete reading behavior.
- An Mb-SIR model is proposed and validated with the actual data.
- The sensitivity and importance of different parameters are evaluated.
- Some management insights and information propagation control measures are derived.

ARTICLE INFO

Article history:
Received 14 November 2013
Received in revised form 13 April 2014
Available online 15 October 2014

Keywords:
Microblog
Information propagation
Incomplete reading behavior
SIR model

ABSTRACT

Microblog is one of the most popular communication channels on the Internet, and has already become the third largest source of news and public opinions in China. Although researchers have studied the information propagation in microblog using the epidemic models, previous studies have not considered the incomplete reading behavior among microblog users. Therefore, the model cannot fit the real situations well. In this paper, we proposed an improved model entitled Microblog-Susceptible–Infected–Removed (Mb-SIR) for information propagation by explicitly considering the user's incomplete reading behavior. We also tested the effectiveness of the model using real data from Sina Microblog. We demonstrate that the new proposed model is more accurate in describing the information propagation in microblog. In addition, we also investigate the effects of the critical model parameters, e.g., reading rate, spreading rate, and removed rate through numerical simulations. The simulation results show that, compared with other parameters, reading rate plays the most influential role in the information propagation performance in microblog.

1. Introduction

Microblogging is a broadcast medium that exists in the form of blogging. A microblog differs from a traditional blog in that its content is typically smaller in both actual and aggregated file size. Microblogs allow users to exchange small elements of content such as short sentences, individual images, or video links [1]. Microblog is an effective tool for social exchanges, real-time news updates and even for advertisements of new products and services.

Microblog users can post news and information easily and the information will be seen by its followers quickly. The followers can reply to the original message or repost the message conveniently. Therefore, hundreds of thousands of instant

* Corresponding author.
E-mail addresses: huangjiajia1023@163.com, jiajiasunny@sina.com (J. Huang).

http://dx.doi.org/10.1016/j.physa.2014.10.042
0378-4371/

messages can be released on microblog every minute. With the widespread applications of the internet, microblogs are playing more important role in information dissemination and social exchanges. When some specific event occurs, even an ordinary person can induce a large number of retweets, leading to a hot topic of discussion in microblog community. On the one hand, if a rumor is released for malicious purposes, it can be disseminated quickly through the microblog without verification, leading to harmful effects to some individuals and/or organizations. In this situation, we should take quick actions to reduce or stop its propagation in order to reduce the harmful effects. On the other hand, we may want to spread the positive news and events to as many followers as possible. In both cases, it is important to understand the mechanisms of information propagation in microblogs, so that we can design and implement measures to effectively boost or restrain the propagation process.

A growing body of literature attempts to understand the propagation dynamics in social interaction networks through the online and offline channels. Many researchers studied the offline propagation dynamics using Galam models [2], epidemic models [3–5], game theory [6] and other theories [7–10]. However, with the increased popularity of Internet, online social networks become the principal information propagation carriers. Cheong et al. [11] reviewed research on social network and divided related research into two domains: the user domain and the message domain. The user domain studies focus on usage behaviors [12–15] such as forwarding, commenting and posting. The message domain studies focus on the content and propagation behaviors of the messages. For instance, Zhao et al. [16] adopted quantitative methods, with consideration of users, media and content in microblog to analyze the characteristics of hot topics propagation and describe the statistics result such as frequent words and emotional factors in hot topics. Lloyd et al. [17] reviewed the application of epidemic models in virus spread among computers and people. Moreover, some researchers discussed the spreading of malware [18], news [19] and hot topics [20–22] on Twitter.

Compared with other approaches, the epidemic model is more suitable for describing the microblog information spreading. As people get infection of epidemic diseases through the contact with other infected persons, the microblog users get notice of message through reading microblogs. As the outbreak of epidemic disease, the microblog information will be spread out by the informed users and make a range of the public involved rapidly. Due to the similar propagation mechanism between epidemic diseases and online social networks, epidemic models have been widely used in revealing the propagation dynamics of online social networks. Yan et al. [23] proposed an extended Susceptible–Infected (SI) propagation model to incorporate bursty human behavior and limited attention on information propagation in online social networks. Zhao et al. [24,25] constructed a Susceptible–Infected–Removed (SIR) model to discuss information propagation dynamics in new media considering the forgetting mechanism. Zhao et al. [26] analyzed online social network spread by proposing a Susceptible–Infected–Hibernator–Removed (SIHR) model, making a direct link from ignorants to stiflers or hibernators. Xiong et al. [27] established an SCIR diffusion model with four possible states: susceptible, contacted, infected and refractory. Their propagation model ends up with states of infected and refractory. Cheng et al. [28] introduced a stochastic epidemic model to describe rumor diffusion, with the infectious probability depending on the strength of ties.

Although a number of studies have investigated the information propagation using the epidemic theories, none of the studies considered the incomplete reading behavior in microblog. More specifically, only a proportion of the poster's followers can find and read the microblog message. This reading rate will influence the microblog propagation significantly. Nevertheless, the current existing epidemic models do not take this factor into consideration. Under this circumstance, we propose a novel model named Microblog-Susceptible–Infected–Removed (Mb-SIR model). Taking Sina Microblog, one of the largest microblog platforms in China, as an example, a series of numerical simulation experiments are conducted to validate the effectiveness of the Mb-SIR.

The remaining part of this paper is organized as follows. In Section 2, the traditional SIR model is analyzed and the Mb-SIR model is constructed considering the incomplete reading behavior in microblog community. In Section 3, the verification case data are described, and the new proposed Mb-SIR model is validated using the case data. In Section 4, the key parameters on propagation are analyzed through numerical simulations and some managerial insights are derived. Finally, Section 5 outlines the conclusions and provides some directions for further study.

2. The Mb-SIR model considering incomplete reading behavior

2.1. The traditional SIR model

The epidemic model is proposed to explore the mechanism of infections propagation, predict the future course of an outbreak and evaluate strategies to control an epidemic [29]. Generally speaking, two types of models are used to analyze the infection of diseases, i.e. stochastic model and deterministic model. The stochastic model predicts the future outbreak based on risk of exposure and characteristics of the disease and other illness dynamics [30]. The model can be used to predict the propagation of diseases through simulations with sufficient data. The deterministic model is easier to construct and requires less data and describe the evolution at the aggregate level. As one of the most popular deterministic epidemic models, SIR model [31] assumes that individuals are infectious as soon as infected. Meanwhile, there is a chance for the infected to recover and cease to spread disease with permanent immunity. This model categorizes population into three segments: Susceptible, Infected and Removed. The spreading process involves the change of an individual from Susceptible (S) to Infected (I), then to Removed (R).