Physica A 463 (2016) 202–218

Contents lists available at ScienceDirect

Physica A

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

Characterizing super-spreading in microblog: An epidemic-based information propagation model

Yu Liu [a,*], Bai Wang [a], Bin Wu [a], Suiming Shang [b], Yunlei Zhang [a], Chuan Shi [a]

[a] Beijing Key Laboratory of Intelligence Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China

[b] School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China

HIGHLIGHTS

- Information propagation on OSNs is studied considering super-spreading phenomenon.
- A SAIR model is proposed to characterize the information propagation with Weibo data.
- Super-spreaders in information propagation on OSNs are identified and characterized.
- The sensitivity of parameters depicting super-spreading phenomenon is analyzed.

ARTICLE INFO

Article history:
Received 22 January 2016
Received in revised form 25 May 2016
Available online 22 July 2016

Keywords:
Information propagation regularity
Epidemic-based model
Super-spreading
Super-spreader
Social media
Microblog

ABSTRACT

As the microblogging services are becoming more prosperous in everyday life for users on Online Social Networks (OSNs), it is more favorable for hot topics and breaking news to gain more attraction very soon than ever before, which are so-called "super-spreading events". In the information diffusion process of these super-spreading events, messages are passed on from one user to another and numerous individuals are influenced by a relatively small portion of users, a.k.a. super-spreaders. Acquiring an awareness of super-spreading phenomena and an understanding of patterns of wide-ranged information propagations benefits several social media data mining tasks, such as hot topic detection, predictions of information propagation, harmful information monitoring and intervention. Taking into account that super-spreading in both information diffusion and spread of a contagious disease are analogous, in this study, we build a parameterized model, the SAIR model, based on well-known epidemic models to characterize super-spreading phenomenon in tweet information propagation accompanied with super-spreaders. For the purpose of modeling information diffusion, empirical observations on a real-world Weibo dataset are statistically carried out. Both the steady-state analysis on the equilibrium and the validation on real-world Weibo dataset of the proposed model are conducted. The case study that validates the proposed model shows that the SAIR model is much more promising than the conventional SIR model in characterizing a super-spreading event of information propagation. In addition, numerical simulations are carried out and discussed to discover how sensitively the parameters affect the information propagation process.

* Corresponding author.
E-mail addresses: liuyu@bupt.edu.cn (Y. Liu), wangbai@bupt.edu.cn (B. Wang), wubin@bupt.edu.cn (B. Wu), m13474069802@163.com (S. Shang), yunlei0518@126.com (Y. Zhang), shichuan@bupt.edu.cn (C. Shi).

http://dx.doi.org/10.1016/j.physa.2016.07.022
0378-4371/

1. Introduction

Weibo[1], launched by Sina in August 2009, is a Twitter-like microblogging service. On this online social network (OSN), users can obtain breaking news and other trending information, dig their interests and activities, share personal opinions and argue with other users, update statues and establish relationships, etc. During the following years since the foundation of Weibo, there has been an explosive growth of microblog users and the contents (tweets and retweets) generated by users on this microblog platform. Tweets are posted and retweeted by Weibo users, forming cascades [1]. In the way that one microblogger posts a tweet and others repost it, information spreads in a considerable extent on online social networks.

Despite a significant amount of work on models, explanations, and patterns of information propagation on online social media, the dynamic underlying super-spreading events on OSNs remains uncertain. Towards developing a better understanding of super-spreading on OSN, in this paper, we aim to build a mathematical model to discover how to describe and characterize super-spreading of information propagation on OSNs.

According to studies on information propagation on online social networks [2–4], firsthand news reports, bizarre anecdotes as well as rumors awaiting clearance mostly disseminate like super-spreading epidemic events. In the circumstance of follower–followee relationship between microblog users, one user transmits a message to his followers simply by reposting this message, and then the information is distributed instantaneously point-to-surface. After the awareness of some non-ordinary stories, various behaviors of participants' on OSNs lead to many kinds of information diffusion patterns: a small group of users remain silent and just being onlookers; the majority will pass the news on, in the same way that they are influenced by each other and this behavior is mocked, recognized as Herd Behavior [5]; meanwhile, only a very few individuals will transmit the news and influence a very vast coverage of crowds of users than normal users would do, and this minority should be referred to as super-spreaders, and this kind of information propagation should be viewed as a super-spreading event. In most cases, super-spreaders are generally opinion leaders, or those who have quantities of followers on microblog platforms, though sometimes not vice versa.

The diffusion of information on OSNs bears some similarity to an infectious disease spread [6,7]. Although the probability that one individual user would retweet a piece of message relates to his interests and intention, crowds of users would not resist telling others about uncommon stories from a macroscopic perspective. This pattern corresponds to an individual's infection caused by a contagion, who contacted by an infectious person will get infected, even though some are immune to such a contagion. The emerging super-spreaders and the consequent super-spreading phenomenon in a social media scheme are also both analogous to the respective connotations in the field of epidemiology. In epidemiology, the term super-spreader [8,9] refers to a host that infects disproportionally more secondary contacts than others who have also been infected with the same virus or bacteria. Correspondingly in information propagation, a super-spreading story will spread in a notably large range on OSNs through a minor portion of users. Cases in disease spreading mostly conform the 20/80 rule [9], where approximately 20% of infected individuals are responsible for 80% of disease transmissions. Comparably, situations in information diffusion commonly comply the rule of 1% [10] in Internet culture, in which 1% of users contribute to the distribution mostly, while the other 99% spread little.

Online social network, especially microblog, is a hot research focus. Model, analysis and prediction of information propagation on microblog occupies an important part. Till now, there have been some studies on information propagation based on epidemic models.

Yang et al. [11] modeled information diffusion in implicit networks through a simple linear model, which can be utilized to quantify and predict the influence of users, and inspired us. But their model does not fit quantitative analysis of characterizing super-spreading phenomena. Lerman et al. [6] are the first to characterize the spread of news on OSNs by using epidemic models. Then, Abdullah et al. [7] formulated a clear epidemic-based model for news spreading on Twitter, following whom, there have been several studies based on epidemic models. The simple but operational and adoptable epidemic method of modeling is exploited to track trending topics, rumors and generic information on OSNs. These information propagation models are modified based on conventional epidemic models, such as the SIR model, the SEIR model, etc. Based on the SIR model, Li et al. [12] considered various kinds of user behaviors on Weibo and divide users into four groups with different transition rates between groups. Meng et al. [13] added a new compartment into the conventional SIR model in order to catch the phenomenon of rumor spreading in real life. Xia et al. [14] introduced a new class representing authorities that can diffuse authoritative information to prevent rumors from spreading. Zhao et al. [15,16] modified the SIR model to integrate forgetting and remembering mechanisms with variable forgetting rate to catch rumor dissemination characteristics. Zhou et al. [4] studied the characteristics of "information bomb" by leveraging a modified SIR model, which include influence breadth and strength of dynamic of information diffusion. Work by Xu et al. [17] and Zhao et al. [18] took into account complex network features to model information dissemination in mobile social networks and rumor spreading in the new media age, respectively. Zhang et al. [19] introduced a latent group of users who would be infected by an infectious individual, become latent and later then infectious. Although none of them addressed the super-spreading phenomena that benefits the information diffusion cascades on OSNs, previous work has shed light on the modeling of super-spreading.

In the field of disease spreading analysis, super-spreading events have been of great interest among researchers since the global outbreaks of severe acute respiratory syndrome (SARS) in 2003 [20]. Moreover, many infectious diseases were

[1] http://weibo.com/.