Physica A 449 (2016) 53–63

Contents lists available at ScienceDirect

Physica A

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

Modelling information dissemination under privacy concerns in social media

Hui Zhu [a,*], Cheng Huang [a], Rongxing Lu [b], Hui Li [a]

[a] State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an, China
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

HIGHLIGHTS

- We define the states and rules of information diffusion under privacy concerns.
- An information diffusion model is proposed under users' privacy concerns.
- The accuracy of model is validated by the real dataset.
- The statistical analysis of different users' privacy concerns is completed.
- Two classic networks have been simulated and analyzed under privacy concerns.

ARTICLE INFO

Article history:
Received 6 February 2015
Received in revised form 6 October 2015
Available online 6 January 2016

Keywords:
Privacy concerns
Information dissemination
Social media
Dynamic evolution equations

ABSTRACT

Social media has recently become an important platform for users to share news, express views, and post messages. However, due to user privacy preservation in social media, many privacy setting tools are employed, which inevitably change the patterns and dynamics of information dissemination. In this study, a general stochastic model using dynamic evolution equations was introduced to illustrate how privacy concerns impact the process of information dissemination. Extensive simulations and analyzes involving the privacy settings of general users, privileged users, and pure observers were conducted on real-world networks, and the results demonstrated that user privacy settings affect information differently. Finally, we also studied the process of information diffusion analytically and numerically with different privacy settings using two classic networks.

1. Introduction

With the development of mobile and web-based technologies, social media platforms, such as Facebook, Twitter, and LiveJournal, have become an important interactive platform allowing people to share information [1]. However, when increasing numbers of users publish news, ideas, and pictures/videos through social media, these platforms are capable of greatly influencing the real world (e.g., fake financial information may seriously lead to turbulence of financial markets in nations worldwide [2]). Generally, social media users can be divided into two groups: *general users* and *special users*. *General users* transfer information through various social media tools, receive information from others, and have their friends share views and news in LiveJournal or other platforms. *Special users*, such as celebrities, use the same platforms to post information, news, and knowledge; however, their information may be more effectively and authoritatively transmitted

* Corresponding author.
E-mail address: zhuhui@xidian.edu.cn (H. Zhu).

http://dx.doi.org/10.1016/j.physa.2015.12.044
0378-4371/

due to their reputation and authority. For instance, general Twitter users may pay more attention to celebrities based on recognition, making their information and news more attractive and enabling it to spread faster relative to those of general users [3,4]. Therefore, information diffusion in social media is a complex socio-psychological process, the analysis of which poses many challenges.

In order to accurately model information diffusion in social media, we need to know both a correct description of information dissemination and a quantitative formulation of various factors that motivate/restrict users to participate in social media. Based on the epidemic-spreading model [5], many information-dissemination models in social media have been proposed, including the susceptible–infected–susceptible (SIS) model and the susceptible–infected–removed (SIR) model, with several corresponding theoretical approaches for SIS and SIR having been presented, such as an individual-based mean-field approach [6,7], a degree-based mean-field approach [8–10], and a generating-function approach [11,12]. Pastor-Satorras et al. [13] and Vespignani [14] used the degree-based mean-field approach to describe information spreading using the SIS model, where the general methodology could be easily extended to almost all dynamic processes in social networks. Li et al. [15] and Ferreira et al. [16] also developed an SIS model using a degree-based mean-field approach for proposal of their models. Zhao et al. [17] and Wang et al. [18] improved the SIR model to a susceptible-infected-hibernator-removed model and a spreader–ignorant-stifler1-stifler2 model, respectively, using the individual-based mean-field approach to analyze rumour propagation in social networks. Additionally, many models are based on matrix calculations [19] that are used to describe information dissemination in social media. In order to analyze the dynamics of information dissemination thoroughly, the conditions of a given social network are usually changed in order to research different cases of information dissemination [20–22].

Most of the existing models discuss and analyze the process of information dissemination from various perspectives, such as different social media tools, different sources of information, and different user reputations [23]. In this study, we focused on a new perspective, i.e., user privacy concerns, and attempted to reveal their effect on information dissemination in social media. Specifically, we considered whether information dissemination would be influenced by user privacy settings after the information had been produced, posted, and transferred. Users generally classify their information into three types: public information, half-privacy information, and privacy information, which describe whether the information can be viewed by anyone, friends, or no one, respectively. In order to protect their privacy, social media users may define privacy settings for different information using various social media tools; however, the effects of user privacy settings may lead to different dynamics of information spreading. Therefore, in this study and based on user privacy concerns, we proposed a general stochastic model to simulate the process of information diffusion under the constraints of user privacy concerns in social media. There exist some typical approaches [24–26] toward helping characterize user behaviour in social media; however, most of these are based on analysing detailed data obtained from "crawling" through user data from real-world sessions, based on their access to popular social networks. This data cannot be used to analyze the general process associated with information diffusion in social media. Therefore, considering our current model, the effects of individual behaviour on information diffusion are not included in our work and will be discussed in future work.

Here, we make several contributions to the study of information dynamics in social media. First, we provide specific definitions to the state of nodes and transition rules for nodes in social media. As a basis for analysing information diffusion with privacy concerns, we introduce a model of information spreading. Compared with previous models, we provide a simplified, but realistic description of this process. A LiveJournal dataset was used for simulation and analysis with different privacy concerns for general users, privileged users, and pure observers, collectively representing most users in social media. The simulation results showed that the different privacy settings of general users impacted the process of information dissemination, with privileged users having more rigorous privacy settings exerting greater effects on slowing information diffusion relative to general users, and pure observer privacy concerns having almost no influence on information dissemination, even under rigorous privacy settings. Finally, we numerically investigated two classic dynamics graphs, the directed erdos-renyi random (DER) graph and the directed scale-free (DSF) graph, with privacy concerns. With the DER graph, we found that privacy concerns had little influence on information diffusion, and with the DSF graph, the simulation results showed that privacy concerns exerted great influence on information dynamics, either promoting or blocking information diffusion.

This paper is organized as follows: In Section 2, we define the states and rules of information diffusion in social media according to user privacy concerns. Then, our proposed general stochastic model is presented to describe information dissemination with dynamic evolution equations in Section 3, followed by extensive simulations and discussions regarding various user privacy concerns in Section 4. In Section 5, we investigate two classic graphs with different privacy concerns. Finally, we present our conclusions in Section 6.

2. Definitions of privacy concerns in social media

A network formed by social media can be viewed as a graph, $G = (V, E)$, where the nodes (vertices), V, are users and the links (edges), E, between nodes can be described as the interactions or interconnections between users [9]. Specifically, under different user privacy concerns, information may be diffused only via direct links from a user, and the diffused information will only influence some of the neighbouring nodes for this user. In order to accurately describe the relationships between users in social media, we defined the *out-degree* and *in-degree* of a node in G, where the *out-degree* captures how information is transferred to neighbouring nodes via out-links, and the *in-degree* captures how information can be