Information Sciences 288 (2014) 87–107

Contents lists available at ScienceDirect

Information Sciences

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

The relativity of privacy preservation based on social tagging

Baozhen Lee [a,*], Weiguo Fan [b], Anna C. Squicciarini [c], Shilun Ge [a], Yun Huang [d]

[a] Department of Information Management, Jiangsu University of Science and Technology, China
[b] Accounting and Information Systems, Virginia Polytechnic Institute and State University, USA
[c] College of Information Science and Technology, The Pennsylvania State University, USA
[d] Department of Industrial Engineering and Management Sciences, Northwestern University, USA

ARTICLE INFO

Article history:
Received 21 July 2012
Received in revised form 5 May 2014
Accepted 2 August 2014
Available online 10 August 2014

Keywords:
Privacy preservation
Social tagging
Relativity
Hierarchical community

ABSTRACT

Privacy preservation has gained importance with the development of tools for personal information retrieval and social information sharing in Web 2.0 environments. This paper addresses the need for a paradigm shift concerning publicity and privacy together in the process of personal information service and privacy preservation. In addition, we also bridge the conflict between interactivity and independency in the process of social sharing and privacy preservation. By analyzing the paradigm and conflict, we identify the criteria of privacy relativity, such as publicity vs. privacy and interactivity vs. independency. Using hierarchical clustering methods, we analyze the hierarchical community characteristics of tag networks which reflect users' preferences and the features of information resources through social tagging. Based on the hierarchical communities of tag networks and relevant parameters of partitions such as the k of k-clique and the number of communities N at a certain level, we can determine and adjust the degree and the scope of users' preferences and the features of information resources which can be revealed according to users' permissions. The adjustment process can reflect the trade-off between privacy preservation and information access. We believe that this work will prove to be an important first-step toward the personalization and socialization of security policies in Web 2.0 environments.

1. Introduction

Privacy has been investigated in various fields such as law, economics, management, marketing, psychology, and philosophy. And yet, it is widely recognized that as a concept, privacy "is in disarray and nobody can articulate what it means" [32]. Different from conventional data security technologies, privacy preserving technologies apply to prevent released documents from carrying sensitive and user-forbidden information [38,39]. Meanwhile, the challenge of this body of work is that absolute safety does not exist when it comes to user privacy: to enjoy better Web services (e.g. personalized Web services), users have to trade parts of their personal information for the better understanding on each individual preferences from Web applications [29]. It is the problem how to trade-off between better Web services and safer user privacy which can be reflected by the relativity about privacy preservation. This paper discusses the characteristic of relativity of the privacy preservation based on social tagging in the context of online social networks [37].

* Corresponding author.

http://dx.doi.org/10.1016/j.ins.2014.08.015
0020-0255/

Online social networks and Web 2.0 techniques can improve communication and collaboration among users in closed communities [16]. The richness of User-Generated Content (UGC) is the common characteristic of popular Web 2.0 platforms, such as Delicious, YouTube, Facebook, and Blogger. This has led to a new scenario whereby users not only access contents by participating personally, but also create new contents (e.g., blogs and tags) by sharing socially; predominantly, this content is also used as a form of primary communication between users, especially when commenting or evaluating the content created by other users [31]. The collection of digital information by users has created tremendous opportunities for knowledge-based decision making.

However, potential hazards and challenges have arisen from these new forms of collaboration with respect to privacy, trust and reputation of individuals and organizations. Despite the demand for the exchange of information among users driven by mutual benefits, information and related tags typically contain sensitive information about individuals and sharing such information will violate individual privacy [14]. At the same time, with the popularity of web search, the flat tendency in accessing web resources increases the risk of privacy leak with its spread. Social tagging leaks information about who we are and what we care about, therefore, some privacy will be lost in exchange for the benefits of digital services. Online social networks brought the voluntary disclosure of personal information to the mainstream, rendering the potential intrusion of privacy a critical and acute concern.

How to trade off between privacy preservation and information access has been studied using technical analysis [12], management model construction [3], etc. However, these studies mainly focus on the privacy preservation in the traditional Web environment. In Web 2.0 environment, users can show their preferences by social tagging because that they can participate actively in the process of social tagging [15,24]. In addition, the features of information resources can also be revealed comprehensively by social tagging because users can see the annotated results from each other and share relevant resources and tags through learning and referencing each other [10]. This paper mainly focuses on the two points of view, the personal information service and social information sharing, to analyze the relativity of privacy preservation based on social tagging in Web 2.0 environments.

In Web 2.0 environments, users can personally annotate and socially share relevant information resource by social tagging techniques. The results of social tagging can reflect users' preferences, reveal the features of information resource, and form the associated relations among users and information resources by relevant tags. When users retrieve or share relevant information resources based on social tagging, privacy preservation has the characteristic of relativity with the degree and the scope of users' preferences and the features of information resources which can be revealed according to relevant users' permission. By analyzing the relationship between privacy preservation and information access, we argue that privacy preservation should be balanced not only between the publicity and the privacy of a given user's preferences when he or she retrieves and accesses relevant information resources, but also between the interactivity and the independence of users when they create and organize their information resources. This trade-off can be reflected by hierarchical communities of relevant tag networks constructed in social tagging. And then users can adjust the relevant parameter of community partition to define the degree and scope of their privacy which can be revealed according to their needs. We present statistical and visual results, report findings, and suggest future work extending to many real-world applications using the Clique Percolation Method (CPM) [25,26] based on a dataset extracted from typical social tagging website, such as Delicious and BibSonomy.

The rest of this paper is organized as follows. Section 2 summarizes the relevant literatures. In Section 3, we analyze the problem and relevant methodology. Sections 4 and 5 present hierarchical communities of tag networks to reveal the relativity of privacy preservation from two points of view respectively. We conclude and discuss future research in Section 6.

2. Relevant literature

2.1. Privacy preservation and information access

Traditional privacy protection of constructed data belongs to the research about database (DB), while the privacy preservation of sensitive data belongs to the research of information retrieval (IR) or adversarial IR which includes spam filtering, link-bombing, and advertisement blocking.

Traditional network data security and encryption technologies mainly focus on encryption measures in data transmission. Some classical algorithms ensure that the content of data is encrypted for protection and cannot be intercepted or manipulated by unauthorized persons when it is transported from end to end. However, the privacy preservation of sensitive information is not equivalent to data security and encryption technology [39].

With the development of Web technologies, personal and customized services have gradually become reality, for example, personalized search, news recommendation, and intelligent/adaptive e-commerce websites, and so on. To provide personal services, users' preferences and behaviors should be tracked and learned to construct users' profiles and decision models, which include sensitive or secretive personal information [27]. The key issue of user privacy preservation is to define, measure, and control sensitive information in the user models.

The research of privacy preservation focuses on avoiding sensitive personal information in public user models, and balancing between better personal information services and safety of user privacy [13,36]. The precondition of privacy