Information Sciences 463–464 (2018) 214–226

Contents lists available at ScienceDirect

Information Sciences

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

Differentially private graph-link analysis based social recommendation

Taolin Guo*, Junzhou Luo, Kai Dong, Ming Yang

School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 211189, PR China

ARTICLE INFO

Article history:
Received 8 November 2017
Revised 11 April 2018
Accepted 21 June 2018
Available online 22 June 2018

Keywords:
Social recommendation
Online social network
Differential privacy

ABSTRACT

Modern social networks always require a social recommendation system which recommends nodes to a target node based on the existing links originate from this target. This leads to a privacy problem since the target node can infer the links between other nodes by observing the recommendations it received. As a rigorous notion of privacy, differential privacy has been used to define the link privacy in social recommendation. However, existing work shows that the accuracy of applying differential privacy to the recommendation is poor, even under an unreasonable privacy guarantee. In this paper, we find that this negative conclusion is problematic due to an overly-restrictive definition on the sensitivity. We propose a mechanism to achieve differentially private graph-link analysis based social recommendation. We make experiments to evaluate the privacy and accuracy of our proposed mechanism, the results show that our proposed mechanism achieves a better trade-off between privacy and accuracy in comparison with existing work.

1. Introduction

Making recommendations to users in online social networks (OSNs) increase not only their degree of engagement, but also the entire OSN's popularity and connectivity, thus has attracted significant attentions recently. A social recommendation system is one fundamental component of current OSNs. It computes utility of recommending any node (user or item) in the social graph to a target user, and selects those with the highest utilities, based on not only the entities' attributes such as target's prior history, but also the social links include user-to-user (e.g., friend [6,52]), item-to-item (e.g., similarity [2,28]) and user-to-item (e.g., like and dislike [30]) relations.

However, such social recommendation comes with a privacy concern that the social links originate from a user can be disclosed to others, since the presence or absence of these private links may affect the recommending results to the target users. Suppose a simple social graph (as illustrated in Fig. 1), the target user A can infer the existence of the private link between B and C, based on the observation that C is recommended.

An intuitive way to address this problem is to randomize the recommendations at the cost of sacrificing accuracy. Here, a trade-off between privacy and accuracy should be made. Most existing randomization techniques (for a survey, see [7]) lack a solid theoretical foundation on formalizing this trade-off. Differential privacy [9] is a rigorous notion of privacy in data analysis, which ensures that any output is "essentially" equally likely to occur, independent of the presence or absence

* Corresponding author.
E-mail address: guotaolin@seu.edu.cn (T. Guo).

https://doi.org/10.1016/j.ins.2018.06.054
0020-0255/

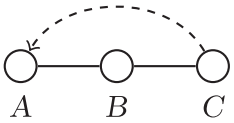

Fig. 1. Recommending C to A discloses the link between B and C. Circles are entities, solid lines are links between entities, and directed dashed lines are recommendations.

of the record of any single individual. It provides reasonable measurements on privacy and accuracy, and has been applied to a wide range of privacy preserving scenarios.

Differential privacy is also used in the area of social recommendation. Machanavajjhala *et al.* [31] suggest that differential privacy in a graph-link analysis based social recommendation requires that modifying an arbitrary social link should have a negligible effect on the recommendations to any target node. They argue that there is an inherent high sensitivity in social recommendation since the presence or absence of a link in a social network affects the recommendations of multiple nodes, result in very poor recommendation accuracy even to ensure an unreasonable privacy guarantee. According to this negative result, many approaches are based on an assumption that differentially private social recommendation is unfeasible [25,40]. This assumption has also been widely recognized in recent approaches [18,19,41,47,55].

However, this result is problematic due to two limitations. The first limitation is that the recommendation function is inconsistent when they define recommendation process and sensitivity. They consider that the recommendation function outputs the node who has the highest utility when defining the recommendation process, but consider this function outputs a utility vector which contains the utilities of recommending all node to the target when defining sensitivity. The latter definition will lead to overestimating the value of sensitivity, and underestimating the recommendation accuracy. The other limitation is that the relaxation is inconsistent when they define privacy and sensitivity. For the knowledge about the edges originate from the attack node, they suppose this knowledge should not be protected in defining privacy, but suppose the opposite in defining sensitivity. The relaxation used in their privacy definition reflects the natural setting in which the attacker already knows whether or not it is connected to other nodes in the graph; while their sensitivity definition results in overestimated sensitivity and underestimated recommendation accuracy.

In this paper, we propose a feasible differentially private social recommendation mechanism by addressing the aforementioned limitations. Firstly, we follow the relaxation that the edges originate from the target node is not privacy for it, and redefine sensitivity using the recommendation function which outputs the node with the highest utility rather than a utility vector. Then using our defined sensitivity, we transform the problem of achieving differentially private social recommendation by the *Exponential* mechanism to that by the *Report One-Sided Noisy Arg-Max* mechanism, since the utility of recommending a node is monotonic. Lastly, we evaluate our method from both privacy and accuracy perspectives, and compare with the commonly recognized existing approach [31], the results show that our method can achieve a much better and feasible trade-off between privacy and accuracy.

The contributions of this paper can be concluded as follows.

- It is commonly accepted that it is infeasible to perform social recommendations that are both differentially private and accurate. This opinion raise from the negative results [31] presented by Machanavajjhala *et al.* We notice and present two limitations in their definition on sensitivity, which lead to underestimating the accuracy. To the best of our knowledge, this paper is the first to point out that this opinion is problematic.
- We design and implement a social recommendation mechanism which ensures differential privacy, by addressing the limitations of exiting mechanisms and redefining the sensitivity in social recommendation. Moreover, we prove that the social recommendation can be modeled as a function with a monotonic utility, thus the *Report One-Sided Noisy Arg-Max* mechanism instead of the traditional *Exponential* mechanism can be used to achieve higher accuracy.
- We perform experiments on two open OSN datasets using three common utility functions. The results show that our mechanism achieves a better trade-off between privacy and accuracy. On the one hand, it achieves higher accuracy than the existing work while still meets a reasonable level of privacy. On the other hand, it greatly enhances accuracy with the increase in the privacy budget.

The rest of this paper is organized as follows. Section 2 introduces related work and Section 3 introduces the problem definition. Our proposed mechanism is detailed in Section 4 and the experiments are performed in Section 5. At last, we conclude this paper in Section 6.

2. Related work

In the past years, various privacy-preserving recommendation approaches have been proposed to protect users' privacy, while also ensuring the recommendation accuracy at the same time. We briefly summarize these approaches into three main categories [21]: anonymization, cryptographic protocols, and differential privacy mechanisms.