Computer Networks 170 (2020) 107106

Contents lists available at ScienceDirect

Computer Networks

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

Privacy preservation for social networks sequential publishing

Safia Bourahla[a,b,*], Maryline Laurent[c], Yacine Challal[a,d]

[a] Laboratoire des Méthodes de Conception des Systémes, Ecole nationale Supérieure d'Informatique, BP 68M, 16309, Oued-Smar, Algiers, Algeria
[b] Université Blida 2, Blida, Algeria
[c] SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France
[d] Centre de Recherche sur l'Information Scientifique et Technique, CERIST, Algiers, Algeria

ARTICLE INFO

Article history:
Received 6 May 2019
Revised 22 December 2019
Accepted 13 January 2020
Available online 14 January 2020

Keywords:
Privacy preserving
Social networks
Anonymization
Sequential releases

ABSTRACT

The proliferation of social networks allowed creating a big quantity of data about users and their relationships. Such data contain much private information. Therefore, anonymization is required before publishing the data for data mining purposes (scientific research, marketing, decision support, etc). Most of the anonymization works in social networks focus on publishing one instance while not considering the need for anonymizing sequential releases. However, many cases show that sequential releases may infer private information even though individual instances are anonymized. This paper studies the privacy issues of sequential releases and proposes a privacy preserving solution for this case. The proposed solution ensures three privacy requirements (users' privacy, groups' privacy and edges' privacy), and it considers the case where many users and groups may share the same profiles. Some experiments over some complex queries show that the utility of the released data is better preserved than other solutions, with regard to the privacy of users, groups and edges.

1. Introduction

Nowadays, social networks are becoming an integral part of daily interactions of modern life. Facebook claims over 1.4 billion monthly and 900 million daily active users [1]. By creating personal profiles, that contain demographic information, social networks allow users to create and join groups which have different interests across political, economic, and geographic borders. The affiliation of users to groups is considered as rich information that can be used by network researchers, sociologists, application designers and others for data mining tasks. For example, authors in [2] propose new metrics, namely the dispersion and the monopoly coefficients to refining the study of bipartite structures, particularly, when there is a community neighborhood overlapping. These two metrics are used to capture the intricate patterns observed in real social networks.

As a result, the full publication of this relationship information meets the need of data miners and allows them to perform good data mining tasks. Social networks include two types of data publishing:

1. One release: when a single instance of the social network is published.
2. Sequential releases: when several instances of the same social network are published over time to reflect its evolution.

However, the full publication of these social network data violates the users' privacy because an adversary can infer the affiliation links that the victim would like to keep private. To overcome this problem, researchers have proposed several techniques [3,9,18,20] to anonymize the data before their publication so that the privacy of users is preserved and the needs of data miners are satisfied. These techniques deal with different privacy risks which are Zheleva and Getoor [26]:

- Identity disclosure: the adversary can identify the victim from the published graph.
- Content disclosure: the sensitive attributes are identified and associated to the victim in the published graph.
- Social link disclosure: a sensitive relationship between two users is revealed.
- Affiliation link disclosure: the adversary can identify whether the victim belongs to a particular group.

One drawback of these techniques is that they consider the "one release case" only. However, to better answer the needs of data analyzers, it is recommended to publish sequential anonymized releases for the same social network to reflect its evolution in time. Applying directly these techniques to publishing

[*] Corresponding author.
E-mail addresses: sa_bourahla@esi.dz (S. Bourahla), maryline.laurent@telecom-sudparis.eu (M. Laurent), y_challal@esi.dz (Y. Challal).
URL: http://www.esi.dz (S. Bourahla), http://www.esi.dz (Y. Challal)

https://doi.org/10.1016/j.comnet.2020.107106
1389-1286/© 2020 Elsevier B.V. All rights reserved.