Information Sciences 275 (2014) 232–256

Contents lists available at ScienceDirect

Information Sciences

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

Identity obfuscation in graphs through the information theoretic lens

Francesco Bonchi [a,*], Aristides Gionis [b], Tamir Tassa [c]

[a] Yahoo! Research, Avinguda Diagonal 177, 08018 Barcelona, Spain
[b] Department of Information and Computer Science, Aalto University, Finland
[c] Department of Mathematics and Computer Science, The Open University, Ra'anana, Israel

ARTICLE INFO

Article history:
Received 13 September 2012
Received in revised form 28 August 2013
Accepted 6 February 2014
Available online 21 February 2014

Keywords:
Anonymity
Data publishing
Graph
Social network
Randomization
Information theory

ABSTRACT

Analyzing the structure of social networks is of interest in a wide range of disciplines. Unfortunately, sharing social-network datasets is often restrained by privacy considerations. One way to address the privacy concern is to anonymize the data before publishing. Randomly adding or deleting edges from the social graph is one of the anonymization approaches that have been proposed in the literature. Recent studies have quantified the level of anonymity that is obtained by random perturbation by means of a posteriori belief probabilities and, by conducting experiments on small datasets, arrived at the conclusion that random perturbation cannot achieve meaningful levels of anonymity without deteriorating the graph properties.

We offer a new information-theoretic perspective on the question of anonymizing a social network by means of random edge additions and deletions. We make an essential distinction between image and preimage anonymity and propose a more accurate quantification, based on entropy, of the anonymity level that is provided by the perturbed network. We explain why the entropy-based quantification, which is global, is more adequate than the previously used local quantification that was based on a posteriori belief probabilities. We also prove that the anonymity level as quantified by means of entropy is always greater than or equal to the one based on a posteriori belief probabilities. In addition, we introduce and explore the method of random sparsification, which randomly removes edges, without adding new ones.

Extensive experimentation on several very large datasets shows that randomization techniques for identity obfuscation are back in the game, as they may achieve meaningful levels of anonymity while still preserving properties of the original graph. As the methods we study add and remove edges, it is natural to ask whether an adversary might use the disclosed perturbed graph structure to reconstruct, even partially, the original graph. We thus study the resilience of obfuscation by random sparsification to adversarial attacks that are based on *link prediction*. Given a general link prediction method, with a predefined level of prediction accuracy, we show how to quantify the level of anonymity that is guaranteed by the obfuscation. We empirically prove that even for very accurate link prediction methods, the level of anonymity guaranteed remains very close to the one before the attack.

Finally, we show how the randomization method may be applied in a distributed setting, where the network data is distributed among several non-trusting sites, and explain why randomization is far more suitable for such settings than other existing approaches.

* Corresponding author.
E-mail addresses: bonchi@yahoo-inc.com (F. Bonchi), aristides.gionis@aalto.fi (A. Gionis), tamirta@openu.ac.il (T. Tassa).

http://dx.doi.org/10.1016/j.ins.2014.02.035
0020-0255/

1. Introduction

A social network is a graph structure holding information on a set of entities and the relations between them. Such information is of interest in a wide range of disciplines, including sociology, psychology, market research, and epidemiology. Very often social-network data cannot be published in their raw form since they contain sensitive information. The immediate first step to respect the privacy of individuals is to remove identifying attributes like names or social security numbers from the data. However, such a naïve anonymization is far from being sufficient. As shown by Backstrom et al. [2], the mere structure of the released graph may reveal the identity of the individuals behind some of the vertices. Hence, one needs to apply a more substantial procedure of sanitization on the graph before its release.

The methods for identity obfuscation in graphs fall into three main categories. The methods in the first category provide k-anonymity in the graph via edge additions or deletions [10,33,51,61,63]. The second category includes methods that add noise to the data in the form of random additions, deletions or switching of edges, in order to prevent adversaries from identifying their target in the network, or from inferring the existence of links between vertices [17,22,34,50,55–57]. The methods in the third category do not alter the graph data like the methods of the two previous categories; instead, they group together vertices into super-vertices of size at least k, where k is the required threshold of anonymity, and then publish the graph data in that coarser resolution [9,12,13,20,21,42,58].

In this paper we focus on the second category: changing the graph structure via random perturbations. Algorithms in this category usually utilize one of two graph perturbation strategies: random addition and deletion of edges, or random switching of edges. In the first strategy one randomly adds h non-existing edges after randomly deleting h existing edges; such techniques preserve the total number of edges in the graph [22,55]. In the second strategy, one selects h quadruples of vertices $\{u, v, x, y\}$ where (u, v) and (x, y) are edges and (u, x) and (v, y) are not, and switches between them, so that the two former edges become non-edges and the two latter non-edges become edges [55–57]; such techniques preserve the degree of all vertices in the graph.

Hay et al. [22] investigated methods of random perturbation in order to achieve identity obfuscation in graphs. They concentrated on re-identification of vertices by their degree. Given a vertex v in the real network, they quantified the level of anonymity that is provided for v by the perturbed graph as $(\max_u\{\Pr(v|u)\})^{-1}$, where the maximum is taken over all vertices u in the released graph and $\Pr(v|u)$ stands for the belief probability that u is in fact the target vertex v. By performing experimentation on the Enron dataset, using various values of h (the number of added and removed edges), they found out that in order to achieve a meaningful level of anonymity for the vertices in the graph, h has to be tuned so high that the resulting features of the perturbed graph no longer reflect those of the original graph.

Those methods were revisited by Ying et al. [54], who compared this perturbation method to the method of k-degree anonymity due to Liu and Terzi [33]. They too used the a posteriori belief probabilities to quantify the level of anonymity. Based on experimentation on two modestly sized datasets (Enron and Polblogs) they arrived at the conclusion that the deterministic approach of k-degree anonymity preserves the graph features better for given levels of anonymity.

Other authors have confirmed that topological features (e.g., clustering coefficient, or the largest eigenvalue of the adjacency matrix, that we also assess in Section 8.2) are "significantly lost in the randomized graph when a medium or high perturbation is applied" [50,55]. Those studies too use the same modestly sized datasets (Enron and Polblogs).

1.1. Our contributions

We provide a new information-theoretic perspective on the strategy of random additions and deletions of edges in graphs. Our main contribution is showing that randomization techniques for identity obfuscation are back in the game, as they may achieve meaningful levels of obfuscation while still preserving characteristics of the original graph. We prove our claim by means of a principled theoretical analysis and a thorough experimental assessment. In particular, we make the following contributions:

Types of obfuscation. We introduce a fundamental distinction between two different forms of privacy and two corresponding measures of the level of anonymity achieved by the perturbed network. One measure is called k-obfuscation, or *k-image obfuscation*, and it is a measure of privacy against an adversary who tries to locate in the perturbed graph the image of a specific individual. The second is called *k-preimage obfuscation*; it is a measure of privacy against an adversary who does not have any particular target individual, but she tries to examine the released graph and deduce the identity of any of the vertices in that graph.

More accurate measures of obfuscation. Our measures are defined by means of the entropy of the probability distributions that are induced on the vertices of the perturbed graph (in the case of k-obfuscation) or those which are induced on the vertices of the original graph (in the case of k-preimage obfuscation). This is in contrast to Hay et al. [22] who based their definition of k-candidate anonymity on a posteriori belief probabilities. While the a posteriori belief probability is a *local* measure that examines, for each vertex in the perturbed graph, the probability that the vertex originated from the target individual in question, the entropy is a *global* measure that examines the entire distribution of those belief probabilities. We explain and exemplify why the entropy-based measure is more accurate than the a posteriori belief probability, where