On the Privacy and Utility of Anonymized Social Networks

Yi Song, Department of Computer Science, School of Computing, National University of Singapore, Singapore

Xuesong Lu, Department of Computer Science, School of Computing, National University of Singapore, Singapore

Sadegh Nobari, Department of Computer Science, School of Computing, National University of Singapore, Singapore

Stéphane Bressan, Department of Computer Science, School of Computing, National University of Singapore, Singapore

Panagiotis Karras, Department of Management Science & Information Systems, Rutgers Business School, Rutgers University, Newark, NJ, USA

ABSTRACT

One is either on Facebook or not. Of course, this assessment is controversial and its rationale arguable. It is nevertheless not far, for many, from the reason behind joining social media and publishing and sharing details of their professional and private lives. Not only the personal details that may be revealed, but also the structure of the networks are sources of invaluable information for any organization wanting to understand and learn about social groups, their dynamics and members. These organizations may or may not be benevolent. It is important to devise, design and evaluate solutions that guarantee some privacy. One approach that reconciles the different stakeholders’ requirement is the publication of a modified graph. The perturbation is hoped to be sufficient to protect members’ privacy while it maintains sufficient utility for analysts wanting to study the social media as a whole. In this paper, the authors try to empirically quantify the inevitable trade-off between utility and privacy. They do so for two state-of-the-art graph anonymization algorithms that protect against most structural attacks, the k-automorphism algorithm and the k-degree anonymity algorithm. The authors measure several metrics for a series of real graphs from various social media before and after their anonymization under various settings.

Keywords: Anonymization, K-Automorphism Algorithm, K-Degree Anonymity Algorithm, Privacy, Social Network, Utility

DOI: 10.4018/jaras.2013040101
INTRODUCTION
From marketers and insurance companies to hackers and terrorists, many are thirsty and hungry for the wealth of information contained in the content and structure of social media. How can this information be made available for legitimate usage such as the academic study of network dynamics (Kumar, Novak, & Tomkins, 2006) without compromising the privacy of members of the social media?

Graph anonymization algorithms have been, and are being, devised that allow the publication of a perturbed network. The perturbation is hoped to provide sufficient privacy for the members while maintaining adequate utility for the analysts wanting to study the social media as a whole. Such algorithms must strike an inevitable trade-off between privacy and utility.

In this paper we try and empirically quantify this trade-off for the $k$-degree anonymity algorithm (Liu & Terzi, 2008) and $k$-automorphism algorithm (Zou, Chen, & Ozsu, 2009). The $k$-degree anonymity algorithm transforms the original graph into one in which, at least, $k$ vertices have the same degree. The transformed graph is $k$-degree anonymous. $k$-automorphism is a state-of-the-art graph that protects against most structural attacks. The algorithm transforms the original graph into one in which, at least, $k$ subgraphs are structurally identical. The transformed graph is $k$-automorphic.

We measure several utility metrics for a series of real graphs from various social media before and after their anonymization under varying settings.

The rest of the paper is organized as follows. In the Related Work Section we briefly review the background and related work on graph anonymization algorithms. Section Anonymization summarizes the anonymization algorithm that we study. Section Utility Metrics presents the graph metrics we use for experiments. In Section Data Set we describe the data sets that we use and in Experiments Section we present and analyze the results of our extensive experiments. Finally we conclude in the Conclusion.

RELATED WORKS
Backstrom et al. (2007), one of the pioneering works on graph anonymization, point out shortcomings of naive graph anonymization, which replaces identity of individuals by synthetic identifiers. Attackers are able to identify their target from remarkable existing (passive attacks) or created (active attacks) structural local properties of the graph.

To prevent such attacks, various anonymization methods are proposed that further modify the graph in order to hide remarkable structural features.

Liu and Terzi (2008) propose to anonymize graphs by inserting new edges in order to make the degree of each vertex undistinguishable from that of, at least, $k - 1$ other vertices. Attacks based on degree information can be prevented. The graph is said to be $k$-degree-anonymous after modification. Zhang et al. (2009) propose to swap and delete edges based on degree information in order to prevent re-identification of sensitive edges in simple graph. Ying and Wu (2008) suggest randomized edge addition, deletion and switching to anonymize graph while controlling the effect on graph spectrum.


Considering labels of vertices in addition to the basic structural information, Zhou and Pei (2008) propose to extract the neighborhoods of all vertices, group vertices and anonymize the neighborhoods of vertices in the same group by generalizing vertex labels and adding edges.

Hay, Miklau, Jensen, Towsley, & Weis (2008) study re-identification risk caused by queries on graph structure on simple graph, and design a grouping method. The main idea is to group vertices into partitions and to publish the...