Prevent Identity Disclosure in Social Network Data Study

Kun Bai, Ying Liu, Peng Liu
The Pennsylvania State University
{Kbai, yliu, pliu }@ist.psu.edu

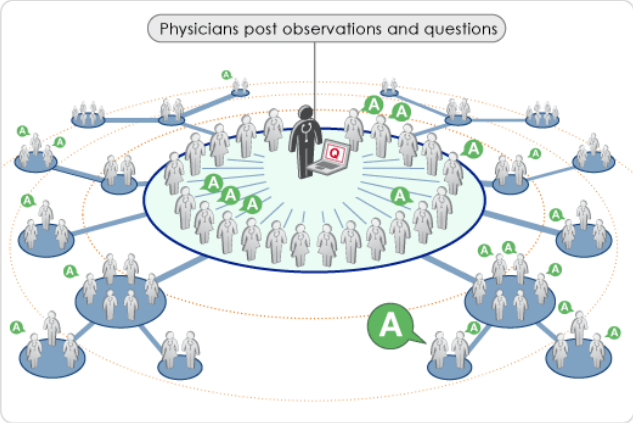
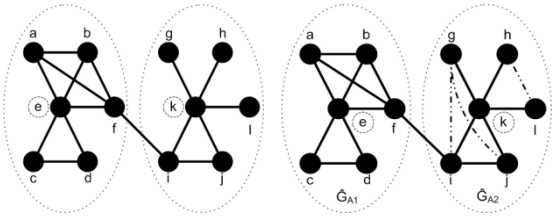


Fig 1. A Social Network Example



An Example of K-Dominated-Sets Anonymization. Three edges are added. (The labels for each vertex are just for the sake of illustration and are not in the graphs). (a) has two dominated sets induced to subgraphs H={k, g, h, i, j, l}, H={e, a, b, c, d, f}, which are isomorphic graphs after the edge operations in (b).

First, we form a degree sequence D from all vertex *v 2 G, then bucket them into m classes each of which contains* only distinct degree value. For each class *ci, if the number* of vertices *jVci j 5 k, we construct a new degree sequence* locally that is *k-anonymous so that each class has at least k* vertices and the information loss is minimized.

Second, we chose those vertices we consider as dominating vertices and then determine the dominated set of each dominating vertex. These dominated sets cover the entire graph *G, and may have overlaps. We than anonymize the domi*nated sets within each equivalent class starting with those dominating vertices with high degrees because of the common power law of degree distribution, high-degree vertices are always minority.

Experimental Results

We show in Figure 2(a), 2(b) the performance of the k-degree anonymization algorithm on different data sets with different settings of parameter *k.*

We show in Figure 2(c), 2(d) the performance of the *k*-dominated sets anonymization algorithm in terms of the number of added ages, and the information loss (graph centrality skewness, equivalently), where the average vertex degree of the data set is 4.

Conclusion

In this paper, we deal with an important problem of pre-serving privacy in non-relational publication data. We model the non-relational data in a graph and propose a practically feasible solution. An extensive empirical study is conducted on real data sets. The experimental results demonstrate that our solution is highly practical to prevent privacy leakage from the identity disclosure attacks.