Structural Diversity for Resisting Community Identification in Published Social Networks

Abstract:

As an increasing number of social networking data is published and shared for commercial and research purposes, privacy issues about the individuals in social networks have become serious concerns. Vertex identification, which identifies a particular user from a network based on backgroundknowledge such as vertex degree, is one of the most important problems that have been addressed. In reality, however, each individual in a social network is inclined to be associated with not only a vertex identity but also a community identity, which can represent the personal privacy information sensitive to the public, such as political party affiliation. This paper first addresses the new privacy issue, referred to as community identification, by showing that the community identity of a victim can still be inferred even though the social network is protected by existing anonymity schemes. For this problem, we then propose the concept of structural diversity to provide the anonymity of the community identities. The k-Structural Diversity Anonymization (k-SDA) is to ensure sufficient vertices with the same vertex degree in at least k communities in a social network. We propose an Integer Programming formulation to find optimal solutions to k-SDA and also devise scalable heuristics to solve large-scale instances of k-SDA from different perspectives. The performance studies on real data sets from various perspectives demonstrate the practical utility of the proposed privacy scheme and our anonymization approaches.