

Protecting Anonymity in Published Social Networks



Gerome Miklau

Michael Hay, Gerome Miklau (PI), David Jensen

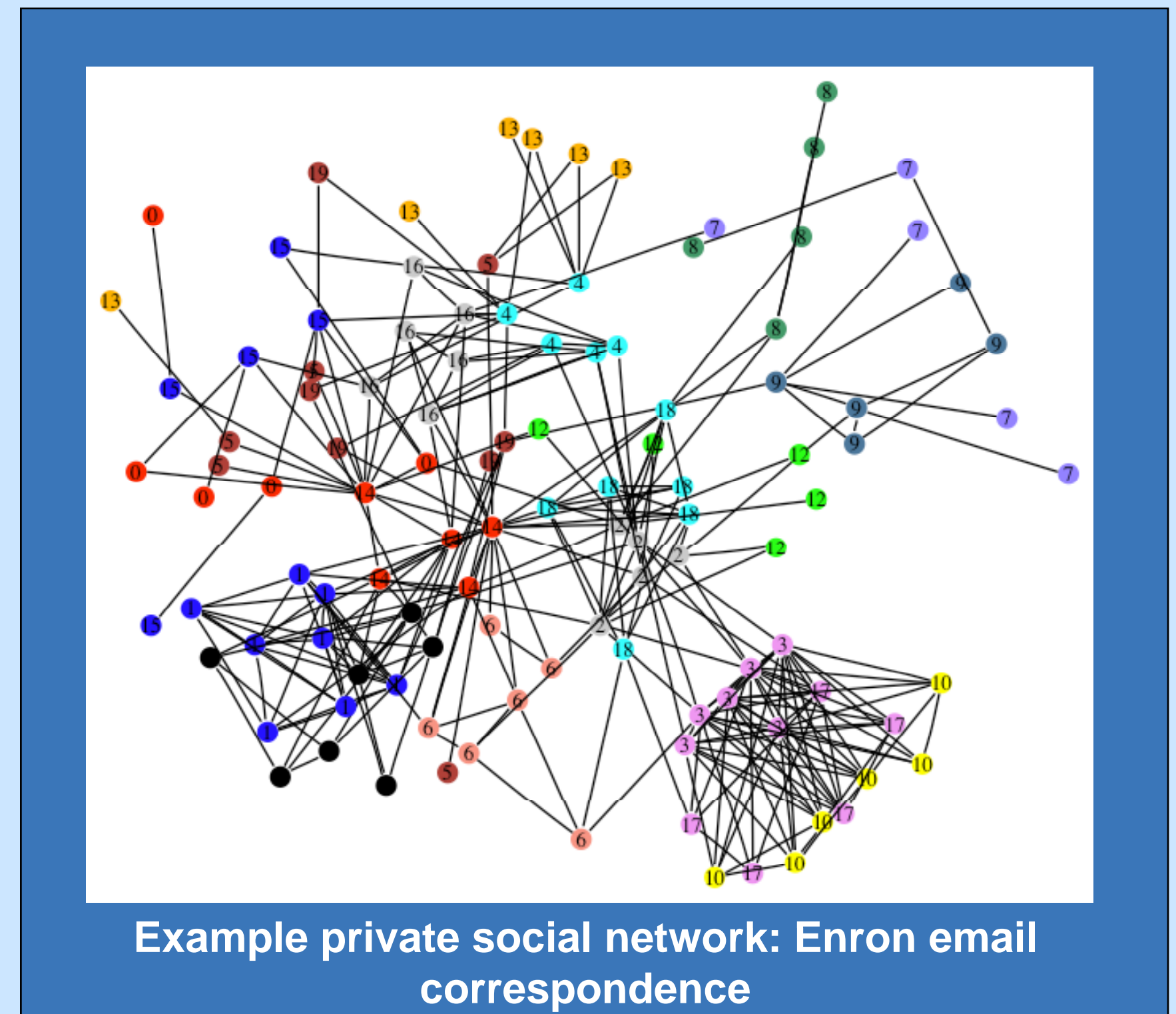
www.cs.umass.edu/~miklau

Objectives

Understand privacy risk of publishing social network data, and design algorithm to transform network to ensure anonymity without distorting important network properties.

Overview Dissemination and analysis of networked data has many benefits, but it is often severely constrained by privacy concerns. An individual's connections (i.e. the graph structure around them) can be distinguishing, and may be used to re-id an anonymous individual.

We present a framework for assessing the privacy risk of sharing anonymized network data. After formalizing models of adversary knowledge, we measure re-identification risk in real social networks drawn from diverse domains. Our results show that common anonymization techniques are inadequate. We propose a novel graph transformation technique in which random edge perturbation is applied to a partitioning of the original graph in order to obscure the distinguishing structural features of individuals while preserving global properties of the graph.



Approach and Impact

New approach

- Formal knowledge models
- Evaluation of disclosure
- Anonymization algorithm

Research Impact

- Demonstrated vulnerabilities of naïve anonymization
- Algorithm protects anonymity via clustering and random perturbation, yet preserves global network structure

Assessing Re-identification Risk We model an adversary with partial knowledge of the local neighborhood around a target node. On real social networks, we simulate adversary attacks. Our findings: adversary can re-identify with a small amount of knowledge (e.g., degree); structural similarity of nodes varies widely across networks; real networks less vulnerable than theory suggests.

Anonymization Algorithm partitions nodes into groups of minimum size k , and then induces probability distribution over random graphs conditioned on partitioning. Uses local search to find partitioning that maximizes the likelihood of input graph. Although local structure is randomly perturbed, algorithm preserves global network properties (e.g., resiliency).

Algorithm Evaluation

