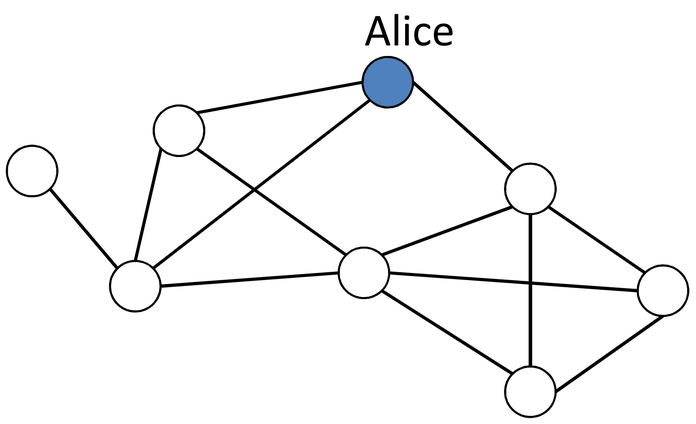
CERIAS

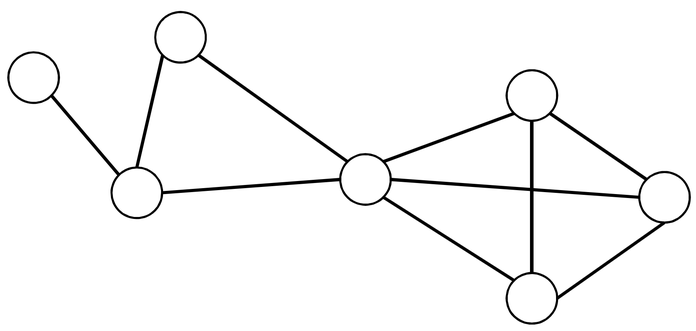
the center for education and research in information assurance and security

ε-Differential Node Privacy in Graph Data Queries

Christine Task, Chris Clifton
Computer Science and Statistics, Purdue University



Friend Network Data:
Nodes = Individuals
Edges = Friendships
Query: What is the smallest number of individuals connecting parts of the graph?



Neighboring Graphs:
Differ in one individual

Min-Cut(G1) = 2

Min-Cut(G2) = 1

Differentially Private Min-cut :
Randomized Query Result could be 1.37, 1.80, 1.46, …

Differential Privacy guarantees sufficient noise that guessing which of the neighboring graphs produced the query result is unlikely. This protects individual privacy: if query results from G1 and G2 are indistinguishable, we cannot learn Alice's data.

Query Sensitivity is a measure of the maximum difference between query results on *any* neighboring graphs. High sensitivity queries require adding so much noise that results are useless – hence, we cannot perform such analyses *and* guarantee privacy.

High Sensitivity Queries:

- Graph Isomorphism
- Average Node Degree
- Graph Diameter
- PageRank
- Connected Components

Open Problems:

- Social Cluster Identification
- Propagation Algorithms (popularity measures)
- Subgraph Counting with unique edges

Low Sensitivity Queries:

- Subgraph Counting with unique nodes
- Degree Distribution[1]
- Min-Cut
- Graph Estimation[2]

[1] Michael Hay, et al., "Accurate Estimation of the Degree Distribution of Private Networks", IEEE International Conference on Data Mining, 2009

[2] Darakhshan J. Mir and Rebecca N. Wright, "A differentially private graph estimator", International Workshop on Privacy Aspects of Data Mining, 2009.

PURDUE UNIVERSITY

CERIAS

Discovery Park e-Enterprise Center