ROADMAP

1. Remind the *formal* definition of the problem.
2. Present previous work on anonymization bounds for random graphs.
3. Share some ideas on practical de-anonymization.


DANGER
НЕБЕЗПЕКА

PROBLEM DEFINITION

Given graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ of size $n = |V|$, a permutation π on V maps vertices from G_1 onto G_2.

Our task is to find the best such permutation, denoted by π_0.

COST FUNCTION

Either graph similarity or node similarity.

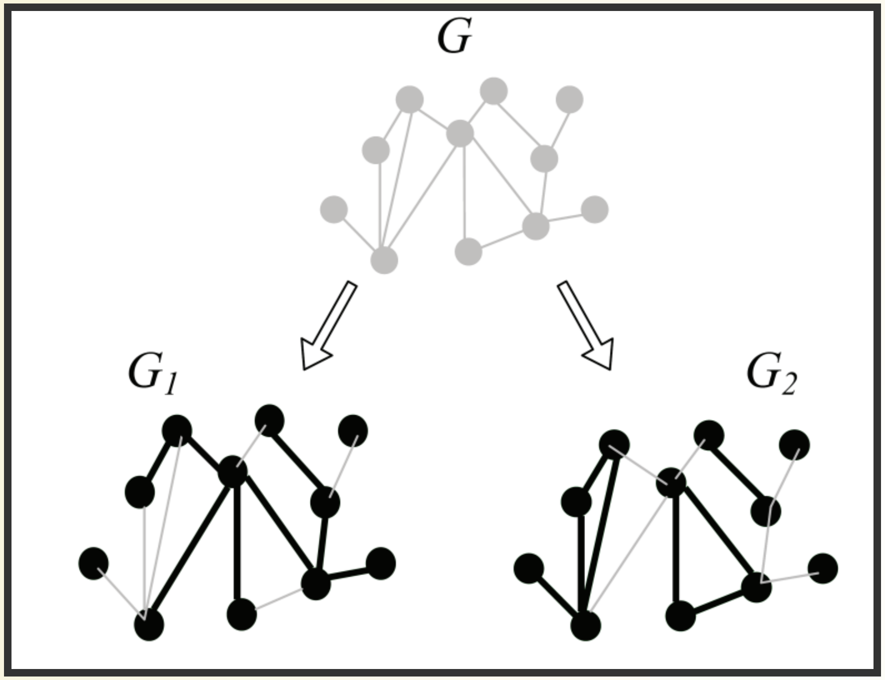
ON THE PRIVACY OF ANONYMIZED NETWORKS

P. Pedarsani, M. Grossglauser, KDD '11

QUESTION

“Can we assume that a sufficiently sparse network is inherently anonymous?”

RANDOM GRAPH MODEL



$$G \sim Erd\ddot{o}sR\acute{e}nyi(n, p)$$

G_1 and G_2 sample edges from G with probability s.

COST FUNCTION

$$\Delta_\pi = \sum_{e \in E_1} 1_{\{\pi(e) \notin E_2\}} + \sum_{e \in E_2} 1_{\{\pi^{-1}(e) \notin E_1\}}$$

THEOREM

For $s = \omega(\frac{1}{n})$ and $p \to 0$, if

$$ps\,\frac{s^2}{2-s} = 8\,\frac{\log n}{n} + \omega(n^{-1})$$

then the identity permutation π_0 minimizes the error criterion Δ_π, yielding perfect matching of the vertex sets of G_1 and G_2.

COROLLARY

For a fixed value of s, if ps is a bit larger than $8\,\frac{\log n}{n}$, with high probability π_0 minimizes the error function and yields the correct matching. Keep in mind that, for G to be connected, $ps > \frac{\log n}{n}$.

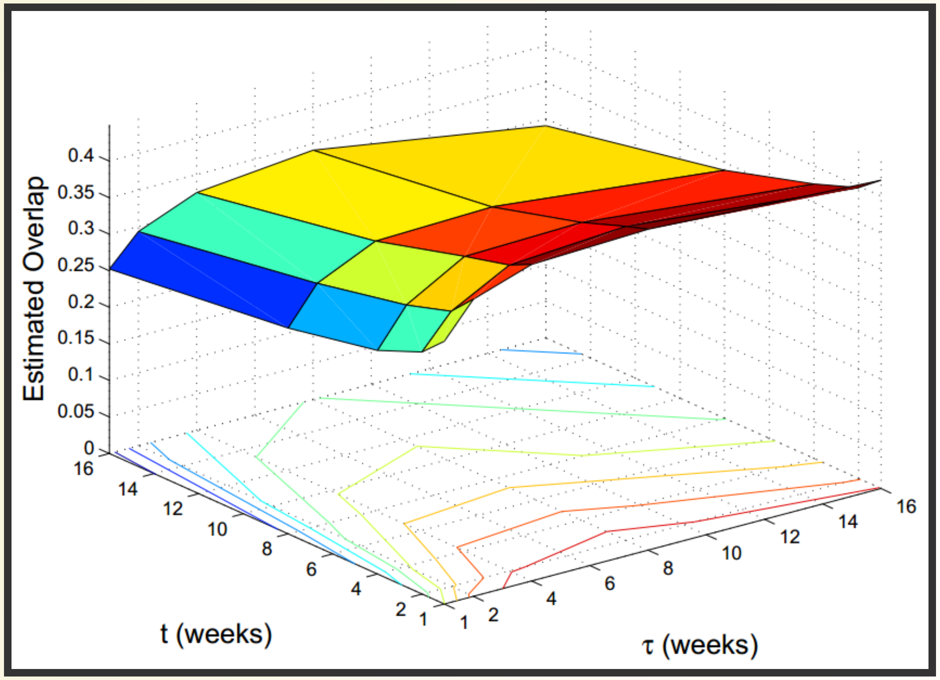
EXPERIMENTS

Email communications in EPFL for 75 weeks.

Investigate similarities between different snapshots.

Check the correlations between random pairs of edges.

EXPERIMENTS



t: duration for graph accumulation

τ: time distance between two graphs

Edge overlap is robust to changes in the parameters.

TAKE-HOME MESSAGE

Even if G_1 and G_2 are *sampled randomly* from G, the set of parameters that avoids de-anonymization is very narrow.

HOWEVER ...

- Graphs may not be generated from the same model.
- We may happen to know the correspondence for some nodes.
- No algorithm for finding π_0.

IDEAS

Project the nodes and compare the pair-wise similarity (embedding).

ELECTRICAL NETWORKS

For the efficient approximation of the effective resistances in an electrical network, *Spielman and Srivastava* create the following $k \times n$ projection matrix:

$$Z = QBL^{\dagger}$$

For details, see

Graph Sparsification by Effective Resistances, D. Spielman and N. Srivastava

ORDERING THE SIMILARITIES

Use matching nodes as beacons (with high probability, highest degree nodes are unique). However, we will need to explore how many beacons are sufficient.

SUMMARY

Theoretical bounds for anonymizing graphs

Ideas for practical de-anonymization:

- Use embeddings and beacons
- Electrical networks
- Low-stretch spanning tree
- Manifold alignment