Many data sets can be naturally represented as network of pairwise relationships (graph):
- Friendship in online social network
- Romantic relationships
- Financial transactions

Goal: Publish info about network structure for use by researchers, govt’ agencies, ...

Privacy concerns: Sensitive information
- De-anonymizing social networks [BDK’07, NS’09]
- Link prediction by de-anonymization [NSR’11]

This work: new privacy-preserving algorithms

Subgraph Counts

For graphs G and H, \( f_H(G) = \# \) of occurrences of H in G

K-star Triangle K-triangle

Subgraph counts are widely used in:
- Exponential random graph models
- Descriptive graph statistics (e.g. clustering coefficient)

Our work: We present d.p. algorithms for accurately releasing subgraph counts

Measuring Privacy

Goal: Rigorous privacy guarantee

Challenges
- Attacker’s prior information unknown
- Attack algorithm / methodology unknown

\( \varepsilon \)-Differential privacy [DMNS’06]: For all neighbors G and G’ (differing in one record):

\[
Pr[A(G) = s] \leq e^\varepsilon \cdot Pr[A(G') = s]
\]

Hides the presence/absence of individual relationships

Techniques

Add noise, proportional to S/\(\varepsilon\) (where S is sensitivity). How to minimize S?

This work: Instance-specific noise
- Local sensitivity (also used in [RHMS’09])

\[
LS_f(G) = \max_{G' \in \mathbb{N}(G)} |f(G) - f(G')|
\]
- Nearly optimal amplitude, but not d.p.

Previous work: Global sensitivity (mostly)

\[
GS_f(G) = \max_G LS_f(G)
\]

Main approaches:
- Smooth sensitivity [NRS’07]
- D.p. approximation to \(LS_f(G)\) (this work).

Results

New algorithms:
- K-stars (linear time algorithms for computing smooth sensitivity)
- K-triangles (d.p. approximation to \(LS_f(G)\))

Bounds on noise in terms of \(LS_f(G)\):

Conditions, when amplitude of noise is close to \(\frac{LS_f(G)}{\varepsilon}\)

Average-case analysis in Gilbert model G(n,p)

\[
np = \omega(\sqrt{n \log n}), \varepsilon = o(\log n)\]

\[
\left| A_H(G) - f_H(G) \right| = O\left(\frac{1}{\varepsilon n^2 p}\right)
\]

Computing smooth sensitivity is NP-hard for some subgraph queries (K-triangles, K-cycles)

Experimental Evaluation

Results of experimental evaluation for \(\varepsilon = 0.5\):

<table>
<thead>
<tr>
<th>Dataset</th>
<th>(H)</th>
<th>(\Delta)</th>
<th>Median Relative Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CitGrqC</td>
<td>\Delta</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>Nodes = 3242</td>
<td>2*</td>
<td>0.0041</td>
<td></td>
</tr>
<tr>
<td>Edges = 2960</td>
<td>3*</td>
<td>0.114</td>
<td></td>
</tr>
<tr>
<td>CitEuiPh</td>
<td>\Delta</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>Nodes = 9577</td>
<td>2*</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Edges = 51971</td>
<td>3*</td>
<td>0.0009</td>
<td></td>
</tr>
<tr>
<td>CitAcaPh</td>
<td>\Delta</td>
<td>0.000043</td>
<td></td>
</tr>
<tr>
<td>Nodes = 18772</td>
<td>2*</td>
<td>0.00023</td>
<td></td>
</tr>
<tr>
<td>Edges = 306160</td>
<td>3*</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td>CitKslPh</td>
<td>\Delta</td>
<td>0.0007</td>
<td></td>
</tr>
<tr>
<td>Nodes = 51971</td>
<td>2*</td>
<td>0.000008</td>
<td></td>
</tr>
<tr>
<td>Edges = 237010</td>
<td>3*</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>CitConMat</td>
<td>\Delta</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td>Nodes = 23133</td>
<td>2*</td>
<td>0.00158</td>
<td></td>
</tr>
<tr>
<td>Edges = 18936</td>
<td>3*</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>CitAcaPh</td>
<td>\Delta</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Nodes = 36692</td>
<td>2*</td>
<td>0.00055</td>
<td></td>
</tr>
<tr>
<td>Edges = 267062</td>
<td>3*</td>
<td>0.0021</td>
<td></td>
</tr>
</tbody>
</table>