GraphProtector: A Visual Interface for Employing and Assessing Multiple Privacy Preserving Graph Algorithms

Xumeng Wang¹, Wei Chen¹, Jia-Kai Chou², Chris Bryan⁴, Huihua Guan³, Wenlong Chen¹, Tianyi Lao¹, Kwan-Liu Ma²

¹: Zhejiang University, State Key Lab of CAD&CG
²: University of California, Davis
³: Alibaba Group
⁴: Arizona State University
Motivation

Could you find the professor?
Structural Features to Identify Nodes

- **Degree**
  # edges connected to a node

- **Hub fingerprint**
  Hub: node with special features
  Fingerprint: connection status with hubs

- **Subgraph**
  A group of connected nodes
K-anonymity

Structure feature should have at least $k$ occurrences.

A higher $k$ →
{ Better protection \quad \text{Worse utility} }

How to set appropriate $k$?
Motivation

Subgraph (circle) 
Degree = 10

Subgraph (path) 

Degree = 1

Hub fingerprint (\(\times \sqrt{\times} \sqrt{\sqrt{\times}}\))

Hub fingerprint (\(\sqrt{\sqrt{\sqrt{\times}}}\))

Hub fingerprint (\(\sqrt{\sqrt{\times}}\))

Subgraph (cluster) 
Degree = 14

Degree = 18

How to set \(k\) for so many features?
Motivation

Privacy Experts
- Identify privacy issues
- Customize schemes
- Evaluate results

Visualization Tools
- Intuitive representations
- Explanation
- Assessment and comparison
Related Work: Privacy Preservation for Graphs

• K-anonymity [ACM SIGMOD 2008, VLDB 2009, ACM SIGMOD 2010]
  Construct similar (structural) features.

• Differential Privacy [ACM SIGKDD 2014, ACM SIGCOMM 2011]
  Make perturbations to data.

• Graph-only Models [ASIACCS 2009, SDM 2008]
  Cluster nodes or randomly edit edges.
Related Work: Evaluating Privacy Preservation

- Privacy preservation
  - Query results of specific features [VLDB 2008, VLDB 2014]

- Utility loss
  - Structure properties [AJS1987, AJS2004]
  - Specific analysis tasks [ACM SIGKDD 2012, ACM WSDM 2013]
Related Work: Privacy-aware Visualizations

Graph Data [IEEE PVIS 2017]

Multi-attribute Tabular Data [IEEE TVCG 2018]
Task Requirements

TR1: Learn the characteristics.

TR2: Guide auto-processing.

TR3: Evaluate and compare schemes.

TR4: Record the provenance.
Workflow & Interface

Original data

Visual specification

Privacy preservation

Processed data
Workflow

Learn About the Characteristics. (TR1)

Original data

Visual specification

Privacy preservation

Processed data

Overview

Distribution
Workflow

Original data

Visual specification

Privacy preservation

Processed data

Specifying utility metrics. (TR3)

Specifying identity priority. (TR2)
Workflow

Original data
Visual specification
Privacy preservation
Processed data

Prioritize these individuals
Try not to modify these individuals
Do not handle these individuals
Visual Design: Priority View

- **Other nodes**
  - Degree: 0.3~0.8
  - Closeness: 0~0.5
  - Betweenness: 0~0.5

- **All nodes**
  - Degree: 0.3~0.8
  - Closeness: 0~0.5
  - Betweenness: 0~0.5
Workflow

Original data

Visual specification

Privacy preservation

Processed data
Visual Design: Protector View

- Amount of feature occurrences
  - Satisfied
  - Unsatisfied

- Distribution changes
Visual Design: Degree Protector

Degree: # edges connected to a node
Visual Design: Hub Fingerprint Protector

Hub: node with special features
Fingerprint: connection status with hubs

Ex. fingerprint: × × √

The amount of occurrences

Hub node

Disconnected

Connected

K line
Hub: node with special features
Fingerprint: connection status with hubs
Visual Design: Subgraph Protector

Subgraph: a group of connected nodes

<table>
<thead>
<tr>
<th>Classic</th>
<th>External</th>
<th>2 detected subgraphs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete graph</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of nodes: 4  Tolerance: 1/6
Workflow

1) Identify risk.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Privacy</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>😊</td>
<td>😞</td>
</tr>
<tr>
<td>S2</td>
<td>😊</td>
<td>😊</td>
</tr>
</tbody>
</table>

2) Specify scheme.

3) Compare schemes. (TR3)

4) Execute scheme. (TR4)
Workflow

Original data

Visual specification

Privacy preservation

Processed data
Visual Design: Provenance View

STEP 1: Degree

- Clustering coefficient: 0.509, 0.2%
- Joint degree (Euclidean): 58.975
- Added edges: 10, 0.4%

Metric value changes

Edge modifications
Workflow

Original data
Visual specification
Privacy preservation
Processed data

Explain the result. (TR4)
Case: Facebook Friendship Data

- Sub-dataset from “Learning to discover social circles in ego networks.” [NIPS2012]
- 333 nodes (users)
- 2519 edges (friendships)
GraphProtector

A visual interface for employing and assessing multiple privacy preserving graph algorithms
Case: Face-to-Face Contacts Dataset

- Collected during the exhibition INFECTIOUS
- [http://konect.uni-koblenz.de/networks/sociopatterns-infectious](http://konect.uni-koblenz.de/networks/sociopatterns-infectious)
- 410 nodes (participants)
- 2765 edges (conversations lasted over 20 seconds)
Case: Face-to-Face Contacts Dataset

Scheme 1
Lock: 0%~2%

Scheme 2
Lock: 98%~100%
Case2: Face-to-Face Contacts Dataset

Degree protector: $k = 2$
Case: Face-to-Face Contacts Dataset

Degree protector: $k = 2$

Scheme 1

Scheme 2
User Reviews

• A live, hands-on demo about 30 minutes
  ✓ All protectors are easy to use
  ✓ Helps interpretation.
  ✓ A “fine-grained data processing” pipeline.
?
  Trouble with the provenance view.
Discussion

• Detailed guidance

Terminals
(privacy preserving goals)

Directions
(processing priorities)

Prioritize these individuals
Try not to modify these individuals
Do not handle these individuals

(processing priorities)
Discussion

• Detailed guidance

• Performance

Pre-computation for metric values

Lazy searches

Ranking: 50%
Value: 45

2 detected subgraphs
Discussion

- Detailed guidance
- Performance
- Extensibility
Thank you

Xumeng Wang, Wei Chen, Jia-Kai Chou, Chris Bryan, Huihua Guan, Wenlong Chen, Rusheng Pan, Kwan-Liu Ma

Acknowledgement

National 973 Program of China (2015CB352503)
National Natural Science Foundation of China ((61772456 and 61761136020)
Alibaba-Zhejiang University Joint Institute of Frontier Technologies
U.S. National Science Foundation (IIS-1320229 and IIS-1741536)
Q&A

GraphProtector: A Visual Interface for Employing and Assessing Multiple Privacy Preserving Graph Algorithms

Xumeng Wang
wangxumeng@zju.edu.cn

Jia-Kai Chou
jkchou@ucdavis.edu
http://vidi.cs.ucdavis.edu/People/ChouJia-Kai

Chris Bryan
cbryan16@asu.edu