GAP: DIFFERENTIALLY PRIVATE GRAPH NEURAL NETWORKS WITH AGGREGATION PERTURBATION

Sina Sajadmanesh  
Idiap & EPFL

Ali Shahin Shamsabadi  
The Alan Turing Institute

Aurélien Bellet  
Inria

Daniel Gatica-Perez  
Idiap & EPFL

32nd USENIX Security Symposium  
August 2023
Graph Neural Networks (GNNs) are state-of-the-art algorithms for learning on graphs

- **Tasks:** node classification, link prediction, ...
- **Applications:** recommendation systems, credit issuing, traffic forecasting, drug discovery, ...

Graph data could be privacy-sensitive
- e.g., users' personal attributes, financial transactions, medical/biological networks, ...

GNNs are vulnerable to privacy attacks
- e.g., link stealing attack [He et al., 2021] or membership inference attack [Olatunji et al., 2021]
Graph Neural Networks (GNNs) are state-of-the-art algorithms for learning on graphs

- **Tasks:** node classification, link prediction, ...
- **Applications:** recommendation systems, credit issuing, traffic forecasting, drug discovery, ...

Graph data could be privacy-sensitive

- e.g., users’ personal attributes, financial transactions, medical/biological networks, ...
Graph Neural Networks (GNNs) are state-of-the-art algorithms for learning on graphs.

- **Tasks:** node classification, link prediction, ...
- **Applications:** recommendation systems, credit issuing, traffic forecasting, drug discovery, ...

Graph data could be privacy-sensitive:
- e.g., users’ personal attributes, financial transactions, medical/biological networks, ...

GNNs are vulnerable to privacy attacks:
- e.g., link stealing attack [He et al., 2021] or membership inference attack [Olatunji et al., 2021]
Graph Neural Networks (GNNs) are state-of-the-art algorithms for learning on graphs

- Tasks: node classification, link prediction, ...
- Applications: recommendation systems, credit issuing, traffic forecasting, drug discovery, ...

Graph data could be privacy-sensitive

- e.g., users’ personal attributes, financial transactions, medical/biological networks, ...

GNNs are vulnerable to privacy attacks

- e.g., link stealing attack [He et al., 2021] or membership inference attack [Olatunji et al., 2021]

How to preserve privacy of individuals when learning over graph data?
Our Contributions

▶ GAP: a novel GNN with differential privacy (DP) guarantees

- Aggregation Perturbation to preserve privacy of graph edges
- Tailored Architecture to maintain privacy budget
- Formal Privacy Analysis for both edge-level and node-level DP
Graph Neural Networks (GNNs) learn node representations based on node features and the graph structure.
Differential Privacy [Dwork et al., 2006]

Randomized algorithm $A$ is $\epsilon$-DP if for all neighboring datasets $G \simeq G'$ and all sets of outputs $S$:

$$\frac{\Pr[A(G) \in S]}{\Pr[A(G') \in S]} \leq e^\epsilon$$

- **Edge-Level DP**: Neighboring graph datasets differ by at most one edge.
- **Node-Level DP**: Neighboring graph datasets differ by at most one node (and all adjacent edges).
Differential Privacy [Dwork et al., 2006]

Randomized algorithm $A$ is $\epsilon$-DP if for all neighboring datasets $G \simeq G'$ and all sets of outputs $S$:

$$\frac{\Pr[A(G) \in S]}{\Pr[A(G') \in S]} \leq e^{\epsilon}$$

- **Edge-Level DP**
  Neighboring graph datasets differ by at most one edge

- **Node-Level DP**
  Neighboring graph datasets differ by at most one node (and all adjacent edges)
Differential Privacy [Dwork et al., 2006]

Randomized algorithm $A$ is $\epsilon$-DP if for all neighboring datasets $G \sim G'$ and all sets of outputs $S$:

$$\frac{\Pr[A(G) \in S]}{\Pr[A(G') \in S]} \leq e^\epsilon$$

- **Edge-Level DP**
  Neighboring graph datasets differ by at most one edge

- **Node-Level DP**
  Neighboring graph datasets differ by at most one node (and all adjacent edges)
Exploding Sensitivity

- With a K-layer GNN, each node affects the embedding of all the nodes in its K-hop neighborhood
- $O(D^K)$ gradient terms change at once ($D$ is maximum degree)
Exploding Sensitivity

- With a K-layer GNN, each node affects the embedding of all the nodes in its K-hop neighborhood.
- \(O(D^K)\) gradient terms change at once (\(D\) is maximum degree).

Inference Privacy

- GNNs query the graph structure during inference.
- Private information leaks at inference, even with a private model.
Challenges of Learning GNNs with DP: Why Not DP-SGD?

- Exploding Sensitivity
  - With a K-layer GNN, each node affects the embedding of all the nodes in its K-hop neighborhood
  - $O(D^K)$ gradient terms change at once ($D$ is maximum degree)

- Inference Privacy
  - GNNs query the graph structure during inference
  - Private information leaks at inference, even with a private model

$DP$-$SGD$ cannot be directly applied to GNNs
**Our Approach: Aggregation Perturbation**

- **Aggregation Perturbation:** adding noise to output of the aggregation step
  - Prevents the exploding sensitivity problem by composing differentially private aggregation steps
  - Ensures inference privacy
Our Approach: Aggregation Perturbation

- **Aggregation Perturbation:** adding noise to output of the aggregation step
  - Prevents the exploding sensitivity problem by composing differentially private aggregation steps
  - Ensures inference privacy

- Applying aggregation perturbation to the conventional GNNs is costly
  - Every forward pass of the model consumes privacy budget
  - The excessive noise results in poor performance
Our Approach: Aggregation Perturbation

- **Aggregation Perturbation**: adding noise to output of the aggregation step
  - Prevents the exploding sensitivity problem by composing differentially private aggregation steps
  - Ensures inference privacy

- Applying aggregation perturbation to the conventional GNNs is **costly**
  - Every forward pass of the model consumes privacy budget
  - The excessive noise results in poor performance

*We decouple the aggregation operations from the model parameters to maintain the privacy budget*
GNN with Aggregation Perturbation (GAP)
✓ Edge-level DP
GAP’s Advantages

✔ Edge-level DP
✔ Node-level DP through combination with DP-SGD
  • For bounded-degree graphs
✓ Edge-level DP
✓ Node-level DP through combination with DP-SGD
  • For bounded-degree graphs
✓ Multi-hop aggregations
✓ Edge-level DP
✓ Node-level DP through combination with DP-SGD
  • For bounded-degree graphs
✓ Multi-hop aggregations
✓ Zero-cost inference privacy
**Task:** Node Classification  
**Baselines:** MLP, GraphSAGE

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Classes</th>
<th>Nodes</th>
<th>Edges</th>
<th>Features</th>
<th>Med. Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>6</td>
<td>26,406</td>
<td>2,117,924</td>
<td>501</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>YEAR</td>
<td>USER</td>
<td>FRIENDSHIP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reddit</td>
<td>8</td>
<td>116,713</td>
<td>46,233,380</td>
<td>602</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>COMMUNITY</td>
<td>POST</td>
<td>MUTUAL USER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amazon</td>
<td>10</td>
<td>1,790,731</td>
<td>80,966,832</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>CATEGORY</td>
<td>PRODUCT</td>
<td>MUTUAL PURCHASE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Comparison of Non-Private Methods

#### Accuracy of Non-Private Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Facebook</th>
<th>Reddit</th>
<th>Amazon</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAP-$\infty$</td>
<td>80.0 ± 0.48</td>
<td>99.4 ± 0.02</td>
<td>91.2 ± 0.07</td>
</tr>
<tr>
<td>SAGE-$\infty$</td>
<td>83.2 ± 0.68</td>
<td>99.1 ± 0.01</td>
<td>92.7 ± 0.09</td>
</tr>
</tbody>
</table>
Edge-Level DP Accuracy-Privacy Trade-Off

Facebook

Reddit

Amazon
Node-Level DP Accuracy-Privacy Trade-Off

Facebook

Reddit

Amazon

Accuracy (%)

Privacy Cost (ε)

GAP-∞  MLP-DP  GAP-NDP  SAGE-NDP
Mean AUC of node-level membership inference attack.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>$\epsilon = 1$</th>
<th>$\epsilon = 2$</th>
<th>$\epsilon = 4$</th>
<th>$\epsilon = 8$</th>
<th>$\epsilon = 16$</th>
<th>$\epsilon = \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>GAP-NDP</td>
<td>50.16</td>
<td>50.25</td>
<td>50.61</td>
<td>51.11</td>
<td>52.66</td>
<td>81.67</td>
</tr>
<tr>
<td>Reddit</td>
<td>GAP-NDP</td>
<td>50.04</td>
<td>50.39</td>
<td>51.20</td>
<td>52.23</td>
<td>52.54</td>
<td>54.97</td>
</tr>
<tr>
<td>Amazon</td>
<td>GAP-NDP</td>
<td>50.06</td>
<td>50.23</td>
<td>50.54</td>
<td>51.53</td>
<td>51.72</td>
<td>66.68</td>
</tr>
</tbody>
</table>
CONCLUSION

- GNNs leak private information
  - They are vulnerable to privacy attacks

- Implementing DP in GNNs is challenging
  - Exploding sensitivity
  - Inference privacy

- Our Differentially Private GNN: GAP
  - Ensures both edge-level and node-level DP
  - Supports multi-hop aggregations
  - Provides inference privacy
THANK YOU!

Questions: +sina.sajadmanesh@epfl.ch
Code: /github.com/sisaman/GAP


Effect of the Number of Hops

**Facebook**

*GAP-EDP Accuracy (%)*  
- $\varepsilon = 4$  
- $\varepsilon = 1$

**Reddit**

*GAP-EDP Accuracy (%)*  
- $\varepsilon = 4$  
- $\varepsilon = 1$

**Amazon**

*GAP-EDP Accuracy (%)*  
- $\varepsilon = 4$  
- $\varepsilon = 1$

---

**Facebook**

*GAP-NDP Accuracy (%)*  
- $\varepsilon = 16$  
- $\varepsilon = 8$

**Reddit**

*GAP-NDP Accuracy (%)*  
- $\varepsilon = 16$  
- $\varepsilon = 8$

**Amazon**

*GAP-NDP Accuracy (%)*  
- $\varepsilon = 16$  
- $\varepsilon = 8$
Effect of the Encoder Module

Facebook

GAP-EDP Accuracy (%)

Privacy Cost (ε)

Facebook

W/ EM

W/O EM

Reddit

GAP-EDP Accuracy (%)

Privacy Cost (ε)

Reddit

W/ EM

W/O EM

Amazon

GAP-EDP Accuracy (%)

Privacy Cost (ε)

Amazon

W/ EM

W/O EM

Facebook

GAP-NDP Accuracy (%)

Privacy Cost (ε)

Facebook

W/ EM

W/O EM

Reddit

GAP-NDP Accuracy (%)

Privacy Cost (ε)

Reddit

W/ EM

W/O EM

Amazon

GAP-NDP Accuracy (%)

Privacy Cost (ε)

Amazon

W/ EM

W/O EM
Effect of the Maximum Degree

Facebook

Reddit

Amazon

GAP-NDP Accuracy (%)

Maximum Degree ($D$)

Facebook

Reddit

Amazon

$\epsilon = 4$  
$\epsilon = 16$