Differentially Private Range Subgraph Counting

Xian Chen

Joint work with

Ruobing Bai Pan Peng

University of Science and Technology of China (USTC)

Range Subgraph Counting (RSC)

Input

- Graph $G = (V, E, \mathbf{a})$ with n vertices and attribute $\mathbf{a}(v) \in \mathbb{R}^d$ for each $v \in V$
- Pattern H with $O(1)$ vertices
- Query set Q of d-dimensional ranges

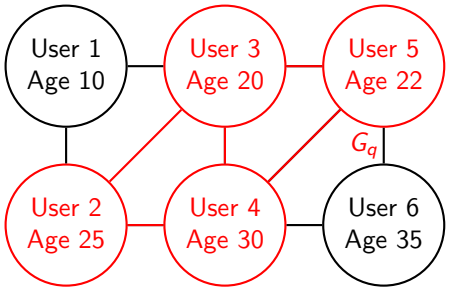
For a query $q = [\ell_1, r_1] \times \cdots \times [\ell_d, r_d]$

- Filter vertices: $V_q = \{v \in V \mid \ell_i \leq a_i(v) \leq r_i,\ \forall i \in [d]\}$
- Induce subgraph: $G_q = G[V_q]$

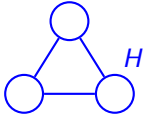
Goal

- Count $f_H(G_q)$: # occurrences of H in G_q

an example social network graph G



range query $q = [2, 5] \times [20, 30]$



$f_H(G_q) = 2$ **when H is a triangle**

DP Range Subgraph Counting (DPRSC)

Definition (Edge DP)

A randomized algorithm $\mathcal{A}$ is (ε, δ)-*DP* if for all events S and all neighboring graphs $G \sim G'$ (i.e., differing by one edge),

$$\Pr[\mathcal{A}(G) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(G') \in S] + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is *pure DP (ε-DP)*; otherwise it is *approximate DP*.

DP Range Subgraph Counting (DPRSC)

Definition (Edge DP)

A randomized algorithm $\mathcal{A}$ is (ε, δ)-*DP* if for all events S and all neighboring graphs $G \sim G'$ (i.e., differing by one edge),

$$\Pr[\mathcal{A}(G) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(G') \in S] + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is *pure DP (ε-DP)*; otherwise it is *approximate DP*.

Additive error lower bounds implied by edge DP:

- Instance-dependent: $\Omega_{\varepsilon,\delta}(\mathrm{LS}_{f_H}(G))$ w.p. $> \frac{2}{3}$, where *local sensitivity*

$$\mathrm{LS}_{f_H}(G) = \max_{G' : G \sim G'} |f_H(G) - f_H(G')|$$

- Worst-case: $\Omega_{\varepsilon,\delta}(\mathrm{GS}_{f_H})$ w.p. $> \frac{2}{3}$, where *global sensitivity*

$$\mathrm{GS}_{f_H} = \max_{G} \mathrm{LS}_{f_H}(G) = \max_{G \sim G'} |f_H(G) - f_H(G')|$$

.

The upper bound

- We present the first efficient algorithms for DPRSC with small additive error.
- We reduces RSC to weighted orthogonal range counting via a subgraph projection.
- We incorporate the techniques of noisy range trees [Dwork et al., 2015] and local sensitivity estimation [Nguyen et al., 2023] to achieve accurate private query answering.

Theorem (Approximate DPRSC, Informal)

For any $\varepsilon > 0$ and $\delta \in (0, 1)$, there exists a (ε, δ)-DP algorithm for RSC which satisfy

$$\max_{q \in Q} \left| f_H(G_q) - \widetilde{f}_H(G_q) \right| = O\left(\frac{\widetilde{\mathrm{HS}}_{f_H}(G) \cdot \sqrt{(\varepsilon + \log(1/\delta)) \log(n|Q|)} \cdot \log^{2d} n}{\varepsilon} \right)$$

with probability at least $1 - \frac{1}{n}$. The quantity $\widetilde{\mathrm{HS}}_{f_H}(G)$ can be viewed as an approximation of $\mathrm{LS}_{f_H}(G)$ up to $\mathrm{poly}(\frac{1}{\varepsilon}, \log(1/\delta))$ factors [Nguyen et al., 2023].

- For pure DP, we obtain a corresponding error bound $O\left(\frac{\mathrm{GS}_{f_H} \cdot \sqrt{\log(n|Q|)} \cdot \log^{3d} n}{\varepsilon} \right)$.

The lower bound

- We complement our algorithms with **nearly matching lower bounds**.
- We obtain the bounds by reducing reconstruction attacks to DPRSC and leveraging discrepancy theory [Muthukrishnan and Nikolov, 2012].

Theorem (Lower Bound of DPRSC, Informal)

Assume that $|Q| \geq n^c$ for any sufficiently small constant $c > 0$.
For any H, let $\mathcal{A}$ be an (ε, δ)-DP algorithm for RSC with constants ε, δ and additive error $\eta = \max_{q \in Q} |f_H(G_q) - \widetilde{f}_H(G_q)|$ with a sufficiently large constant success probability.

- If $d = O(1)$, then $\eta = \Omega(\log^{d-1} n \cdot \widetilde{\mathrm{HS}}_{f_H}(G))$;
- If $d = O(\log n)$, then $\eta = 2^{\Omega(d)} \cdot \widetilde{\mathrm{HS}}_{f_H}(G)$;
- If $d = \Omega(\log n)$, then $\eta = n^{\Omega(1)} \cdot \widetilde{\mathrm{HS}}_{f_H}(G)$.

- The quantity $\widetilde{\mathrm{HS}}_{f_H}(G)$ in the hard instance is, with high constant probability, close to GS_{f_H}.
- We further provide a lower bound $\Omega(\log n \cdot \widetilde{\mathrm{HS}}_{f_H}(G))$ for the case $d = 1$, under the additional assumption that $\delta = n^{-\Omega(1)}$.

Other extensions & Experiments

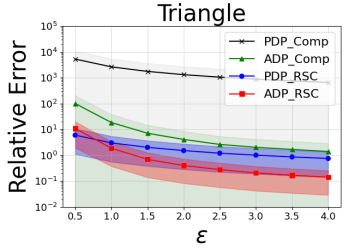
- More details can be found in our paper and poster.

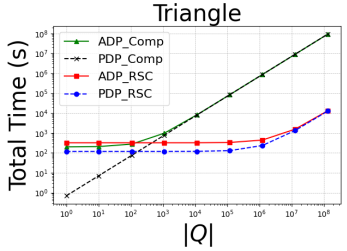
Other extensions

- We give a polynomial-time ε-DP algorithm based on randomized response, with additive error $O(n\sqrt{\log(n|Q|)}\mathrm{GS}_{f_H})$ for any d and constant ε, which improves for $d = \Omega(\log n / \log\log n)$.
- We extend the techiniques in Eliáš et al. [2020] to obtain a stronger lower bound $\Omega(n\mathrm{GS}_{f_H})$ for constant ε, δ when d is sufficiently large.

Experiments

- Evaluate on three real-world datasets with three patterns.
- Achieve lower additive error and faster runtime compared to standard baselines.





Summary & open question

Summary

- We study the DPRSC problem and design the first efficient algorithms with small additive error.
- We prove lower bounds showing an exponential dependence on the attribute dimension, which nearly matches our upper bounds.

Open question

- General algorithms with nearly optimal utility guarantees under attribute-level privacy?
- Tighter upper and lower bounds?
- Extensions of RSC to more challenging settings (e.g., dynamic graphs and alternative privacy models)?

Summary & open question

Summary

- We study the DPRSC problem and design the first efficient algorithms with small additive error.
- We prove lower bounds showing an exponential dependence on the attribute dimension, which nearly matches our upper bounds.

Open question

- General algorithms with nearly optimal utility guarantees under attribute-level privacy?
- Tighter upper and lower bounds?
- Extensions of RSC to more challenging settings (e.g., dynamic graphs and alternative privacy models)?

Thank you for listening!

References

C. Dwork, M. Naor, O. Reingold, and G. N. Rothblum. Pure differential privacy for rectangle queries via private partitions. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 735–751. Springer, 2015.

M. Eliáš, M. Kapralov, J. Kulkarni, and Y. T. Lee. Differentially private release of synthetic graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 560–578. SIAM, 2020.

S. Muthukrishnan and A. Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1285–1292, 2012.

D. Nguyen, M. Halappanavar, V. Srinivasan, and A. Vullikanti. Faster approximate subgraph counts with privacy. *Advances in Neural Information Processing Systems*, 36, 2023.