Benchmarking Triangle Counting Algorithms Under Local Differential Privacy with PLUTO

Long Pham
pham_l3@denison.edu

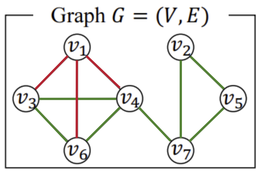
Stacey Truex
truexs@denison.edu

Privacy in Triangle Counting

Privacy compromise from Triangle Counting queries

Scenario[2]**: Consider the graph representing a social network**.

- Person corresponding to v_1 considers their network (adjacency list) private
- v_7 knows the friend list of v_2, v_3, v_4, v_5, v_6 exception: if they are connected to v_1
- **Q:** Does output of triangle counting algorithm $\mathcal{A}$ compromise v_1's privacy w.r.t. v_7?

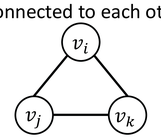


Observe: $\mathcal{A}(G) = 5$

- $\Delta\, v_6 - v_3 - v_4, \Delta v_2 - v_7 - v_5$ are known
- v_1 is not connected to v_7 is known

Q: Can 5 triangles be created in this graph with no $v_1 - v_6$ edge?
No! ⇒ v_1's connection to v_6 leaks!

Triangle in graphs: a triangle is formed if three distinct nodes v_i, v_j, v_k have edges connected to each other



Risk: sensitive connections can be inferred

v_7 takes advantage of **background knowledge** (collusion with other members of the network / public knowledge of connections) and **results from a statistical query** ($\mathcal{A}(G) = 5$ triangles in G) to infer private information (v_1 is connected with v_6)

Prob: How to guarantee privacy of individual connections within a social network?

Sol: Add controlled randomness to $\mathcal{A}$

1. Users perturb sensitive data
2. Users transmit noisy data
3. Server aggregates noisy reports
4. Statistical estimation recovers patterns

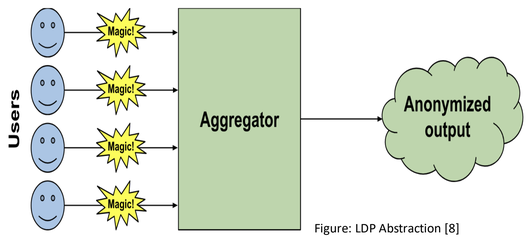


Figure: LDP Abstraction [8]

Definition: ϵ-edge LDP[6]. Let ε in $\mathbb{R}_{\geq 0}$. For any $i \in [n]$, let $\mathcal{R}_i$ with domain $\{0,1\}^n$ be a randomized algorithm of user v_i. $\mathcal{R}_i$ provides ε-edge LDP if for any two neighbor lists a_i, a_i' in $\{0,1\}^n$ that differ in one bit and any $S \subseteq Range(\mathcal{R}_i)$,

$$\Pr[\mathcal{R}_i(a_i) \in S] \leq \Pr[\mathcal{R}_i(a_i') \in S] \cdot e^{\varepsilon}$$

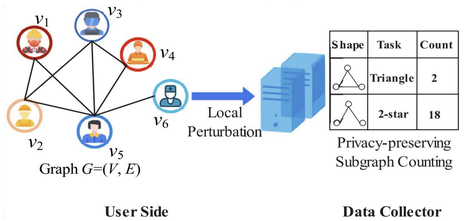
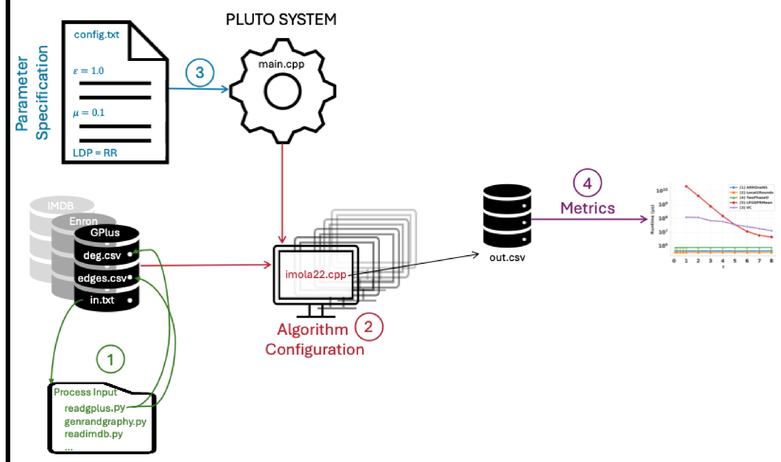


Figure: Privacy-Preserving Graph Analysis Overview [9]

System Architecture

PLUTO: Privacy-preserving benchmarking for Local and Unbiased Triangle Obfuscation



Components

1. Data Processing

- Real world graph networks from SNAP [7]
  - Facebook, Gplus, etc.
  - Template data processing script, support for custom scripts
- Synthetic random graphs
  - Erdös-Rényi
  - Barabási-Albert
- Support for custom pre-defined graph networks

2. Algorithm Configuration

- Add self-contained module & update build configurations

3. Parameter specification

- General parameters:
  Ex: ε: controlling privacy budget
  n: number of nodes controlling input sizes
- Algorithm-specific tuning parameter:
  - μ: edge sampling to reduce noise
  - Privacy budget allocation: allocate epsilon across different perturbation steps
- Choice of LDP Mechanism
  Ex: Randomized Response, Optimized Unary Encoding

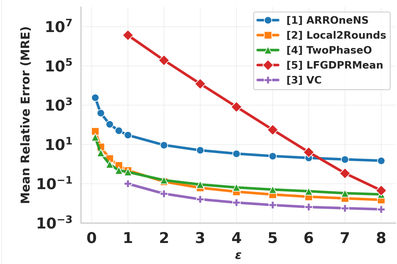
4. Metrics

- Relative Error & L_2-Loss of triangle counts
- Runtime in microseconds
- Communication costs in bytes: upload & download cost between users and server

Empirical Evaluations

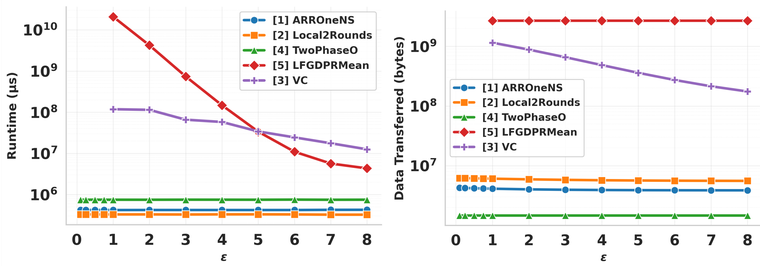
- Algorithms: Local2Rounds [2]:- 2-round protocol with edge clipping
  ARROneNS [1]: communication-efficient extension of Local2Rounds
  TwoPhaseO [4]: an approach leveraging user's extended local views
  LFGDPRMean [5]: 1-round protocol, uses Δs to compute clustering coeff
  VC [3]: 3-round protocol based on pairs of vertices
- Dataset: Enron email network, 36,692 nodes
- Machine: Intel Core i9-13900K, 32 threads, 128GB RAM

Results



- VC[3] consistently yields lowest error
- Local2Rounds[2] outperforms ARROneNS[1] which samples edges (data loss) to increase communication efficiency
- TwoPhaseO[4] performs similarly to [2] but has a different adversarial model as friends are assumed to know each other's connections
- LFGDPR has high bias, but sees the most rapid decrease in MRE as ε incr.

Note: smaller ε ⇒ increased probability of bit-flipping in LDP
⇒ sparse graphs become denser ⇒ higher runtime & communication cost



- LFGDPRMean[5] runs in $O(n^3)$ time and sees highest impact from increased density incurred from low ε values (n: number of nodes)
- VC[3] shows improved runtime in ε settings by limiting computation scope to local neighbors
- TwoPhaseO[4] counts triangles prior to noise integration ⇒ takes advantage of sparseness in real world network
- ARROneNS[1] & Local2Rounds[2] run in $O(n^2 + nd_{max}^2)$ perturbing only edges required in the second round

- LFGDPRMean[5] transmits constant amount of data
- VC[3] reduces communication cost as ε increases, since sparser graphs decrease the likelihood of completing triangles (⇒ ↓transmissions)
- ARROneNS[1] & Local2Rounds[2] same pattern as [3] with a less observable rate of change and at a lower overall communication cost
- TwoPhaseO[4] does not transmit adjacency vectors leading to the lowest measured communication cost

Acknowledgements

Research conducted by the TLS lab at Denison University directed by Dr. Stacey Truex

1 J. Imola, T. Murakami, and K. Chaudhuri, "Communication-Efficient triangle counting under local differential privacy," in 31st USENIX Security Symposium(USENIX Security 22). Boston, MA: USENIX Association, Aug. 2022, pp. 537–554. [Online]. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/imola
2 J. Imola, T. Murakami, and K. Chaudhuri, "Locally differentially private analysis of graph statistics," in 30th USENIX Security Symposium (USENIX Security 21). USENIX Association, Aug. 2021, pp. 983–1000. [Online]. Available: https://www.usenix.org/conference/usenixsecurity21/presentation/imola
3 Y. He, K. Wang, W. Zhang, X. Lin, Y. Zhang, and W. Ni, "Robust privacy-preserving triangle counting under edge local differential privacy," Proc. ACM Manag. Data, vol. 3, no. 3, Jun. 2025. [Online]. Available: https://doi.org/10.1145/3725348
4 H. Sun, X. Xiao, I. Khalil, Y. Yang, Z. Qin, H. W. Wang, and T. Yu, "Analyzing subgraph statistics from extended local views with decentralized differential privacy," in Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, ser. CCS '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 703–717.
5 Q. Ye, H. Hu, M. H. Au, X. Meng, and X. Xiao, "Lf-gdpr: A framework for estimating graph metrics with local differential privacy," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 10, pp. 4905–4920, 2022.
6 Z. Qin, T. Yu, Y. Yang, I. Khalil, X. Xiao, and K. Ren. Generating synthetic decentralized social graphs with local differential privacy. In Proc. CCS'17, pages 425– 438, 2017.
7 J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," http://snap.stanford.edu/data, Jun. 2014.
8 D. Desfontaines, 'Local vs. central differential privacy', 06 2019. [Online]. Available: https://desfontain.es/blog/local-global-differential-privacy.html.
9 T. Wang, J. Liang, S. Wang, L. Zhao, and T. Yang, 'Efficient and accurate graph statistics with adaptive personalized local differential privacy', *Neurocomputing*, vol. 639, p. 130224, 2025.