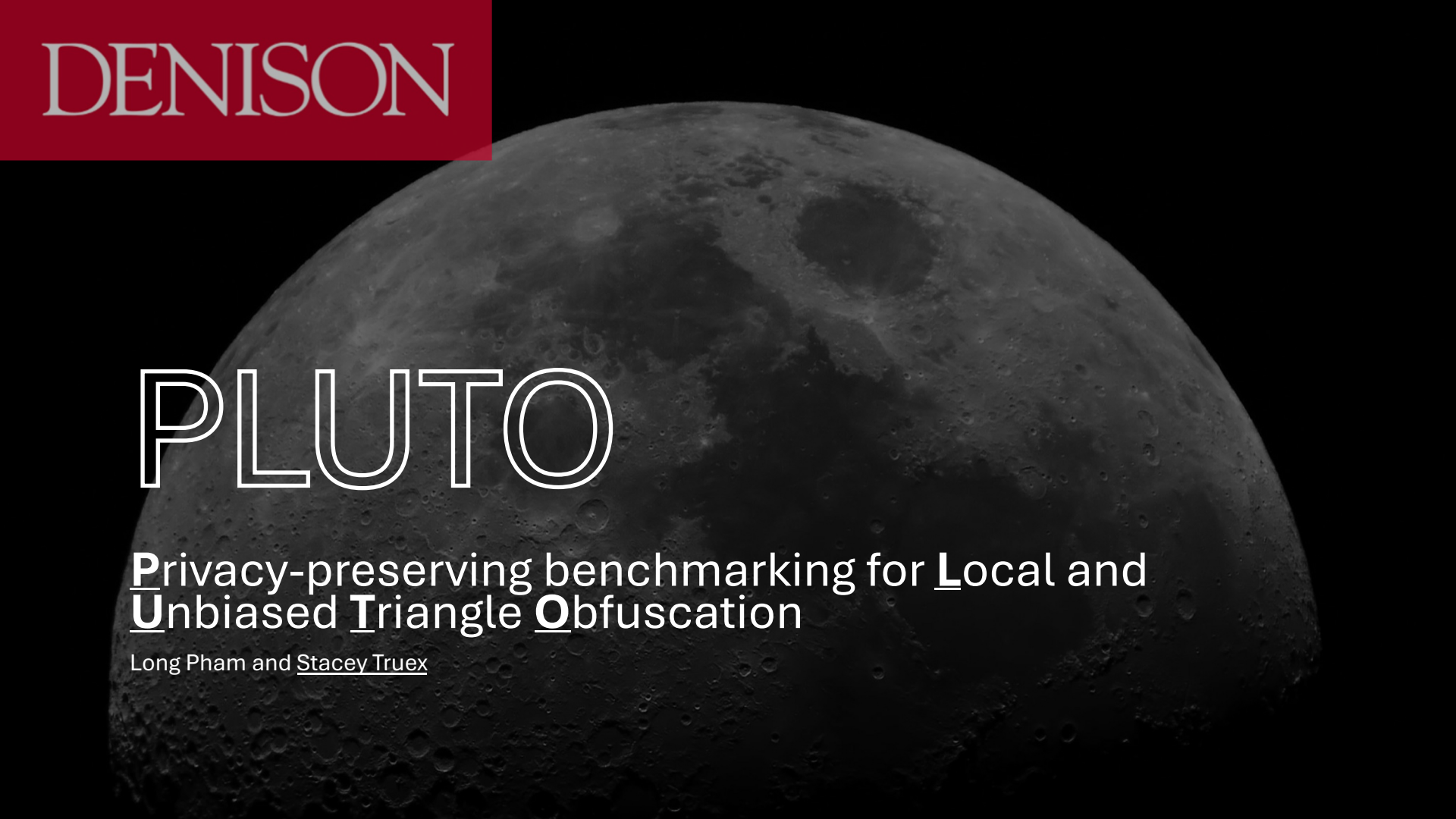
DENISON



PLUTO

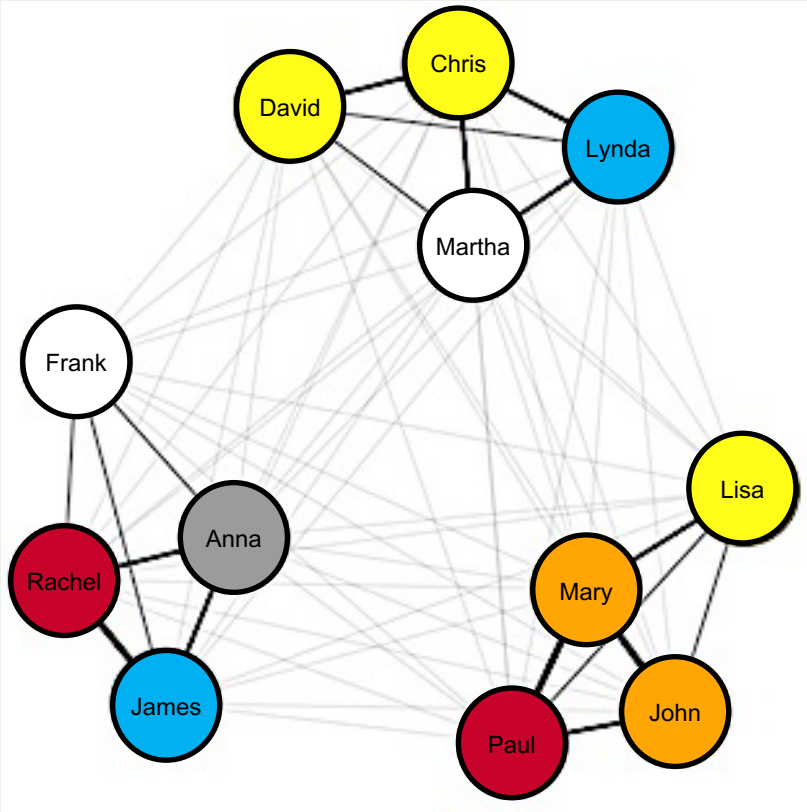
Privacy-preserving benchmarking for **L**ocal and **U**nbiased **T**riangle **O**bfuscation

Long Pham and Stacey Truex

DENISON

Triangle Counting

- Fundamental tool in graph analysis
- Applications:
  - clustering behavior
  - network transitivity
  - community discovery
  - link prediction
  - spam filtering

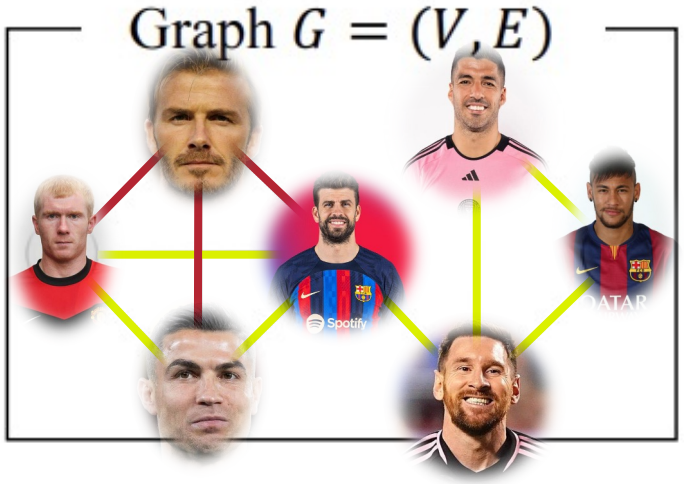


DENISON

Privacy in Triangle Counting

Scenario: Consider the graph representing a social network

- David Beckham considers his network (adjacency list) private
- Lionel Messi knows the friend list of Ronaldo, Suárez, Scholes, Piqué, Neymar and Ronaldo except: if they are connected to Beckham
- **Q:** Does output of triangle counting algorithm $\mathcal{A}$ compromise Beckham's privacy w.r.t. Messi?



Privacy in Triangle Counting

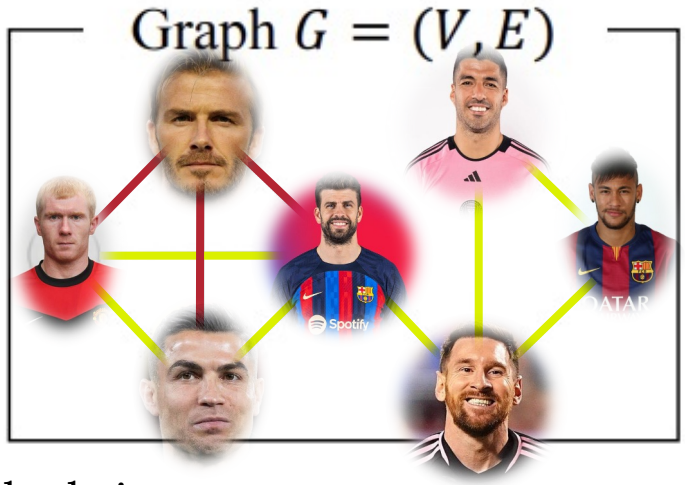
Observe: $\mathcal{A}(G) = 5$

known:

$\Delta\, v_6 - v_3 - v_4$

$\Delta v_2 - v_7 - v_5$

Messi knows he is not connected to Beckham



Q: Can 5 triangles be created in this graph with no connection between Ronaldo and Beckham?

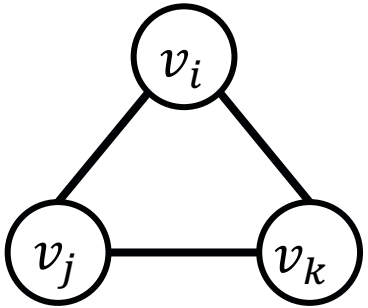
No! ⇒ Beckham's connection to Ronaldo leaks!

DENISON

Privacy in Triangle Counting

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

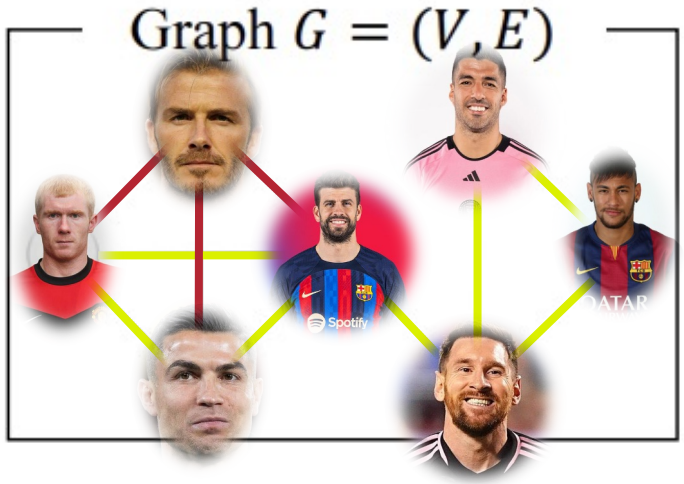
Results from a statistical query

ex: use $\mathcal{A}(G) = 5$ triangles in G to infer private information (Beckham is connected with Ronaldo)

DENISON

Privacy in Triangle Counting

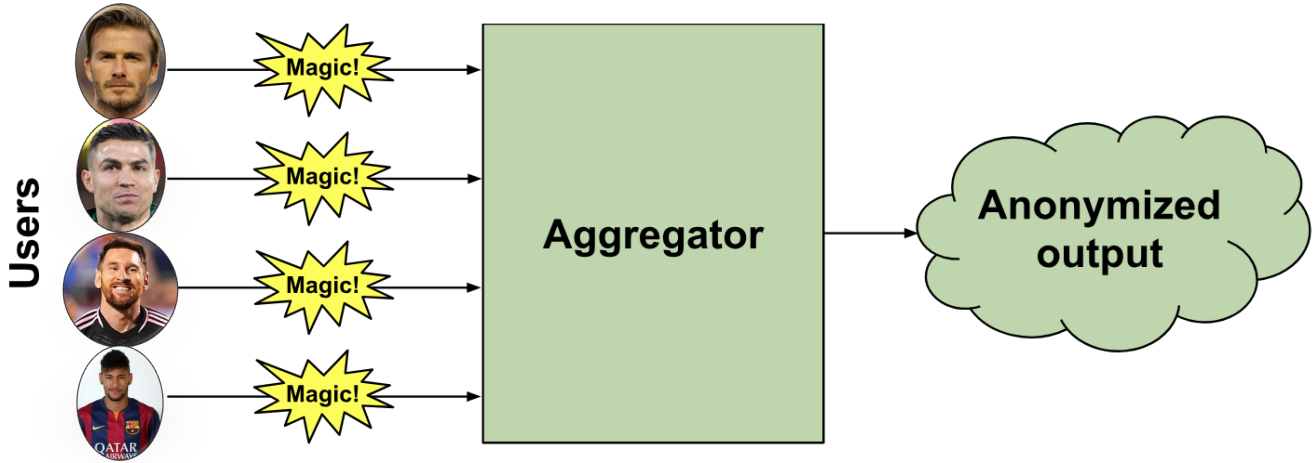
Problem: How to guarantee privacy of individual connections within a social network?



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

Local Differential Privacy (LDP)

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

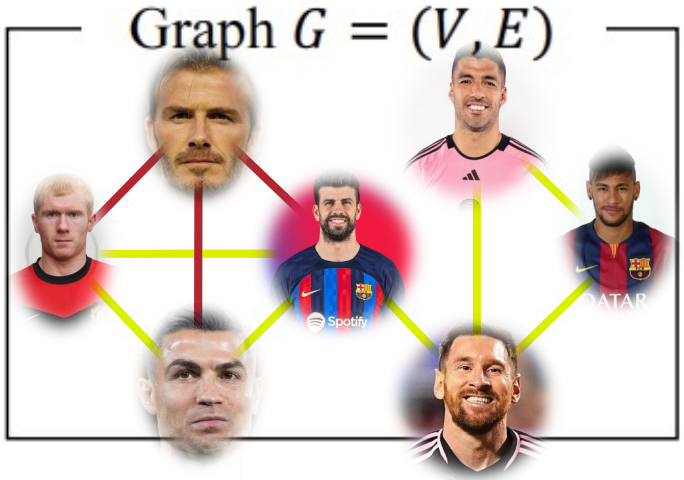


DENISON

Applying LDP to Graph Analysis

Definition: ϵ-edge LDP. 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 $S \subseteq Range(\mathcal{R}_i)$,

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



Local Perturbation

Shape	Task	Count
	Triangle	2
	2-star	18

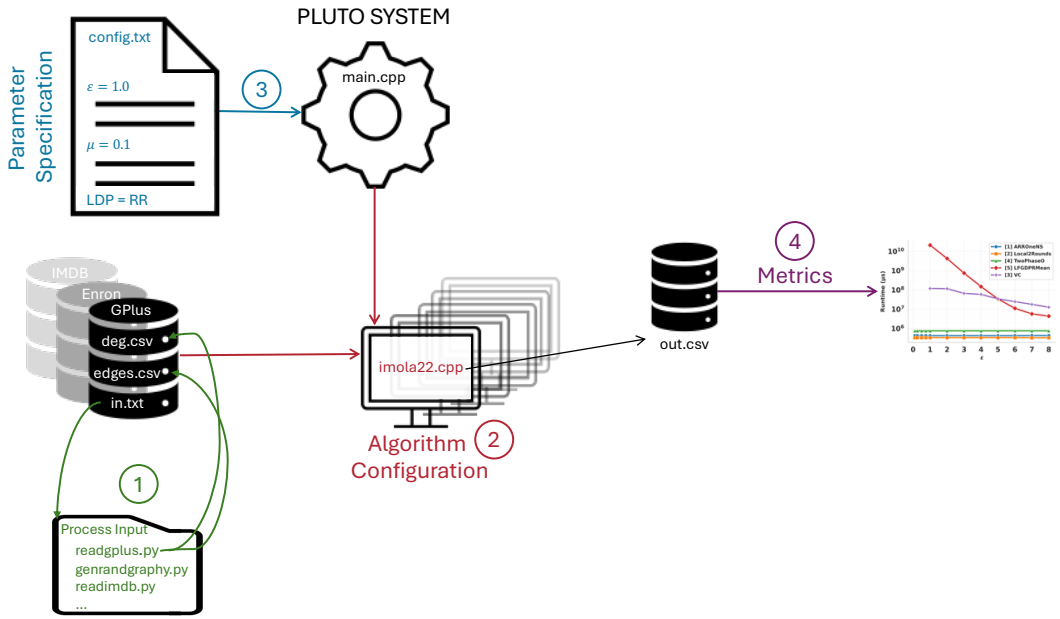
Privacy-preserving Subgraph Counting

User Side **Data Collector**

PLUTO System

1. Addresses the lack of a standardized experimental pipeline

2. Provides a unified & extensible benchmarking framework for evaluating LDP triangle counting algorithms



DENISON

PLUTO System

1. Data Processing

- SNAP networks
- Synthetic graphs: Erdös-Rényi & Barabási-Albert

2. Algorithm Configuration

- Each algorithm is implemented as a self-contained module
- Add self-contained module & update build configurations

DENISON

PLUTO System

3. Parameter Specification

- General parameters:
  - ε: controlling privacy budget, n: number of nodes control input size
- Algorithm-specific tuning parameters:
  - μ: edge sampling to reduce noise
  - privacy-budget allocation: allocate ε non-uniformly across steps
- Choice of LDP mechanism: RR, OUE

4. Metric Evaluation

- Accuracy: Relative Error & L2-Loss
- Runtime in microseconds
- Communication overhead: upload & download cost

DENISON

Empirical Evaluations

Algorithms:

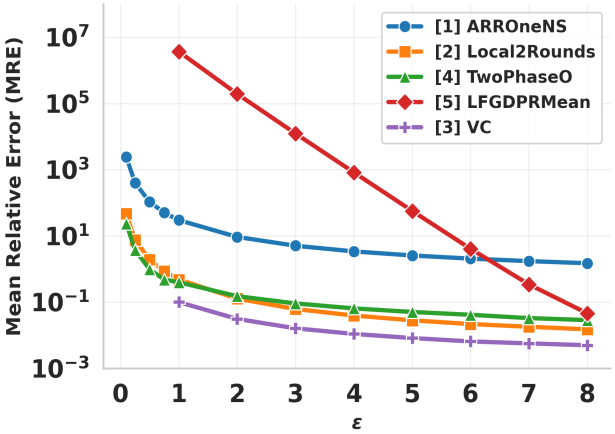
1. Local2Rounds:- 2-round protocol with edge clipping
2. ARROneNS: communication-efficient extension of Local2Rounds
3. TwoPhaseO: an approach leveraging user's extended local views
4. LFGDPRMean: 1-round protocol, uses Δs to compute clustering coefficient
5. VC: 3-round protocol based on pairs of vertices

Experiments:

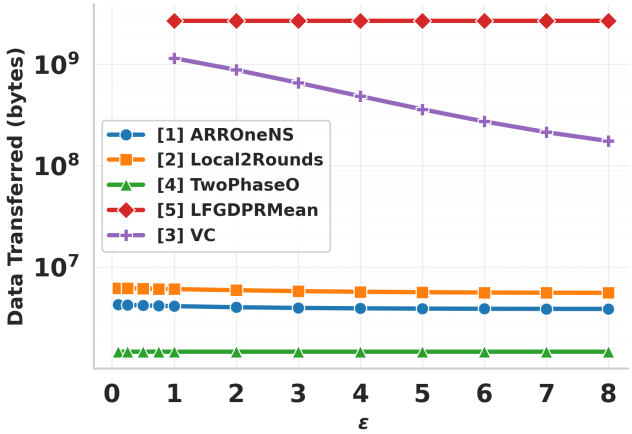
1. Real-world graph evaluation with Enron email communication graph
2. Synthetic graph evaluations with Erdös-Rényi and Barabási-Albert graphs

DENISON

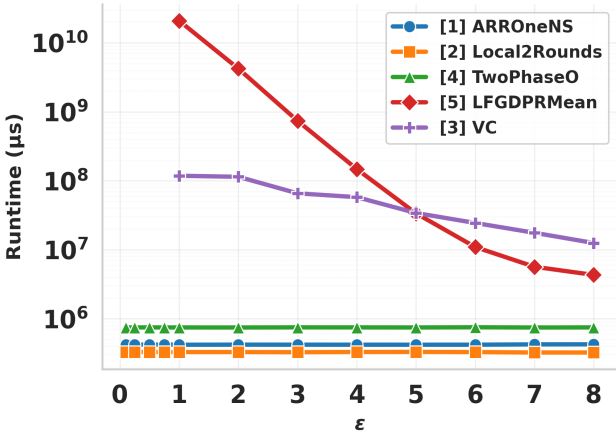
Varying ε: Real-World Network-Enron



- VC lowest
- LFGDPRMean only effective with high privacy budget ($\varepsilon \geq 5$)
- Local2Rounds more accurate than ARROneNS, at the cost of more communication overhead



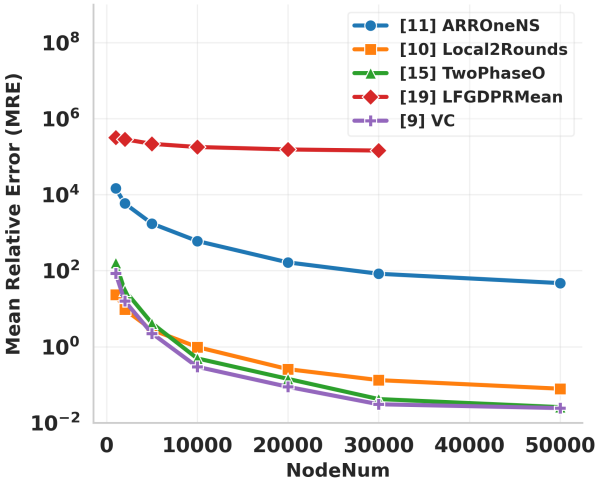
- LFGDPRMean incurs high computation because triangle counting is performed over the entire noisy graph.
- VC limits computation scope to the local neighborhood
- ARROneNS and Local2Rounds have low complexity



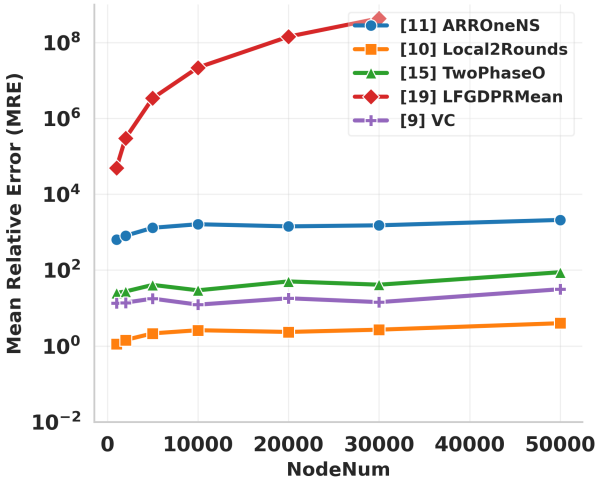
- TwoPhaseO lowest: avoid transferring adjacency vectors, count triangle on unperturbed graph
- LFGDPRMean highest: each participant transmits roughly half of an adjacency bit vector
- The other three saw decreasing runtime at different rates

DENISON

Varying Num Nodes: Synthetic Graphs



ER graph, p = 0.01, $\varepsilon = 1$

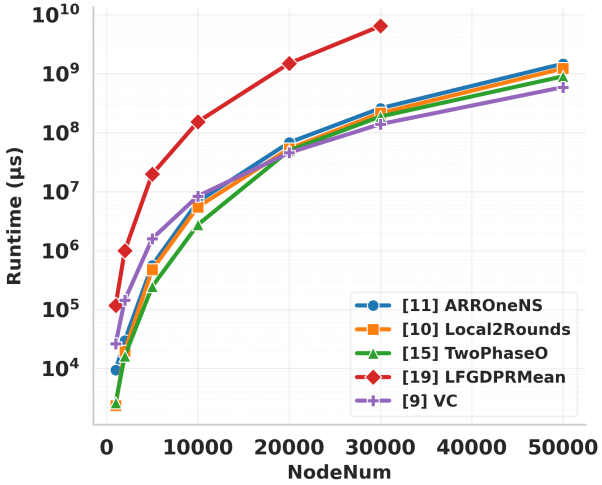


BA graph, m = 5, $\varepsilon = 1$

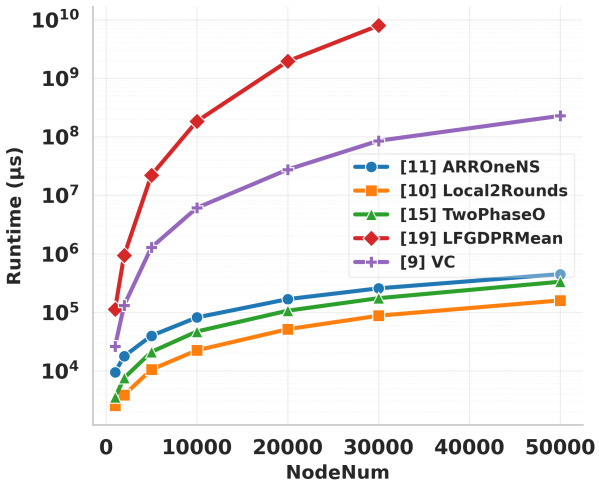
- Accuracy trends and ranking remain relatively stable
- Error increases in BA graphs. As n increases, ER density stays fixed while BA gets sparser.

DENISON

Varying Num Nodes: Synthetic Graphs



ER graph, p = 0.01, $\varepsilon = 1$

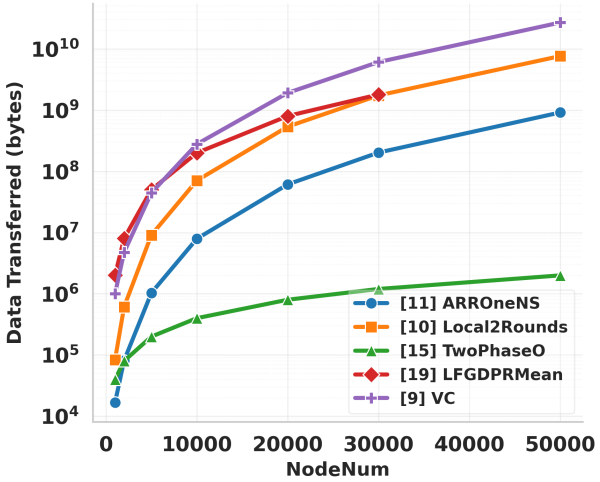


BA graph, m = 5, $\varepsilon = 1$

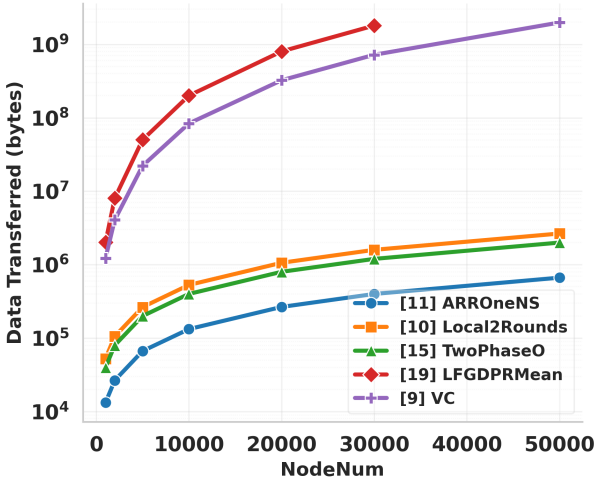
- Runtime different in ER: Local2Rounds, ARROneNS, and TwoPhaseO saw increased runtime, while in Enron and BA they are 1~2 order of magnitudes faster
- Computational efficiency vary substantially across graph topologies, even when node counts are similar

DENISON

Varying Num Nodes: Synthetic Graphs



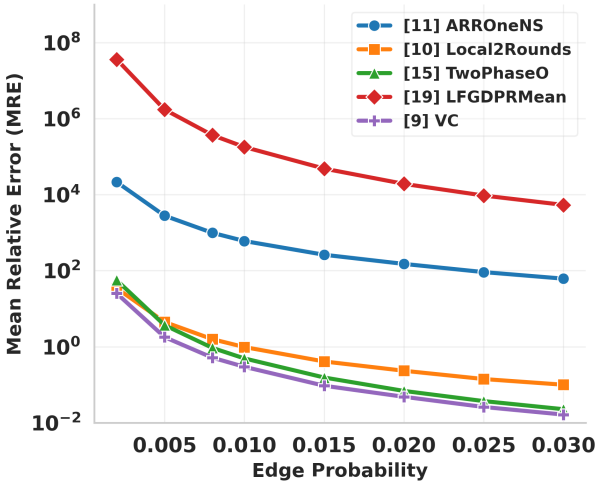
ER graph, p = 0.01, $\varepsilon = 1$



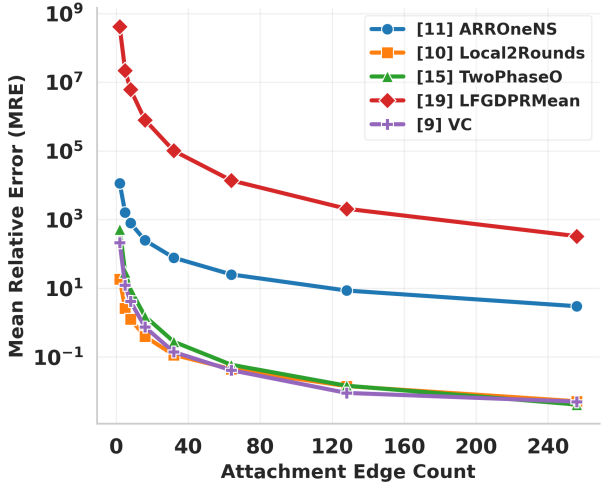
BA graph, m = 5, $\varepsilon = 1$

- ER: Overhead of VC exceeds LFGDPRMean at NodeNum ≥ 10000 , unlike in Enron where VC always incurs lower overhead
- BA: ARROneNS, Local2Rounds, TwoPhaseO are clearly separated from the other two. However, their communication costs are much closer to each other than in ER.

Varying Density: Synthetic Graphs



ER graph, n=10000, $\varepsilon = 1$

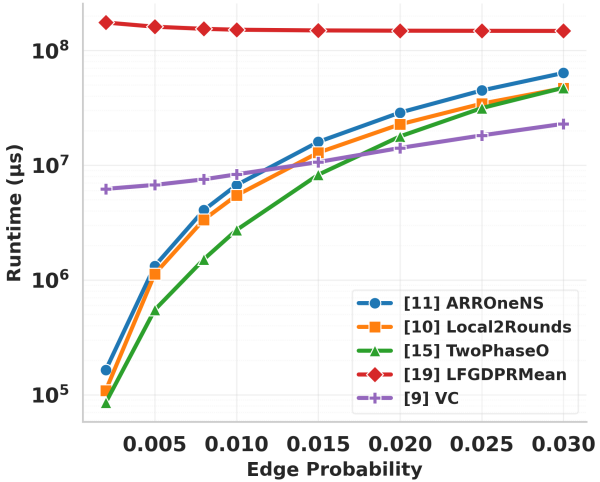


BA graph, n=10000, $\varepsilon = 1$

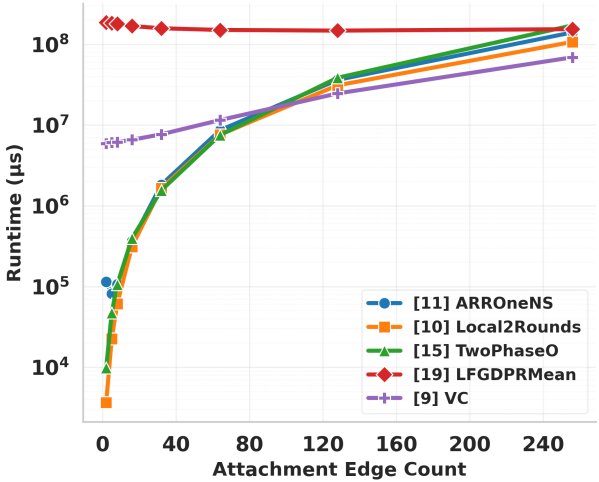
- ER: Increasing p exhibits similar trends as increasing n
- BA: fixing n and increasing m lead to increasing density → trends similar to in ER

DENISON

Varying Density: Synthetic Graphs



ER graph, n=10000, $\varepsilon = 1$



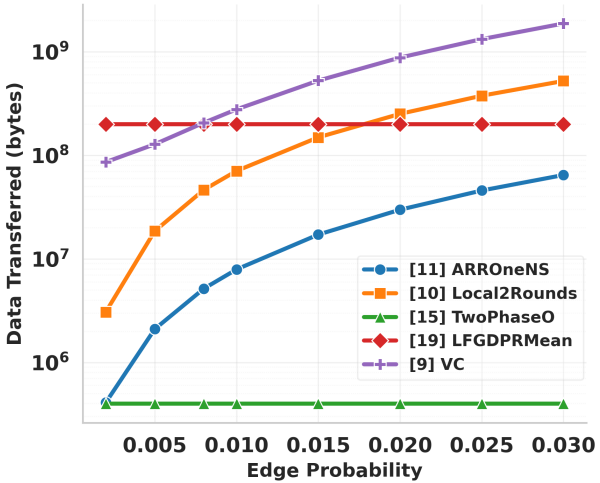
BA graph, n=10000, $\varepsilon = 1$

- Slight decrease in runtime for LFGDPRMean: As the underlying graph becomes denser, randomized response under a fixed privacy budget ($\epsilon = 1$) produces comparatively sparser noisy graphs, reducing triangle counting cost.

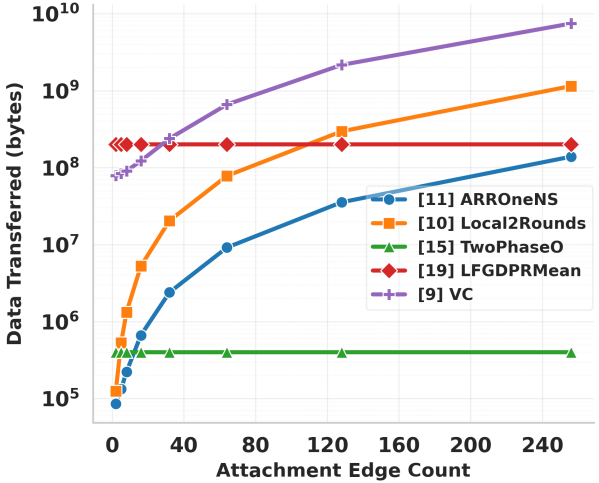
- VC's runtime grows at a slower rate compared to the other three
  → achieves lowest runtime at $p \geq 0.02$ and $m \geq 128$

DENISON

Varying Density: Synthetic Graphs



ER graph, n=10000, $\varepsilon = 1$



BA graph, n=10000, $\varepsilon = 1$

- Communication of LFGDPRMean and TwoPhaseO depends on number of nodes rather than graph density → communication remains relatively stable
- LFGDPRMean exhibits lower communication cost than VC and Local2Rounds at certain thresholds of p and m, unlike in Enron graph where it is consistently higher

DENISON

Insights

1. Graph topology substantially influences algorithm behavior
   Algorithm rankings are not invariant to graph structure, even if the graph sizes are similar
2. Communication-efficient methods incur predictable utility trade-offs
   Reducing communication overhead leads to increased error.
   Runtime difference between these ARROneNS and Local2Rounds remains mostly unchanged
   → accuracy-communication tradeoff can be tuned independently of computational cost.
3. TwoPhaseO is effective under permissive privacy assumptions
   Effective when exposing neighbor lists is acceptable and communication budgets are limited
   (transmission cost only scales with n)

DENISON

Insights

4. LFGDPRMean performs best under weaker privacy requirements
    - Low accuracy and high runtime except at large ϵ
    - Overhead is only dependent on n

5. VC consistently achieves the strongest accuracy
    - Highest accuracy among evaluated methods
        - at relatively high communication cost
    - Extra round to report noisy degrees contributes to communication overhead

DENISON

THANK YOU!

truexs@denison.edu