Differentially Private Community Detection with Privacy Amplification by Shuffling

Xincheng Xu[1] M.A.P. Chamikara[2] Thilina Ranbaduge[2]
Thierry Rakotoarivelo[2] David Smith[2] Qing Wang[1]

[1] Australian National University, Australia

[2] CSIRO, Australia

21st International Workshop on Data Privacy Management (DPM 2026)

Background

Community Detection

- Social group discovery
- Biological analysis
- Communication analysis

Privacy Risks

- Private social connections
- Political affiliations
- Medical associations

Goal

Preserve community-level utility while protecting individual edges.

Differential Privacy on Graphs

Differential Privacy (DP)

A randomized mechanism $\mathcal{M}$ is (ϵ, δ)-DP if for any neighboring inputs D, D':

$$\Pr[\mathcal{M}(D) \in \mathcal{S}] \leq e^{\epsilon} \Pr[\mathcal{M}(D') \in \mathcal{S}] + \delta.$$

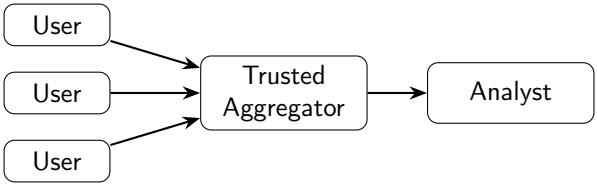
Edge Differential Privacy

For graph data, we define neighboring graphs by one edge change:

$$G \sim G' \quad \Longleftrightarrow \quad G \text{ and } G' \text{ differ in one undirected edge.}$$

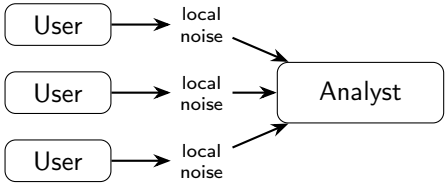
Central vs. Local Edge DP

Central Differential Privacy



- Often better utility.
- Requires a trusted data aggregator.

Local Differential Privacy

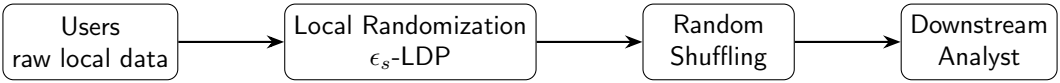


- No original graph is collected.
- Strong local noise can hurt utility.

Motivation

Can we reduce the noise of local privacy without giving a aggregator access to the original graph?

Privacy Amplification by Shuffling



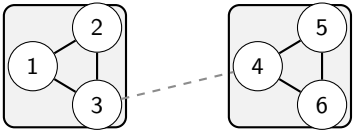
Amplification Effect

After anonymous shuffling, the analyst obtains a stronger $(\epsilon_{\text{amp}}, \delta)$ privacy guarantee, where typically

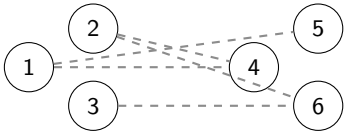
$$\epsilon_{\text{amp}} < \epsilon_s.$$

- Shuffling breaks the link between a report and its source.
- Larger shuffle sets provide stronger privacy amplification.

Why graph shuffling is non-trivial

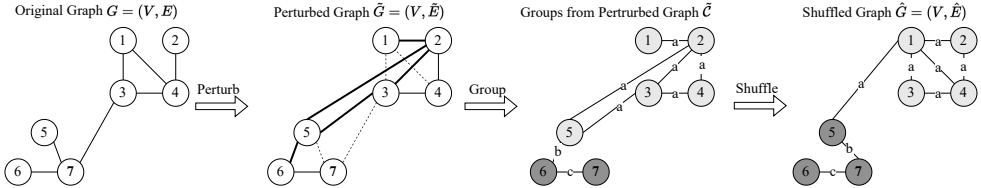


community structure preserved



global shuffling loses locations

SDPCD protocol overview



Stage 1: grouping

$$A \to \tilde{A}^{(1)} \to \mathcal{C}$$

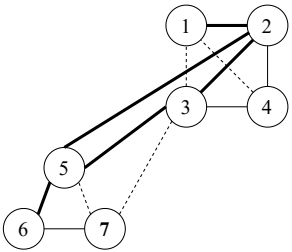
Identify shuffle groups.

Stage 2: shuffling

$$A \to \tilde{A}^{(2)} \to \hat{A}$$

Shuffle within fixed groups.

Stage 1.a: perturbation via RABV



	User 1	User 2	User 3	User 4	User 5	User 6	User 7
User 1	0	1	0	0	0	0	0
User 2	1	0	1	1	0	0	0
User 3	0	1	0	1	1	0	0
User 4	0	1	1	0	0	0	0
User 5	0	0	1	0	0	1	0
User 6	0	0	0	0	1	0	1
User 7	0	0	0	0	0	1	0

Randomized Adjacency Bit Vector (RABV)

- Each user perturbs selected adjacency bits using randomized response.
- Each undirected edge is perturbed exactly once.

Stage 1.b: grouping

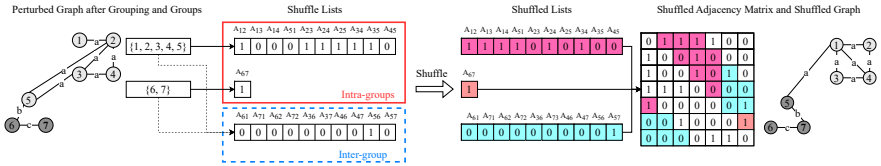
From noisy graph to a fixed partition

$$\mathcal{C}_{\text{tmp}} \leftarrow \text{Louvain}(\tilde{A}^{(1)}), \qquad |C_i| \geq k.$$

Small communities are iteratively merged with their most similar neighbour until the minimum group size is satisfied.

The partition $\mathcal{C}$ is fixed before Stage 2 starts.

Stage 2: shuffling



Shuffle lists

$$Sin_i = \{\tilde{A}^{(2)}_{uv} : u < v,\ u, v \in C_i\}, \qquad Sout_{ij} = \{\tilde{A}^{(2)}_{uv} : u \in C_i,\ v \in C_j\}.$$

Each list is shuffled independently and reconstructed into $\hat{A}$.

Privacy Guarantee

Privacy Budgets

$$\epsilon_{s1} \quad + \quad \epsilon_{s2} \xrightarrow{\text{shuffling}} \epsilon_{\text{amp}}$$

ϵ_{s1} : grouping, ϵ_{s2} : fresh Stage-2 reports, $\epsilon_{\text{amp}} < \epsilon_{s2}$.

Main Guarantee

With minimum group size k, SDPCD achieves

$$\boxed{(\epsilon_{s1} + \epsilon_{\text{amp}}, \delta)\text{-Edge DP}}$$

against the downstream analyst.

Experimental Setup

Datasets

Dataset	$\lvert V\rvert$	$\lvert E\rvert$	$\lvert\mathcal{C}\rvert$
Facebook	4,039	88,234	17
Government	7,057	89,455	29
AstroPh	18,772	198,110	334
Enron	36,692	183,831	1,363

Metric: Modularity

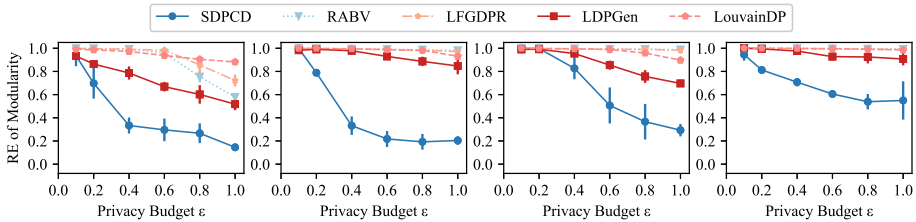
$$q = \sum_{i=1}^{r} \left[\frac{\ell_{C_i}}{|E|} - \left(\frac{d_{C_i}}{2|E|} \right)^2 \right]$$

Higher q indicates stronger community structure.

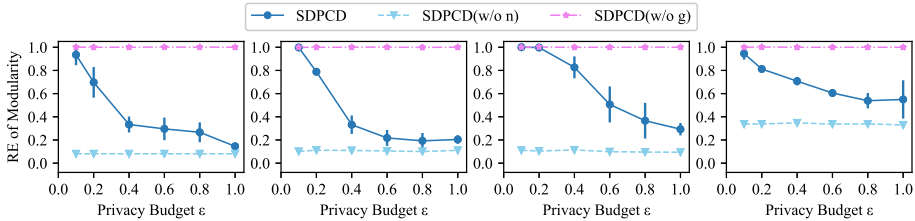
Privacy Settings

$$\delta = 10^{-5}, \quad \epsilon_{\text{total}} \in \{0.1, 0.2, 0.4, 0.6, 0.8, 1.0\}.$$

Modularity relative error



Ablation Study



Conclusion

1. SDPCD brings shuffle amplification to graph community detection.
2. Two-stage privatization gives an amplified privacy proof.
3. Experiments show better privacy–utility trade-offs than prior baselines.

Future work: adaptive budget split, adaptive group size k, overlapping communities, and distributed shuffling.

Thank You!

Xincheng Xu
xincheng.xu@anu.edu.au
Australian National University, Australia