DISTRIBUTED RANDOMIZED SAMPLING FOR LOCALLY PRIVATE INDEPENDENT SETS

Anonymous authors
Paper under double-blind review

ABSTRACT

Many randomized distributed graph algorithms rely on simple local rules and few communication rounds, making them natural starting points for locally private protocols. We explore this connection through the maximum independent set problem under local edge differential privacy. Previous work shows that an edge-private algorithm that always publishes a valid independent set can select at most one vertex. We therefore allow edges between selected vertices and count each such edge as a defect. We develop a noninteractive ε-local edge-private algorithm that samples vertices independently using privately perturbed degrees. For fixed $\varepsilon > 0$ and n-vertex graphs with bounded average degree or degeneracy, the expected output size of our algorithm is $\Omega_\varepsilon(n/\log n)$ and the expected defect count is $O_\varepsilon(n/\log^2 n)$. Moreover, we also show that the ratio of defects to output size tends to zero with high probability. For a fixed ε and sufficiently large n, we also prove that any ε-edge-DP or ε-edge-LDP algorithm achieving an expected approximation ratio $r \leq (n+1)/4$ on every n-vertex graph with at most n edges must incur $\Omega(n/r^2)$ expected defects on some graph in this class. For $r = \Theta(\log n)$, the lower and upper bounds on the expected defect count match asymptotically.

1 INTRODUCTION

Differential privacy (DP) limits how much an algorithm's output distribution can change when its input changes in one protected entry (Dwork et al., 2006). In the central model, a trusted curator accesses the data and releases a privatized result. Local differential privacy (LDP) removes this trust requirement: users randomize their own records before communicating with the curator (Kasiviswanathan et al., 2011; Duchi et al., 2013). For graph data held as adjacency lists, we use *local edge privacy*, which protects a one-bit change in one user's adjacency vector rather than replacement of the entire list (Qin et al., 2017; Imola et al., 2021). Privacy guarantees also depend on which outputs are visible. In the *billboard model*, users combine a differentially private public transcript with their own records to compute personalized outputs that remain privately held (Kearns et al., 2014; Hsu et al., 2016; Dinitz et al., 2025).

In this paper, we investigate how randomized distributed graph algorithms can be adapted to local differential privacy. Randomized distributed graph algorithms form a well-established research area, with algorithms for combinatorial optimization problems including independent set, matching, and dominating set (Luby, 1986; Alon et al., 1986; Boppana et al., 2018; Lotker et al., 2015; Jia et al., 2002). Many use simple local rules and require few communication rounds, making them promising starting points for locally private protocols. In LDP, repeated private messages from the same user can accumulate privacy loss through composition; reducing the number of such messages can therefore help control the total privacy loss for a fixed per-message budget. Moreover, their use of randomization offers a natural starting point for incorporating the randomized mechanisms required by LDP. These shared features motivate exploring the connection, although neither few communication rounds nor randomization alone guarantees local privacy.

We study this question through *maximum independent set*. For a graph $G = (V, E)$, an independent set is a vertex subset containing no adjacent pair, and the objective is to maximize its size, denoted at optimum by $\alpha(G)$. Classical randomized algorithms construct maximal independent sets (Luby, 1986; Alon et al., 1986), while Boppana's random-priority rule gives a simple local approximation to the maximum independent set (Boppana et al., 2018). Under DP, a relevant obstruction follows

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

from work on the complementary vertex-cover problem: Gupta et al. (2010) show that any pure edge-DP algorithm that always releases a valid vertex cover must include at least $n - 1$ vertices with probability one. Taking complements implies that an always-independent set released under the same privacy notion can contain at most one vertex.

1.1 Our contributions

Boppana's algorithm in the billboard model. We observe that Boppana's random-priority algorithm satisfies local edge privacy in the billboard model when the curator publishes graph-independent priorities and users keep their membership decisions private (Section 2.3). The resulting set is independent and has expected size $\sum_{v \in V} 1/(d_v + 1)$, where d_v denotes the degree of vertex v. This gives an $O(1 + \Delta)$ approximation ratio, where $\Delta = \max_v d_v$. This matches (up to polylog factors in Δ) the $\tilde{\Omega}(\Delta)$ approximation barrier for polynomial-time maximum independent set algorithms even without locality or privacy constraints, assuming $\mathrm{NP} \not\subseteq \mathrm{BPP}$ (Chan, 2013).

This illustrates how the billboard model can offer stronger utility guarantees than a model requiring the selected set to be publicly released. It is therefore well suited to applications in which users can act on privately held outputs without revealing their decisions. For example, suppose each agent has a job to perform and edges represent resource conflicts between agents' jobs. If each agent knows its conflicts and can execute its job without revealing its participation to other agents, the billboard model allows the agents to select a conflict-free set of jobs without publishing the schedule.

Locally private sampling with defects. Without the billboard model, the vertex-cover obstruction of Gupta et al. (2010) rules out nontrivial always-independent, publicly released sets under edge DP. We therefore relax independence and allow *defects*, counting each edge induced by the released set as one defect. Inspired by Boppana's degree-sensitive inclusion probabilities, we develop a noninteractive edge-LDP algorithm that independently samples vertices using privately perturbed degrees (Section 3). Let S be the output node set and $e(S)$ be defect count in S. For fixed $\varepsilon > 0$ and graphs of bounded average degree or degeneracy, Corollary 3.3 proves $\mathbb{E}[|S|] = \Omega_\varepsilon(n/\log n)$ and $\mathbb{E}[e(S)] = O_\varepsilon(n/\log^2 n)$. Consequently, the ratio of expected defects to expected solution size satisfies $\mathbb{E}[e(S)]/\mathbb{E}[|S|] = O_\varepsilon(1/\log n)$ and tends to zero as n grows. Moreover, when $\mathbb{E}[|S|] \to \infty$, Proposition A.1 also implies $e(S)/|S| = o_\varepsilon(1)$ with high probability. Sparsity assumptions, such as bounded degree or degeneracy, are also commonly used to obtain stronger guarantees for graph algorithms under LDP, including degree-sensitive subgraph counting (Imola et al., 2021; Hillebrand et al., 2025a) and cycle counting on degeneracy-bounded graphs (Hillebrand et al., 2025b).

Defect-count lower bounds. For any fixed privacy budget and sufficiently large n, an edge-DP or edge-LDP algorithm with expected approximation ratio $r \le (n + 1)/4$ on every graph with at most n edges must incur $\Omega(n/r^2)$ expected defects on some graph in this class, where the hidden constant depends on the privacy budget (Section 4). The unrestricted result follows immediately because a guarantee on every graph includes this sparse class. Thus, for fixed $\varepsilon > 0$ and $r = \Theta(\log n)$, the edge-LDP lower bound $\Omega_\varepsilon(n/\log^2 n)$ matches our defect upper bound on bounded-average-degree graphs up to constants depending on ε.

Experimental evaluation. Experiments on Barabási-Albert and public graphs demonstrate a bicriteria size–defect tradeoff (Section 5). On the sparser inputs, Algorithm 1 generally releases larger sets than randomized response followed by greedy selection, although it accepts more defects; the baseline becomes relatively more competitive as graph density increases. In the $BA(n, 10)$ scaling experiment at $\varepsilon = 1$, for example, our mean output grows from 11.94 vertices at $n = 500$ to 60.07 at $n = 3000$, whereas the baseline grows from 15.56 to 21.95. The one-bit sampler runs in under 10^{-3} seconds on every tested input and remains feasible on PubMed, where the randomized response algorithm is impractical.

1.2 Related work

Combinatorial optimization under central DP. Combinatorial optimization and graph problems have been widely studied in the central DP model. Gupta et al. (2010) study private approximation algorithms for problems including vertex cover, set cover, minimum cut, and facility location. Other work addresses densest subgraphs under central edge privacy (Dinitz et al., 2025). Particularly relevant to our work is graph colouring with defects. Christiansen et al. (2026) study edge-private

colourings that allow each vertex to have a bounded number of neighbours of the same colour, and establish tradeoffs between the number of colours and defectiveness. Xie et al. (2026) also develop central edge-private colouring algorithms that control this maximum per-vertex defect.

Combinatorial optimization under LDP. In the local model, edge-private algorithms address densest subgraphs (Dhulipala et al., 2022; Dinitz et al., 2025; Dhulipala et al., 2025), core decomposition (Dhulipala et al., 2022; 2025; Mundra et al., 2025), and low out-degree ordering (Dhulipala et al., 2022; 2025). Related work studies subgraph counting (Imola et al., 2021; Hillebrand et al., 2025a;b; Suppakitpaisarn et al., 2025), including communication-efficient protocols and algorithms with degeneracy-sensitive error guarantees.

LDP in the billboard model. The billboard approach separates a differentially private public transcript from personalized outputs computed using users' own records, as in work on private games, matchings, and allocations (Kearns et al., 2014; Hsu et al., 2016). In an LDP implementation, the transcript must be generated through locally private communication rather than by a trusted curator; keeping the final outputs private does not alone ensure LDP. For graphs, Dinitz et al. (2025) develop local edge-private algorithms for implicit matchings.

Facility location and output visibility. Facility location has been studied under LDP both with a *superset assumption* (Cohen-Addad et al., 2022) and without it (Pfisterer et al., 2025). Under this assumption, the mechanism releases candidate facilities, clients privately determine their assignments, and only facilities serving clients incur opening costs. This resembles the billboard model's use of a differentially private public transcript and privately computed outputs (Kearns et al., 2014; Hsu et al., 2016), but the superset assumption determines which facilities are counted in the objective, whereas the billboard model determines which outputs are visible.

Node and weight privacy. Node-DP protects a vertex's entire incident-edge set, rather than a single edge (Kasiviswanathan et al., 2013). Its local analogue, node-LDP, imposes this stronger protection on locally generated messages. The impossibility results of Raskhodnikova et al. (2026) highlight the severe utility limitations of node-LDP for graph analysis. These limitations motivate our decision not to consider node-LDP in this work and to focus instead on local edge privacy. We use these results as motivation for our choice of privacy model, rather than as a direct impossibility theorem for the independent-set guarantees studied here.

Weight-DP instead protects edge weights on a publicly known graph topology, as in private shortest-path and distance release (Sealfon, 2016; Fan et al., 2022). Weight-LDP provides a local counterpart in which users privatize their weight information before communicating; recent work studies below-threshold triangle counts under this notion (Pfisterer et al., 2026). These models concern sensitive numerical weights, whereas our edge-LDP formulation protects the presence or absence of the conflicts that define an independent set.

2 Preliminaries

2.1 Notations

Let $G = (V, E)$ be a simple undirected graph on a public vertex set V of size n. Write $N(v)$ for the neighbours of v, $d_v = |N(v)|$, $\Delta = \max_v d_v$, and $\bar{d} = 2|E|/n$. Let $\alpha(G)$ denote the independence number. For $S \subseteq V$, let $G[S]$ denote the subgraph induced by S: its vertex set is S, and its edge set $E(G[S])$ consists of all edges of G with both endpoints in S. Write $e(S) = |E(G[S])|$ for the number of these edges. We call $e(S)$ the *defect count*: each induced edge is one violated independence constraint. It is not the number of conflicting vertices or the maximum induced degree. All logarithms are natural. For a released set S, we study the two quantities $\mathbb{E}[|S|]$ and $\mathbb{E}[e(S)]$ separately. An approximation ratio $r \geq 1$ in expectation means $\mathbb{E}[|S|] \geq \alpha(G)/r$. Each user v holds an adjacency vector $a_v \in \{0,1\}^n$. On graph inputs, $(a_v)_u = (a_u)_v$ and $(a_v)_v = 0$.

2.2 Differential Privacy

Definition 2.1 (Edge privacy (Imola et al., 2021, Definition 1))**.** Graphs G, G' on the same vertex set are neighbouring if they differ in one undirected edge. A mechanism M satisfies *ε-edge differential*

privacy (ε-edge-DP) if, for every such pair and every measurable events F, $P(M(G) \in F) \leq e^{\varepsilon} P(M(G') \in F)$.

In this paper, we consider only noninteractive local edge privacy and refer to it simply as local edge privacy.

Definition 2.2 (Local edge privacy)**.** Following the local edge-privacy formulation of Imola et al. (2021), a local randomizer R is *ε-edge-LDP* if, for all adjacency vectors a, a' differing in one coordinate and all measurable events F, $P(R(a) \in F) \leq e^{\varepsilon} P(R(a') \in F)$. An *$\varepsilon$-edge-LDP protocol* fixes one such randomizer R_v per user before observing any messages. Each user applies a randomizer R_v to its own input using independent private randomness and sends the resulting message to the curator. The protocol has the form $M(a_1, \ldots, a_n) = \mathcal{A}(R_1(a_1), \ldots, R_n(a_n))$, where $\mathcal{A}$ is the curator's aggregation function.

Definition 2.3 (LEDP in the billboard model, adapted from (Hsu et al., 2016; Dinitz et al., 2025))**.** A protocol is *ε-edge-LDP in the billboard model* if it constructs a public transcript B using an ε-edge-LDP protocol and each user v computes a privately held output $o_v = f_v(a_v, B)$. The final outputs are not sent back to the curator. Privacy is defined on adjacency lists, under a one-bit change in one user's adjacency vector.

2.3 Boppana's algorithm

Boppana's randomized algorithm selects vertices according to a uniformly random ordering (Boppana et al., 2018). Let $\pi : V \to \{1, \ldots, n\}$ be a uniformly random bijection. A vertex joins the set if it precedes every neighbour in this ordering, so $I = \{v \in V : \pi(v) < \pi(u) \text{ for all } u \in N(v)\}$. For every ordering, I is independent, because the later endpoint of any edge cannot be selected. Each vertex is first among itself and its d_v neighbours with probability $1/(d_v + 1)$. Thus linearity of expectation and convexity give $\mathbb{E}[|I|] = \sum_{v \in V} \frac{1}{d_v + 1} \geq \frac{n}{\bar{d}+1}$. Since $\alpha(G) \leq n$, the algorithm has an $O(1 + \bar{d})$ approximation ratio in expectation.

We can implement the same selection rule under local edge privacy under the billboard model. After calculating o_v, no user sends a message or reports its output.

Proposition 2.4 (Boppana in the billboard model)**.** *The preceding implementation is 0-edge-LDP in the billboard model.*

Proof. The public transcript contains only π, whose distribution is independent of all adjacency lists. Since no user releases information, its local message can be taken to be a constant randomizer. The transcript is therefore 0-private. □

Even without locality or privacy constraints, polynomial-time algorithms for maximum independent set face a $\tilde{\Omega}(\Delta)$ approximation barrier, assuming NP $\not\subseteq$ BPP for randomized algorithms (Chan, 2013); the tilde suppresses polylogarithmic factors in Δ. Since $\bar{d} \leq \Delta$, Boppana's algorithm in the billboard model achieves an $O(1 + \Delta)$ approximation ratio in expectation. It is therefore almost optimal. In contrast, Gupta et al. (2010) imply that any pure edge-DP algorithm that publicly releases an always-valid independent set can select at most one vertex. Thus, keeping membership decisions private in the billboard model enables substantially stronger utility than publicly releasing them under edge DP.

3 Locally private randomized sampling

Now we describe our sampling algorithm (Algorithm 1). Inspired by Boppana et al. (2018)'s decisions, let each user sample a candidate bit with a probability that decreases with their degree. Unlike the selections of Boppana et al. (2018), which depend on a shared random ordering, each user samples their candidate bit independently of all other users. We add Laplace noise to each user's degree to make their message locally private, and then add a positive shift to suppress excessive sampling caused by negative noise.

Proposition 3.1 (Privacy and communication)**.** *Algorithm 1 is a noninteractive ε-edge-LDP protocol and sends exactly one bit per user to the curator. The curator may release S publicly. On consistent graph inputs its transcript also satisfies central 2ε-edge-DP.*

Algorithm 1: One-bit locally private sampling without conflict removal

Input : Adjacency list a_v at each user; public $n \geq 2$ and $\varepsilon > 0$
Output: The curator-held set $S = \{v : z_v = 1\}$, possibly with defects

1 $c_\varepsilon \leftarrow \lceil \frac{4}{\varepsilon} \log n \rceil$;
2 **for** *each user v in parallel* **do**
3 $\quad L_v \sim \mathrm{Lap}(1/\varepsilon)$ independently;
4 $\quad p_v \leftarrow 1/(\max(d_v + c_\varepsilon + L_v, 0) + 1)$;
5 $\quad$ sample $z_v \sim \mathrm{Bernoulli}(p_v)$ independently and send z_v to the curator;
6 **end**

Proof. The degree $d_v = \sum_u (a_v)_u$ has sensitivity one under a one-bit change. Thus $d_v + L_v$ is an ε-private Laplace mechanism. Adding the public shift, truncating, and sampling a Bernoulli bit are randomized post-processing. Each local message therefore satisfies Definition 2.2, and constructing or publishing S is post-processing of the messages. Changing one undirected edge changes one bit in each of two endpoint records, so composing their privacy bounds gives 2ε for the full transcript. The noisy degrees and sampling probabilities themselves are not transmitted. □

We now analyze the utility of Algorithm 1. Our goal is to select many vertices while inducing few edges; these are the two components of the approximation guarantee defined in Section 2. Throughout this section, assume $n \geq 8$ and $\varepsilon > 0$, and write $c = c_\varepsilon$ for the shift defined in Line 1 of Algorithm 1. Define the degree-sensitive quantity

$$\Sigma := \sum_{v \in V} \frac{1}{d_v + c + 1/\varepsilon + 1}. \tag{3.1}$$

Theorem 3.2 (Released size and defect)**.** *The set S released by Algorithm 1 satisfies*

$$\mathbb{E}[|S|] \geq \Sigma, \qquad \mathbb{E}[e(S)] \leq \min\left\{ \frac{9}{16}\Sigma, \frac{|E|}{(c/2+1)^2} \right\} + \frac{1}{n}. \tag{3.2}$$

The first defect bound relates the expected number of violated constraints to the degree-sensitive size guarantee. The second is sharper on sparse graphs. We defer the full proof to Appendix A.2, and show the resulting approximation bounds for sparse graphs in Corollary 3.3.

Proof sketch of Theorem 3.2. For each vertex, let $D_v = \max(d_v + c + L_v, 0)$, $p_v = (D_v + 1)^{-1}$, and $q_v = \Pr(v \in S) = \mathbb{E}[p_v]$. The membership bits are independent, so $\mathbb{E}[|S|] = \sum_{v \in V} q_v$ and $\mathbb{E}[e(S)] = \sum_{uv \in E} q_u q_v$. Since $D_v \leq d_v + c + |L_v|$ and $\mathbb{E}[|L_v|] = 1/\varepsilon$, Jensen's inequality gives $q_v \geq (d_v + c + 1/\varepsilon + 1)^{-1}$, which yields the size bound after summing over vertices. For the defect bound, the Laplace tail and the choice of c imply $q_v \leq h_v + \eta$, where $h_v = (d_v + c/2 + 1)^{-1}$ and $\eta = (2n^2)^{-1}$. Expanding $\sum_{uv \in E}(h_u + \eta)(h_v + \eta)$ shows that the terms involving η contribute at most $1/n$. Moreover,

$$\sum_{uv \in E} h_u h_v \leq \min\left\{ \frac{1}{2} \sum_{v \in V} \frac{1}{d_v + c + 1}, \frac{|E|}{(c/2+1)^2} \right\}.$$

Finally, $\varepsilon(c+1) > 8$ implies $(d_v + c + 1)^{-1} \leq (9/8)(d_v + c + 1/\varepsilon + 1)^{-1}$. Combining these estimates gives the factor $9/16$ and proves (3.2).

The degree-sensitive guarantee translates into an approximation ratio depending on average degree, and therefore also on degeneracy.

Corollary 3.3 (Sparsity-sensitive bounds)**.** Algorithm 1 has an approximation ratio in expectation of $O(1 + \bar{d} + \varepsilon^{-1} \log n)$, or $O(1 + D + \varepsilon^{-1} \log n)$ on D-degenerate graphs. For fixed $\varepsilon > 0$ and bounded average degree or degeneracy, as n grows,

$$\mathbb{E}[|S|] = \Omega_\varepsilon(n/\log n), \qquad \mathbb{E}[e(S)] = O_\varepsilon(n/\log^2 n).$$

Consequently, $\mathbb{E}[e(S)]/\mathbb{E}[|S|] = O_\varepsilon(1/\log n)$. Moreover, Proposition A.1 implies that $e(S)/|S| \to 0$ with high probability as n grows.

Proof. By convexity and Theorem 3.2, $\mathbb{E}[|S|] \geq \Sigma \geq \frac{n}{\bar{d}+c+1/\varepsilon+1}$. Using $\alpha(G) \leq n$ and $c \leq 4\varepsilon^{-1}\log n+1$ proves the average-degree approximation ratio. On a D-degenerate graph, $|E| \leq Dn$ and therefore $\bar{d} \leq 2D$, giving the degeneracy bound. The edge-count estimate in Theorem 3.2 also yields $\mathbb{E}[e(S)] \leq \frac{n\bar{d}}{2(c/2+1)^2} + \frac{1}{n}$. For fixed ε, we have $c = \Theta_\varepsilon(\log n)$. With bounded $\bar{d}$, substitution gives the asserted size and defect estimates. Dividing the defect upper bound by the size lower bound proves that $\mathbb{E}[e(S)]/\,\mathbb{E}[|S|] = O_\varepsilon(1/\log n)$. Finally, as $\mathbb{E}[|S|] \to \infty$, Proposition A.1 implies $e(S)/|S| \to 0$ with probability at least $1-(\log n)^{-1/2} - e^{-\mathbb{E}[|S|]/8} = 1 - o_\varepsilon(1)$. □

Logarithmic sparsity. When $\bar{d} = o(\log n)$, the same proof implies that $e(S)/|S| \to 0$ with probability $1-(\frac{\bar{d}}{\log n})^{-\frac{1}{2}} - e^{-\frac{\mathbb{E}[|S|]}{8}} = 1-o_\varepsilon(1)$. The analogous implication also holds when $D = o(\log n)$.

Tuning the expected defect budget. The curator can reduce defects further without accessing the graph by independently retaining each selected vertex with a public probability θ: this post-processing preserves ε-edge-LDP, scales expected size by θ, and scales expected defects by θ^2. Thus, given a public upper bound $B > 0$ on $\mathbb{E}[e(S)]$ and a target budget $b \geq 0$, choosing $\theta = \min\{1, \sqrt{b/B}\}$ guarantees $\mathbb{E}[e(S_\theta)] \leq b$ and $\mathbb{E}[|S_\theta|] \geq \min\{1, \sqrt{b/B}\}\Sigma$. The graph-independent choice $B = n/(2(c+1)) + 1/n$ requires no public degree bound.

4 Lower bounds in terms of defect count

Throughout this section, let $n \geq 2$. We now bound what a private *released* set can achieve under an expected defect budget. For a mechanism M and graph G, write $m_M(G) = \mathbb{E}[|M(G)|]$ and $b_M(G) = \mathbb{E}[e_G(M(G))]$, where $e_G(S) = |E(G[S])|$ and the expectation is over the mechanism. A uniform expected defect budget $b \geq 0$ requires $b_M(G) \leq b$ for every graph in the input class under consideration. This counts all violated edges, rather than introducing a probability of any violation. We prove the lower bound directly for graphs with at most n edges, equivalently average degree at most 2; the unrestricted version follows immediately because a guarantee on every graph includes this sparse class. We first work with central edge privacy, with budget ρ, and then substitute $\rho = 2\varepsilon$ for local protocols.

Theorem 4.1 (Sparse size–defect lower bound)**.** *Suppose M is central ρ-edge-DP and satisfies $m_M(G) \geq \alpha(G)/r$ on every n-vertex graph G with $|E(G)| \leq n$, for some $r \leq (n+1)/2$. Let $a = \frac{n+1}{2r}$ and $\eta_n = \binom{n}{2}e^{-(n-1)/6}$. Then there exists an n-vertex graph G with $|E(G)| \leq n$ such that*

$$b_M(G) \geq \frac{a(a-1)}{2(1+(n-1)e^\rho)} - \eta_n. \tag{4.1}$$

In particular, for fixed ρ and all sufficiently large n, if $r \leq (n+1)/4$, then some graph of average degree at most 2 satisfies $b_M(G) \geq \frac{n}{32e^\rho r^2}$.

Proof sketch of Theorem 4.1. Draw $G \sim G(n, 1/n)$, write $Z = |E(G)|$, and let $\mathcal{A}$ be the event $Z \leq n$. Applying the approximation guarantee on $\mathcal{A}$, using the elementary bound $\alpha(G) \geq n - Z$ and the nonnegativity of the output size outside $\mathcal{A}$, yields $\mathbb{E}[|M(G)|] \geq a = (n+1)/(2r)$. To relate output size to defects, Lemma B.1 considers each unordered pair $\{u, v\}$ while conditioning on all other edges. If q_0 and q_1 are the probabilities that both endpoints are selected when uv is absent and present, respectively, edge privacy gives $q_0 \leq e^\rho q_1$. Since the edge is present with probability $1/n$, the pair contributes q_1/n in expectation to the defect count, whereas its probability of being jointly selected is $(1-1/n)q_0 + q_1/n$. Comparing these quantities and summing over all pairs yields

$$\mathbb{E}[e_G(M(G))] \geq \frac{1}{1+(n-1)e^\rho}\,\mathbb{E}\binom{|M(G)|}{2}.$$

Jensen's inequality then gives the lower bound $a(a-1)/(2(1+(n-1)e^\rho))$. A Chernoff bound shows that $\Pr[Z > n] \leq e^{-(n-1)/6}$, so graphs outside the sparse class contribute at most η_n to the expected defect count. Consequently, some graph with at most n edges satisfies (4.1). Finally, when $r \leq (n+1)/4$, the leading term is at least $n/(16e^\rho r^2)$, while η_n is eventually at most half this quantity. The full proof appears in Appendix B.1.

Corollary 4.2 (Local protocols)**.** Fix $\varepsilon > 0$. For all sufficiently large n, an ε-edge-LDP protocol with approximation ratio $r \le (n+1)/4$ and uniform expected defect budget b on every n-vertex graph with at most n edges satisfies $b \ge \frac{n}{32e^{2\varepsilon}r^2}$. The same conclusion holds, in particular, if the approximation and defect guarantees are required on every n-vertex graph.

Proof. One undirected edge changes one bit in each of two endpoint records. The full local transcript, and hence its released set, is central 2ε-edge-DP. Apply Theorem 4.1 with $\rho = 2\varepsilon$. □

For fixed $\varepsilon > 0$, Corollary 4.2 gives the necessary tradeoff $b = \Omega_\varepsilon(n/r^2)$ even when both guarantees are restricted to graphs of average degree at most 2. Setting $r = \Theta_\varepsilon(\log n)$ yields $b = \Omega_\varepsilon(n/\log^2 n)$, matching the dependence on n in our upper guarantee for bounded-average-degree graphs in Corollary 3.3. Thus the defect scaling is asymptotically tight at this approximation scale within that class, up to constants depending on ε.

5 Experimental evaluation

5.1 Methods and evaluation metrics

To the best of our knowledge, this is the first work to study independent-set algorithms under LDP, so we are unaware of a problem-specific LDP benchmark. We therefore use randomized response, the classical privacy mechanism introduced by Warner (1965) and previously applied to graph statistics under local edge privacy by Imola et al. (2021). The baseline applies ε-randomized response independently to every edge bit, flipping each bit with probability $1/(1+e^\varepsilon)$. Releasing the randomized bits is ε-edge-LDP. The curator then runs a greedy independent-set algorithm on the obfuscated graph (Halldórsson and Radhakrishnan, 1997); we call the resulting method RR+Greedy.

We evaluate both methods on the original input graph using the selected-set size $|S|$ and the defect count $e(S)$. Thus, defects produced by RR+Greedy are edges of the original graph, not of the obfuscated graph on which its greedy step operates. The communication requirements differ substantially: Algorithm 1 sends one bit per user, whereas RR+Greedy releases a bit for every possible edge and therefore uses $\Theta(n^2)$ bits. We use RR+Greedy as a utility baseline rather than a communication-competitive method.

5.2 Inputs and experimental protocol

For synthetic experiments, we use Barabási-Albert preferential-attachment graphs (Barabási and Albert, 1999). These graphs are useful for maximum independent set because they are sparse but have heterogeneous degrees and hubs: selecting a hub can exclude many low-degree alternatives, making degree-sensitive selection consequential. We use the standard star-initialized model, starting from a star on $m+1$ vertices and adding each later vertex with m preferential-attachment edges. We first vary ε at $n = 3000$, and then fix $\varepsilon = 1.0$ while varying n for $m \in \{10, 100\}$.

We also evaluate on the fixed public graphs PolBlogs, Cora, and CiteSeer. PubMed is used only for Algorithm 1, because randomized response over all unordered pairs would require a prohibitively large obfuscated graph. Table 1 summarizes the inputs and repetition counts.

Table 1: Experimental inputs. Barabási-Albert rows use independent graph draws from the star-initialized preferential-attachment model; public-graph rows are fixed input graphs. PolBlogs is from Adamic and Glance (2005); Cora, CiteSeer, and PubMed are standard citation-network benchmarks as used by Yang et al. (2016) and Fey and Lenssen (2019).

Input	n	Parameter	Graph draws	Repetitions per draw
$BA(3000, 5)$	3000	attachment $m = 5$	10	10
$BA(3000, 10)$	3000	attachment $m = 10$	10	10
$BA(n, m)$ scaling	500–3000	$m \in \{10, 100\}$, $\varepsilon = 1.0$	10 per n, m	10
PolBlogs	1490	fixed graph	1	50
Cora	2708	fixed graph	1	50
CiteSeer	3327	fixed graph	1	50
PubMed	19717	fixed graph, ours only	1	50

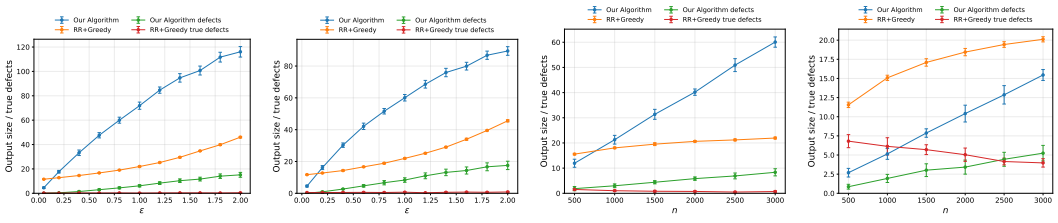
For the ε-sweep experiments, we ran our Algorithm 1 and RR+Greedy (where applicable), with $\varepsilon \in \{0.05, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ and used the same grid for the one-method PubMed experiment. For Barabási-Albert inputs, error bars summarize variability across graph-draw means: for each graph draw and method, we first average the randomized repetitions and then compute the standard deviation across the 10 graph draws. For fixed public graphs, error bars show one empirical standard deviation over randomized repetitions conditional on the fixed input. All code was run using Python on a 16-core M4 MacBook laptop.

5.3 Synthetic-graph results

Privacy-budget sweep. Panels (a)–(b) of Figure 5.1 show the ε-sweep results for $BA(3000, 5)$ and $BA(3000, 10)$. Across the tested budgets, RR+Greedy has fewer defects, while Algorithm 1 releases larger sets. At $\varepsilon = 2$, for example, Algorithm 1 released 116.08 ± 4.31 vertices with 15.14 ± 1.99 defects on $BA(3000, 5)$, compared with 46.05 ± 0.60 vertices and 0.42 ± 0.25 defects for RR+Greedy. On $BA(3000, 10)$, the corresponding means were 89.46 ± 2.72 vertices and 17.54 ± 2.62 defects for Algorithm 1, versus 45.58 ± 0.67 vertices and 0.85 ± 0.35 defects for RR+Greedy. The advantage in output size therefore comes with more violations in the original graph.

Scaling with graph size. In panels (c)–(d) of Figure 5.1, we fix $\varepsilon = 1.0$ and vary $n \in \{500, 1000, 1500, 2000, 2500, 3000\}$. For $m = 10$, the mean output size of Algorithm 1 increased from 11.94 ± 1.60 at $n = 500$ to 60.07 ± 2.04 at $n = 3000$, while its defect count increased from 1.85 ± 0.84 to 8.34 ± 1.42. RR+Greedy increased from 15.56 ± 0.31 vertices and 1.52 ± 0.35 defects to 21.95 ± 0.37 vertices and 0.70 ± 0.28 defects. Thus, Algorithm 1 overtakes RR+Greedy in size as n grows on the sparser $m = 10$ graphs.

For $m = 100$, Algorithm 1 increased from 2.69 ± 0.56 vertices and 0.86 ± 0.31 defects at $n = 500$ to 15.45 ± 0.71 vertices and 5.24 ± 1.01 defects at $n = 3000$. RR+Greedy increased from 11.55 ± 0.35 vertices and 6.81 ± 0.85 defects to 20.11 ± 0.34 vertices and 3.97 ± 0.55 defects. The contrast between $m = 10$ and $m = 100$ is consistent with the sparsity dependence in Corollary 3.3: our approximation ratio is $O(1 + \bar{d} + \varepsilon^{-1} \log n)$, and the average degree of these graphs is approximately $2m$, so the guarantee weakens as m increases. RR+Greedy exhibits the opposite relative trend on the denser $m = 100$ inputs. A likely explanation is that randomized response flips every possible edge bit. On a very sparse graph, false-positive edges created from the much larger set of nonedges can dominate the obfuscated graph and obscure its original structure; as the input becomes denser, retained true edges form a larger share of the reported adjacencies, leaving the greedy step with a more informative graph (Mukherjee and Suppakitpaisarn, 2025).



(a) $BA(3000, 5)$ (b) $BA(3000, 10)$ (c) $BA(n, 10)$ (d) $BA(n, 100)$

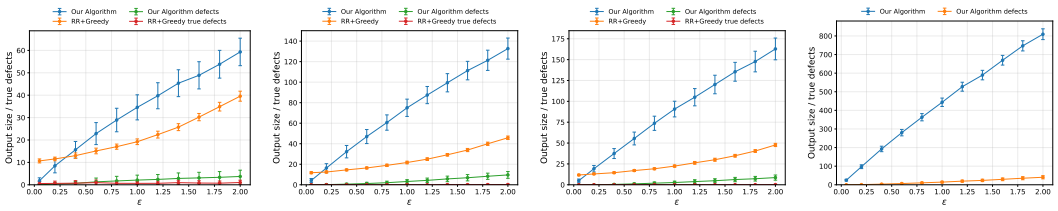
Figure 5.1: Experimental results on Barabási-Albert graphs

5.4 Public-graph results

Figure 5.2 reports the ε sweeps on the public graphs. On PolBlogs, Algorithm 1 produces larger mean sets after the smallest privacy budgets, while RR+Greedy maintains fewer defects. At $\varepsilon = 1$, for example, the respective mean sizes are 34.58 and 19.22, and the mean defect counts are 2.10 and 0.32. Cora and CiteSeer show the same qualitative pattern over most of the tested range: the one-bit sampler releases larger sets but accepts more original-graph defects.

On PubMed, randomized response over all unordered vertex pairs would require approximately $1.94 \cdot 10^8$ edge bits and, for small ε, would produce a dense obfuscated graph. That is too large for our computational environment. We therefore report only Algorithm 1. Its mean output starts at

24.76 vertices with 0.02 defects at $\varepsilon = 0.05$, and both quantities increase steadily over the tested range. This experiment shows that the one-bit sampler remains computationally feasible on a graph for which our RR+Greedy implementation is not.



(a) PolBlogs (b) Cora (c) CiteSeer (d) PubMed

Figure 5.2: Experimental results on public graphs

5.5 Runtime Comparisons

Table 2 reports the mean and empirical standard deviation of wall-clock algorithm time over the randomized runs for the listed epsilon-sweep inputs in Section 5. For the Barabási-Albert rows, the average also includes the 10 independent graph draws. These measurements reflect our vectorized centralized simulation on the stated hardware, not distributed round complexity.

The one-bit sampler is below 10^{-3} seconds on every listed input. Its mean time is about $8.6 \cdot 10^{-5}$ to $1.15 \cdot 10^{-4}$ seconds on the fixed public graphs, $3.67 \cdot 10^{-4}$ seconds on PubMed, and about $8.7 \cdot 10^{-4}$ to $9.4 \cdot 10^{-4}$ seconds on the $BA(3000, m)$ inputs. RR+Greedy is much slower in these implementations: about 0.12 seconds on PolBlogs, 0.41 seconds on Cora, 0.63 seconds on CiteSeer, and about 4.6 to 4.8 seconds on the Barabási-Albert inputs. A closer inspection shows that this cost is dominated by randomized-response graph generation rather than the greedy pass itself, especially for privacy levels at which the obfuscated graph is dense. RR+Greedy was not run on PubMed because the randomized-response graph over all unordered pairs is prohibitively large. Thus the timing measurements are consistent with the communication comparison: that Algorithm 1 sends one bit per user, while RR+Greedy materializes a quadratic-size privatized edge representation.

Table 2: Computation times of our method and RR+Greedy

Graph	Ours (s)	RR+Greedy (s)
$BA(3000, 5)$	0.0008733 ± 0.0017	4.841 ± 23
$BA(3000, 10)$	0.0009198 ± 0.0017	4.749 ± 23
$BA(3000, 20)$	0.0009382 ± 0.0016	4.643 ± 22
PolBlogs	$8.555 \cdot 10^{-5} \pm 2.4 \cdot 10^{-5}$	0.1191 ± 0.12
Cora	$0.0001047 \pm 2.5 \cdot 10^{-5}$	0.4086 ± 0.35
CiteSeer	$0.0001149 \pm 2.5 \cdot 10^{-5}$	0.6275 ± 0.51
PubMed	$0.0003672 \pm 2.3 \cdot 10^{-5}$	–

6 Conclusion and Future Work

We studied how randomized distributed algorithms can be adapted to local edge differential privacy through maximum independent set. The form of the output is important: Boppana's rule gives a private independent-set solution when users keep their membership decisions to themselves, whereas publicly releasing the selected set requires relaxing independence. Our main algorithm independently samples vertices using privately perturbed degrees and sends one bit per user. On graphs of average degree or degeneracy of the scale $o(\log n)$, the realized ratio of number of defects to output size tends to zero as the graph grows. Our lower bounds show that algorithms guaranteeing larger expected outputs must allow more expected defects on some graph. This remains true when both the guarantees and the witness are restricted to graphs of average degree at most 2. At approximation ratio $\Theta_\varepsilon(\log n)$, the resulting $\Omega_\varepsilon(n/\log^2 n)$ lower bound matches the defect scaling of our upper guarantee for bounded-average-degree graphs.

As future work, we plan to adapt randomized distributed algorithms for matching (Lotker et al., 2015) and dominating set (Jia et al., 2002). We also plan to use the more communication- and computation-efficient realization of randomized response for sparse graph data proposed by Hillebrand et al. (2026) as a scalable RR+Greedy baseline in future comparisons.

REPRODUCIBILITY STATEMENT

The privacy definitions and notation are specified in Section 2, and the proposed sampling mechanism is given in Algorithm 1. Section 3 provides the privacy analysis and sparse-graph consequences. Complete proofs of the size and defect guarantees and the defect-count lower bound appear in Appendices A and B, respectively. Sections 5.1 and 5.2 describe the baseline, evaluation metrics, synthetic graph generation, public benchmark datasets, parameter settings, repetition counts, error-bar definitions, and computational environment. Table 1 summarizes the experimental inputs, and Table 2 reports runtime measurements. The supplementary material provides an anonymized implementation, dependency specifications, dataset preparation instructions, random seeds, and scripts for reproducing the reported figures and tables.

AI USE STATEMENT

Generative AI tools assisted with formatting, rewriting, organization, and mathematical consistency checks, including development of defect-count bounds and proof details. The authors are responsible for verifying the arguments, references, and final manuscript.

REFERENCES

L. A. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD*, pages 36–43, 2005.

N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

R. B. Boppana, M. M. Halldórsson, and D. Rawitz. Simple and local independent set approximation. In *SIROCCO*, 2018. arXiv:1803.00786.

Y. Caro. New results on the independence number. Technical Report, Tel Aviv University, 1979.

S. O. Chan. Approximation resistance from pairwise independent subgroups. In *STOC*, pages 447–456, 2013.

A. B. G. Christiansen, E. Rotenberg, T. A. Steiner, and J. Vlieghe. Private graph colouring with limited defectiveness. In *SOFSEM*, volume 16448 of *Lecture Notes in Computer Science*, pages 404–415. Springer, 2026.

V. Cohen-Addad, Y. Esencayi, C. Fan, M. Gaboardi, S. Li, and D. Wang. On facility location problem in the local differential privacy model. In *AISTATS*, volume 151 of *Proceedings of Machine Learning Research*, pages 3914–3929, 2022.

L. Dhulipala, M. Henzinger, G. Z. Li, Q. C. Liu, A. R. Sricharan, and L. Zhu. Near-optimal differentially private graph algorithms via the multidimensional AboveThreshold mechanism. In *ESA*, volume 351 of *LIPIcs*, pages 91:1–91:20, 2025.

L. Dhulipala, Q. C. Liu, S. Raskhodnikova, J. Shi, J. Shun, and S. Yu. Differential privacy from locally adjustable graph algorithms: k-core decomposition, low out-degree ordering, and densest subgraphs. In *FOCS*, 2022. arXiv:2211.10887.

M. Dinitz, S. Kale, S. Lattanzi, and S. Vassilvitskii. Almost tight bounds for differentially private densest subgraph. In *SODA*, pages 2908–2950, 2025.

M. Dinitz, G. Z. Li, Q. C. Liu, and F. Zhou. Differentially private matchings. arXiv:2501.00926, 2025.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and minimax bounds: sharp rates for probability estimation. In *Advances in Neural Information Processing Systems*, volume 26, 2013. arXiv:1305.6000.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

A. Epasto, V. Mirrokni, B. Perozzi, A. Tsitsulin, and P. Zhong. Differentially private graph learning via sensitivity-bounded personalized PageRank. In *Advances in Neural Information Processing Systems*, volume 35, 2022. arXiv:2207.06944.

M. Fey and J. E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. arXiv:1903.02428.

C. Fan, P. Li, and X. Li. Private graph all-pairwise-shortest-path distance release with improved error rate. In *Advances in Neural Information Processing Systems*, volume 35, pages 17844–17856, 2022.

A. M. Frieze. On the independence number of random graphs. *Discrete Mathematics*, 81(2):171–175, 1990.

A. Gupta, K. Ligett, F. McSherry, A. Roth, and K. Talwar. Differentially private combinatorial optimization. In *SODA*, pages 1106–1125, 2010. arXiv:0903.4510.

M. M. Halldórsson and J. Radhakrishnan. Greed is good: Approximating independent sets in sparse and bounded-degree graphs. *Algorithmica*, 18(1):145–163, 1997.

Q. Hillebrand, P. Manurangsi, V. Suppakitpaisarn, and P. Vajanopath. Decisive margins in differentially private voting. arXiv:2608.18772, 2026.

Q. Hillebrand, V. Suppakitpaisarn, and T. Shibuya. Communication cost reduction for subgraph counting under local differential privacy via hash functions. *Transactions on Machine Learning Research*, 2025a.

Q. Hillebrand, V. Suppakitpaisarn, and T. Shibuya. Cycle counting under local differential privacy for degeneracy-bounded graphs. In *STACS*, volume 327 of *LIPIcs*, pages 49:1–49:22, 2025b.

Q. Hillebrand, V. Suppakitpaisarn, and T. Shibuya. Communication-efficient publication of sparse vectors under differential privacy via Poisson private representation. In *AsiaCCS*, 2026. arXiv:2405.01951.

R. Hladík and J. Tětek. Near-universally-optimal differentially private minimum spanning trees. In *FORC*, volume 329 of *LIPIcs*, pages 6:1–6:19, 2025.

J. Hsu, Z. Huang, A. Roth, T. Roughgarden, and Z. S. Wu. Private matchings and allocations. *SIAM Journal on Computing*, 45:1953–1984, 2016. Conference version in STOC 2014. arXiv:1311.2828.

J. Imola, T. Murakami, and K. Chaudhuri. Locally differentially private analysis of graph statistics. In *USENIX Security*, 2021. arXiv:2010.08688.

L. Jia, R. Rajaraman, and T. Suel. An efficient distributed algorithm for constructing small dominating sets. *Distributed Computing*, 15:193–205, 2002.

S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011. arXiv:0803.0924.

S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith. Analyzing graphs with node differential privacy. In *TCC*, 2013.

M. Kearns, M. Pai, A. Roth, and J. Ullman. Mechanism design in large games: incentives and privacy. In *ITCS*, 2014. arXiv:1207.4084.

Z. Lotker, B. Patt-Shamir, and S. Pettie. Improved distributed approximate matching. *Journal of the ACM*, 62(5):38:1–38:17, 2015.

M. Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing*, 15(4):1036–1053, 1986.

S. Mukherjee and V. Suppakitpaisarn. Local differential privacy-preserving spectral clustering for general graphs. *Transactions on Machine Learning Research*, 2025. arXiv:2309.06867.

P. Mundra, C. Papamanthou, J. Shun, and Q. C. Liu. Practical and accurate local edge differentially private graph algorithms. *Proceedings of the VLDB Endowment*, 18(11):4199–4213, 2025.

K. Pfisterer, Q. Hillebrand, and V. Suppakitpaisarn. Facility location problem under local differential privacy without super-set assumption. In *DBSec*, volume 15722 of *Lecture Notes in Computer Science*, pages 293–310, 2025. arXiv:2506.15224.

K. Pfisterer, Q. Hillebrand, and V. Suppakitpaisarn. Publishing below-threshold triangle counts under local weight differential privacy. arXiv:2601.01710, 2026.

Z. Qin, T. Yu, Y. Yang, I. Khalil, X. Xiao, and K. Ren. Generating synthetic decentralized social graphs with local differential privacy. In *CCS*, pages 425–438, 2017.

S. Raskhodnikova, A. Smith, C. Wagaman, and A. Zavyalov. Local node differential privacy. arXiv:2602.15802, 2026.

A. Sealfon. Shortest paths and distances with differential privacy. In *PODS*, pages 29–41, 2016. arXiv:1511.04631.

V. Suppakitpaisarn, D. Ponnoprat, N. Hirankarn, and Q. Hillebrand. Counting graphlets of size k under local differential privacy. In *AISTATS*, volume 258 of *Proceedings of Machine Learning Research*, pages 5005–5013, 2025.

S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

V. K. Wei. A lower bound on the stability number of a simple graph. Bell Labs Technical Memorandum, 1981.

M. Xie, J. Wu, D. Nguyen, and A. Srinivasan. Differentially private graph coloring. arXiv:2602.13460, 2026.

Z. Yang, W. W. Cohen, and R. Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In *ICML*, volume 48 of *Proceedings of Machine Learning Research*, pages 40–48, 2016.

Z. Zou, C. Fan, M. Dinitz, J. Liu, and J. Upadhyay. A generalized binary tree mechanism for private approximation of all-pair shortest distances. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

A Proofs for locally private randomized sampling

A.1 From a Ratio of Expectations to a High-Probability Realized Ratio

Proposition A.1. *Let S be a random vertex set whose membership indicators are independent, and suppose $\mathbb{E}[e(S)] \le f_\varepsilon(n)\,\mathbb{E}[|S|]$ for some $f_\varepsilon(n) \to 0$ as $n \to \infty$. Then, with probability at least $1 - \sqrt{f_\varepsilon(n)} - e^{-\mathbb{E}[|S|]/8}$, we have $e(S) \le 2\sqrt{f_\varepsilon(n)}|S|$. When $\mathbb{E}[|S|] \to \infty$ as $n \to \infty$, this probability is $1 - o_\varepsilon(1)$.*

Proof. If $f_\varepsilon(n) = 0$, then $e(S) = 0$ almost surely and the claim is immediate. Otherwise, by Markov inequality,

$$\Pr\left(e(S) > \sqrt{f_\varepsilon(n)}\,\mathbb{E}[|S|]\right) \le \frac{\mathbb{E}[e(S)]}{\sqrt{f_\varepsilon(n)}\,\mathbb{E}[|S|]} \le \sqrt{f_\varepsilon(n)}.$$

Since $|S|$ is a sum of independent Bernoulli random variables, by the multiplicative Chernoff bound,

$$\Pr\left(|S| < \frac{\mathbb{E}[|S|]}{2}\right) \le e^{-\mathbb{E}[|S|]/8}.$$

Putting these bounds together, we have that

$$e(S) \le \sqrt{f_\varepsilon(n)}\,\mathbb{E}[|S|] \le 2\sqrt{f_\varepsilon(n)}|S|$$

with probability at least $1 - \sqrt{f_\varepsilon(n)} - e^{-\mathbb{E}[|S|]/8}$ by the union bound, completing the proof. □

A.2 Full proof of Theorem 3.2

For each vertex, write

$$D_v = \max(d_v + c + L_v, 0), \qquad p_v = \frac{1}{D_v + 1}, \qquad q_v = \Pr(v \in S) = \mathbb{E}[p_v].$$

Each membership bit depends only on its user's independent noise and Bernoulli randomness. Thus the bits are independent even after averaging over the noise, and

$$\mathbb{E}[|S|] = \sum_{v \in V} q_v, \qquad \mathbb{E}[e(S)] = \sum_{uv \in E} q_u q_v.$$

We therefore need a lower bound on each marginal q_v to control size, and an upper bound to control defects.

Lemma A.2 (Lower bound on inclusion)**.** *For every vertex v, $q_v \ge (d_v + c + 1/\varepsilon + 1)^{-1}$.*

Proof. Since $D_v \le d_v + c + |L_v|$ and $\mathbb{E}[|L_v|] = 1/\varepsilon$, we have $\mathbb{E}[D_v] \le d_v + c + 1/\varepsilon$. Jensen's inequality gives

$$q_v = \mathbb{E}\left[\frac{1}{D_v + 1}\right] \ge \frac{1}{\mathbb{E}[D_v] + 1} \ge \frac{1}{d_v + c + 1/\varepsilon + 1}.$$

□

Lemma A.3 (Upper bound on inclusion)**.** *For every vertex v, $q_v \le (d_v + c/2 + 1)^{-1} + 1/(2n^2)$.*

Proof. On the event $L_v \ge -c/2$, we have $p_v \le (d_v + c/2 + 1)^{-1}$. On its complement, use $p_v \le 1$. The Laplace tail and $\varepsilon c \ge 4 \log n$ give

$$\Pr(L_v < -c/2) = \tfrac{1}{2} e^{-\varepsilon c/2} \le \frac{1}{2n^2},$$

proving the claim. □

The positive shift makes unusually large inclusion probabilities rare. The next lemma sums their effect over edges, rather than bounding only the probability of any conflict.

Lemma A.4 (Expected defect count)**.** *The released set satisfies*

$$\mathbb{E}[e(S)] \le \min\left\{\frac{1}{2}\sum_{v\in V}\frac{1}{d_v+c+1}, \frac{|E|}{(c/2+1)^2}\right\} + \frac{1}{n}.$$

Proof. Set $h_v = (d_v + c/2 + 1)^{-1}$ and $\eta = 1/(2n^2)$. Independence and Lemma A.3 imply

$$\begin{aligned}\mathbb{E}[e(S)] &\le \sum_{uv\in E}(h_u+\eta)(h_v+\eta)\\ &= \sum_{uv\in E} h_u h_v + \eta\sum_{v\in V} d_v h_v + \eta^2|E|\\ &\le \sum_{uv\in E} h_u h_v + \frac{1}{2n} + \frac{1}{8n^2} \le \sum_{uv\in E} h_u h_v + \frac{1}{n},\end{aligned}$$

where $d_v h_v \le 1$ and $|E| \le n^2/2$. To bound the remaining sum by vertex degrees, use $h_u h_v \le (h_u^2 + h_v^2)/2$:

$$\sum_{uv\in E} h_u h_v \le \frac{1}{2}\sum_{v\in V} d_v h_v^2 = \frac{1}{2}\sum_{v\in V}\frac{d_v}{(d_v+c/2+1)^2} \le \frac{1}{2}\sum_{v\in V}\frac{1}{d_v+c+1}.$$

The last inequality follows from $(d+c/2+1)^2 - d(d+c+1) = d + (c/2+1)^2 \ge 0$. Alternatively, $h_v \le 1/(c/2+1)$ gives $\sum_{uv\in E} h_u h_v \le |E|/(c/2+1)^2$. Combining the two estimates proves the lemma. □

Proof of Theorem 3.2. Summing Lemma A.2 over vertices gives $\mathbb{E}[|S|] \ge \Sigma$. For the defect bound, observe that $\varepsilon(c+1) \ge 4\log n \ge 4\log 8 > 8$. Hence, for every $d \ge 0$,

$$\frac{1}{d+c+1} = \left(1 + \frac{1}{\varepsilon(d+c+1)}\right)\frac{1}{d+c+1/\varepsilon+1} \le \frac{9}{8}\frac{1}{d+c+1/\varepsilon+1}.$$

Substituting into Lemma A.4 gives the factor $9/16$; its edge-count bound supplies the other term in (3.2). □

B Proofs for defect-count lower bounds

Lemma B.1 (Random-graph defect inequality)**.** *Let M be central ρ-edge-DP, and let G have independently present edges with probability $p \in (0,1)$. Expectations in this lemma include both G and the mechanism. Then*

$$\mathbb{E}[e_G(M(G))] \ge \frac{p}{p+(1-p)e^{\rho}}\,\mathbb{E}\binom{|M(G)|}{2}.$$

Proof. For each unordered pair $\{u,v\}$, define

$$X_{uv} = \mathbf{1}\{uv \in E(G)\}, \qquad Y_{uv} = \mathbf{1}\{u,v \in M(G)\}.$$

Thus, $X_{uv}Y_{uv}$ indicates that the pair $\{u,v\}$ contributes a defect, while Y_{uv} indicates that both vertices are selected, regardless of whether they are adjacent.

Fix one pair $\{u,v\}$ and fix whether every other possible edge is present or absent. Let H be the graph formed by the edges fixed as present, so $uv \notin E(H)$. The only remaining graph randomness is whether uv is present: conditionally, $G = H$ with probability $1-p$ and $G = H + uv$ with probability p. Define

$$q_0 = \Pr[\{u,v\} \subseteq M(H)], \qquad q_1 = \Pr[\{u,v\} \subseteq M(H+uv)],$$

where these probabilities are over the randomness of M. Because H and $H+uv$ differ in one edge, ρ-edge privacy gives $q_0 \le e^{\rho} q_1$. Therefore,

$$\begin{aligned}\mathbb{E}[Y_{uv} \mid H] &= (1-p)q_0 + pq_1\\ &\le \big((1-p)e^{\rho} + p\big)q_1,\end{aligned}$$

whereas

$$\mathbb{E}[X_{uv}Y_{uv} \mid H] = pq_1.$$

Combining these relations yields

$$\mathbb{E}[X_{uv}Y_{uv} \mid H] \geq \frac{p}{p + (1-p)e^{\rho}} \, \mathbb{E}[Y_{uv} \mid H].$$

Averaging over the fixed states of the other edges gives the same inequality without conditioning. Finally, sum it over all unordered pairs. Since, for every realization of G and M,

$$e_G(M(G)) = \sum_{\{u,v\}} X_{uv}Y_{uv} \quad \text{and} \quad \binom{|M(G)|}{2} = \sum_{\{u,v\}} Y_{uv},$$

linearity of expectation proves the claim. □

B.1 Proof of Theorem 4.1

Proof. Draw $G \sim G(n, 1/n)$, write $Z = |E(G)|$, and let $\mathcal{A}$ be the event $Z \leq n$. Set $q = \Pr(\mathcal{A})$. We use the approximation guarantee only after conditioning on $\mathcal{A}$. By the law of total expectation and the nonnegativity of the output size,

$$\begin{aligned}
\mu := \mathbb{E}[|M(G)|] &= q\,\mathbb{E}[|M(G)| \mid \mathcal{A}] + (1-q)\,\mathbb{E}[|M(G)| \mid \mathcal{A}^c] \\
&\geq q\,\mathbb{E}[|M(G)| \mid \mathcal{A}] \\
&\geq \frac{q}{r}\,\mathbb{E}[\alpha(G) \mid \mathcal{A}] \\
&\geq \frac{q}{r}\,\mathbb{E}[n - Z \mid \mathcal{A}] \\
&\geq \frac{1}{r}\,\mathbb{E}[n - Z] = \frac{n+1}{2r} = a.
\end{aligned}$$

To justify the last inequality, decompose the unconditional expectation according to $\mathcal{A}$:

$$\mathbb{E}[n - Z] = q\,\mathbb{E}[n - Z \mid \mathcal{A}] + (1-q)\,\mathbb{E}[n - Z \mid \mathcal{A}^c].$$

On $\mathcal{A}^c$ we have $Z > n$, and hence the second term is nonpositive. Therefore $\mathbb{E}[n-Z] \leq q\,\mathbb{E}[n-Z \mid \mathcal{A}]$, as claimed. Substituting $p = 1/n$ into Lemma B.1 gives

$$\mathbb{E}[b_M(G)] \geq \frac{1}{1 + (n-1)e^{\rho}} \, \mathbb{E}\binom{|M(G)|}{2}.$$

The function $f(x) = x(x-1)/2$ is convex, so Jensen's inequality gives

$$\mathbb{E}\binom{|M(G)|}{2} = \mathbb{E}[f(|M(G)|)] \geq f(\mathbb{E}[|M(G)|]) = \frac{\mu(\mu-1)}{2}.$$

Since $r \leq (n+1)/2$, we have $a \geq 1$ and hence $\mu(\mu-1) \geq a(a-1)$. Combining these inequalities yields

$$\mathbb{E}[b_M(G)] \geq L := \frac{a(a-1)}{2(1 + (n-1)e^{\rho})}.$$

It remains to show that graphs outside $\mathcal{A}$ cannot account for much of this expectation. The binomial variable Z has mean $(n-1)/2$. A standard multiplicative Chernoff bound gives

$$\Pr[Z > n] \leq \Pr[Z \geq 2\,\mathbb{E}\,Z] \leq e^{-\mathbb{E}[Z]/3} = e^{-(n-1)/6}.$$

Moreover, $b_M(G) \leq |E(G)| \leq \binom{n}{2}$ for every graph. Hence

$$(1-q)\,\mathbb{E}[b_M(G) \mid \mathcal{A}^c] \leq \binom{n}{2}(1-q) \leq \eta_n.$$

Since

$$\mathbb{E}[b_M(G)] = q\,\mathbb{E}[b_M(G) \mid \mathcal{A}] + (1-q)\,\mathbb{E}[b_M(G) \mid \mathcal{A}^c],$$

it follows that $q\,\mathbb{E}[b_M(G) \mid \mathcal{A}] \geq L - \eta_n$. If $L - \eta_n > 0$, the conditional average of $b_M(G)$ over $\mathcal{A}$ is therefore at least $(L - \eta_n)/q \geq L - \eta_n$, so some graph in $\mathcal{A}$ satisfies (4.1). If $L - \eta_n \leq 0$, the claim follows immediately from $b_M(G) \geq 0$.

If $r \leq (n+1)/4$, then $a \geq 2$, so $a(a-1) \geq a^2/2$. Together with $1 + (n-1)e^{\rho} \leq ne^{\rho}$, this gives

$$L \geq \frac{a^2}{4ne^{\rho}} = \frac{(n+1)^2}{16ne^{\rho}r^2} \geq \frac{n}{16e^{\rho}r^2}.$$

For fixed ρ, the exponentially small term η_n is at most half this lower bound for all sufficiently large n, uniformly over the stated range of r. This proves the final claim. $\square$