MINIMAX OPTIMAL PAGERANK UNDER ROW-LEVEL DIFFERENTIAL PRIVACY

Anonymous authors
Paper under double-blind review

ABSTRACT

PageRank turns directed links into node-importance scores. But a node's outgoing-link profile can encode sensitive information which can be leaked through PageRank's release. We study differential privacy under replacement of one entire outgoing row on a fixed public node set. We propose two differential private PageRank mechanisms under row-level privacy, i.e., the full outgoing-link profile. Both proposed mechanisms share a regularized query that combines degree flooring with a cap on the mass transmitted by each node. We prove global row-sensitivity bounds and quantify regularization bias. DP PageRank–Full Vector adds calibrated Gaussian noise to release all ranking scores, while DP PageRank–Top k uses Gumbel perturbations to release only the set of top nodes. We provide their utility guarantees with upper bounds of the deviation from original non-private PageRank. We also develop lower bounds. The lower bounds match the upper bounds of our algorithms for dense networks, under both score estimation and the top-k set recovery. Numerical studies on three real datasets and simulated networks further support the overwhelming superiority of our mechanisms.

1 INTRODUCTION

Directed graphs are a standard abstraction for relational data, where nodes represent entities and directed edges encode asymmetric relations from one entity to another. Ranking algorithms on directed graphs convert this link structure into global scores of importance, influence, or authority and hence give the rankings of entities. A classical algorithm is PageRank (Page et al., 1999). Let A denote the directed adjacency matrix. PageRank ranks nodes through the stationary behavior of a random walk on the row-normalized transition matrix from A. Thus, it turns the observed directed links into a global ranking vector. These rankings are often used as compact summaries of the graph in search, recommendation, citation analysis, and network-based decision making.

Outgoing-link-profile privacy for node ranking. The link pattern used for ranking may encode sensitive behavior. In a directed graph, the outgoing links of a single entity are recorded as one row of the adjacency matrix A. The entire row can reveal preferences, strategies, or private relationships of one entity. Hence, the entity would like to keep the whole row private, not only on one individual edge. Releasing a PageRank ranking vector therefore raises a row-level privacy concern: can one infer how a particular entity links to the rest of the graph?

Differential privacy (DP) aims to have a stable distribution of outputs on neighboring datasets. To protect the privacy of one entity on all its outgoing links, we define the neighboring datasets as two adjacency matrices that differ only on one row. Hence, we refer to this notion as *row-level DP*. Previous private ranking literature focuses on edge-level DP Arora & Upadhyay (2019); Epasto et al. (2022); Wei et al. (2024). In this paper, we aim to study row-level DP for directed graph ranking.

Perturbation-guided regularization for private ranking. Replacing a row can redirect the contribution of one node to many recipients. Nodes with few outgoing edges can put large weight on a single transition, while a node with a large PageRank score can transmit substantial mass. We control both effects: degree flooring bounds each transition weight, and mass clipping bounds each source's total contribution. These operations induce bias, so the analysis continues to compare against original PageRank rather than treating the regularized fixed point as the utility target.

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.

The two release mechanisms share this query. DP PageRank–Full Vector uses its global ℓ_2 sensitivity for Gaussian calibration. DP PageRank–Top k uses its coordinatewise sensitivity for Gumbel selection. They have different output spaces and privacy calibrations; a comparison between them must retain those distinctions and any explicit output postprocessing.

1.1 Our Contributions

First, we analyze the common degree-flooring and mass-clipping query under outgoing-row replacement. The sensitivity bounds hold globally, including for dangling rows. The bias bound accounts for the interaction between the degree floor and the source contribution cap.

Second, we provide privacy and utility guarantees for DP PageRank–Full Vector and DP PageRank–Top k. The Gaussian guarantee separates raw full-vector squared error into bias and noise, while the ordered top-k guarantee includes regularization bias in an exact-set-recovery condition.

Third, we study limits for the original PageRank target under uniform restart. A neighboring-graph construction gives a constant relative-risk obstruction on unrestricted inputs. On a public dense minimum-out-degree class, a row-preserving empirical-mean embedding yields matching full-vector relative-risk rates, and a packing construction gives a necessary gap for exact top-k recovery. The sufficient top-1 gap matches this order in the stated nonsaturated regime. Section 3 presents the shared query and full-vector mechanism, Section 4 presents top-k release, and Section 4.3 states the scope of the lower bounds.

Lastly, we implement our proposed methods on three real datasets from SNAP in Section 5 (Email-Eu-core, CollegeMsg, and Wiki-Vote). We demonstrate that our methods have overwhelmingly superior performance over existing methods and other benchmarks.

1.2 Related works

Differential privacy for graph data is commonly studied under either edge-level or node-level neighboring relations. As we have mentioned, edge-level DP protects the presence/absence of a single edge, which is the dominant setting in private graph analysis (Hay et al., 2009; Nissim et al., 2007; Zhang et al., 2015). The dyad-wise randomized response approach and edge-flipping approach protect the whole released graph, but the utility can substantially degrade in sparse networks (Hehir et al., 2022; Karwa et al., 2017; Mülle et al., 2015; Warner, 1965). Node-level DP provides a stronger guarantee, usually by hiding an entire node together with its edges, and has motivated bounded-degree projection and Lipschitz-extension techniques (Kasiviswanathan et al., 2013; Raskhodnikova & Smith, 2016). However, it does not match the goal of the ranking problem.

When it comes to the ranking problem, existing formal guarantees are mostly on edge-DP. PageRank and personalized PageRank are especially sensitive to edge perturbations because changing a single edge can substantially modify a transition probability; this effect is most pronounced around low-degree nodes. Epasto et al. (2022) study personalized PageRank under edge-level and joint edge-level DP by designing sensitivity-bounded push-flow approximations, with utility guarantees that improve on high-degree graphs. Wei et al. (2024) instead view PageRank as a graph diffusion process, inject noise during the diffusion iterations, and use degree-based thresholding together with privacy amplification by iteration to control the privacy loss caused by low-degree nodes. These works protect individual edges, whereas our setting protects an entire outgoing-link profile. This changes the perturbation geometry: replacing one row can simultaneously alter many transition probabilities and therefore requires mechanisms different from edge-level private PageRank.

A recurring way to avoid overly pessimistic global sensitivity is to regularize the problem before privatization, so that noise is calibrated to a regularized domain rather than to arbitrary worst-case inputs (Chaudhuri et al., 2011). In private learning, recent work on differentially private ridge regression introduces an extra regularization parameter to control the global sensitivity (Wang & Banjernee, 2026).

Clipping and flow control already appear in private PageRank and graph diffusion (Epasto et al., 2022; Wei et al., 2024). We do not claim clipping itself as a new technique. Our focus is the analysis of an entire outgoing-row replacement, the interaction of degree flooring and a source contribution cap, and lower bounds for the original PageRank target under this privacy relation.

2 PROBLEM SETUP

Row-level adjacency. We consider directed graphs on a fixed public node set $V = [n]$, represented by an adjacency matrix $A \in \{0,1\}^{n\times n}$, where $A_{ij} = 1$ indicates a directed edge from node i to node j. Let $A_{i,*}$ denote the i-th row of A, and let $d_i = \sum_{j=1}^{n} A_{ij}$ be the out-degree of node i. Our privacy notion is based on *row-level adjacency*. Two directed graphs A, A' are called adjacent, denoted $A \sim A'$, if there exists a node $i \in [n]$ such that A and A' differ only on the i-th row. This captures the requirement that the released output should hide the entire set of outgoing links of any single node.

Differential privacy. Following the definition of differential privacy Dwork et al. (2006), a randomized mechanism $\mathcal{M}$ satisfies row-level (ε, δ)-differential privacy if for any $A \sim A'$ and measurable set of outputs S,

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

PageRank. Let $P(A)$ be the row-stochastic transition matrix associated with A. For $d_i > 0$, we set $P(A)_{ij} = A_{ij}/d_i$, and for $d_i = 0$, we use the standard dangling-node convention $P(A)_{i,*} = \frac{1}{n}\mathbf{1}^\top$. Given a damping parameter $\beta \in (0,1)$ and a fixed public teleportation vector μ as a distribution on the node set, the PageRank vector $\boldsymbol{\pi}_{\mathrm{PR}}(A)$ is the unique solution

$$\boldsymbol{\pi}_{\mathrm{PR}}(A) = (1-\beta)\mu + \beta P(A)^\top \boldsymbol{\pi}_{\mathrm{PR}}(A). \tag{1}$$

For global PageRank, we take $\mu = \frac{1}{n}\mathbf{1}$ as the uniform distribution. Equivalently, $\boldsymbol{\pi}_{\mathrm{PR}}(A) = (1-\beta)(I - \beta P^\top)^{-1}\mu$. Thus, π_i is the stationary probability of being at node i under a random walk that follows outgoing edges with probability β and teleports with probability $1-\beta$.

Ranking targets. We study the accuracy of private PageRank scores and the recovery of highly ranked nodes. For ranking, practical interest often centers on a few leading nodes rather than the exact ordering of all nodes. In web search, for example, the first page receives most attention (Granka et al., 2004) and provides quick access to query-relevant information; users searching deeper for a particular item may care less about the precise ordering of later results. This emphasis on leading nodes is also reflected in previous work on private PageRank (Epasto et al., 2022; Wei et al., 2024).

3 PRIVATE PAGERANK: FULL VECTOR

In Section 3.1, we first establish the perturbation analysis to identify the structural sources of instability under row-DP. The instability comes from nodes with low out-degrees and from nodes that transmit a large amount of PageRank mass. It guides us to design the DP mechanism based on regularization. In Section 3.2, we establish the theoretical guarantee of (ε, δ)-DP of our new mechanism and demonstrate its utility. The utility can be divided into the regularization bias and the privacy noise. It guides the selection of the tuning parameters κ and m in Section 3.4.

3.1 PERTURBATION-GUIDED DP MECHANISM ON PAGERANK

Let $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ denote the PageRank vectors from A and A', respectively, where $A \sim A'$ are row-neighboring adjacency matrices. If they differ in row i, the change in that row's contribution, evaluated at $\boldsymbol{\pi}'$, is $\beta\pi'_i(P_{i,:} - P'_{i,:})^\top$. Low out-degrees allow large transition weights, while a large source score increases the mass affected by a row replacement. These two effects motivate degree flooring and mass clipping, respectively.

We consider an augmented transition with a uniform lower bound on the out-degrees. Given a public tuning parameter $m \ge 1$, define the augmented matrix Q as

$$Q_{ij} = A_{ij} / \max\{d_i, m\}. \tag{2}$$

This transition artificially raises all out-degrees to a uniform level m, which is set manually.

Next, we limit the mass transmitted by each node. Given a public tuning parameter $\kappa > 0$, let $\mathrm{clip}_\kappa(\boldsymbol{z})_i = \min\{z_i, \kappa\}$ for nonnegative $\boldsymbol{z}$. Let $\hat{\boldsymbol{\pi}}$ denote the regularized PageRank vector defined by

$$\hat{\boldsymbol{\pi}} = (1-\beta)\mu + \beta Q^\top \mathrm{clip}_\kappa(\hat{\boldsymbol{\pi}}). \tag{3}$$

The contribution from node i to node j is $\beta Q_{ij} \min\{\hat{\pi}_i, \kappa\}$. Thus each source transmits total mass at most $\beta\kappa$, and each edge carries mass at most $\beta\kappa/m$. The cap applies to outgoing mass; the receiving

score $\hat{\pi}_j$ may exceed κ. Dangling rows of Q are zero. No normalization is applied to $\hat{\boldsymbol{\pi}}$, whose total mass may be below one.

The defining map is a β-contraction in ℓ_1, so $\hat{\boldsymbol{\pi}}$ exists uniquely, is nonnegative and has total mass at most one. The argument is included in the sensitivity proof in Appendix A.1. The mechanism computes $\hat{\boldsymbol{\pi}}$ and then performs a single private release. Numerical solution is described in Appendix B.

We hope the regularized PageRank vector $\hat{\boldsymbol{\pi}}$ stabilizes the ranking. The uniform perturbation bounds of $\hat{\boldsymbol{\pi}}$ under $A \sim A'$ are demonstrated in Lemma 1.

Lemma 1 (Global sensitivity). *Let $A \sim A'$ be two $n \times n$ adjacency matrices that differ only in the ith row. Let $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\pi}}'$ be the corresponding regularized PageRank vectors. Then we have*

$$\begin{aligned} \|\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}'\|_2 &\le \Delta_{\mathrm{reg},2} := \frac{\beta\kappa}{\sqrt{m}}\left(\sqrt{2} + \frac{2\beta}{1-\beta}\right), \\ \|\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}'\|_\infty &\le \Delta_{\mathrm{reg},\infty} := \frac{\beta\kappa}{m}\frac{1+\beta}{1-\beta}. \end{aligned} \tag{4}$$

These bounds hold globally, including for dangling rows, without a spectral-norm condition on P.

Combining the global sensitivity $\Delta_{\mathrm{reg},2}$ in Lemma 1 with the analytic Gaussian mechanism (AGM) in Balle & Wang (2018), we immediately have a DP mechanism for this regularized PageRank, which is summarized as Algorithm 1.

For $\varepsilon \ge 0$ and $0 < \delta < 1$, let Φ denote the standard normal distribution function and choose the smallest $\sigma > 0$ satisfying

$$\Phi\left(\frac{\Delta_{\mathrm{reg},2}}{2\sigma} - \frac{\varepsilon\sigma}{\Delta_{\mathrm{reg},2}}\right) - e^{\varepsilon}\Phi\left(-\frac{\Delta_{\mathrm{reg},2}}{2\sigma} - \frac{\varepsilon\sigma}{\Delta_{\mathrm{reg},2}}\right) \le \delta. \tag{5}$$

This calibration is valid for any $\varepsilon \ge 0$ (Balle & Wang, 2018, Theorem 8).

Algorithm 1 DP PageRank–Full Vector

Require: Adjacency matrix A; public outdegree level m, mass cap κ, β, μ and privacy parameters (ε, δ)

Ensure: A private PageRank release $\tilde{\boldsymbol{\pi}}$

1: Let $Q_{ij} = \frac{A_{ij}}{\max\{m, d_i\}}$.
2: Let $\hat{\boldsymbol{\pi}}$ solve $\hat{\boldsymbol{\pi}} = (1-\beta)\mu + \beta Q^\top \mathrm{clip}_\kappa(\hat{\boldsymbol{\pi}})$.
3: Let σ be calibrated to $\Delta_{\mathrm{reg},2}$ by (5).
4: Output $\tilde{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} + \sigma\zeta$, where $\zeta \sim \mathcal{N}(0, I_n)$.

3.2 Utility analysis

The protection of privacy inevitably brings a worse accuracy in estimation. For our regularization DP mechanism, the error comes from two sources: the bias caused by regularization and the added noise. In the utility analysis, we quantify both sources. The following Theorem 1 quantifies the bias between the regularized PageRank vector $\hat{\boldsymbol{\pi}}$ and the original PageRank vector $\boldsymbol{\pi}$.

Theorem 1 (Regularization bias). *Let A be an $n \times n$ adjacency matrix, and let $\boldsymbol{\pi}$ and $\hat{\boldsymbol{\pi}}$ denote its original and regularized PageRank vectors, respectively. Define* $\mathrm{bias}^2 := \|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2^2$. *Then*

$$\mathrm{bias}^2 \le \left[\frac{\beta}{1-\beta}\sum_i \left\{\pi_i - \frac{d_i}{\max\{d_i, m\}}\min\{\pi_i, \kappa\}\right\}\right]^2. \tag{6}$$

If $\min_i d_i \ge m$ *and* $\|\boldsymbol{\pi}\|_\infty \le \kappa$, *then* $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}$.

Combining the sensitivity bound in Lemma 1 with the bias in Theorem 1 yields the following utility guarantee for Algorithm 1.

Corollary 1 (DP and utility guarantee). *Algorithm 1 is (ε, δ)-row-level differentially private for $\varepsilon \ge 0$ and $0 < \delta < 1$. Furthermore, at each fixed graph,*

$$\mathbb{E}\|\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2^2 = \mathrm{bias}^2 + n\sigma^2. \tag{7}$$

3.3 Lower bounds and matching

In this subsection, we discuss how tight is the upper bound. For simplicity, we consider loop-free directed binary graphs on the public node set and use uniform restart $\mu = \mathbf{1}/n$.

Theorem 2 (Dense-graph matching). *Fix $\beta \in (0,1)$ and public $0 < \rho \le 1/4$. For every integer $n \ge 4$, put $m = \max\{1, \lfloor \rho n \rfloor\}$, $\kappa = (1-\beta)/n + \beta/m$, $\delta = n^{-2}$ and $0 < \varepsilon \le 256$. The infima below range over mechanisms M that are globally (ε, δ)-row-DP. Then*

$$\inf_{M} \sup_{A:\, d_{\min}(A) \ge m} \mathbb{E} \frac{\|M(A) - \boldsymbol{\pi}(A)\|_2^2}{\|\boldsymbol{\pi}(A)\|_2^2} \asymp_{\beta,\rho,256} \min\left\{1, \frac{\log n + \varepsilon}{n\varepsilon^2}\right\}. \tag{8}$$

Algorithm 1 has zero regularization bias on these graphs and attains expected squared relative error $O_{\beta,\rho}((\log n + \varepsilon)/(n\varepsilon^2))$, and expected relative error at most its square root.

The lower bound uses a row-preserving encoding of empirical column means and the fingerprinting-code primitive of Steinke & Ullman (2016, Definition 3.2 and Theorem 3.3). Appendix A.5 supplies the complete argument, including the finite-n case.

3.4 Choice of κ and m

The degree floor m and source mass cap κ jointly control the regularization bias and privacy noise. The threshold m plays the role of a regularization parameter. A larger value of m reduces the noise required for privacy, because it improves the worst-case sensitivity bound for fixed κ. At the same time, a larger value of m modifies more rows and therefore increases the bias bound in Theorem 1. The choice of m is therefore a direct privacy–bias tradeoff. For fixed m, sensitivity is proportional to κ and the calibrated Gaussian variance is proportional to κ^2; decreasing κ can increase clipping bias.

The bias expression in Theorem 1 is especially interpretable. Rows with $d_i \ge m$ have no degree-flooring loss, but can still contribute clipping bias if $\pi_i > \kappa$. Under uniform restart and $d_{\min} \ge m$, the PageRank equation gives $\pi_j \le (1-\beta)/n + \beta/m$. Therefore the public choice $\kappa = (1-\beta)/n + \beta/m$ gives zero bias. When $m = \rho n$, this choice is $\kappa = (1 - \beta + \beta/\rho)/n$, as in the dense regime of Section 3.3. More generally, $\kappa = c/n$ sets the source cap at c times the average PageRank mass $1/n$; the experimental value $c = 4$ is a public setting, not an optimality claim.

From a practical perspective, there are several natural ways to choose m. If domain knowledge or external information suggests a lower bound on the out-degree, then m can be fixed in advance as a public tuning parameter. One may also select m based on a separate public dataset or simulation study. If one wishes to adapt m to the private graph itself, then this selection must be privatized and its privacy cost must be accounted for separately. In the present paper, we treat m as a fixed public parameter and focus on the utility guarantees conditional on that choice. The same requirements apply to κ.

4 Private PageRank: Top k

In many applications of PageRank, the objective is to identify the highest-ranked group of k nodes. The separation between that group and the remaining nodes determines the difficulty of recovery. We study this set-identification objective for original PageRank. The mechanism releases an ordered list of k indices, so its privacy guarantee must protect that entire list even though the recovery guarantee concerns its underlying set.

4.1 DP mechanism for top-k selection

Let $1 \le k < n$ be public. DP PageRank–Top k uses the exact regularized fixed point $\hat{\boldsymbol{\pi}}$ in (3). The perturbation analysis in Section 3.1, specifically Lemma 1, supplies its global coordinatewise sensitivity $\Delta_{\mathrm{reg},\infty}$. Adding independent $\mathrm{Gumbel}(0, b)$ noise to the coordinates and taking the largest k in order is equivalent to successive exponential-mechanism selections without replacement (McSherry & Talwar, 2007; Durfee & Rogers, 2019).

Here, b is the scale parameter of the Gumbel noise and controls the magnitude of the perturbation. It plays a role analogous to σ in the Gaussian mechanism of Algorithm 1. For total privacy budget

$\varepsilon > 0$, $0 < \delta < 1$, choose

$$b = \Delta_{\mathrm{reg},\infty} \min \left\{ \frac{2k}{\varepsilon}, \quad \frac{\sqrt{k/2}}{\sqrt{\log(1/\delta) + \varepsilon} - \sqrt{\log(1/\delta)}} \right\}. \tag{9}$$

The first scale uses pure-DP composition; the second uses the bounded range of each selection's privacy loss and zero-concentrated DP (zCDP) composition (Durfee & Rogers, 2019; Cesar & Rogers, 2021; Bun & Steinke, 2016). Both are valid for every $\varepsilon > 0$. Their minimum selects the smaller valid scale using public parameters only. Each output length k receives its own full-budget calibration.

Algorithm 2 DP PageRank–Top k

Require: A; public β, μ, m, κ, k; total budget (ε, δ)
Ensure: A private ordered list $(\hat{\sigma}_1, \ldots, \hat{\sigma}_k)$
1: Compute $\hat{\boldsymbol{\pi}}$, the unique fixed point in (3)
2: Compute $\Delta_{\mathrm{reg},\infty}$ and b from (4) and (9)
3: Draw independent $G_j \sim \mathrm{Gumbel}(0, b)$ for $j \in [n]$
4: Output the indices of the k largest $\hat{\pi}_j + G_j$, in decreasing order

Neither the perturbed values nor the unreleased portion of the ordering are published.

4.2 Utility analysis

Write the top k (unordered) indices as:

$$\mathcal{T}_k = \mathcal{T}(\boldsymbol{\pi}; k), \qquad \hat{\mathcal{T}}_k = \mathcal{T}(\hat{\boldsymbol{\pi}}; k), \qquad \tilde{\mathcal{T}}_k = \mathcal{T}(\hat{\boldsymbol{\pi}} + \boldsymbol{G}; k),$$

where $\boldsymbol{G} = (G_1, \ldots, G_n)$ is the Gumbel perturbation in Algorithm 2, and all sets use one fixed public tie rule. The set $\hat{\mathcal{T}}_k$ reflects the change due to regularization alone, while $\tilde{\mathcal{T}}_k = \{\hat{\sigma}_1, \ldots, \hat{\sigma}_k\}$ is the set of released indices. Define the boundary gap of original PageRank as $\mathrm{Gap}_k(\boldsymbol{\pi}) = \pi_{(k)} - \pi_{(k+1)}$, where $\pi_{(j)}$ is its jth largest coordinate.

Theorem 3 (Privacy). *For public $1 \le k < n$, $\varepsilon > 0$ and $0 < \delta < 1$, Algorithm 2 is globally (ε, δ)-row-level differentially private for its ordered output.*

Theorem 4 (Exact set recovery). *For $\eta \in (0, 1)$, if*

$$\mathrm{Gap}_k(\boldsymbol{\pi}) > 2\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_\infty + 2b \log \frac{2n}{\eta}, \tag{10}$$

then Algorithm 2 satisfies $\Pr(\tilde{\mathcal{T}}_k = \mathcal{T}_k) \ge 1 - \eta$.

The boundary gap must dominate the actual coordinatewise regularization error and the selection noise. An explicit upper bound on $\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_\infty$ follows from the proof of Theorem 1 in Appendix A; it is distinct from the squared ℓ_2 bias of Section 3.2. Theorem 4 guarantees recovery of the set, not its internal order.

Meanwhile, a wider target group gives a related guarantee for the fraction of released nodes belonging to the true top-L.

Corollary 2 (Selection within the true top-L). *For public integers $k \le L < n$ and $\eta \in (0, 1)$, if*

$$\pi_{(k)} - \pi_{(L+1)} > 2\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_\infty + 2b \log \frac{2n}{\eta}, \tag{11}$$

then $\Pr\big(\tilde{\mathcal{T}}_k \subseteq \mathcal{T}(\boldsymbol{\pi}; L)\big) \ge 1 - \eta$. On this event, $|\tilde{\mathcal{T}}_k \cap \mathcal{T}(\boldsymbol{\pi}; L)|/k = 1$.

Here L is an evaluation range; the mechanism still selects k indices using (9). This is a sufficient inclusion condition, not an optimality statement for partial overlap. Proofs of the privacy, recovery and inclusion results are in Appendix A.6 and the following subsections.

The public parameter choices in Section 3.4 also apply here. For top-k recovery, their effect is measured by coordinatewise bias and selection noise relative to the boundary gap. Full-vector error instead uses ℓ_2 bias and noise across all coordinates, so suitable settings of m and κ need not coincide for the two objectives. Selecting them from the private graph requires accounting for that selection's privacy cost.

4.3 LOWER BOUNDS AND MATCHING

For this subsection, use uniform restart, fixed $\beta \in (0,1)$, public $0 < \rho \le 1/4$, $m = \max\{1, \lfloor \rho n \rfloor\}$ and $\delta = n^{-2}$. Utility is evaluated on loop-free directed binary graphs with $d_{\min} \ge m$, while privacy is required on the full input domain. The lower bound concerns identifying the original top-k set, including when the published output is ordered.

Theorem 5 (Necessary top-k gap on dense inputs)**.** *Fix an integer $k \ge 1$. Let $n = 4N$, where N is even, $N \ge \max\{64, 2k\}$ and $n/k \ge 16 \cdot 4^{2/k}$. For $0 < \varepsilon \le \frac{k}{2}(\log n - \log k)$, every globally (ε, n^{-2})-row-DP mechanism releasing a k-element set or an ordered list of k distinct indices has exact-set-recovery failure probability at least $2/3$ on some graph A with $d_{\min}(A) \ge m$ and*

$$\mathrm{Gap}_k(\boldsymbol{\pi}(A)) \ge \frac{3\beta}{16} \min\left\{\frac{1}{n}, \frac{\log n}{n^2 \varepsilon}\right\}. \tag{12}$$

For each fixed k, the size conditions hold for all sufficiently large n along the stated divisibility convention. The proof gives a finite packing inequality for every $\varepsilon > 0$ and tracks the accumulated group-privacy δ; the displayed budget range ensures a constant failure bound.

Corollary 3 (Fixed-k matching)**.** *Fix k and $\eta \in (0, 1/2)$, and set $\kappa = (1-\beta)/n + \beta/m$. For every $n \ge 4$ with $k < n$ and every $\varepsilon > 0$, Algorithm 2 recovers the original top-k set with probability at least $1 - \eta$ on every graph with $d_{\min} \ge m$ satisfying*

$$\mathrm{Gap}_k(\boldsymbol{\pi}) > \frac{4k\Delta_{\mathrm{reg},\infty}}{\varepsilon} \log \frac{2n}{\eta}. \tag{13}$$

Under the size and budget conditions of Theorem 5, in the nonsaturated regime $\log n/(n\varepsilon) \to 0$, the necessary and sufficient gap orders match at $\log n/(n^2\varepsilon)$, with constants depending on β, ρ, k, η.

On these dense graphs, $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}$ and $\Delta_{\mathrm{reg},\infty} = O_{\beta,\rho}(n^{-2})$. In the saturated regime, the lower bound remains an obstruction at gaps of order $1/n$; the upper sufficient condition may be unattainable there. The matching statement does not determine optimal dependence on growing k, nor does it concern sparse graphs, $\delta = 0.1$, internal ordering, Average Rank, or top-L overlap.

5 EXPERIMENTS ON REAL DATA

Due to page limit, we present the simulation results in the Appendix E and discuss only the real data results here.

Datasets. We evaluate PageRank on three directed networks from SNAP. Email-Eu-core records email communication within a European research institution (SNAP, n.d.b), CollegeMsg records private sender-to-recipient messages on a university online social network (SNAP, n.d.a), and Wiki-Vote records Wikipedia administrator elections (SNAP, n.d.c). The networks have respectively 1,005, 1,899, and 7,115 nodes, and 24,929, 20,296, and 103,689 directed edges. Their median out-degrees are 14, 3, and 2.

Experimental setup. We use $\beta = 0.8$, uniform public restart, and $\delta = 0.1$. The public regularization parameters $m = 16$ and $\kappa = 4/n$ remain fixed across privacy budgets; they are not selected to optimize the observed results. Numerical implementation details are in Appendix B. Each method is evaluated on 50 releases on each graph, with means and standard errors computed from the per-release metrics. The reference is non-private PageRank $\boldsymbol{\pi}$ on the original graph.

Tables 2–3 record the relative ℓ_2 errors and top-k overlap. In the tables, PR Full Vector and PR Top k denote DP PageRank–Full Vector and DP PageRank–Top k, respectively. We consider Edge Flipping, the uniform vector, and PageRank with naive Gaussian and Gumbel noise as comparisons.

Vector error and ranking. The first five columns report relative ℓ_2 error overall and within four disjoint groups in the fixed non-private ranking. The group boundaries are $\lceil 0.05n \rceil$, $\lceil 0.3n \rceil$, and $\lceil 0.5n \rceil$. For a group S, the error is $\|(\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi})_S\|_2 / \|\boldsymbol{\pi}\|_2$; every denominator uses the full reference vector. We average these norms across releases. The four group errors therefore decompose the squared overall error for each release, not the square of its displayed mean.

For average top-k rank, we retain only the first k nodes in each released vector's ordering. These nodes receive ranks $1, \ldots, k$ and all other nodes receive $(n + k + 1)/2$. We then average the assigned

ranks over $\mathcal{T}(\boldsymbol{\pi}; k)$ (defined in in (36)). This evaluation uses the top-k list rather than the complete vector ordering. Independent random tie breaking for the uniform vector gives the exact expected rank $(n+1)/2$. The main table reports $k = 10$; Appendix D contains all results for $\varepsilon = 32, 64, 128$ and $k = 1, 5, 10$.

Table 1: Relative ℓ_2 error and average top-10 rank at $\varepsilon = 64$; lower values are better. Best and second-best values are bold and italic, respectively, within each dataset and metric; ranking uses unrounded means.

		Relative ℓ_2 error					
Data Set	Method	Overall	Top 5%	5–30%	30–50%	Bottom 50%	Average Rank of Top 10 nodes
Email-Eu-core	PR Full Vector	.736	**.345**	.414	.286	.410	**311.6**
	Edge Flipping	**.643**	*.474*	**.293**	*.043*	**.318**	*494.9*
	Uniform	*.644*	.475	*.294*	**.042**	*.318*	503.0
	Naive Gaussian	85.056	19.405	42.530	37.736	60.138	502.0
CollegeMsg	PR Full Vector	.884	**.581**	.489	.267	.365	**703.9**
	Edge Flipping	*.735*	*.605*	*.288*	*.080*	*.290*	947.4
	Uniform	**.735**	.605	**.288**	**.080**	**.290**	950.0
	Naive Gaussian	143.503	32.820	72.050	63.748	101.233	*945.5*
Wiki-Vote	PR Full Vector	.915	**.731**	.430	.184	.290	**2281.7**
	Edge Flipping	*.844*	*.749*	*.280*	*.142*	*.230*	*3555.9*
	Uniform	**.844**	.749	**.280**	**.142**	**.230**	3558.0
	Naive Gaussian	423.633	95.266	210.759	189.488	300.049	3563.0

At $\varepsilon = 64$, PR Full Vector has a larger overall error than Uniform on all three networks, but a smaller top-5% error and a lower average top-10 rank. On Email-Eu-core, the overall errors are 0.736 versus 0.644, while the top-5% errors are 0.345 versus 0.475 and the average top-10 ranks are 311.6 versus 503.0. The top-5% error improvements are smaller on CollegeMsg (0.581 versus 0.605) and Wiki-Vote (0.731 versus 0.749). Low-ranked nodes have similar small PageRank values, so a nearly uniform vector can have small numerical errors without retaining the head of the ranking. The rank column shows this distinction; the results do not establish uniform superiority across datasets or error groups.

Top-L overlap. Table 2 evaluates the fraction of K selected nodes that belong to $\mathcal{T}(\boldsymbol{\pi}; L)$, for $K = 1, 5$. Here K denotes the output length k of the mechanism, while L specifies the reference set used for evaluation. For vector methods, we take the first K nodes of the same raw release. Each Gumbel mechanism uses its full (ε, δ) budget separately for each K; the $K = 1$ output is not the prefix of a release calibrated for $K = 5$. Naive Gumbel applies the selection rule in (9) to original PageRank with coordinatewise sensitivity bound β. Uniform selection has expected overlap L/n. When $K = L = 1$, overlap is the top-1 hit rate; for $K > 1$, mean overlap is not the probability of recovering the entire set.

At $\varepsilon = 32$, PR Top k selects the true top node in 96%, 82%, and 94% of releases on Email-Eu-core, CollegeMsg, and Wiki-Vote, respectively, compared with 6%, 8%, and 24% for PR Full Vector. For $K = L = 5$ at the same budget, PR Full Vector has the higher mean overlap on all three networks (18.8%, 10.4%, 18.4% versus 17.2%, 4.4%, 10.0%). At $\varepsilon = 128$, this comparison reverses: PR Top k reaches 77.2%, 71.2%, and 60.4%, versus 53.6%, 39.6%, and 53.6% for PR Full Vector. Thus the observed advantage depends on both the budget and the output length. The smaller budget $\varepsilon = 8$ gives substantially weaker overlap. Increasing L improves coverage of high-ranked nodes, but it relaxes the target and does not establish strict top-K set recovery. All percentages for randomized methods are empirical averages; observing 0 or 100% over 50 releases does not establish a population probability of zero or one.

6 Conclusion and Discussion

We study PageRank under privacy for a node's entire outgoing-link profile. Degree flooring limits individual transition weights, while mass clipping limits each source's contribution budget. A shared

Table 2: Top-L overlap (%) on real networks. Each cell lists $\varepsilon = 8 \,/\, 32 \,/\, 128$ in that order; larger values are better. Bold and italic identify the best and second-best means within each dataset, column and privacy budget.

		Top-1 node			Top-5 node		
Data Set	Method	$L = 1$	$L = 3$	$L = 5$	$L = 5$	$L = 15$	$L = 25$
Email-Eu-core	PR Full Vector	0/*6*/*46*	*2*/*26*/*76*	*4*/*32*/*88*	**4/18.8**/*53.6*	**7.6/35.2**/*84.8*	**9.6/42**/*91.6*
	PR Top k	**18/96/100**	**32/100/100**	**36/100/100**	.8/*17.2*/**77.2**	3.2/*30.4*/**98.8**	5.2/*36*/**99.2**
	Edge Flipping	0/0/0	0/2/0	0/4/0	.8/.8/2	2.8/2/6	4/4/8
	Uniform	.10/.10/.10	.30/.30/.30	.50/.50/.50	.50/.50/.50	1.49/1.49/1.49	2.49/2.49/2.49
	Naive Gaussian	*2*/0/0	*2*/0/0	2/0/0	*2*/.4/.8	*4*/1.2/2	5.2/2/2.8
	Naive Gumbel	0/0/0	0/0/0	0/0/0	1.6/1.2/.8	3.6/1.6/2	*5.6*/2.8/3.6
College Msg	PR Full Vector	0/*8*/*28*	*2*/*18*/*34*	*2*/*24*/*52*	**.4/10.4**/*39.6*	**2/18.8**/*69.6*	**4.4/23.6**/*75.2*
	PR Top k	**8/82/100**	**10/90/100**	**12/98/100**	**.4**/*4.4*/**71.2**	*1.6*/*8*/**96**	*1.6*/*10*/**98**
	Edge Flipping	0/0/0	0/0/0	*2*/0/0	**.4**/0/0	.4/0/.8	1.2/.4/1.6
	Uniform	*.05*/.05/.05	.16/.16/.16	.26/.26/.26	*.26*/.26/.26	.79/.79/.79	1.32/1.32/1.32
	Naive Gaussian	0/2/0	0/2/0	0/2/0	0/.4/1.2	1.2/.4/1.6	1.2/.8/1.6
	Naive Gumbel	0/0/0	0/0/0	0/0/0	0/0/.4	.8/.4/.8	*1.6*/.8/1.6
Wiki-Vote	PR Full Vector	0/*24*/*66*	*2*/*34*/*90*	*4*/*36*/*98*	**2.4/18.4**/*53.6*	**3.6/29.2**/*80.4*	**3.6/32.4**/*86.4*
	PR Top k	**26/94/100**	**36/100/100**	**36/100/100**	*.4*/*10*/**60.4**	*1.6*/*14.4*/**94.8**	*2.4*/*15.6*/**97.6**
	Edge Flipping	0/0/0	0/0/0	0/0/0	0/0/0	0/0/.4	0/0/.4
	Uniform	*.01*/.01/.01	.04/.04/.04	.07/.07/.07	.07/.07/.07	.21/.21/.21	.35/.35/.35
	Naive Gaussian	0/0/0	0/0/0	0/0/0	0/.4/.4	0/.4/.4	0/.4/.4
	Naive Gumbel	0/0/0	0/0/0	0/0/0	*.4*/.4/.4	.8/.4/.4	.8/.4/.4

regularized fixed point supports full-vector Gaussian release and ordered top-k Gumbel release, with explicit sensitivity and regularization-bias bounds relative to original PageRank.

The lower bounds show why uniform accuracy over unrestricted graphs is impossible and identify a dense utility class on which full-vector relative-risk rates match. The top-1 recovery gap has a matching order in the stated dense, nonsaturated regime. These results do not imply sparse-graph minimax optimality, which can be future research problems. The privacy model retains a public node set and protects outgoing-row replacement; extending it to full node removal remains a separate question. Simulations and experiments on three real datasets from SNAP have demonstrated the superiority of our proposed methods over existing ones.

AI Use Statement

We used generative AI tools, including OpenAI Codex, to assist with theoretical development, implementation of the methods and synthetic graph experiments, data preparation and implementation. These tools also assisted with literature search, manuscript editing, and preparation of figures and tables. We have verified the mathematical proofs and final claims. We have read the suggested references. We have checked the final code. The authors are responsible for the final manuscript and its mathematical claims, experimental results, and references.

Ethics Statement

This work studies privacy protection for PageRank releases on directed networks. Our empirical evaluation uses publicly available network datasets from SNAP, with sources identified in Section 5. We did not recruit participants or collect new personal information. The privacy guarantees concern replacement of one outgoing-link profile on a fixed public node set; they do not protect node participation or the removal of all incident edges. Their interpretation depends on the stated privacy parameters and adjacency relation. The reported experiments evaluate utility under these specified conditions.

Reproducibility Statement

The mechanisms and their assumptions are specified in Sections 3 and 4, with proofs provided in Appendix A. Appendix B describes the numerical computation of the regularized PageRank vector. Section 5 identifies the real datasets and reports the experimental settings, evaluation metrics, and number of repeated releases. Additional real-data results are provided in Appendix D.

References

Raman Arora and Jalaj Upadhyay. On differentially private graph sparsification and applications. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019.

Borja Balle and Yu-Xiang Wang. Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 394–403. PMLR, 2018. URL https://proceedings.mlr.press/v80/balle18a.html.

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds, 2016. URL https://arxiv.org/abs/1605.02065.

Mark Cesar and Ryan Rogers. Bounding, concentrating, and truncating: Unifying privacy loss composition for data analytics. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pp. 421–457. PMLR, 2021. URL https://proceedings.mlr.press/v132/cesar21a.html.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109, 2011.

David Durfee and Ryan M. Rogers. Practical differentially private top-k selection with pay-what-you-get composition. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/b139e104214a08ae3f2ebcce149cdf6e-Abstract.html.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006.

Alessandro Epasto, Vahab Mirrokni, Bryan Perozzi, Anton Tsitsulin, and Peilin Zhong. Differentially private graph learning via sensitivity-bounded personalized pagerank. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, 2022. URL https://arxiv.org/abs/2207.06944v3. Corrected version: arXiv:2207.06944v3, 2024.

Laura A. Granka, Thorsten Joachims, and Geri Gay. Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2004. URL https://www.cs.cornell.edu/People/tj/publications/granka_etal_04a.pdf.

Michael Hay, Chao Li, Gerome Miklau, and David Jensen. Accurate estimation of the degree distribution of private networks. In *Proceedings of the 9th IEEE International Conference on Data Mining (ICDM)*, pp. 169–178. IEEE, 2009. doi: 10.1109/ICDM.2009.11.

Jonathan Hehir, Aleksandra Slavković, and Xiaoyue Niu. Consistent spectral clustering of network block models under local differential privacy. *Journal of Privacy and Confidentiality*, 12(2), 2022. doi: 10.29012/jpc.785.

Vishesh Karwa, Pavel N. Krivitsky, and Aleksandra B. Slavković. Sharing social network data: Differentially private estimation of exponential-family random graph models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3):481–500, 2017. doi: 10.1111/rssc.12185.

Shiva Prasad Kasiviswanathan, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Analyzing graphs with node differential privacy. In *Theory of Cryptography Conference*, pp. 457–476. Springer, 2013.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science*, 2007. URL https://www.microsoft.com/en-us/research/publication/mechanism-design-via-differential-privacy/.

Yvonne Mülle, Chris Clifton, and Klemens Böhm. Privacy-integrated graph clustering through differential privacy. In *Proceedings of the Workshops of the EDBT/ICDT 2015 Joint Conference*, volume 1330 of *CEUR Workshop Proceedings*, pp. 247–254. CEUR-WS.org, 2015.

Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 75–84. ACM, 2007. doi: 10.1145/1250790.1250803.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999. URL http://ilpubs.stanford.edu:8090/422/.

Sofya Raskhodnikova and Adam Smith. Lipschitz extensions for node-private graph statistics and the generalized exponential mechanism. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 495–504. IEEE, 2016.

SNAP. CollegeMsg temporal network. https://snap.stanford.edu/data/CollegeMsg.html, n.d.a. Accessed September 26, 2026.

SNAP. Email-Eu-core network. https://snap.stanford.edu/data/email-Eu-core.html, n.d.b. Accessed September 26, 2026.

SNAP. Wikipedia vote network. https://snap.stanford.edu/data/wiki-Vote.html, n.d.c. Accessed September 26, 2026.

Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *Journal of Privacy and Confidentiality*, 7(2):3–22, 2016. doi: 10.29012/jpc.v7i2.648.

Wanjie Wang and Tathagata Banjernee. Decoupling regularization and privacy in differentially private ridge regression and ERM. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=ueWWbqp0rc.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. doi: 10.1080/01621459.1965.10480775.

Rongzhe Wei, Eli Chien, and Pan Li. Differentially private graph diffusion with applications in personalized pageranks. In *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024.

Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. Private release of graph statistics using ladder functions. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pp. 731–745. ACM, 2015. doi: 10.1145/2723372.2737785.

A Proofs

A.1 Proof of Lemma 1

For every row $q_i^\top$ of Q, nonnegativity and the degree floor give $\sum_j Q_{ij} = d_i/\max\{d_i, m\} \le 1$ and $0 \le Q_{ij} \le 1/m$. Therefore $\|q_i\|_1 \le 1$, $\|q_i\|_\infty \le 1/m$, and $\|q_i\|_2^2 \le (1/m)\sum_j Q_{ij} \le 1/m$. In particular, for any signed vector $\boldsymbol{v}$, the triangle inequality gives

$$\|Q^\top \boldsymbol{v}\|_1 \le \sum_i |v_i| \|q_i\|_1 \le \|\boldsymbol{v}\|_1, \quad \|Q^\top \boldsymbol{v}\|_2 \le \sum_i |v_i| \|q_i\|_2 \le \frac{\|\boldsymbol{v}\|_1}{\sqrt{m}}, \quad \|Q^\top \boldsymbol{v}\|_\infty \le \frac{\|\boldsymbol{v}\|_1}{m}.$$

Coordinatewise clipping is 1-Lipschitz in ℓ_1. Thus the map $F(\boldsymbol{z}) = (1-\beta)\mu+\beta Q^\top \operatorname{clip}_\kappa(\boldsymbol{z})$ satisfies $\|F(\boldsymbol{z}) - F(\boldsymbol{z}')\|_1 \le \beta\|\boldsymbol{z} - \boldsymbol{z}'\|_1$. It maps the closed set of nonnegative vectors with mass at most one into itself: $\|F(\boldsymbol{z})\|_1 \le (1-\beta) + \beta\|\boldsymbol{z}\|_1 \le 1$. This set is complete, so the contraction theorem gives a unique fixed point there. Every nonnegative fixed point satisfies $\|\hat{\boldsymbol{\pi}}\|_1 \le (1-\beta) + \beta\|\hat{\boldsymbol{\pi}}\|_1$, hence has mass at most one. This also proves uniqueness on the entire nonnegative orthant.

For adjacent inputs differing in row i, let $w = q_i - q_i'$. Since both rows are nonnegative, $\langle q_i, q_i'\rangle \ge 0$, so $\|w\|_2^2 \le \|q_i\|_2^2 + \|q_i'\|_2^2 \le 2/m$. Also $\|w\|_1 \le 2$ and $\|w\|_\infty \le 1/m$. Put $\boldsymbol{c} = \operatorname{clip}_\kappa(\hat{\boldsymbol{\pi}})$ and $\boldsymbol{c}' = \operatorname{clip}_\kappa(\hat{\boldsymbol{\pi}}')$. Subtracting the two fixed-point equations and using $Q - Q' = e_i w^\top$ gives

$$\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}' = \beta Q'^\top(\boldsymbol{c} - \boldsymbol{c}') + \beta c_i w, \qquad 0 \le c_i \le \kappa.$$

Consequently, $\|\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}'\|_1 \le \beta\|\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}'\|_1 + 2\beta\kappa$. Rearrangement, followed by the other row bounds, proves all three claims:

$$\begin{aligned}
\|\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}'\|_1 &\le \frac{2\beta\kappa}{1-\beta} = \Delta_{\mathrm{reg},1},\\
\|\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}'\|_2 &\le \frac{\beta}{\sqrt{m}}\|\boldsymbol{c} - \boldsymbol{c}'\|_1 + \beta\kappa\sqrt{2/m} \le \frac{\beta\Delta_{\mathrm{reg},1}}{\sqrt{m}} + \beta\kappa\sqrt{2/m} = \Delta_{\mathrm{reg},2},\\
\|\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}'\|_\infty &\le \frac{\beta}{m}\|\boldsymbol{c} - \boldsymbol{c}'\|_1 + \frac{\beta\kappa}{m} \le \frac{\beta\Delta_{\mathrm{reg},1}}{m} + \frac{\beta\kappa}{m} = \Delta_{\mathrm{reg},\infty}.
\end{aligned}$$

A dangling row is zero in Q and satisfies every inequality used above.

A.2 Proof of Theorem 1

The original transition matrix P has row sums one, including its uniform dangling rows. Since $P - Q$ is entrywise nonnegative, the PageRank equation gives the nonnegative residual

$$\boldsymbol{\pi} - F(\boldsymbol{\pi}) = \beta\big[(P-Q)^\top\boldsymbol{\pi} + Q^\top(\boldsymbol{\pi} - \operatorname{clip}_\kappa(\boldsymbol{\pi}))\big] \ge 0.$$

Summing its coordinates and using the row sums of P and Q yields

$$\|\boldsymbol{\pi} - F(\boldsymbol{\pi})\|_1 = \beta\sum_i \left[\pi_i - \frac{d_i}{\max\{d_i, m\}}\min\{\pi_i, \kappa\}\right].$$

At $d_i = 0$, the corresponding summand is π_i, so the identity also includes dangling rows. By the contraction inequality,

$$\|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_1 \le \|\boldsymbol{\pi} - F(\boldsymbol{\pi})\|_1 + \|F(\boldsymbol{\pi}) - F(\hat{\boldsymbol{\pi}})\|_1 \le \|\boldsymbol{\pi} - F(\boldsymbol{\pi})\|_1 + \beta\|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_1.$$

Dividing by $1-\beta$ and applying $\|\boldsymbol{v}\|_2 \le \|\boldsymbol{v}\|_1$ gives the bound on $\mathrm{bias}^2 = \|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2^2$.

For the coordinatewise bound used in Section 4, subtraction gives

$$\boldsymbol{\pi} - \hat{\boldsymbol{\pi}} = \beta(P-Q)^\top\boldsymbol{\pi} + \beta Q^\top[\boldsymbol{\pi} - \operatorname{clip}_\kappa(\boldsymbol{\pi})] + \beta Q^\top[\operatorname{clip}_\kappa(\boldsymbol{\pi}) - \operatorname{clip}_\kappa(\hat{\boldsymbol{\pi}})].$$

The ℓ_∞ row bound and clipping's ℓ_1 Lipschitz property imply

$$\begin{aligned}
\|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_\infty &\le \beta\|(P-Q)^\top\boldsymbol{\pi}\|_\infty + \frac{\beta}{m}\left[\sum_i(\pi_i - \kappa)_+ + \|\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}\|_1\right]\\
&\le \beta\|(P-Q)^\top\boldsymbol{\pi}\|_\infty + \frac{\beta}{m}\sum_i(\pi_i - \kappa)_+\\
&\quad + \frac{\beta^2}{m(1-\beta)}\sum_i\left\{\pi_i - \frac{d_i}{\max\{d_i, m\}}\min\{\pi_i, \kappa\}\right\},
\end{aligned}$$

where the final step substitutes the explicit ℓ_1 bound just proved. This bound is not the actual ℓ_2 bias. If $d_{\min} \ge m$ and $\|\boldsymbol{\pi}\|_\infty \le \kappa$, then $Q = P$ and $\operatorname{clip}_\kappa(\boldsymbol{\pi}) = \boldsymbol{\pi}$. Hence $F(\boldsymbol{\pi}) = \boldsymbol{\pi}$, and uniqueness gives $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}$.

A.3 Proof of Corollary 1

Lemma 1 bounds the global ℓ_2 sensitivity by $\Delta_{\mathrm{reg},2}$. The analytic Gaussian condition (5) therefore ensures (ε, δ)-row-DP by Balle & Wang (2018, Theorem 8); the monotonicity in mean separation from their Lemma 7 permits using a sensitivity upper bound. Its hypotheses allow every $\varepsilon \geq 0$, with $0 < \delta < 1$ here.

At a fixed graph, $\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}$ is deterministic and the release is $\tilde{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} + \sigma\zeta$, with $\mathbb{E}\zeta = 0$ and $\mathbb{E}\|\zeta\|_2^2 = n$. Expanding the square gives

$$\mathbb{E}\|\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2^2 = \|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2^2 + 2\sigma\langle\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}, \mathbb{E}\zeta\rangle + \sigma^2\mathbb{E}\|\zeta\|_2^2 = \mathrm{bias}^2 + n\sigma^2.$$

The deterministic probability vector $\boldsymbol{\pi}$ has positive norm, so division by $\|\boldsymbol{\pi}\|_2^2$ gives the squared relative-error identity. Finally, Cauchy–Schwarz yields $\mathbb{E}\|\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2 \leq \sqrt{\mathbb{E}\|\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2^2}$; division by $\|\boldsymbol{\pi}\|_2$ proves the non-squared bound.

A.4 A lower bound

Theorem 6 (Unrestricted-graph lower bounds). *Fix $\beta \in (0,1)$, $n \geq 3$, $\varepsilon \geq 0$ and $0 \leq \delta < 1$. Every globally (ε, δ)-row-DP vector mechanism M satisfies*

$$\sup_A \mathbb{E}\frac{\|M(A) - \boldsymbol{\pi}(A)\|_2}{\|\boldsymbol{\pi}(A)\|_2} \geq \frac{\beta^2}{\sqrt{2}(1+\beta)}\frac{1-\delta}{1+e^{\varepsilon}}, \tag{14}$$

$$\sup_A \mathbb{E}\frac{\|M(A) - \boldsymbol{\pi}(A)\|_2^2}{\|\boldsymbol{\pi}(A)\|_2^2} \geq \frac{\beta^4}{2(1+\beta)^2}\frac{1-\delta}{1+e^{\varepsilon}}. \tag{15}$$

For fixed β, $0 < \varepsilon \leq 256$ and $\delta \leq 1/2$, these bounds do not vanish as n increases. Their constants can be small; the result is a worst-case obstruction to uniform relative consistency.

We first give a two-point argument that will also be used for the finite-size part of Corollary 2. For adjacent graphs with PageRank vectors $\boldsymbol{p}_0, \boldsymbol{p}_1$, let $D = \|\boldsymbol{p}_0 - \boldsymbol{p}_1\|_2$ and $B = \max\{\|\boldsymbol{p}_0\|_2, \|\boldsymbol{p}_1\|_2\}$. Decode the output of an arbitrary private vector mechanism by choosing its nearer target in ℓ_2, resolving an equality by a fixed rule. Let a be its probability of choosing target 1 on input 0 and b its probability of choosing target 0 on input 1. Applying privacy to the two decision events, in the two directions, gives $1 - a \leq e^{\varepsilon}b + \delta$ and $1 - b \leq e^{\varepsilon}a + \delta$. Adding and rearranging yields $(a+b)/2 \geq (1-\delta)/(1+e^{\varepsilon})$.

On a decoding error, the triangle inequality forces the distance from the true target to be at least $D/2$: if the other target is no farther from the output, their mutual distance is at most twice the error. The relative error on that event is therefore at least $D/(2B)$. Averaging the two inputs and then bounding the average by the maximum proves separately

$$\max_{j\in\{0,1\}} \mathbb{E}\frac{\|M(A_j) - \boldsymbol{p}_j\|_2}{\|\boldsymbol{p}_j\|_2} \geq \frac{D}{2B}\frac{1-\delta}{1+e^{\varepsilon}},$$

$$\max_{j\in\{0,1\}} \mathbb{E}\frac{\|M(A_j) - \boldsymbol{p}_j\|_2^2}{\|\boldsymbol{p}_j\|_2^2} \geq \frac{D^2}{4B^2}\frac{1-\delta}{1+e^{\varepsilon}}.$$

For the unrestricted domain, take node 3 as a hub. Every other node has its only edge to node 3. In A_0 the hub points only to node 1, and in A_1 it points only to node 2. These graphs exist for every $n \geq 3$, have no self-loops or dangling rows, and differ in one outgoing row. Put $t = (1-\beta)/n$. If H is the hub's PageRank mass, its equation is $H = t + \beta(1-H)$, hence $H = (\beta + t)/(1+\beta)$. The chosen neighbor has score $t + \beta H$; all other non-hub nodes have score t. These scores sum to $(1+\beta)H + (n-1)t = \beta + nt = 1$, and satisfy every coordinate equation. Their squared norms and separation are

$$B^2 = H^2 + (t + \beta H)^2 + (n-2)t^2, \qquad D^2 = 2\beta^2 H^2.$$

Because both targets are probability vectors, $B \leq 1$, while $H \geq \beta/(1+\beta)$. Thus $D/(2B) = \beta H/(\sqrt{2}B) \geq \beta^2/[\sqrt{2}(1+\beta)]$. Substitution into the two inequalities proves the stated bounds, without an asymptotic restriction on n. At $0 < \varepsilon \leq 256$ and $\delta \leq 1/2$, the remaining multiplier is at least $1/[2(1+e^{256})]$, explaining both the nonvanishing obstruction and its potentially small constant.

A.5 Proof of Corollary 2

We prove the upper bounds first. For any $\varepsilon > 0$, put $L = \log(1/\delta)$ temporarily. Gaussian perturbation at sensitivity $\Delta_{\mathrm{reg},2}$ is private at the sufficient scale

$$\sigma = \frac{\Delta_{\mathrm{reg},2}}{\sqrt{2}\,\varepsilon}(\sqrt{L+\varepsilon} + \sqrt{L}).$$

Indeed, for two means at distance $D \le \Delta_{\mathrm{reg},2}$, the log density ratio under the first Gaussian law has distribution $\mathcal{N}(a, 2a)$, where $a = D^2/(2\sigma^2)$. This follows by expanding the two squared distances in their densities. The displayed scale gives $a \le (\sqrt{L+\varepsilon} - \sqrt{L})^2$, and hence $a + 2\sqrt{aL} \le \varepsilon$. The Gaussian Chernoff bound therefore bounds the probability that the log density ratio exceeds ε by $e^{-L} = \delta$. On its complement, the first density is at most e^{ε} times the second; integration over any measurable event proves privacy. The same argument applies in the reverse direction, and $D = 0$ is immediate. Since AGM selects the smallest valid scale, its variance satisfies

$$\sigma^2_{\mathrm{AGM}} \le \frac{\Delta^2_{\mathrm{reg},2}}{2\varepsilon^2}(\sqrt{L+\varepsilon} + \sqrt{L})^2 \le \frac{\Delta^2_{\mathrm{reg},2}(2L+\varepsilon)}{\varepsilon^2}. \tag{16}$$

The last inequality is $2\sqrt{L(L+\varepsilon)} \le 2L + \varepsilon$. This is only an upper bound on the analytic calibration used in Algorithm 1, and imposes no small-budget restriction.

For the rounded public degree floor in the corollary, $\rho n/2 \le m \le N := \lfloor n/4 \rfloor$. To see the lower inequality, if $\rho n < 2$ then $m \ge 1 \ge \rho n/2$, and otherwise $\lfloor \rho n \rfloor \ge \rho n/2$; the upper inequality follows from $\rho \le 1/4$ and $N \ge 1$. On every graph with $d_{\min} \ge m$, $Q = P$, and the PageRank equation gives $\pi_j \le (1-\beta)/n + \beta/m = \kappa$. Theorem 1 gives $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}$. In particular,

$$\Delta_{\mathrm{reg},2} \le \beta\sqrt{\frac{2}{\rho}}\left(1 - \beta + \frac{2\beta}{\rho}\right)\left(\sqrt{2} + \frac{2\beta}{1-\beta}\right) n^{-3/2}.$$

Since $\|\boldsymbol{\pi}\|_2^2 \ge 1/n$, the exact raw Gaussian identity and (16), now with $\delta = n^{-2}$, give

$$\begin{aligned}\sup_{A:\, d_{\min}(A) \ge m} \mathbb{E}\frac{\|\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2^2}{\|\boldsymbol{\pi}\|_2^2} &\le n^2 \sigma^2_{\mathrm{AGM}} \\ &\le \frac{2\beta^2}{\rho}\left(1 - \beta + \frac{2\beta}{\rho}\right)^2\left(\sqrt{2} + \frac{2\beta}{1-\beta}\right)^2 \frac{4\log n + \varepsilon}{n\varepsilon^2}.\end{aligned}$$

Cauchy–Schwarz proves the corresponding expected relative-error bound. To establish the capped minimax upper bounds, the constant output $\mathbf{1}/n$ is globally private and has squared relative error $1 - 1/(n\|\boldsymbol{\pi}\|_2^2) \le 1$, and thus relative error at most one. Choose between this output and Algorithm 1 using the public number $n^2\sigma^2_{\mathrm{AGM}}$. This gives both claimed capped rates, since $4\log n + \varepsilon \le 4(\log n + \varepsilon)$. The cap is a bound for this public choice; it is not a cap on the raw Gaussian release.

We next construct the hard graphs for the lower bounds. Partition $4N$ nodes into N variable nodes R, N anchor nodes H, and $2N$ target nodes v_j^+, v_j^-, $j \in [N]$, and put $U = R \cup H$. The remaining $n - 4N$ nodes are fillers. Given $X \in \{-1, +1\}^{N \times N}$, variable node i points to v_j^+ when $X_{ij} = +1$ and to v_j^- otherwise, for each j. Every anchor points to all $2N$ targets. Every target and filler points to all $2N$ nodes in U. No edges point to a filler. The graph is binary and loop-free, every degree is at least $N \ge m$, and changing one database row changes exactly one outgoing graph row.

Write $s_j = N^{-1}\sum_i X_{ij}$, $t = (1-\beta)/n$, and use $\pi(U) = \sum_{u \in U}\pi_u$ for the total mass of U. All mass outside U flows uniformly into U, so $\pi(U) = 2Nt + \beta(1 - \pi(U))$, or $\pi(U) = (\beta + 2Nt)/(1+\beta)$. The individual PageRank equations are therefore

$$\pi_u = \frac{\pi(U)}{2N}\ (u \in U), \qquad \pi_{v_j^\pm} = t + \frac{\beta\pi(U)}{2N} \pm \frac{\beta\pi(U)}{4N}s_j, \qquad \pi_w = t\ (w \text{ a filler}). \tag{17}$$

For example, the variable nodes contribute $\pi(U)(1 \pm s_j)/(4N)$ to target $v_j^\pm$, and anchors contribute $\pi(U)/(4N)$, giving the middle formula after multiplying by β and adding t. These coordinates satisfy all equations and sum to one, so uniqueness identifies them with original PageRank. Since $n \ge 4N$, we have $\beta/(1+\beta) \le \pi(U) \le 1/2$. Coordinates in U are at most $1/(4N)$; a target coordinate is

at most $(1-\beta)/(4N) + 3\beta/(8N) \le 3/(8N)$, and a filler coordinate is at most $1/(4N)$. Hence $\|\boldsymbol{\pi}\|_2^2 \le \max_i \pi_i \le 3/(8N)$.

Given any globally private vector mechanism M, estimate s_j by multiplying $M_{v_j^+} - M_{v_j^-}$ by $2N/(\beta\pi(U))$ and clipping to $[-1,1]$. Denote the result by $\hat{s}_j$. The factor is public and independent of X. Clipping cannot increase its error, and Cauchy–Schwarz over the N pairs gives the pointwise inequality

$$\frac{1}{N}\sum_{j=1}^{N}|\hat{s}_j - s_j| \le \frac{2\sqrt{2N}}{\beta\pi(U)}\|M(A) - \boldsymbol{\pi}(A)\|_2 \le \frac{\sqrt{3}(1+\beta)}{\beta^2}\frac{\|M(A) - \boldsymbol{\pi}(A)\|_2}{\|\boldsymbol{\pi}(A)\|_2}. \tag{18}$$

Thus this decoder is a row-private empirical-mean mechanism on all sign databases, with a fixed-input error bound. We give its required lower bound explicitly.

The fingerprinting primitive of Steinke & Ullman (2016, Definition 3.2 and Theorem 3.3) supplies a universal $C_{\mathrm{fp}} \ge 1$ as follows. For q users and N coordinates, if $N \ge C_{\mathrm{fp}} q^2 \log(1/\xi)$, there is a random code $C \in \{-1,+1\}^{q\times N}$ and a tracing procedure Trace with completeness error $1/100$, robustness radius $N/8$, and per-user soundness ξ. Specifically, the probability that an output is within $N/8$ in ℓ_1 of the code's empirical mean but traces no user is at most $1/100$. For each fixed user i, if its code row is replaced by a fixed public row before producing the output, the probability that the tracer names i is at most ξ. These guarantees hold for arbitrary output procedures taking values in $[-1,1]^N$. Randomized procedures are included by conditioning on their independent coins and then averaging. We use this primitive directly, rather than assigning a numerical range to a small-ε marginal-estimation theorem.

Let α be the supremum, over deterministic $N \times N$ sign databases, of the expected average ℓ_1 error of any (ε,δ)-private empirical-mean mechanism, with $\delta = n^{-2}$. Set $L = 2\log n$ and $T = \log(n^2/(8N))$. Since $4N \le n \le 7N$, we have $L/4 \le T \le L/2$: the first inequality is equivalent to $n^{3/2} \ge 8N$, and the second to $n \le 8N$. For now assume $N \ge 4C_{\mathrm{fp}}(L+\varepsilon)$.

First suppose $\varepsilon \le T/2$. Choose $q = \lfloor\sqrt{N/(C_{\mathrm{fp}}L)}\rfloor$ and $r = \min\{\lfloor N/q\rfloor, \lfloor T/\varepsilon\rfloor\}$. Then $q \ge 2$, $q \ge \frac{1}{2}\sqrt{N/(C_{\mathrm{fp}}L)}$, and $q \le N/2$. Both quantities rounded in the definition of r are at least two, so $r \ge \frac{1}{2}\min\{N/q, T/\varepsilon\}$, $rq \le N$, and $r\varepsilon \le T$. Take the fingerprinting code with $\xi = \delta$, which exists because $C_{\mathrm{fp}}q^2L \le N$. Form $X(C)$ by repeating every code row r times and padding its remaining $N - rq$ rows with the all-$+1$ row. The empirical means obey $\overline{X} = (rq/N)\overline{C} + (1 - rq/N)\mathbf{1}$.

Apply the proposed mean mechanism to $X(C)$, subtract the known padding term, multiply by $N/(rq)$, and clip to $[-1,1]^N$; call the result $B(C)$. Its expected average error relative to $\overline{C}$ is at most $N\alpha/(rq)$. A code-row replacement changes r input rows. Iterating the privacy inequality along those replacements gives group privacy parameters $r\varepsilon$ and $\delta_r = \delta\sum_{j=0}^{r-1} e^{j\varepsilon} \le r\delta e^{r\varepsilon}$.

If $\alpha \le rq/(100N)$, Markov's inequality gives probability at least $1 - 8/100 = 0.92$ that $B(C)$ is within the robustness radius. Completeness then gives probability at least 0.91 of tracing a nonempty set. Some user i is therefore traced with probability at least $0.91/q$. For each realization of the code and tracing randomness, apply group privacy to the event of tracing i, comparing C to the code with row i replaced by the public row, and then average. Soundness bounds the replacement experiment by $\xi = \delta$, and so

$$\Pr[i \in \mathrm{Trace}(B(C))] \le e^{r\varepsilon}\delta + \delta_r \le (1+r)\delta e^{r\varepsilon} \le \frac{1+r}{8N} \le \frac{1}{4q}.$$

The penultimate step uses $e^{r\varepsilon} \le e^T = 1/(8N\delta)$; the final step uses $q + rq \le 2N$. This contradicts $0.91/q$, so $\alpha > rq/(100N)$. The rounding and $T \ge L/4$ give

$$\frac{rq}{N} \ge \frac{1}{2}\min\left\{1, \frac{qT}{N\varepsilon}\right\} \ge \frac{1}{16\sqrt{C_{\mathrm{fp}}}}\min\left\{1, \frac{\sqrt{L/N}}{\varepsilon}\right\}.$$

Here $\varepsilon \le T/2 \le L/4$, so $L+\varepsilon \le 5L/4$; reducing the constant from $1/1600$ to $1/2000$ yields

$$\alpha \ge \frac{1}{2000\sqrt{C_{\mathrm{fp}}}} \min\left\{1, \frac{\sqrt{(L+\varepsilon)/N}}{\varepsilon}\right\}. \tag{19}$$

If instead $\varepsilon > T/2$, choose $\xi = \delta e^{-\varepsilon}$, $q = \lfloor \sqrt{N/(C_{\mathrm{fp}}(L+\varepsilon))} \rfloor$, and $r = 1$. Again $q \ge 2$ and $q \ge \frac{1}{2}\sqrt{N/(C_{\mathrm{fp}}(L+\varepsilon))}$. The code exists because $\log(1/\xi) = L + \varepsilon$. Pad its rows to N and undo the padding exactly as above. If $\alpha \le q/(100N)$, some user is traced with probability at least $0.91/q$. Privacy and soundness now bound this probability by $e^{\varepsilon}\xi + \delta = 2\delta$, a contradiction because $q \le N$ and $2N\delta \le 1/(8N) < 0.91$. It follows that

$$\alpha \ge \frac{1}{200\sqrt{C_{\mathrm{fp}} N (L+\varepsilon)}} \ge \frac{1}{1800\sqrt{C_{\mathrm{fp}}}} \frac{\sqrt{(L+\varepsilon)/N}}{\varepsilon}.$$

The second inequality uses $\varepsilon > T/2 \ge L/8$, hence $\varepsilon/(L+\varepsilon) > 1/9$. It implies (19) also in this case. Throughout, the supremum defining α is over fixed databases; the random code serves only to establish existence of a hard input.

Apply (19) to the decoder in (18). Since $(L+\varepsilon)/N \ge (\log n + \varepsilon)/n$, every globally private M obeys, whenever $N \ge 4C_{\mathrm{fp}}(2\log n + \varepsilon)$,

$$\sup_{A:\, d_{\min}(A) \ge m} \mathbb{E}\frac{\|M(A) - \boldsymbol{\pi}(A)\|_2}{\|\boldsymbol{\pi}(A)\|_2} \ge \frac{\beta^2}{2000\sqrt{3C_{\mathrm{fp}}}(1+\beta)} \min\left\{1, \frac{\sqrt{\log n + \varepsilon}}{\sqrt{n}\,\varepsilon}\right\}, \tag{20}$$

$$\sup_{A:\, d_{\min}(A) \ge m} \mathbb{E}\frac{\|M(A) - \boldsymbol{\pi}(A)\|_2^2}{\|\boldsymbol{\pi}(A)\|_2^2} \ge \left[\frac{\beta^2}{2000\sqrt{3C_{\mathrm{fp}}}(1+\beta)}\right]^2 \min\left\{1, \frac{\log n + \varepsilon}{n\varepsilon^2}\right\}. \tag{21}$$

For the second inequality, Cauchy–Schwarz at each fixed input gives expected squared relative error at least the square of expected relative error, and taking suprema preserves this inequality. This proves a squared lower bound from a non-squared one, without reversing that implication. Taking the infimum over M preserves both bounds.

It remains to remove the extra condition on N. In the same construction, compare the all--1 database with the database obtained by replacing its first row by all $+1$. The graphs are neighbors, and every s_j changes by $2/N$. Formula (17) gives separation $D^2 = \beta^2\pi(U)^2/(2N^3)$. Both target squared norms are at most $3/(8N)$, so, with B as in the two-point argument above, $D/(2B) \ge \beta\pi(U)/(\sqrt{3}N) \ge \beta^2/[\sqrt{3}(1+\beta)N]$. That argument proves for every $n \ge 4$

$$\sup_{A:\, d_{\min}(A) \ge m} \mathbb{E}\frac{\|M(A) - \boldsymbol{\pi}(A)\|_2}{\|\boldsymbol{\pi}(A)\|_2} \ge \frac{\beta^2}{\sqrt{3}(1+\beta)N} \frac{1-n^{-2}}{1+e^{\varepsilon}}, \tag{22}$$

$$\sup_{A:\, d_{\min}(A) \ge m} \mathbb{E}\frac{\|M(A) - \boldsymbol{\pi}(A)\|_2^2}{\|\boldsymbol{\pi}(A)\|_2^2} \ge \frac{\beta^4}{3(1+\beta)^2N^2} \frac{1-n^{-2}}{1+e^{\varepsilon}}. \tag{23}$$

For completeness an explicit threshold is $\lceil 256C_{\mathrm{fp}}(256 + \log(256C_{\mathrm{fp}}))\rceil$. At and above this threshold, $n \ge 28C_{\mathrm{fp}}(2\log n + 256)$, and hence $N \ge n/7 \ge 4C_{\mathrm{fp}}(2\log n + 256)$. To check the first inequality at the unrounded threshold, use

$$\log\big[256C_{\mathrm{fp}}(256 + \log(256C_{\mathrm{fp}}))\big] \le 256 + 2\log(256C_{\mathrm{fp}}).$$

This follows from $\log z \le z$ with $z = 256 + \log(256C_{\mathrm{fp}})$. The right side of the desired inequality is then at most $28C_{\mathrm{fp}}(768 + 4\log(256C_{\mathrm{fp}}))$, which is smaller than $256C_{\mathrm{fp}}(256 + \log(256C_{\mathrm{fp}}))$. Above that point, the derivative of $n - 28C_{\mathrm{fp}}(2\log n + 256)$ is $1 - 56C_{\mathrm{fp}}/n > 0$. Rounding upward therefore preserves the inequality.

Below this threshold, use (22)–(23), $1 - n^{-2} \ge 15/16$, $\varepsilon \le 256$, and $N < \lceil 256C_{\mathrm{fp}}(256 + \log(256C_{\mathrm{fp}}))\rceil$. The resulting positive constant for expected relative error is at least

$$\frac{15\beta^2}{16\sqrt{3}(1+\beta)(1+e^{256})\lceil 256C_{\mathrm{fp}}(256 + \log(256C_{\mathrm{fp}}))\rceil}.$$

For expected squared relative error, square the factor $\beta^2/(\sqrt{3}(1+\beta))$ and the reciprocal threshold, retaining the same multiplier $15/[16(1+e^{256})]$. Each capped rate is at most one. Taking the smaller of these finite-size constants and the respective constants in (20)–(21) proves both lower rates for all $n \geq 4$ and $0 < \varepsilon \leq 256$. These uniform constants can be very small; the explicit large-graph bounds are stronger whenever their condition holds. Combining with the upper bounds completes the proof.

A.6 Proof of Theorem 3

We first verify the ordered-output representation. The distribution function of $G_j \sim \mathrm{Gumbel}(0, b)$ is $\Pr(G_j \leq t) = \exp[-\exp(-t/b)]$. Consequently, for $t \geq 0$, the independent transformed variables satisfy

$$\Pr\left(\exp[-(\hat{\pi}_j + G_j)/b] > t\right) = \exp[-e^{\hat{\pi}_j/b} t].$$

They are therefore independent exponential waiting times with rates $e^{\hat{\pi}_j/b}$. Decreasing order of $\hat{\pi}_j + G_j$ is increasing order of these waiting times. For any remaining collection, the probability that node j has its smallest waiting time is the integral of its exponential density times the survival probabilities of the others, giving its rate divided by the sum of the remaining rates. After the smallest waiting time is removed, the exponential memoryless property leaves independent waiting-time residuals with their original rates. Induction gives the same conditional selection rule after each previous choice, establishing equality in distribution of the entire ordered list with peeling exponential mechanisms, as in Durfee & Rogers (2019, Lemma 4.2).

Fix a step $\ell \in \{1, \ldots, k\}$ and the same preceding ordered choices $i_1, \ldots, i_{\ell-1}$ on two adjacent inputs A, A'. Every distinct history has positive probability on both inputs. In the following displays all probabilities are conditional on that fixed history, and sums range only over nodes outside $\{i_1, \ldots, i_{\ell-1}\}$. Then

$$\Pr_A(\hat{\sigma}_\ell = j) = \int_0^\infty e^{\hat{\pi}_j/b} \exp\left[-t \sum_{a \notin \{i_1,\ldots,i_{\ell-1}\}} e^{\hat{\pi}_a/b}\right] dt = \frac{e^{\hat{\pi}_j/b}}{\sum_{a \notin \{i_1,\ldots,i_{\ell-1}\}} e^{\hat{\pi}_a/b}},$$

and the neighboring-input log probability ratio is

$$\log \frac{\Pr_A(\hat{\sigma}_\ell = j)}{\Pr_{A'}(\hat{\sigma}_\ell = j)} = \frac{\hat{\pi}_j - \hat{\pi}'_j}{b} + \log \frac{\sum_{a \notin \{i_1,\ldots,i_{\ell-1}\}} e^{\hat{\pi}'_a/b}}{\sum_{a \notin \{i_1,\ldots,i_{\ell-1}\}} e^{\hat{\pi}_a/b}}. \tag{24}$$

Lemma 1 gives $|\hat{\pi}_j - \hat{\pi}'_j| \leq \Delta_{\mathrm{reg},\infty}$ for every remaining node. Termwise comparison of the sums bounds their log ratio in $[-\Delta_{\mathrm{reg},\infty}/b, \Delta_{\mathrm{reg},\infty}/b]$. Thus each conditional selection is ε_0-DP, where $\varepsilon_0 = 2\Delta_{\mathrm{reg},\infty}/b$, by exactly the normalization argument of McSherry & Talwar (2007, Theorem 6). Moreover, subtracting (24) at any two remaining outputs cancels the normalization term. The range of its values is at most $2\Delta_{\mathrm{reg},\infty}/b = \varepsilon_0$, the bounded-range property of Durfee & Rogers (2019, Lemma 4.3). No monotonicity of the PageRank coordinates under row replacement is assumed.

Here is the precise concentration consequence of that range bound. Hoeffding's bound for a centered variable of range length ε_0 gives log moment-generating function at most $\lambda^2\varepsilon_0^2/8$ for all real λ, conditionally on the history; see Cesar & Rogers (2021, Lemma 17). The exponential of the negative log probability ratio has conditional expectation one, because summing $\Pr_A(\hat{\sigma}_\ell = j) \Pr_{A'}(\hat{\sigma}_\ell = j)/\Pr_A(\hat{\sigma}_\ell = j)$ gives one. Applying the centered bound at $\lambda = -1$ therefore bounds the mean log probability ratio by $\varepsilon_0^2/8$. Applying it again at any $\lambda > 0$ yields

$$\log \sum_j \Pr_A(\hat{\sigma}_\ell = j) \left[\frac{\Pr_A(\hat{\sigma}_\ell = j)}{\Pr_{A'}(\hat{\sigma}_\ell = j)}\right]^\lambda \leq \frac{(\lambda + \lambda^2)\varepsilon_0^2}{8}. \tag{25}$$

This is $\varepsilon_0^2/8$-zCDP: dividing by λ bounds the Rényi divergence of order $1 + \lambda$ by $(1 + \lambda)\varepsilon_0^2/8$. Reversing A, A' gives the same bound in the other direction.

For the whole list, the probability ratio is the product of the k conditional ratios. The pure-DP bounds multiply to at most $e^{k\varepsilon_0}$. For concentration, repeatedly condition on the first $k-1, k-2, \ldots, 0$ choices in the moment of that product. Each application of (25) contributes at most $\exp[(\lambda+\lambda^2)\varepsilon_0^2/8]$, uniformly over its history. The resulting log moment bound is $k(\lambda + \lambda^2)\varepsilon_0^2/8$, proving $k\varepsilon_0^2/8$-zCDP

for the ordered transcript. This is the adaptive composition rule in Bun & Steinke (2016, Lemma 2.3), with its fixed-history hypotheses verified here.

If the first scale in (9) is selected, then $k\varepsilon_0 = \varepsilon$ and the transcript is pure ε-DP. If the second is selected, then

$$\frac{k\varepsilon_0^2}{8} = \left(\sqrt{\log(1/\delta) + \varepsilon} - \sqrt{\log(1/\delta)}\right)^2.$$

By Bun & Steinke (2016, Proposition 1.3), $k\varepsilon_0^2/8$-zCDP implies approximate DP with budget $k\varepsilon_0^2/8 + 2\sqrt{(k\varepsilon_0^2/8)\log(1/\delta)} = \varepsilon$ and failure parameter δ. The identity follows by expanding the square, and holds for every $\varepsilon > 0$, $0 < \delta < 1$. The scale minimum uses only public inputs, so the selected mechanism has one of these two certified guarantees; it does not compose their costs. This proves global privacy for the published ordered indices. It provides no additional release of their numerical values or of the remaining ordering.

A.7 Proof of Theorem 4

For every $t \geq 0$, the Gumbel distribution gives

$$\Pr(G_j > t) = 1 - \exp[-e^{-t/b}] \leq e^{-t/b}, \qquad \Pr(G_j < -t) = \exp[-e^{t/b}] \leq e^{-t/b}.$$

The first inequality uses $1 - e^{-u} \leq u$ for $u \geq 0$; the second uses $e^{t/b} \geq t/b$. With $t = b\log(2n/\eta) > 0$, a union bound over both tails and all n coordinates gives $\Pr(\max_j |G_j| > t) \leq 2ne^{-t/b} = \eta$. Location zero in $\mathrm{Gumbel}(0, b)$ is used through this distribution function, not as a zero-mean assertion.

On the event $\max_j |G_j| \leq t$, every $i \in \mathcal{T}_k$ and $j \notin \mathcal{T}_k$ satisfy $\pi_i - \pi_j \geq \mathrm{Gap}_k(\boldsymbol{\pi})$ and

$$(\hat{\pi}_i + G_i) - (\hat{\pi}_j + G_j) \geq \mathrm{Gap}_k(\boldsymbol{\pi}) - 2\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_\infty - 2t > 0.$$

Thus all k members of $\mathcal{T}_k$ precede every nonmember in the perturbed ordering, proving $\tilde{\mathcal{T}}_k = \mathcal{T}_k$ on that event. Its probability is at least $1 - \eta$. No comparison between two members of $\mathcal{T}_k$ is used, so no guarantee of their internal order follows.

The actual error $\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_\infty$ is bounded by the expanded expression in the proof of Theorem 1. That argument first bounds the residual from $P - Q$ and clipped mass and then substitutes the proved ℓ_1 bound into the final propagation term. The resulting expression is an upper bound on the coordinatewise error, not a new definition of it or of bias^2.

A.8 Proof of Corollary 2

Use the same event $\max_j |G_j| \leq t$ with $t = b\log(2n/\eta)$, whose probability is at least $1 - \eta$ by the two tail bounds just proved. For $i \in \mathcal{T}_k$ and $j \notin \mathcal{T}(\boldsymbol{\pi}; L)$, the ordered coordinates give $\pi_i \geq \pi_{(k)}$ and $\pi_j \leq \pi_{(L+1)}$. Therefore

$$(\hat{\pi}_i + G_i) - (\hat{\pi}_j + G_j) \geq \pi_{(k)} - \pi_{(L+1)} - 2\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_\infty - 2t > 0.$$

The public tie rule gives $\mathcal{T}_k \subseteq \mathcal{T}(\boldsymbol{\pi}; L)$. Hence at least k nodes inside the true top-L precede every node outside it, so no outside node can be among the first k perturbed indices. This proves $\tilde{\mathcal{T}}_k \subseteq \mathcal{T}(\boldsymbol{\pi}; L)$, and its cardinality k implies the stated intersection fraction equals one. The argument uses the original k-calibrated release; it neither releases L nodes nor incurs a second selection budget.

A.9 Proof of Theorem 5

We first establish the finite packing inequality without a budget restriction. Use the graph construction in Appendix A.5 with $n = 4N$ and no fillers. There are N variable nodes R, N anchors H and the $2N$ targets v_j^+, v_j^-, with $U = R \cup H$. Each variable chooses one target in each pair according to its sign row, anchors point to all targets, and targets point to all of U. Every variable out-degree is N, every anchor and target out-degree is $2N$, every node in U has in-degree $2N$, and every target has between N and $2N$ in-neighbors. Thus the graphs are loop-free and lie in the utility domain since $m \leq N$.

For completeness, both sides of this bipartition have total PageRank mass $1/2$: their masses sum to one, and the mass of U solves $\sum_{u\in U} \pi_u = (1-\beta)/2 + \beta(1 - \sum_{u\in U} \pi_u)$. All nodes in U have

identical incoming flow, so $\pi_u = 1/n$. For column mean $s_j = N^{-1}\sum_i X_{ij}$, the variable nodes contribute $(1+s_j)/(2n)$ to v_j^+ and the anchors contribute $1/(2n)$; the minus target replaces s_j by $-s_j$. Adding restart gives

$$\pi_u = \frac{1}{n} \quad (u \in U), \qquad \pi_{v_j^\pm} = \frac{1}{n} \pm \frac{\beta}{2n} s_j. \tag{26}$$

Because N is even, start with a database whose first $N/2$ rows are all $+1$ and whose last $N/2$ rows are all -1; call its graph A_0. For an integer $1 \le r \le N/2$, fix r negative rows. For each k-subset $J \subseteq [N]$, replace their entries by $+1$ precisely in columns J, producing A_J. It differs from A_0 in exactly r outgoing rows, and $s_j = (2r/N)\mathbf{1}_{\{j\in J\}}$. Equation (26) then implies

$$\mathcal{T}(\boldsymbol{\pi}(A_J); k) = \{v_j^+ : j \in J\}, \qquad \mathrm{Gap}_k(\boldsymbol{\pi}(A_J)) = \frac{\beta r}{nN}. \tag{27}$$

Exactly those k coordinates lie above $1/n$; nodes in U and unselected pairs have value $1/n$, and the opposite members of selected pairs lie below it. The top-k set is unique even though its members tie with each other.

If a mechanism publishes an ordered list, discard its order. This is postprocessing, preserves privacy, and leaves its exact-set success probability unchanged. We may therefore write M for a set-valued mechanism. Applying the privacy inequality successively along the r row replacements gives

$$\Pr\big(M(A_J) = \{v_j^+ : j \in J\}\big) \le e^{r\varepsilon} \Pr\big(M(A_0) = \{v_j^+ : j \in J\}\big) + \delta \sum_{h=0}^{r-1} e^{h\varepsilon}.$$

Indeed, one replacement adds δ; if the error term after s replacements is $\delta\sum_{h=0}^{s-1} e^{h\varepsilon}$, one further privacy inequality multiplies it by e^ε and adds δ, giving the sum through $h = s$. This proves the displayed formula by induction. Each intermediate graph is obtained by replacing another one of the fixed variable rows, so the required adjacency path consists of valid inputs. For $\varepsilon > 0$, the accumulated term also equals $\delta(e^{r\varepsilon} - 1)/(e^\varepsilon - 1)$.

The $\binom{N}{k}$ events $M(A_0) = \{v_j^+ : j \in J\}$ are disjoint. Summing the last inequality over them and dividing by $\binom{N}{k}$ bounds average success by $e^{r\varepsilon}/\binom{N}{k} + \delta\sum_{h=0}^{r-1} e^{h\varepsilon}$. At least one graph thus has failure at least

$$\left[1 - \frac{e^{r\varepsilon}}{\binom{N}{k}} - \delta \sum_{h=0}^{r-1} e^{h\varepsilon}\right]_+. \tag{28}$$

This inequality is valid for every $\varepsilon > 0$, every even $N \ge 2$, $1 \le k \le N-1$, and $1 \le r \le N/2$. It uses privacy on the full input domain and imposes no additive-noise or sensitivity assumption on M.

Now suppose $N \ge \max\{64, 2k\}$ and $e^\varepsilon \le \binom{N}{k}/4$. Set

$$r = \begin{cases} \lfloor \min\{N/2, \log(N/8)/\varepsilon\} \rfloor, & \varepsilon \le \log(N/8), \\ 1, & \varepsilon > \log(N/8). \end{cases} \tag{29}$$

In the first case, both entries of the minimum are at least one, so $1 \le r \le N/2$ and the floor is at least half the unrounded minimum. Also $e^{r\varepsilon} \le N/8$. The combinatorial bound $\binom{N}{k} \ge N$ follows for $1 \le k \le N-1$ by symmetry and the ratio $\binom{N}{j+1}/\binom{N}{j} = (N-j)/(j+1) \ge 1$ up to the middle index. With $\delta = (4N)^{-2}$,

$$\delta \sum_{h=0}^{r-1} e^{h\varepsilon} \le \delta r e^{r\varepsilon} \le \frac{1}{16N^2} \frac{N}{2} \frac{N}{8} = \frac{1}{256}.$$

Equation (28) gives failure at least $1 - 1/8 - 1/256 = 223/256 > 2/3$.

For $N \ge 64$, $\log(N/8) \ge (3/8)\log(4N) = (3/8)\log n$: after collecting logarithms, this is exactly $\log N \ge 6\log 2$. Consequently the integer chosen in this first case obeys

$$\frac{3}{16} \min\left\{N, \frac{\log n}{\varepsilon}\right\} \le \frac{1}{2} \min\left\{\frac{N}{2}, \frac{3\log n}{8\varepsilon}\right\} \le r \le \min\left\{N, \frac{\log n}{\varepsilon}\right\}.$$

Multiplication by $\beta/(nN)$ and $n = 4N$ shows

$$\frac{3\beta}{16}\min\left\{\frac{1}{n}, \frac{\log n}{n^2\varepsilon}\right\} \le \mathrm{Gap}_k(\boldsymbol{\pi}(A_J)) \le 4k\beta\min\left\{\frac{1}{n}, \frac{\log n}{n^2\varepsilon}\right\}. \tag{30}$$

For the upper inequality here the stronger constant 4β suffices, since $\min\{a, 4b\} \le 4\min\{a, b\}$; replacing it by $4k\beta$ is valid for $k \ge 1$.

In the second case $r = 1$, so the group-privacy term is exactly δ. The cardinality condition makes (28) at least $3/4 - \delta > 2/3$. Since $\varepsilon > \log(N/8) \ge (3/8)\log n$ and $n \ge 256$, the minimum in (30) is its second term. The ratio of the actual gap to $\beta\log n/(n^2\varepsilon)$ is $4\varepsilon/\log n > 3/2$. On the other hand, $e^\varepsilon \le \binom{N}{k}/4$ and $\binom{N}{k} \le N^k$ imply $\varepsilon < k\log N < k\log n$, so this ratio is less than $4k$. Thus (30) holds in this case as well. In both cases the graph has the required gap and failure, with a positive integer r and the full group-δ term accounted for.

We finally verify the theorem's stated budget range. For $N \ge k$, $\binom{N}{k} = \prod_{j=0}^{k-1}(N-j)/(k-j) \ge (N/k)^k$, since $(N-j)/(k-j) \ge N/k$ term by term. Thus the finite sufficient condition

$$\varepsilon \le k[\log n - \log(4k)] - \log 4 \tag{31}$$

ensures $e^\varepsilon \le \binom{N}{k}/4$. The size condition $n/k \ge 16 \cdot 4^{2/k}$ gives $\log(n/k) \ge 2\log 4 + (2/k)\log 4$. Multiplication by $k/2$ and rearrangement give

$$\frac{k}{2}\log(n/k) \le k\log(n/k) - k\log 4 - \log 4.$$

Hence the theorem's upper budget $\frac{k}{2}(\log n - \log k)$ implies (31). This establishes every stated parameter condition and proves the theorem. The exact bound (28) remains valid outside this range, but can be zero.

There is also a finite-output reason to retain a size/budget restriction. Compute the true top-k set with public tie breaking. Among all $\binom{n}{k}$ sets, publish the true set with probability $e^\varepsilon/(e^\varepsilon + \binom{n}{k} - 1)$ and each other set with probability $1/(e^\varepsilon + \binom{n}{k} - 1)$, returning a chosen set in public canonical order. For any two graphs and any output, its probability ratio is at most e^ε, so summing over outputs proves pure row-DP. Failure on every unique-top-k input is $(\binom{n}{k} - 1)/(e^\varepsilon + \binom{n}{k} - 1)$, which is at most η when $\varepsilon \ge \log[((1-\eta)/\eta)(\binom{n}{k} - 1)]$. This is an information-theoretic counterexample to a positive-gap, constant-failure lower bound at arbitrary finite n and large ε, regardless of computational efficiency. It does not modify Algorithm 2.

A.10 Proof of Corollary 3

For $d_{\min} \ge m$, degree flooring is inactive and $Q = P$. Uniform restart and $P_{ij} \le 1/m$ give $\pi_j = (1-\beta)/n + \beta\sum_i P_{ij}\pi_i \le (1-\beta)/n + \beta/m = \kappa$. Therefore $\boldsymbol{\pi}$ solves the regularized fixed-point equation, and uniqueness gives $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}$. The coordinatewise sensitivity remains a global bound, even though this zero-bias identity is asserted only on the utility domain.

The public floor satisfies $m \ge \rho n/2$: if $\rho n < 2$, $m \ge 1 \ge \rho n/2$, and otherwise $\lfloor \rho n \rfloor \ge \rho n/2$. Substitution in Lemma 1 yields the explicit bound

$$\Delta_{\mathrm{reg},\infty} = \frac{\beta\kappa}{m}\frac{1+\beta}{1-\beta} \le \frac{2\beta}{\rho}\left(1 - \beta + \frac{2\beta}{\rho}\right)\frac{1+\beta}{1-\beta}\frac{1}{n^2}. \tag{32}$$

Because (9) gives $b \le 2k\Delta_{\mathrm{reg},\infty}/\varepsilon$, condition (13) implies $\mathrm{Gap}_k(\boldsymbol{\pi}) > 2b\log(2n/\eta)$. Theorem 4 with zero coordinatewise bias then proves the success probability for every $\varepsilon > 0$. For fixed k, η, β, ρ, its sufficient gap is at most a constant multiple of $\log n/(n^2\varepsilon)$, since $\log(2n/\eta) = \log n + \log(2/\eta)$.

Under the lower theorem's conditions, if $\log n/(n\varepsilon) \to 0$, its minimum eventually equals $\log n/(n^2\varepsilon)$. Equations (30) and (28) then exhibit, for every globally private mechanism, a graph with gap of this order and failure above $2/3$. The upper theorem gives success at least $1 - \eta > 1/2$ uniformly above a sufficiently large constant multiple of the same order. These are the asserted necessary and sufficient orders for exact set identification, with k fixed.

To verify that the sufficient-gap domain is nonempty in this comparison, the explicit finite condition

$$\varepsilon \ge \frac{32k}{\rho}\left(1 - \beta + \frac{2\beta}{\rho}\right)\frac{1+\beta}{1-\beta}\frac{\log(2n/\eta)}{n} \tag{33}$$

combines with (32) to make the right side of (13) at most $\beta/(4n)$. The packing construction with $r = N/2$ has gap $\beta/(2n)$, strictly exceeds this sufficient threshold, and has $d_{\min} \geq m$. Condition (33) eventually holds along any fixed-parameter sequence with $\log n/(n\varepsilon) \to 0$, because $\log(2n/\eta)/\log n \to 1$. Thus the matching assertion does not rely on an empty utility domain. At smaller budgets no such conclusion is inferred: the saturated lower bound still applies, while the sufficient threshold can exceed the available gaps. No claim about optimal dependence on growing k, within-set order, Average Rank, or top-L overlap follows from these exact-set arguments.

A.11 Proof of Proposition 1

The notation in this proof is the historical spectral-query notation of Appendix C. For adjacent inputs differing in row i, set $D = Q - Q' = e_i d^\top$ and $E = T - T'$. Each row of Q has squared norm at most $1/m$. Since the rows are nonnegative, $\|d\|_2 \leq \sqrt{2/m}$. The projection Π onto a closed convex set is firmly nonexpansive in Frobenius norm: its variational inequalities imply $\|E\|_F^2 \leq \langle E, D\rangle_F$. Consequently,

$$\|E - D/2\|_F \leq \|D\|_F/2, \qquad \|E^\top v\|_2 \leq \frac{\|d\|_2}{2}\big(|v_i| + \|v\|_2\big)$$

for every vector v.

The fixed-point equation and $\|T\|_2 \leq 1$ give $\|x - (1-\beta)\mu\|_2 \leq \beta\|x\|_2$. Completing the square yields the public ball

$$\left\| x - \frac{\mu}{1+\beta} \right\|_2 \leq \frac{\beta L}{1+\beta}.$$

It also contains x', has diameter $2\beta L/(1+\beta)$, and implies $\|x\|_2 \leq L$. For a vector v in this ball, write $\|v\|_2 + |v_i| = \max_{\|z\|_2 \leq 1,\, a \in \{-1,1\}} (z + ae_i)^\top v$. Maximizing over the ball gives

$$\frac{(z + ae_i)^\top \mu + \beta L \|z + ae_i\|_2}{1+\beta}.$$

Reflecting the ith coordinate shows that the maximum with $a = -1$ cannot exceed the one with $a = 1$, because $\mu_i \geq 0$. A maximum can be taken on the unit sphere. With $t = z_i$, the other coordinates align with μ_{-i}, giving the expression defining $LH_\beta(s_i)$. Replacing negative t by $|t|$ cannot decrease it. Thus $\|v\|_2 + |v_i| \leq LH_\beta(s_i)$; for $n = 1$ the same expression is still an upper bound.

Finally, the resolvent identity gives

$$x - x' = \beta(I - \beta T^\top)^{-1} E^\top x', \qquad \|(I - \beta T^\top)^{-1}\|_2 \leq (1-\beta)^{-1}.$$

Combining the preceding inequalities proves the middle term of (35). The first term follows from $\|E\|_2 \leq \|D\|_F$ and $\|x'\|_2 \leq L$; the last is the ball diameter. Taking their minimum completes the proof. When $\beta = 0$, the query is constant.

B Numerical solution and accuracy records

The mathematical mechanisms compute the unique regularized fixed point $\hat{\boldsymbol{\pi}}$ and then make one Gaussian or Gumbel release. The solver variables $\boldsymbol{x}^{(t)}$ approximate $\hat{\boldsymbol{\pi}}$; iteration is a numerical way to solve that equation; intermediate values are not published. The existing implementation starts at $\boldsymbol{x}^{(0)} = \mu$ and applies

$$\boldsymbol{x}^{(t+1)} = (1-\beta)\mu + \beta Q^\top \operatorname{clip}_\kappa(\boldsymbol{x}^{(t)}). \tag{34}$$

There is no normalization inside the solver. A public fixed count T gives the deterministic approximation bound

$$\|\boldsymbol{x}^{(T)} - \hat{\boldsymbol{\pi}}\|_1 \leq \beta^T \|\mu - \hat{\boldsymbol{\pi}}\|_1 \leq 2\beta^T.$$

The configured $\beta = 0.8$, $T = 127$ gives a bound below 10^{-12}. The actual count, tolerance, residuals and public parameters must remain in the experimental manifest even though the main mechanisms are defined at $\hat{\boldsymbol{\pi}}$.

The same global sensitivities in Lemma 1 hold for this public fixed-count solver in exact arithmetic. The adjacent-iterate difference obeys

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}'^{(t+1)}\|_1 \le \beta\|\boldsymbol{x}^{(t)} - \boldsymbol{x}'^{(t)}\|_1 + 2\beta\kappa, \qquad \boldsymbol{x}^{(0)} - \boldsymbol{x}'^{(0)} = 0.$$

Induction gives $\Delta_{\mathrm{reg},1}$ at each step, and the row-norm inequalities in Appendix A.1 give $\Delta_{\mathrm{reg},2}$ and $\Delta_{\mathrm{reg},\infty}$. Thus the recorded solver output can be calibrated without silently assuming exact convergence. A data-dependent stopping rule is not justified by this argument.

For implementation checks, the raw Gaussian identity uses the actual deterministic solver output, and its distance from the fixed point is recorded separately. The difference between solver bias and regularization bias is at most $2\beta^T$ in each of the ℓ_1, ℓ_2 and ℓ_∞ norms. A numerical top-k recovery check correspondingly accounts for at most $4\beta^T$ in the pairwise gap. These are solver-accuracy records, not additional terms in the exact-query theorems. Floating-point residual checks do not themselves constitute a finite-precision privacy proof.

C HISTORICAL SPECTRAL-PROJECTION BASELINE

This appendix records the former spectral-projection query and its valid public calibration for baseline comparisons. It is not DP PageRank–Full Vector and is not used in its sensitivity or utility proof. The original spectral manuscript and experiment outputs are archived separately. No old conditional perturbation or sharpness claim is transferred to the mass-clipped query.

Let $Q_{ij} = A_{ij}/\max\{d_i, m\}$, let $T = \Pi(Q)$ be the Frobenius projection onto the spectral-norm unit ball, and write the raw query as

$$x = (1-\beta)(I - \beta T^\top)^{-1}\mu.$$

For $L = \|\mu\|_2$ and $s_i = \mu_i/L$, define the public function

$$H_\beta(s) = \frac{1}{1+\beta}\max_{0\le t\le 1}\left\{s(1+t) + \sqrt{1-s^2}\sqrt{1-t^2} + \beta\sqrt{2+2t}\right\}.$$

Proposition 1. For fixed public $m \ge 1$, $0 \le \beta < 1$, and probability vector μ, the raw regularized PageRank query has outgoing-row sensitivity at most

$$\Delta_{\mathrm{ref}} = \min\left\{\frac{\sqrt{2}\beta L}{(1-\beta)\sqrt{m}}, \quad \frac{\beta L \max_i H_\beta(s_i)}{(1-\beta)\sqrt{2m}}, \quad \frac{2\beta L}{1+\beta}\right\}. \tag{35}$$

Its proof is in Appendix A.11.

For this baseline, Gaussian noise is calibrated to the raw spectral query with Δ_{ref}; any nonnegative clipping and normalization is separately labeled postprocessing. The exact-query analysis is distinct from numerical residual checks and from the finite-iteration mass-clipping proof. Historical experiment tables and figures are not reproduced as evidence for the two updated algorithms.

D ADDITIONAL REAL-DATA RESULTS

Table 3: Full results for Table 1 at $\varepsilon = 32, 64, 128$, including average top-$k$ rank for $k = 1, 5, 10$. Entries use the same 50 saved releases, raw outputs, fixed settings and full $\|\boldsymbol{\pi}\|_2$ denominator as in Section 5. Uniform errors and expected ranks are independent of ε and are shown once per dataset. Lower values are better; standard errors are retained in the result CSV.

		Relative ℓ_2 error					Average top-k rank		
ε	Method	Overall	Top 5%	5–30%	30–50%	Bottom 50%	$k = 1$	$k = 5$	$k = 10$
Email-Eu-core									
32	PR Full Vector	0.954	0.371	0.513	0.391	0.596	473.4	410.9	389.1
64		0.736	0.345	0.414	0.286	0.410	362.8	308.3	311.6
128		0.610	0.328	0.361	0.222	0.289	272.4	235.8	213.0
32	Edge Flipping	0.644	0.474	0.294	0.043	0.318	503.5	501.5	501.9
64		0.643	0.474	0.293	0.043	0.318	503.5	493.4	494.9
128		0.642	0.473	0.293	0.043	0.317	503.5	495.4	485.9
—	Uniform	0.644	0.475	0.294	0.042	0.318	503.0	503.0	503.0
32	Naive Gaussian	125.158	28.304	62.033	56.325	88.439	503.5	503.5	505.0
64		85.056	19.405	42.530	37.736	60.138	503.5	505.5	502.0
128		58.747	12.852	29.147	26.289	41.720	503.5	501.5	500.0
CollegeMsg									
32	PR Full Vector	1.034	0.589	0.558	0.359	0.531	874.5	853.7	839.0
64		0.884	0.581	0.489	0.267	0.365	760.6	728.2	703.9
128		0.805	0.575	0.454	0.211	0.259	684.6	576.4	510.2
32	Edge Flipping	0.734	0.605	0.288	0.080	0.290	950.5	952.5	953.1
64		0.735	0.605	0.288	0.080	0.290	931.5	941.1	947.4
128		0.734	0.605	0.288	0.080	0.290	950.5	952.5	951.2
—	Uniform	0.735	0.605	0.288	0.080	0.290	950.0	950.0	950.0
32	Naive Gaussian	210.342	45.975	105.282	94.544	148.571	931.5	948.7	951.2
64		143.503	32.820	72.050	63.748	101.233	950.5	944.9	945.5
128		98.330	21.698	49.324	44.071	69.394	950.5	941.1	947.4
Wiki-Vote									
32	PR Full Vector	1.009	0.738	0.480	0.265	0.417	2704.7	2905.8	2908.2
64		0.915	0.731	0.430	0.184	0.290	1566.3	2151.4	2281.7
128		0.869	0.728	0.406	0.131	0.206	1210.5	1653.3	1847.5
32	Edge Flipping	0.844	0.749	0.280	0.142	0.230	3558.5	3560.5	3563.0
64		0.844	0.749	0.280	0.142	0.230	3558.5	3560.5	3555.9
128		0.844	0.749	0.280	0.142	0.230	3558.5	3560.5	3555.9
—	Uniform	0.844	0.749	0.280	0.142	0.230	3558.0	3558.0	3558.0
32	Naive Gaussian	622.662	140.401	310.816	277.852	440.568	3558.5	3546.3	3555.9
64		423.633	95.266	210.759	189.488	300.049	3558.5	3560.5	3563.0
128		290.745	65.149	144.990	130.180	205.679	3558.5	3546.3	3555.9

E SIMULATION

We generate directed Chung–Lu graphs with $n = 2000$ and deterministic Lomax quantile weights: shape 1.2, with incoming and outgoing (scale, floor) pairs $(.85, 30)$ and $(.55, 20)$. Edges are independent, with probabilities proportional to outgoing–incoming weight products, capped at .95, and no self-loops. Density D is the tenth percentile of expected out-degree. We fix $D = 500$, $\beta = .8$, uniform public restart, $\delta = .1$, $\kappa = 4/n$, $\varepsilon = 64$, and $k = 10$. Each configuration uses 50 independent graphs and one release per graph; methods and m values share graphs. Error bars are standard errors (sample standard deviation divided by $\sqrt{50}$).

Let $\mathcal{T}_k = \mathcal{T}(\boldsymbol{\pi}; k)$ be the true top-k set. Vector error is $\|(v - \boldsymbol{\pi})_{\mathcal{T}_k}\|_2 / \|\boldsymbol{\pi}_{\mathcal{T}_k}\|_2$, computed per graph from the raw release before averaging. Unlike the real-data metric, its denominator is the top-k subvector norm. For Average Rank, the released top-k list receives ranks $1, \ldots, k$; omitted nodes receive $(n + k + 1)/2$. Writing these ranks as $\hat{r}(i)$,

$$\text{AverageRank}\,@k = \frac{1}{k} \sum_{i \in \mathcal{T}_k} \hat{r}(i). \tag{36}$$

Vector ties use a fixed public node order. Strict overlap is $|\widetilde{\mathcal{T}}_k \cap \mathcal{T}_k|/k$ for released set $\widetilde{\mathcal{T}}_k$. Uniform random ranking has expected Average Rank $(n + 1)/2$ and overlap k/n; the non-private ideal gives $(k + 1)/2$ and 1. PR Top k releases indices only and has no vector error.

Figure 1 compares both methods with Edge Flipping, Uniform, the non-private ideal, and Naive Gaussian. We omit GraphDiffusion (Wei et al., 2024) because its published privacy guarantees for undirected graphs under edge-level adjacency do not directly cover the directed, row-level setting considered here. Additional density and privacy-budget experiments appear in Appendix F.

Varying m reveals the perturbation–regularization tradeoff: PR Full Vector's error falls from .069 at $m = 200$ to .046 at $m = 450$, then rises to .758 at $m = 1000$. Its best observed Average Rank occurs at $m = 500$. PR Top k preserves overlap near one throughout the scan, reaching .992 at $m = 1000$ versus .714 for PR Full Vector.

F ADDITIONAL SIMULATION RESULTS

We supplement Section E by varying the privacy budget and release size, using the same graph model, metrics, and uncertainty estimates. Figure 2 fixes $D = m = 500$ and compares PR Full Vector and PR Top k for $\varepsilon \in \{8, 16, 32, 64\}$ and $k \in \{10, 20, 50\}$. Each PR Full Vector release is evaluated at all three values of k; PR Top k uses a separate full-budget release for each k. These comparisons reuse the 50 graphs from the fixed-D regularization scan.

Increasing the privacy budget generally improves utility, while recovering a larger top-k set is more difficult. PR Top k attains better ranking utility than PR Full Vector throughout these comparisons. For $k = 10$ and $\varepsilon = 8$, their mean overlaps are .982 and .756, respectively. At $\varepsilon = 64$, these increase to 1.000 and .972, with Average Rank 5.50 and 33.38. The overlap of 1.000 records recovery of the full top-10 set in all 50 observed releases.

The density scan in Figure 3 varies $D = 50, 100, \ldots, 800$ with $m = D$, fixing $\varepsilon = 64$ and $k = 10$. Each density uses 50 independent graphs, shared across methods; these samples are independent of the fixed-D regularization scan. This experiment illustrates the distinction between vector error and head-ranking utility. PR Full Vector's relative error decreases from .201 at $D = 50$ to .052 at $D = 800$, yet ranking utility is best at intermediate densities. At $D = 800$, PR Full Vector and PR Top k retain mean overlaps of .554 and .738, respectively. Edge Flipping and Naive Gaussian remain close to random ranking. In the regularization scan in Figure 1, PR Full Vector's minimum relative error occurs at $m = 450$, whereas its best observed Average Rank occurs at $m = 500$. Thus the preferred regularization level can depend on the utility measure. PR Top k preserves head-ranking utility over a broader range of m in this experiment.

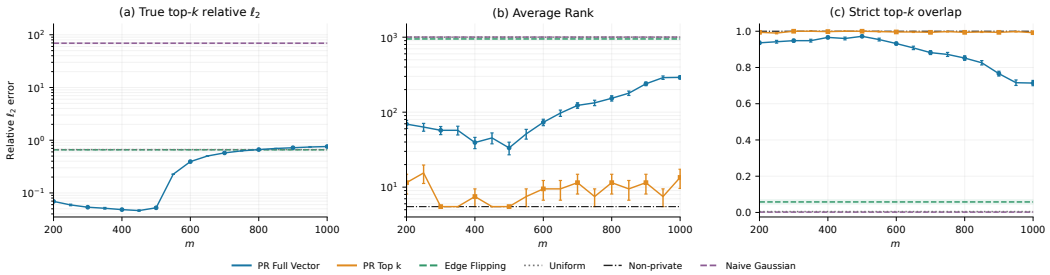

Figure 1: Regularization $m = 200, 250, \ldots, 1000$ at $D = 500$, $\varepsilon = 64$, and $k = 10$. Panels show top-10 relative ℓ_2 error, Average Rank (both logarithmic), and strict overlap (linear). Points and error bars are means $\pm$ standard errors over 50 graphs.

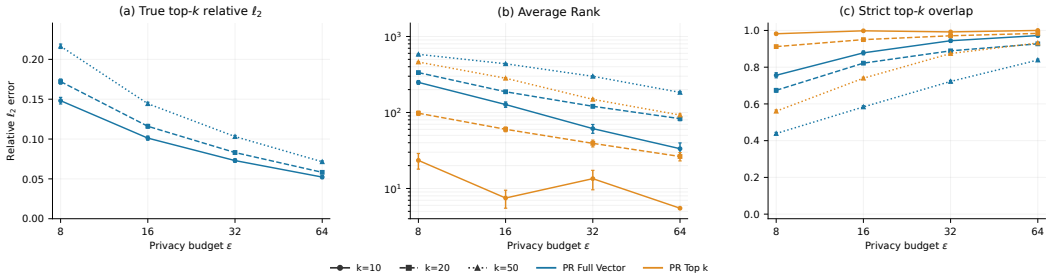

Figure 2: Privacy budget and release size with $D = m = 500$. Panels show true top-k relative ℓ_2 error, Average Rank, and strict top-k overlap for $k = 10, 20, 50$. Points and error bars represent means and standard errors over 50 independent graphs. Average Rank uses a logarithmic axis; the other axes are linear. PR Top k releases indices only and has no vector error.

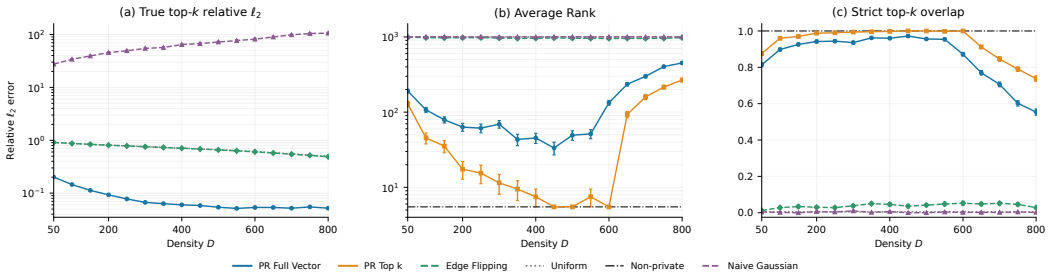

Figure 3: Density $D = 50, 100, \ldots, 800$ with $m = D$, $\varepsilon = 64$, and $k = 10$. Panels show top-10 relative ℓ_2 error, Average Rank (both logarithmic), and strict overlap (linear). Points and error bars are means $\pm$ standard errors over 50 graphs per density.