DIFFERENTIALLY PRIVATE HIERARCHICAL CLUSTERING FOR WELL-CLUSTERABLE GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

We present NOISY-DEGREE HIERARCHY, a pure edge-differentially private algorithm for hierarchical clustering on well-clusterable graphs. It constructs a full hierarchy from noisy vertex degrees, without recovering the clusters, in $O(m + n \log n)$ time. Under the spectral well-clusterability conditions of Manghiuc & Sun (2021), it achieves an $O(k^6/(\lambda_{k+1}^2 \epsilon))$ approximation to Dasgupta's cost with high probability, provided that the minimum weighted degree is at least one and $0 < \epsilon \leq 1$. Here k is the number of clusters, and λ_{k+1} is the $(k+1)$st smallest eigenvalue of the input graph's normalized Laplacian. The private algorithm builds on our simpler non-private DEGREE-ORDERED HIERARCHY, whose $O(k^6/\lambda_{k+1}^2)$ approximation improves the previous $O(k^{22}/\lambda_{k+1}^{10})$ bound under the same conditions. With additional degree-regularity and conductance assumptions, the private approximation factor improves to $O(k/\epsilon)$. We also prove an $\Omega(k)$ lower bound on disjoint regular expanders that matches this dependence on k for fixed, sufficiently small $\epsilon > 0$. Experiments on synthetic and real-world graphs demonstrate the private algorithm's utility and efficiency.

1 INTRODUCTION

Hierarchical clustering (HC) represents data by a tree of nested clusters, supporting applications such as network analysis and gene expression analysis (Ward, 1963; Clauset et al., 2008; Eisen et al., 1998). For a similarity graph $G = (V, E, w_G)$ with nonnegative weights, an HC tree T has one leaf per vertex. Dasgupta's objective (Dasgupta, 2016) is

$$\mathrm{cost}_G(T) := \sum_{\{u,v\} \in E} w_G(u,v) |\operatorname{leaves}(T[u \vee v])|, \tag{1}$$

where $T[u \vee v]$ is the subtree rooted at the lowest common ancestor of u, v. Thus, the objective favors placing highly similar vertices together in small clusters. Write $n = |V|$, $m = |E|$ and T^* for an optimum tree. On general graphs, the best known polynomial-time approximation factor is $O(\sqrt{\log n})$, obtained by recursive sparsest cut (Charikar & Chatziafratis, 2017; Arora et al., 2009); no constant-factor approximation is possible in polynomial time under the Small-Set Expansion Hypothesis (Charikar & Chatziafratis, 2017; Roy & Pokutta, 2017).

Differentially Private Hierarchical Clustering. In many networks, the edges encode sensitive relationships. For unweighted graphs, edge differential privacy (DP) protects these relationships by requiring that adding or removing a single edge changes the distribution of the released output only slightly (Dwork & Roth, 2014). We consider *pure ϵ-edge DP* on weighted graphs with a public vertex set; see Definition 2.1 for the formal definition.

For general weighted graphs, Imola et al. (2023) give a polynomial-time (ϵ, δ)-DP algorithm that achieves cost $O(\sqrt{\log n})\mathrm{cost}_G(T^*) + \widetilde{O}(n^{5/2}/\epsilon)$ with constant probability. They also prove that every pure-DP algorithm has expected cost $\Omega(n^2/\epsilon)$ on some input with optimal cost $O(n/\epsilon)$. Thus, any pure edge-DP algorithm with an $o(n)$ multiplicative approximation factor must have an expected additive error of $\Omega(n^2/\epsilon)$ in the worst case. In contrast, Deng et al. (2025) study the setting where the graph topology is public and obtain an $O(\log^{3/2} n/\epsilon)$ approximation in polynomial time for connected graphs with every edge weight at least one. Their privacy requirement is weaker: it protects only edge weights, not edge presence, so their guarantee does not directly apply to our setting.

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.

To obtain a multiplicative guarantee without the large additive error of worst-case private HC, we study hierarchical clustering on *well-clusterable graphs*: graphs with strong connections within clusters and sparse connections between them. Consider an undirected web graph in which pages within the same website are densely connected, with relatively few edges between websites. Websites give a natural flat partition, while a hierarchy can organize related websites into broader topical groups. Well-clusterability also provides a beyond-worst-case setting in which to explain why simple clustering procedures can produce low-cost trees. This structure has already proved useful in non-private HC: Manghiuc & Sun (2021) and Laenen et al. (2023) obtain approximation guarantees governed by the number of clusters and structural parameters rather than by n. In private HC, Imola et al. (2023) obtain a $(1 + o(1))$ approximation under (ϵ, δ)-DP for (hierarchical) stochastic block models, which provide a distributional counterpart to well-clusterable graphs.

To formalize the cluster structure, let $d_G(v) = \sum_u w_G(u, v)$, $d_{\min} = \min_v d_G(v)$, and $\mathrm{vol}_G(S) = \sum_{v \in S} d_G(v)$. Define conductance by

$$\Phi_G(S) := \frac{w_G(S, V \setminus S)}{\mathrm{vol}_G(S)}, \qquad \Phi_G := \min_{0 < \mathrm{vol}_G(S) \le \mathrm{vol}_G(V)/2} \Phi_G(S), \tag{2}$$

where $w_G(S, V \setminus S)$ is the total weight leaving S. The inner conductance of a cluster is measured in its induced graph.

Definition 1.1 (Well-clusterable graph)**.** A graph is $(k, \phi_{in}, \phi_{out})$-*clusterable* if it admits a partition $\{C_1, \ldots, C_k\}$ such that $\Phi_{G[C_i]} \ge \phi_{\text{in}}$ and $\Phi_G(C_i) \le \phi_{\text{out}}$ for every i.

When ϕ_{in} is much larger than ϕ_{out}, a cluster has a sparse boundary relative to its volume, but cuts within the cluster have comparatively large boundaries. The groups are therefore easy to separate from one another and harder to split further.

The eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ of the normalized Laplacian offer a spectral view of this distinction. Informally, a small λ_k indicates roughly k weakly connected regions, while λ_{k+1} controls the possibility of further sparse splits. Lemma 2.2 makes this precise when λ_k is sufficiently small as a function of k and $\lambda_{k+1} > 0$: it yields at most k clusters with sparse boundaries and an inner-conductance guarantee that strengthens as λ_{k+1} grows.

Well-clusterable graphs offer a practically motivated setting beyond worst-case inputs. Yet provable private HC algorithms can involve substantially more computation: the general-graph approach of Imola et al. (2023), for example, releases a private graph before applying a non-private HC routine and has a polynomial-time guarantee. This motivates a simpler, more scalable approach. We ask:

> *Can well-clusterability yield a* practical *multiplicative approximation algorithm for a full hierarchy under pure edge DP, ideally in nearly linear time?*

1.1 Our Results

Noisy-Degree Hierarchy. We propose Noisy-Degree Hierarchy, which groups vertices by noisy degree and constructs a full HC tree from the resulting buckets without recovering the cluster partition. The next theorem summarizes its privacy, approximation, and running-time guarantees.

Theorem 1.2 (Noisy-Degree Hierarchy, informal)**. Noisy-Degree Hierarchy** *is pure ϵ-edge DP on every input. If $d_{\min} \ge 1$ and the spectral conditions of Lemma 2.2 hold, then, for $0 < \epsilon \le 1$, with probability at least $1 - n^{-10}$, its output satisfies*

$$\mathrm{cost}_G(T_{\text{noisy}}) \le O\left(\frac{k^6}{\lambda_{k+1}^2 \epsilon}\right) \mathrm{cost}_G(T^*). \tag{3}$$

Its running time is $O(m + n \log n)$.

To our knowledge, this is the first nearly linear time pure edge-DP algorithm with a provable multiplicative guarantee for HC on well-clusterable graphs. Its construction uses noisy degrees and sorting, without solving an SDP or a convex program. Our analysis begins with a simpler non-private construction, Degree-Ordered Hierarchy (Theorem 3.4). It orders vertices by nonincreasing degree and merges them successively, achieving an $O(k^6/\lambda_{k+1}^2)$ approximation in $O(m + n \log n)$ time. Under the same spectral conditions, the PruneMerge algorithm of Manghiuc & Sun (2021)

achieves an $O(k^{22}/\lambda_{k+1}^{10})$ approximation in polynomial time. NOISY-DEGREE HIERARCHY makes the degree-based construction private by replacing the exact ordering with noisy degree buckets.

The main analytical challenge is to bound the cost of a hierarchy built from degrees alone, without knowing the cluster memberships. We use expansion within the clusters to relate the cost of the degree-based hierarchy to the optimal HC cost. Privacy introduces a further difficulty: degree noise can change both vertex order and bucket membership. Rather than requiring the noisy ordering to preserve the true one, we charge these changes to the total absolute noise, incurring only an additional $O(1/\epsilon)$ factor in the approximation guarantee.

Degree-Regular Regime. Our algorithm admits a stronger guarantee when vertices within each cluster have comparable degrees. This condition only limits degree variation within each cluster, allowing different clusters to have different sizes and degree scales. For a partition $\{S_i\}_{i=1}^k$, we measure this variation by $\eta_S = \max_i\big(\max_{v\in S_i} d_G(v)/\min_{v\in S_i} d_G(v)\big)$, where $d_G(v)$ includes edges both within and outside the cluster. When η_S is bounded, a cluster's volume is comparable to its size times the degree of any of its vertices. Together with strong internal connectivity, this allows a tighter analysis of the same degree-based construction.

Theorem 1.3 (Degree-regular private HC, informal)**.** *Let G be a $(k, \phi_{in}, \phi_{out})$-clusterable graph with a corresponding partition $\{S_i\}_{i=1}^k$, where $d_{\min} \geq 1$, $\phi_{in} > 0$ and $\phi_{out} \leq 1/2$. For $0 < \epsilon \leq 1$, with probability at least $1 - n^{-10}$* NOISY-DEGREE HIERARCHY *attains*

$$\mathrm{cost}_G(T_{\mathrm{noisy}}) \leq O\left(\frac{k\eta_S}{\phi_{in}\epsilon}\right)\mathrm{cost}_G(T^*). \tag{4}$$

The running time is $O(m + n\log n)$. This guarantee does not require a small λ_k.

For constant inner conductance and bounded degree ratios, the approximation factor becomes $O(k/\epsilon)$ (Corollary 3.6). To illustrate how this structural guarantee improves on general-graph bounds, Appendix E considers a sparse graph family consisting of two expanders with unequal component sizes and different edge weights. On this family, the optimal cost is $\Theta(n^2)$. The polynomial-time guarantee of Imola et al. (2023) gives an approximation factor of $O(\sqrt{\log n}) + \widetilde{O}(\sqrt{n}/\epsilon)$, whereas our algorithm achieves $O(1/\epsilon)$ in $O(n\log n)$ time.

Privacy Lower Bound. We show that a linear dependence on k is necessary, even when the graph consists of disjoint regular expanders.

Theorem 1.4 (Lower bound, informal)**.** *For fixed sufficiently small $\epsilon > 0$ and sufficiently large k, every pure ϵ-edge-DP HC algorithm $\mathcal{A}$ incurs, on some unweighted graph G of $n = ks$ vertices consisting of k disjoint regular expanders of a fixed size s, with probability $1 - e^{-\Omega(n)}$,*

$$\mathrm{cost}_G(\mathcal{A}(G)) \geq \Omega(k)\mathrm{cost}_G(T^*). \tag{5}$$

These graphs satisfy both the spectral conditions and the degree-regular conditions with constant inner conductance. Thus, for fixed sufficiently small $\epsilon > 0$, the lower bound matches the $O(k)$ guarantee of Theorem 1.3, with both bounds holding with high probability. The proof adapts the packing argument of Imola et al. (2023) to regular expander components; Theorem 4.1 gives the precise statement, and Appendix C gives the proof.

Experiments. Our experiments evaluate the utility and running time of our method. At $\epsilon = 1$, NOISY-DEGREE HIERARCHY achieves lower cost than Imola-DP in all tested comparisons, with cost reductions of 39.8%–54.1% across the three datasets. We also evaluate the non-private version against AverageLinkage. In running-time experiments, NOISY-DEGREE HIERARCHY is faster than all three competing methods and exhibits approximately linear growth with graph size. Section 5 presents these results. We further test whether privately querying edge weights between degree buckets improves utility over merging the buckets by degree order. Appendices B.1–B.3 study two methods that use this additional information, and Appendix D.3 reports the comparison.

1.2 OTHER RELATED WORK

Other HC and clustering results. Cohen-Addad et al. (2019) study axioms for HC objectives and obtain near-optimal guarantees for hierarchical stochastic block models; Agarwal et al. (2022) consider sublinear-query and streaming models. For well-clusterable graphs, Peng et al. (2015)

establish spectral clustering guarantees, and Peng (2020); Gluch et al. (2021) develop sublinear-time cluster-membership oracles. Kapralov et al. (2025) estimate the optimal HC cost using labeled samples, while He et al. (2024) privately recover nearly balanced flat clusters. These latter results do not construct a private HC tree.

A further privacy lower bound. When the graph topology is public and zero edge weights are allowed, Imola et al. (2026) prove an $\Omega(n^2/\epsilon)$ expected additive-error lower bound under approximate DP for sufficiently small δ.

2 PRELIMINARIES

2.1 GRAPHS AND HIERARCHICAL CLUSTERING

Throughout the paper, we consider finite undirected graphs without self-loops. For a weighted graph $G = (V, E, w_G)$ with $n = |V| \geq 1$, let $w_G(u, v) \geq 0$ denote the weight of $\{u, v\}$, with zero weight for absent edges. Thus, $E = \{\{u, v\} \in \binom{V}{2} : w_G(u, v) > 0\}$. Define $d_G(v) := \sum_{u \in V} w_G(u, v)$ and $w_G(S, T) := \sum_{u \in S, v \in T} w_G(u, v)$. Volumes and conductances are defined using these weights, as in (2). For every nonempty $S \subseteq V$, write $\Delta_G(S) := \max_{v \in S} d_G(v)$ and $\delta_G(S) := \min_{v \in S} d_G(v)$. The average degree is $d_{\text{avg}}(G) := \text{vol}_G(V)/|V|$.

For graphs with no isolated vertices, let A_G and D_G denote the weighted adjacency matrix and the diagonal degree matrix, respectively. The normalized Laplacian $\mathcal{L}_G := I - D_G^{-1/2} A_G D_G^{-1/2}$ has eigenvalues $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$.

An HC tree T on G is a rooted full binary tree whose leaves correspond bijectively to V. For $u, v \in V$, let $T[u \vee v]$ denote the subtree rooted at their lowest common ancestor (LCA). We use the Dasgupta cost in (1), and write T^* for an optimal HC tree.

2.2 DIFFERENTIAL PRIVACY

Fix a public vertex set V. Two nonnegatively weighted graphs G, G' on V are *edge-adjacent*, written $G \sim G'$, if

$$\sum_{\{u,v\} \in \binom{V}{2}} |w_G(u, v) - w_{G'}(u, v)| \leq 1.$$

Absent edges have weight zero, and the edge sets may differ.

Definition 2.1 (Edge differential privacy)**.** A randomized algorithm $\mathcal{A}$ is *ϵ-edge differentially private* if, for every $G \sim G'$ and every event $\mathcal{O}$ in its output space,

$$\Pr[\mathcal{A}(G) \in \mathcal{O}] \leq e^{\epsilon} \Pr[\mathcal{A}(G') \in \mathcal{O}].$$

All privacy guarantees we establish are pure edge DP.

We use the Laplace mechanism (Dwork & Roth, 2014). For a query f taking values in $\mathbb{R}^d$, define its ℓ_1 sensitivity by

$$\Delta_1(f) := \max_{G \sim G'} \|f(G) - f(G')\|_1.$$

The *Laplace mechanism* releases $f(G) + (Y_1, \ldots, Y_d)$, where the Y_i are independent $\text{Lap}(\Delta_1(f)/\epsilon)$ random variables. It satisfies ϵ-DP.

We also use composition, group privacy, and post-processing.

2.3 SPECTRAL CONDITIONS AND EXPANDER DECOMPOSITION

The following decomposition lemma relates spectral conditions to the cluster structure used in our analysis. Our algorithms do not need to compute the partition guaranteed by the lemma.

Lemma 2.2 (Strong expander decomposition; Oveis Gharan & Trevisan, 2014; Laenen et al., 2023)**.** *There is an absolute constant $c_0 > 1$ such that the following holds. Let G have no isolated vertices*

and let $2 \le k < n$ be an integer. If $\lambda_{k+1} > 0$ and

$$\lambda_k < \left(\frac{1}{270 c_0 (k+1)^6} \right)^2,$$

then V admits a partition $\{C_1, \dots, C_\ell\}$ with $\ell \le k$. For every $1 \le i \le \ell$ and every $u \in C_i$, the following properties hold:

$$\begin{aligned} &\text{(P1)} \quad \Phi_G(C_i) = O(k^6 \sqrt{\lambda_k}), \\ &\text{(P2)} \quad \Phi_{G[C_i]} = \Omega(\lambda_{k+1}^2 / k^4), \\ &\text{(P3)} \quad w_G(u, V \setminus C_i) \le 6(k+1) d_{G[C_i]}(u). \end{aligned}$$

In particular, (P3) implies $d_G(u) \le (6k+7) d_{G[C_i]}(u)$ for every $u \in C_i$.

3 Degree Profiles and Private Hierarchies

In this section, we use degree profiles to analyze Degree-Ordered Hierarchy and Noisy-Degree Hierarchy. We first relate the cost of degree-based constructions to the degree profile, then use cluster structure to obtain the non-private approximation guarantee. We next bound the effect of degree noise to establish the private guarantee, and derive a sharper bound for degree-regular clusters. Full proofs appear in Appendix A.

3.1 Degree Profiles and Bucket Comparison

For a nonnegative vector d of length N, define its degree profile by

$$Q(d) := \sum_{j=1}^{N} j d_{(j)}, \qquad d_{(1)} \ge \cdots \ge d_{(N)} \ge 0. \tag{6}$$

We use $Q(d_G)$ to bound the cost of degree-based constructions, and then use cluster structure to show that $Q(d_G)$ is at most a bounded factor times $\mathrm{cost}_G(T^*)$. This functional and its connection to degree-based HC appear in Kapralov et al. (2025, full version, Lemmas B.1–B.2).

Let $\mathcal{B}$ be a collection of nonempty buckets that partition V. For each bucket $B \in \mathcal{B}$, fix an HC tree T_B whose leaves are the vertices in B. To combine these trees, take an HC tree whose leaves are the buckets and replace each leaf B with T_B. We call the resulting tree on V a lift. Every lift has the same internal cost $I := \sum_{B \in \mathcal{B}} \mathrm{cost}_{G[B]}(T_B)$. Let W be the minimum cost contribution of edges between buckets over all such lifts. Then $\mathcal{O}_{\mathcal{B}} := I + W$ is the minimum total cost among HC trees that contain each T_B as a subtree. This minimum can exceed $\mathrm{cost}_G(T^*)$.

Lemma 3.1 (Profile and bucket comparison)**.** *Let G be a graph with nonnegative edge weights. Suppose $R \ge 1$ and*

$$d_G(u) \le R d_G(v) \qquad \text{for every } B \in \mathcal{B} \text{ and } u, v \in B. \tag{7}$$

For every choice of the internal trees,

$$I + W \le 2RQ(d_G). \tag{8}$$

Proof idea. For each bucket B, let $a_B = \max_{v \in B} d_G(v)$, and set $a_v = a_B$ for every $v \in B$. Order the buckets by decreasing a_B and merge their internal trees successively, joining the current tree and the next bucket's tree under a new root. Let $p_j = \sum_{i \le j} |B_i|$. Each edge between B_j and an earlier bucket then has LCA subtree size p_j. Including the internal edges, the resulting tree has cost at most $\sum_j p_j \mathrm{vol}_G(B_j) \le 2Q(a)$. Since $a \le R d_G$ coordinatewise, this is at most $2RQ(d_G)$. The full proof appears in Appendix A.1.2. □

We next introduce properties of the degree profile that will help relate $Q(d_G)$ to the optimal HC cost. Their proof appears in Appendix A.1.1.

Lemma 3.2 (Degree-profile properties)**.** *If a graph F has conductance at least $h \in (0, 1]$, then every HC tree T on F satisfies*

$$Q(d_F) \le (1 + 2/h)\mathrm{cost}_F(T). \tag{9}$$

For nonnegative vectors $d^{(1)}, \dots, d^{(\ell)}$, their concatenation satisfies

$$Q(d^{(1)} \parallel \cdots \parallel d^{(\ell)}) \le \ell \sum_i Q(d^{(i)}). \tag{10}$$

Here $\parallel$ denotes vector concatenation. Moreover, Q is coordinatewise monotone and positively homogeneous.

3.2 Pointwise Certificate and Degree-Ordered Hierarchy

Theorem 3.3 (Pointwise cluster certificate)**.** *Let G be a graph with nonnegative edge weights. Suppose V has a partition $\{C_1, \dots, C_\ell\}$, with $\ell \le k$, such that*

$$\Phi_{G[C_i]} \ge h > 0, \qquad d_G(v) \le \kappa d_{G[C_i]}(v) \quad (v \in C_i), \tag{11}$$

where $h \le 1$ and $\kappa \ge 1$. Then

$$Q(d_G) \le \ell\kappa(1 + 2/h)\mathrm{cost}_G(T^*). \tag{12}$$

For every bucket partition satisfying (7) *with $R \ge 1$, and every choice of its internal trees,*

$$I + W \le 2R\ell\kappa(1 + 2/h)\mathrm{cost}_G(T^*) \le \frac{6Rk\kappa}{h}\mathrm{cost}_G(T^*). \tag{13}$$

The proof combines Lemma 3.2 with the pointwise degree condition in (11). See Appendix A.2.1 for details.

Theorem 3.4 (Degree-Ordered Hierarchy)**.** Degree-Ordered Hierarchy *orders vertices by nonincreasing degree, breaking ties by public labels. It starts with the first vertex and successively joins the current tree and the next vertex under a new root. The resulting tree T_{degree} satisfies $\mathrm{cost}_G(T_{\mathrm{degree}}) \le Q(d_G)$. Under the spectral conditions of Lemma 2.2,*

$$\mathrm{cost}_G(T_{\mathrm{degree}}) \le O\left(\frac{k^6}{\lambda_{k+1}^2}\right)\mathrm{cost}_G(T^*). \tag{14}$$

The construction takes $O(m + n \log n)$ time and does not require k as input.

Proof idea. An edge of weight w between ranks $i < j$ contributes wj to the tree cost and $w(i+j)$ to $Q(d_G)$. Summing over edges gives $\mathrm{cost}_G(T_{\mathrm{degree}}) \le Q(d_G)$. Applying Theorem 3.3 with $\kappa = 6k+7$ and $h = \Omega(\lambda_{k+1}^2/k^4)$ from Lemma 2.2 gives (14). The full proof appears in Appendix A.3.1. □

Theorem 3.4 improves the $O(k^{22}/\lambda_{k+1}^{10})$ approximation guarantee for PruneMerge of Manghiuc & Sun (2021, Theorem 4) to $O(k^6/\lambda_{k+1}^2)$ under the same spectral conditions, using a simpler construction.

3.3 Noisy-Degree Hierarchy

Noisy-Degree Hierarchy groups vertices into buckets according to their noisy degrees, using a low-degree bucket B_0 and geometric intervals with ratio $\beta > 1$.

Algorithm 1: Noisy degree bucketing

Input: A nonnegatively weighted graph G; $\epsilon_d > 0$; $\beta > 1$
Output: Nonempty buckets with their original interval indices

For each $v \in V$, independently sample $Z_v \sim \mathrm{Lap}(2/\epsilon_d)$ and set $\widehat{d}(v) \leftarrow d_G(v) + Z_v$
$B_0 \leftarrow \{v \in V : \widehat{d}(v) < 1\}$
$B_{j+1} \leftarrow \{v \in V : \beta^j \le \widehat{d}(v) < \beta^{j+1}\}$ for each occupied interval $j \ge 0$
return $\mathcal{B} = \{B_j : B_j \ne \emptyset\}$, retaining each index j

The degree vector has ℓ_1-sensitivity two, so the release is ϵ_d-edge DP. The buckets are obtained by post-processing this release.

Algorithm 2 constructs the hierarchy from these buckets. Within each bucket, it orders vertices by public labels and recursively splits the sequence into halves to form a balanced tree T_B. It then merges the bucket trees in decreasing interval order, without further access to the graph.

Algorithm 2: NOISY-DEGREE HIERARCHY

Input: A nonnegatively weighted graph G; $\epsilon > 0$; $\beta > 1$
Output: An HC tree on V

$\mathcal{B} \leftarrow$ NOISYDEGREEBUCKETS(G, ϵ, β), retaining each nonempty bucket's interval index
Construct the public-order balanced tree T_B for each $B \in \mathcal{B}$
Order the buckets $B^{(1)}, \ldots, B^{(r)}$ by decreasing interval index
$T \leftarrow T_{B^{(1)}}$
for $j = 2, \ldots, r$ **do**
  $T \leftarrow$ a new root with children T and $T_{B^{(j)}}$
return T

Theorem 3.5 (Noisy-Degree Hierarchy)**.** *Let T_{noisy} be the tree returned by Algorithm 2. The algorithm is pure ϵ-edge DP on every input graph. There is a constant C_β, depending only on β, such that on every input with $d_{\min} \geq 1$, with probability at least $1 - n^{-10}$,*

$$\mathrm{cost}_G(T_{\mathrm{noisy}}) \leq C_\beta(1 + 1/\epsilon)Q(d_G). \tag{15}$$

Consequently, its approximation factor is at most $C_\beta(1 + 1/\epsilon)k\kappa(1 + 2/h)$ under (11)*. Under the spectral conditions of Lemma 2.2 it is $O_\beta((1 + 1/\epsilon)k^6/\lambda_{k+1}^2)$. The running time is $O(m + n \log n)$ for fixed $\beta > 1$.*

Proof idea. Privacy follows from the degree release and post-processing. To bound the cost, we account for changes in degree order caused by noise. For pairs of vertices placed in the same bucket or in reversed degree order, the analysis uses both their true degrees and their noise magnitudes. Summing these bounds gives, when $d_{\min} \geq 1$,

$$\mathrm{cost}_G(T_{\mathrm{noisy}}) \leq (\beta + 1)Q(d_G) + \beta(n - 1)\sum_v |Z_v|. \tag{16}$$

The noise magnitudes $|Z_v|$ are independent exponential random variables with mean $2/\epsilon$, and their sum is $O(n/\epsilon)$ with probability at least $1-n^{-10}$. Together with $Q(d_G) \geq n^2/2$, this yields (15). The structural and spectral bounds follow from Theorem 3.3. The full proof appears in Appendix A.3.2. □

3.4 A DEGREE-REGULAR REFINEMENT

We now refine the approximation guarantee using cluster conductance and degree regularity. For a partition $\{S_i\}_{i=1}^k$, let $\eta_S = \max_i \Delta_G(S_i)/\delta_G(S_i)$. This measures the variation in global degrees within each cluster. Under the conditions below, Theorem A.1 gives $Q(d_G) \leq 18k\eta_S \mathrm{cost}_G(T^*)/\phi_{\mathrm{in}}$. Combining this bound with Theorem 3.5 gives the following refinement.

Corollary 3.6 (Degree-regular Noisy-Degree Hierarchy)**.** *Let G be a $(k, \phi_{in}, \phi_{out})$-clusterable graph with a corresponding partition $\{S_i\}_{i=1}^k$, where $d_{\min} \geq 1$, $\phi_{in} > 0$, and $\phi_{out} \leq 1/2$.* NOISY-DEGREE HIERARCHY *is pure ϵ-edge DP and, with probability at least $1 - n^{-10}$, satisfies*

$$\mathrm{cost}_G(T_{\mathrm{noisy}}) \leq \frac{18C_\beta(1 + 1/\epsilon)k\eta_S}{\phi_{in}} \mathrm{cost}_G(T^*). \tag{17}$$

Its running time is $O(m + n \log n)$ for fixed $\beta > 1$.

The proof appears in Appendix A.3.3. For constant inner conductance, bounded degree ratios, and fixed β, the approximation factor is $O(k/\epsilon)$ for $0 < \epsilon \leq 1$. For fixed sufficiently small $\epsilon > 0$, this matches the dependence on k in the lower bound of Section 4.

Remark 3.7 (Relation between the guarantees). The spectral guarantee in Theorem 3.5 and the degree-regular guarantee in Corollary 3.6 apply to the same NOISY-DEGREE HIERARCHY algorithm. The spectral guarantee is expressed in terms of k and λ_{k+1}, whereas Corollary 3.6 uses the inner conductance and degree ratio of the clusters. The latter can give a sharper bound when these parameters are favorable and does not require the small-λ_k condition.

4 LOWER BOUNDS

For sufficiently small $\epsilon > 0$, every pure ϵ-edge-DP algorithm has an approximation factor of $\Omega(k)$ on some graph consisting of k disjoint regular expanders. Formally, we prove the following theorem:

Theorem 4.1. *There exist constants $\epsilon_0, \phi, c_1, c_2 > 0$ and an integer $s \geq 100$ such that the following holds. For every sufficiently large k, let $n = ks$. For every HC algorithm A satisfying ϵ-edge DP with $0 < \epsilon \leq \epsilon_0$, there is an unweighted n-vertex graph G consisting of k disjoint regular ϕ-expanders for which*

$$\Pr[\mathrm{cost}_G(A(G)) \geq c_1 k \mathrm{cost}_G(T^*)] \geq 1 - 2e^{-c_2 n}.$$

The graphs in Theorem 4.1 have $d_{\min} \geq 1$, $\lambda_k = 0$ and $\lambda_{k+1} = \Omega(1)$, so they satisfy the spectral conditions of Lemma 2.2. Thus at least linear dependence on k is unavoidable even with perfectly separated clusters. Furthermore, each cluster in these graphs is regular, so these graphs satisfy the conditions of both Theorem 3.5 and Corollary 3.6. For fixed sufficiently small $\epsilon > 0$, the upper and lower bounds therefore match in order on this constant-parameter degree-regular family, both with probability at least $1 - n^{-10}$ for sufficiently large n.

The proof uses a packing argument (Hardt & Talwar, 2010), following Imola et al. (2023). We construct a large family of well-clusterable graphs, each with optimal HC cost $O(n^2/k)$, and show that, for any two distinct graphs in this family, every HC tree has cost $\Omega(n^2)$ on at least one of them. Group privacy then implies that any ϵ-edge-DP algorithm incurs cost $\Omega(n^2)$ with high probability on some graph in the family. See Appendix C for details.

5 EXPERIMENTS

We evaluate utility and running time on unweighted graphs. Utility is measured by Dasgupta cost, with lower cost indicating better utility. Unless otherwise stated, NOISY-DEGREE HIERARCHY uses $\beta = 2$ and allocates the entire privacy budget $\epsilon = 1$ to the degree release.

5.1 UTILITY

Datasets. We generate synthetic graphs from an SBM with five equal blocks of 200 vertices ($n = 1000$). Edges are sampled independently, with within-block probabilities ranging from approximately 0.03 to 0.20 and a common between-block probability of 0.0025. The average expected degree is 22. Appendix D.1 gives further details. We also test two real-world graphs, Email-Eu-core ($n = 1005$) and Political Blogs ($n = 1490$). We retain all vertices, including 19 and 266 isolated vertices, respectively, so these real-world graphs do not satisfy the $d_{\min} \geq 1$ assumption in our utility guarantees.

Compared methods. We compare NOISY-DEGREE HIERARCHY with the private SparseCut baseline of Imola et al. (2023), which we refer to as Imola-DP. Both methods use the same privacy budget $\epsilon = 1$ under single-edge addition or removal. We also compare with deterministic AverageLinkage, which has shown strong practical performance in previous experiments (Laenen et al., 2023). For this comparison, we use a non-private version that replaces noisy degrees with exact degrees while keeping the same bucketing and tree construction. Our theoretical analysis also applies to this version.

Evaluation protocol. We generate five independent SBM graphs and run each private method five times per graph. On each real-world graph, we run each private method ten times. For each repetition on the same graph, we compute the cost of NOISY-DEGREE HIERARCHY divided by that of Imola-DP. Table 1(a) reports the median cost ratio, its 95% bootstrap confidence interval, and the number of pairs in which NOISY-DEGREE HIERARCHY has lower cost. For the non-private comparison in

Table 1(b), we compute one cost ratio between the exact-degree version and AverageLinkage for each graph. We report the median and range across the five SBM graphs, and a single ratio for each real-world graph. Appendix D.1 provides implementation and statistical details.

Table 1: Utility comparisons on synthetic and real-world graphs: (a) NOISY-DEGREE HIERARCHY versus Imola-DP; (b) the exact-degree version versus AverageLinkage.

(a) Private utility

Input	NOISY-DEGREE / Imola-DP	Wins
SBM	0.5959 [0.5894, 0.6081]	25/25
Email	0.6022 [0.5959, 0.6110]	10/10
Political Blogs	0.4590 [0.4433, 0.4710]	10/10

(b) Non-private utility

Input	Exact degrees / AverageLinkage
SBM	1.6799 [1.6474, 1.7060]
Email	1.2138
Political Blogs	1.1679

Results. NOISY-DEGREE HIERARCHY achieves lower cost than Imola-DP in all 45 comparisons, with cost reductions of 39.8%–54.1% across the three datasets. The exact-degree version has costs about 17–21% higher than those of AverageLinkage on the real-world graphs. The gap is larger on SBM, where its cost is approximately 1.68 times that of AverageLinkage.

5.2 RUNNING-TIME COMPARISON

We compare the running time of NOISY-DEGREE HIERARCHY, Imola-DP, practical SpecWRSC (Laenen et al., 2023) and AverageLinkage on sparse SBM graphs with five equal-sized blocks and a fixed expected degree of 22. For each $n \in \{240, 480, 960, 1920, 3840\}$, we generate five independent graphs and evaluate all four methods on each graph. We measure the running time of each complete HC algorithm on the same server using a single thread. Further experimental details are provided in Appendix D.2.

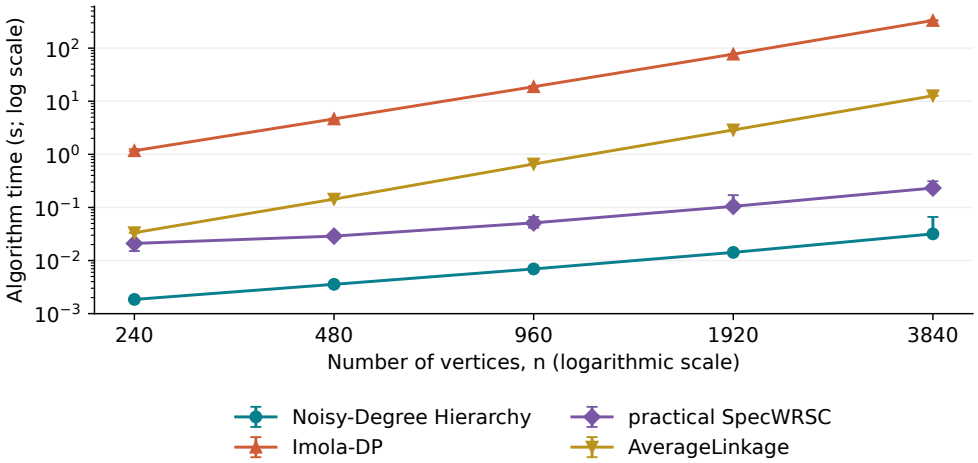


Figure 1: Running times of NOISY-DEGREE HIERARCHY, Imola-DP, practical SpecWRSC and AverageLinkage. Points show medians over five graph instances; error bars show observed minima and maxima.

As shown in Figure 1, our NOISY-DEGREE HIERARCHY is the fastest method at every tested graph size. Its running time increases from 1.85 ms at $n = 240$ to 31.75 ms at $n = 3840$, growing approximately linearly with the number of vertices. The consistently low running time compared with the other methods demonstrates the efficiency of NOISY-DEGREE HIERARCHY as the graph size increases. Together with the utility results, these measurements demonstrate the practical effectiveness of our method.

AI USE STATEMENT

Generative AI tools, including ChatGPT and Codex, were used to assist with manuscript drafting and editing, theoretical analysis and proof development, and experiments. The authors reviewed and revised the AI-assisted material and take full responsibility for the final content of the paper.

REPRODUCIBILITY STATEMENT

This paper is primarily theoretical. All assumptions, algorithmic details, and complete proofs of our theoretical results are provided in the main paper and the appendices. For the empirical evaluation, we describe the datasets, baselines, evaluation metrics, parameter settings, and experimental protocols in Section 5 and Appendix D. The code and instructions for reproducing the experiments are available at `https://anonymous.4open.science/r/dphc-code/`.

REFERENCES

Arpit Agarwal, Sanjeev Khanna, Huan Li, and Prathamesh Patil. Sublinear algorithms for hierarchical clustering. *Advances in Neural Information Processing Systems*, 35:3417–3430, 2022. Full version: `https://arxiv.org/abs/2206.07633`.

Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM*, 56(2):1–37, 2009.

Moses Charikar and Vaggos Chatziafratis. Approximate hierarchical clustering via sparsest cut and spreading metrics. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 841–854. SIAM, 2017.

Aaron Clauset, Cristopher Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008. doi: 10.1038/nature06830.

Vincent Cohen-Addad, Varun Kanade, Frederik Mallmann-Trenn, and Claire Mathieu. Hierarchical clustering: Objective functions and algorithms. *Journal of the ACM*, 66(4), 2019. doi: 10.1145/3321386.

Sanjoy Dasgupta. A cost function for similarity-based hierarchical clustering. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, pp. 118–127, 2016.

Chengyuan Deng, Jie Gao, Jalaj Upadhyay, Chen Wang, and Samson Zhou. On the price of differential privacy for hierarchical clustering. In *The Thirteenth International Conference on Learning Representations*, 2025.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Michael B. Eisen, Paul T. Spellman, Patrick O. Brown, and David Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25):14863–14868, 1998. doi: 10.1073/pnas.95.25.14863.

Grzegorz Gluch, Michael Kapralov, Silvio Lattanzi, Aida Mousavifar, and Christian Sohler. Spectral clustering oracles in sublinear time. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms*, pp. 1598–1617. SIAM, 2021. doi: 10.1137/1.9781611976465.97.

Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, pp. 705–714. ACM, 2010.

Weiqiang He, Hendrik Fichtenberger, and Pan Peng. A differentially private clustering algorithm for well-clustered graphs. In *The Twelfth International Conference on Learning Representations*, 2024.

Jacob Imola, Alessandro Epasto, Mohammad Mahdian, Vincent Cohen-Addad, and Vahab Mirrokni. Differentially private hierarchical clustering with provable approximation guarantees. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 14353–14375. PMLR, 2023.

Jacob Imola, Rasmus Pagh, and Lukas Retschmeier. Lower bounds for private graph optimization problems using reconstruction attacks. *arXiv preprint arXiv:2609.10877*, 2026. URL https://arxiv.org/abs/2609.10877v1.

Michael Kapralov, Akash Kumar, Silvio Lattanzi, Aida Mousavifar, and Weronika Wrzos-Kaminska. Approximating Dasgupta cost in sublinear time from a few random seeds. In *52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025)*, pp. 103:1–103:19. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2025. Full version: https://arxiv.org/abs/2207.02581v3.

Steinar Laenen, Bogdan Adrian Manghiuc, and He Sun. Nearly-optimal hierarchical clustering for well-clustered graphs. In *International Conference on Machine Learning*, pp. 18207–18249. PMLR, 2023. Full version: arXiv:2306.09950v1.

Bogdan-Adrian Manghiuc and He Sun. Hierarchical clustering: $O(1)$-approximation for well-clustered graphs. *Advances in Neural Information Processing Systems*, 34:9278–9289, 2021. Full version: arXiv:2112.09055v1.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science*, pp. 94–103. IEEE, 2007. doi: 10.1109/FOCS.2007.66.

Shayan Oveis Gharan and Luca Trevisan. Partitioning into expanders. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1256–1266. SIAM, 2014.

Pan Peng. Robust clustering oracle and local reconstructor of cluster structure of graphs. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2953–2972. SIAM, 2020.

Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs: Spectral clustering works! In *Conference on Learning Theory*, pp. 1423–1455. PMLR, 2015.

Aurko Roy and Sebastian Pokutta. Hierarchical clustering via spreading metrics. *Journal of Machine Learning Research*, 18(88):1–35, 2017. URL https://jmlr.org/papers/v18/17-081.html.

Joe H. Ward, Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 1963. doi: 10.1080/01621459.1963.10500845.

A Proofs for Degree Profiles and Structural Certificates

This appendix provides the proofs deferred from Section 3.

A.1 Profile and Bucket Comparisons

A.1.1 Expansion and Merging

We give a size-threshold proof using the cost representation of Charikar & Chatziafratis (2017, Claim 2.1). The same $1/h$ dependence follows from the two dense-branch bounds of Manghiuc & Sun (2021, full version, Lemma 3.5).

Proof of Lemma 3.2. Expansion bound. Let $N = |V(F)|$ and $U = \mathrm{vol}_F(V(F))$, and write $d_{(1)} \geq \cdots \geq d_{(N)}$ for the degrees $d_F(v)$ in nonincreasing order. In particular, $\sum_{j=1}^{N} d_{(j)} = U$.

We decompose the tree cost by subtree size. For each integer $1 \leq t < N$, consider the nodes of T whose subtrees have at most t leaves and are maximal under ancestry: their parents have more than t leaves. Leaf nodes are included among the candidates. Every leaf has a unique highest ancestor with at most t leaves. Moreover, the subtrees of two selected nodes are disjoint, since otherwise one would be an ancestor of the other, contradicting maximality. Their leaf sets therefore form a partition $\mathcal{P}_t$ of $V(F)$, and every part has at most t vertices. Let e_t be the total weight of unordered edges with endpoints in different parts of $\mathcal{P}_t$.

For an edge $\{u, v\}$, let s be the number of leaves below its LCA in T. The endpoints lie in the same part of $\mathcal{P}_t$ exactly when $s \le t$: in that case their LCA is contained in a selected subtree, whereas for $s > t$ no subtree with at most t leaves can contain both endpoints. Thus this edge contributes its weight to e_t for precisely $t = 1, \dots, s-1$. Summing over edges counts each weight $s-1$ times, instead of the s times appearing in the HC cost. Since the total edge weight of F is $U/2$, we obtain

$$\sum_{t=1}^{N-1} e_t = \mathrm{cost}_F(T) - U/2. \tag{18}$$

This is the size-threshold cost representation of Charikar & Chatziafratis (2017, Claim 2.1), with the total edge weight subtracted because the threshold sum starts at one.

We next use conductance to show that each e_t controls the degree sum outside the first t ranks. Fix t. If every part $P \in \mathcal{P}_t$ has volume at most $U/2$, the conductance assumption gives $w_F(P, V(F)\backslash P) \ge h\mathrm{vol}_F(P)$ for every part. Summing counts each crossing edge twice, so

$$2e_t = \sum_{P \in \mathcal{P}_t} w_F(P, V(F) \setminus P) \ge h \sum_{P \in \mathcal{P}_t} \mathrm{vol}_F(P) = hU \ge h \sum_{j=t+1}^{N} d_{(j)}.$$

Otherwise, there is a unique part P_* of volume greater than $U/2$. Every other part has volume at most $U/2$, so we apply conductance to those parts only. In this sum, an edge between two such parts is counted twice, while an edge joining one of them to P_* is counted once. The sum of their boundary weights is therefore at most $2e_t$. Also, $|P_*| \le t$, so its volume cannot exceed the sum of the t largest degrees. Consequently,

$$\begin{aligned} 2e_t &\ge \sum_{\substack{P \in \mathcal{P}_t \\ P \ne P_*}} w_F(P, V(F) \setminus P) \\ &\ge h \sum_{\substack{P \in \mathcal{P}_t \\ P \ne P_*}} \mathrm{vol}_F(P) = h\big(U - \mathrm{vol}_F(P_*)\big) \\ &\ge h \left(U - \sum_{j=1}^{t} d_{(j)} \right) = h \sum_{j=t+1}^{N} d_{(j)}. \end{aligned}$$

In both cases we have proved $2e_t \ge h \sum_{j=t+1}^{N} d_{(j)}$.

Summing these inequalities over t recovers the rank weights in the profile: the degree $d_{(j)}$ appears once for each $t < j$, hence

$$\sum_{t=1}^{N-1} \sum_{j=t+1}^{N} d_{(j)} = \sum_{j=1}^{N} (j-1) d_{(j)} = Q(d_F) - U.$$

Combining this identity with (18) gives

$$Q(d_F) \le U + \frac{2}{h}(\mathrm{cost}_F(T) - U/2) \le \frac{2}{h}\mathrm{cost}_F(T) \le (1 + 2/h)\mathrm{cost}_F(T). \tag{19}$$

For the middle inequality, the first upper bound equals $(2/h)\mathrm{cost}_F(T) + (1 - 1/h)U$, and the latter term is nonpositive because $h \le 1$ and $U \ge 0$. This proves (9).

Merging. For any nonnegative vector $d = (d_1, \dots, d_M)$ and $x \ge 0$, let $N_d(x) := |\{j : d_j > x\}|$ count the entries exceeding x. After sorting, these entries occupy precisely the first $N_d(x)$ ranks, whose weights sum to $1 + \cdots + N_d(x) = N_d(x)(N_d(x)+1)/2$. Writing each entry as a threshold integral and interchanging the finite sum with the integral gives

$$\begin{aligned} Q(d) &= \sum_{j=1}^{M} j \int_0^\infty \mathbf{1}_{\{x < d_{(j)}\}}\, dx \\ &= \int_0^\infty \sum_{j=1}^{N_d(x)} j\, dx = \int_0^\infty \frac{N_d(x)^2 + N_d(x)}{2}\, dx. \end{aligned} \tag{20}$$

Now let $d = d^{(1)} \parallel \cdots \parallel d^{(\ell)}$ and write $N_i(x) = N_{d^{(i)}}(x)$. Concatenation gives $N_d(x) = \sum_{i=1}^{\ell} N_i(x)$. For every $x \geq 0$, Cauchy–Schwarz and $\ell \geq 1$ yield

$$\frac{N_d(x)^2 + N_d(x)}{2} \leq \frac{\ell \sum_{i=1}^{\ell} N_i(x)^2 + \sum_{i=1}^{\ell} N_i(x)}{2} \leq \ell \sum_{i=1}^{\ell} \frac{N_i(x)^2 + N_i(x)}{2}.$$

Integrating and applying (20) to each vector gives

$$Q(d) \leq \ell \sum_{i=1}^{\ell} \int_0^{\infty} \frac{N_i(x)^2 + N_i(x)}{2}\, dx = \ell \sum_{i=1}^{\ell} Q(d^{(i)}),$$

which is (10).

Monotonicity and homogeneity. If d and e have the same length and $d_j \leq e_j$ for every j, then $N_d(x) \leq N_e(x)$ for every $x \geq 0$. Since $(z^2 + z)/2$ is nondecreasing for $z \geq 0$, (20) implies $Q(d) \leq Q(e)$, even if the two vectors have different sorting orders. For a scalar $c \geq 0$, the sorted entries of cd are $cd_{(j)}$. Thus $Q(cd) = \sum_j j\, cd_{(j)} = cQ(d)$, proving homogeneity. □

A.1.2 Bucket Comparison

Proof of Lemma 3.1. For each bucket put $a_B = \max_{v \in B} d_G(v)$ and $a_v = a_B$ for $v \in B$. Order $B_1, \ldots, B_r$ by nonincreasing a_B, breaking ties by a fixed rule. Write $s_j = |B_j|$, $p_j = \sum_{i \leq j} s_i$. Set $U_1 = T_{B_1}$ and $U_j = \mathrm{join}(U_{j-1}, T_{B_j})$ for $j = 2, \ldots, r$, where join combines two trees under a new root. The resulting tree U_r is a lift of the fixed bucket trees, so $I + W \leq \mathrm{cost}_G(U_r)$.

Let m_j be the internal edge weight of $G[B_j]$ and w_{ij} the edge weight between B_i, B_j. The internal cost is at most $s_j m_j$; crossing edges with $i < j$ have cost $p_j w_{ij}$. In contrast,

$$\sum_j p_j \mathrm{vol}_G(B_j) = 2 \sum_j p_j m_j + \sum_{i<j} (p_i + p_j) w_{ij}.$$

Termwise comparison therefore bounds the *sum* of internal and crossing costs by this one expression. It follows that

$$I + W \leq \mathrm{cost}_G(U_r) \leq \sum_j p_j \mathrm{vol}_G(B_j) \leq \sum_j p_j s_j a_{B_j}.$$

In the nonincreasing ordering of a, bucket B_j occupies ranks $p_j - s_j + 1, \ldots, p_j$. Their sum is $s_j(2p_j - s_j + 1)/2 \geq s_j p_j/2$, because $s_j \leq p_j + 1$. Consequently $\sum_j p_j s_j a_{B_j} \leq 2Q(a)$. Condition (7) implies $a_v \leq R d_G(v)$ for every vertex, including zero degrees, so $Q(a) \leq RQ(d_G)$. This proves (8) for arbitrary internal trees. □

A.2 The Two Structural Certificates

A.2.1 Pointwise Certificate

Proof of Theorem 3.3. Let $d^{(i)}$ be the induced degree vector on $G[C_i]$ and let $d^{[i]}$ be the restriction of the global vector to C_i. By (11) and Lemma 3.2,

$$Q(d_G) \leq \ell \sum_i Q(d^{[i]}) \leq \ell\kappa \sum_i Q(d^{(i)}) \leq \ell\kappa(1 + 2/h) \sum_i \mathrm{OPT}_{G[C_i]}.$$

Restrict T^* to C_i by deleting other leaves and suppressing unary nodes. Each internal edge's LCA can only lose leaves, so summing the restricted costs over i gives $\sum_i \mathrm{OPT}_{G[C_i]} \leq \mathrm{cost}_G(T^*)$. This proves (12). Under the bucket assumptions, Lemma 3.1 then gives (13). □

A.2.2 Aggregate Degree-Regular Certificate

The following statement isolates the structural bound used in Section 3.4 and its general bucket consequence.

Theorem A.1 (Aggregate degree-regular certificate)**.** *Let G be a $(k, \phi_{in}, \phi_{out})$-clusterable graph with positive weighted degrees and a corresponding partition $\{S_i\}_{i=1}^{k}$, where $\phi_{in} > 0$ and $\phi_{out} \leq 1/2$. Put*

$$\eta_S := \max_i \frac{\Delta_G(S_i)}{\delta_G(S_i)}. \tag{21}$$

Then

$$Q(d_G) \leq \frac{18k\eta_S}{\phi_{in}} \mathrm{cost}_G(T^*). \tag{22}$$

For every bucket partition satisfying (7)*, and every fixed choice of its internal trees,*

$$I + W \leq \frac{36Rk\eta_S}{\phi_{in}} \mathrm{cost}_G(T^*). \tag{23}$$

Proof of Theorem A.1. We first bound $|S|\mathrm{vol}_G(S)$ in terms of the optimal cost within each cluster, then combine these bounds using the merging inequality for Q. Fix a cluster S of size N. Write $U = \mathrm{vol}_{G[S]}(S)$, $U_G = \mathrm{vol}_G(S)$, and $D = \max_{v \in S} d_{G[S]}(v)$. The outer conductance assumption gives $U \geq U_G/2 > 0$, since global degrees are positive. The ratio in (21) gives

$$D \leq \Delta_G(S) \leq \eta_S \delta_G(S) \leq \frac{\eta_S U_G}{N} \leq \frac{2\eta_S U}{N}.$$

For any nonnegative vector with sum U and entries at most D,

$$Q(d) \geq \frac{U^2}{2D}. \tag{24}$$

Since the coefficient of an entry in Q increases with its rank, the minimum is obtained by filling the earliest ranks up to D before assigning mass to later ranks. Writing $U = D(q + \theta)$ with integer $q \geq 0$ and $0 \leq \theta < 1$, a minimizing vector therefore has q entries equal to D, followed by an entry θD and zeros. Its profile is

$$D\big[q(q+1)/2 + \theta(q+1)\big] \geq D(q+\theta)^2/2 = U^2/(2D).$$

Using (19) in the induced graph yields

$$\mathrm{OPT}_{G[S]} \geq \frac{\phi_{\text{in}}}{2} Q(d_{G[S]}) \geq \frac{\phi_{\text{in}} U^2}{4D} \geq \frac{\phi_{\text{in}} N U_G}{16\eta_S}.$$

Restricting T^* to the clusters and summing gives a degree-mass comparison (Laenen et al., 2023, Lemma A.8):

$$\sum_i |S_i|\mathrm{vol}_G(S_i) \leq \frac{16\eta_S}{\phi_{\text{in}}} \mathrm{cost}_G(T^*) \leq \frac{18\eta_S}{\phi_{\text{in}}} \mathrm{cost}_G(T^*). \tag{25}$$

For the restrictions $d^{[i]}$ of the global degrees, $Q(d^{[i]}) \leq |S_i|\mathrm{vol}_G(S_i)$. Thus

$$Q(d_G) \leq k \sum_i Q(d^{[i]}) \leq \frac{18k\eta_S}{\phi_{\text{in}}} \mathrm{cost}_G(T^*),$$

and Lemma 3.1 completes the proof. □

A.3 Hierarchy Guarantees

A.3.1 Degree-Ordered Hierarchy

Proof of Theorem 3.4. An edge of weight w between ranks $i < j$ contributes wj to cost and $w(i+j)$ to $Q(d_G)$. Sum over edges and apply (12). For (14), substitute $\kappa = 6k + 7$ and $h = \Omega(\lambda_{k+1}^2/k^4)$ from the spectral decomposition. Computing weighted degrees, sorting, and constructing $2n - 1$ tree nodes give the time bound, with m the number of positive-weight edges. □

The decomposition in Lemma 2.2 is the weighted result of Laenen et al. (2023, Lemma A.10). Its existence does not require their polynomial bound on the ratio of positive edge weights. In the termination argument of their Claim A.5, the number of parts increases only finitely often, and

the core sets maintained by their decomposition algorithm shrink only finitely often between such increases. Once the number of parts and the core sets are fixed, each remaining update strictly decreases the total weight of edges between parts. Since there are only finitely many partitions, these updates cannot continue indefinitely. Thus the search terminates for arbitrary positive edge weights. The weight ratio controls the running time of that search, which our construction does not perform.

A.3.2 Noisy-Degree Hierarchy

We give a cost bound for the interval ordering that also applies to the quotient methods' shared buckets. For $0 < \rho < 1$, define

$$F_\beta(\epsilon_d, \rho) := \beta + 1 + \frac{4\beta}{\epsilon_d}\left(1 + \sqrt{\frac{2\log(1/\rho)}{n}} + \frac{\log(1/\rho)}{n}\right). \tag{26}$$

Lemma A.2 (Cumulative cost of noisy intervals)**.** *Let G have $d_{\min} \geq 1$, and form noisy buckets with independent $Z_v \sim \mathrm{Lap}(2/\epsilon_d)$. Choose any HC tree within each bucket and merge these trees successively in decreasing interval order, obtaining T_{ord}. For every $0 < \rho < 1$, with probability at least $1 - \rho$, simultaneously for all choices of the internal trees,*

$$\mathrm{cost}_G(T_{\mathrm{ord}}) \leq F_\beta(\epsilon_d, \rho) Q(d_G).$$

Proof. Write the nonempty buckets in merge order as $B^{(1)}, \ldots, B^{(r)}$, put $s_j = |B^{(j)}|$ and $p_j = \sum_{a \leq j} s_a$. Let m_j be the total weight of edges inside $B^{(j)}$, and w_{ij} the weight between $B^{(i)}$ and $B^{(j)}$. Every internal edge has LCA size at most s_j, and an edge from $B^{(i)}$ to $B^{(j)}$, $i < j$, has LCA size p_j. Hence

$$\begin{aligned}\mathrm{cost}_G(T_{\mathrm{ord}}) &\leq \sum_j s_j m_j + \sum_{i<j} p_j w_{ij} \\ &\leq 2\sum_j p_j m_j + \sum_{i<j}(p_i + p_j) w_{ij} = \sum_j p_j \mathrm{vol}_G(B^{(j)}) =: S. \end{aligned} \tag{27}$$

Let $J(v)$ be the original increasing interval index, and write $z^+ = \max\{z, 0\}$ and $z^- = \max\{-z, 0\}$. Suppose $J(v) \geq J(u)$. If $u \in B_0$, then $\widehat{d}(u) < 1$, so

$$d_G(u) = \widehat{d}(u) - Z_u < 1 - Z_u \leq 1 + Z_u^- \leq \beta d_G(v) + Z_u^-.$$

The last inequality uses $d_G(v) \geq 1$. If $u \notin B_0$, both vertices lie in geometric buckets and $\widehat{d}(u) < \beta\widehat{d}(v)$, including when they share a bucket. Subtracting Z_u gives $d_G(u) < \beta d_G(v) + \beta Z_v - Z_u \leq \beta d_G(v) + \beta Z_v^+ + Z_u^-$. Thus, in both cases,

$$d_G(u) \leq \beta d_G(v) + \beta Z_v^+ + Z_u^-. \tag{28}$$

Order the vertices by true degree, $d_i = d_G(v_i)$ with $d_1 \geq \cdots \geq d_n$, breaking ties arbitrarily, and put $U = \sum_i d_i$. Since buckets are merged in decreasing interval order, the prefix containing a vertex u consists of all vertices v with $J(v) \geq J(u)$. Thus

$$S = \sum_{u \in V} d_G(u)\big|\{v \in V : J(v) \geq J(u)\}\big|.$$

The term $v = u$ contributes $d_G(u)$, so subtracting U removes these self-contributions. We can then group the remaining terms by unordered pairs of distinct vertices.

For a pair $i < j$, its contribution to $S - U$ is

$$\begin{cases} d_j, & J(v_i) > J(v_j), \\ d_i + d_j, & J(v_i) = J(v_j), \\ d_i, & J(v_i) < J(v_j). \end{cases}$$

The first case follows the true degree order. In the other two cases, (28) applies with $u = v_i$ and $v = v_j$, bounding the larger degree d_i by $\beta d_j + \beta Z_{v_j}^+ + Z_{v_i}^-$. Consequently, in all three cases the

contribution is at most $(\beta+1)d_j + \beta Z^+_{v_j} + Z^-_{v_i}$. Since $Q(d_G) = U + \sum_{i<j} d_j$, summing over all pairs gives

$$\mathrm{cost}_G(T_{\mathrm{ord}}) \le (\beta+1)Q(d_G) - \beta U + X,$$
$$X := \sum_{i=1}^{n} [\beta(i-1)Z^+_{v_i} + (n-i)Z^-_{v_i}] \le \beta(n-1)\sum_v |Z_v|. \tag{29}$$

This proves the pathwise bound (16), uniformly over the internal trees.

Set $\sigma = 2/\epsilon_d$. The variables $Y_v = |Z_v|/\sigma$ are independent exponentials of mean one. For $0 < a < 1$,

$$\log \mathbb{E} e^{a(\sum_v Y_v - n)} = n[-\log(1-a) - a] \le \frac{na^2}{2(1-a)}.$$

The Chernoff bound with $a = \sqrt{2t/n}/(1+\sqrt{2t/n})$ gives, for every $t > 0$,

$$\Pr\left[\sum_v |Z_v| > \sigma(n + \sqrt{2nt} + t)\right] \le e^{-t}. \tag{30}$$

Since $d_{\min} \ge 1$, we have $Q(d_G) \ge n(n+1)/2 \ge n^2/2$. Taking $t = \log(1/\rho)$ in (30) and substituting the resulting bound into (29) gives the claimed factor. □

Proof of Theorem 3.5. For adjacent weighted graphs,

$$\|d_G - d_{G'}\|_1 \le 2 \sum_{\{u,v\} \in \binom{V}{2}} |w_G(u,v) - w_{G'}(u,v)| \le 2.$$

Thus the degree release is ϵ-edge DP on all nonnegative weighted inputs, and all subsequent steps are post-processing. For utility, apply Lemma A.2 with $\epsilon_d = \epsilon$ and $\rho = n^{-10}$. Since $\log n/n \le 1/e$, one possible choice is

$$C_\beta = \beta + 1 + 4\beta\left(1 + \sqrt{20/e} + 10/e\right), \qquad F_\beta(\epsilon, n^{-10}) \le C_\beta(1 + 1/\epsilon).$$

The stated structural factor follows from (12). Substituting $\kappa = 6k+7$ and $h = \Omega(\lambda^2_{k+1}/k^4)$ gives the spectral bound. For each vertex with $\widehat{d}(v) \ge 1$, compute $j = \lfloor \log_\beta \widehat{d}(v) \rfloor$ and assign it to B_{j+1}; vertices with $\widehat{d}(v) < 1$ belong to B_0. Only occupied buckets are stored. In the arithmetic model, including unit-cost evaluation of these indices, computing weighted degrees and keys, sorting, and constructing a pointer tree take $O(m + n\log n)$ time on every noise outcome. At most n intervals are occupied, regardless of the magnitudes of the edge weights. □

A.3.3 Degree-Regular Consequences

Proof of Corollary 3.6. Combine (22) with the high-probability cost bound of Theorem 3.5:

$$\mathrm{cost}_G(T_{\mathrm{noisy}}) \le C_\beta(1+1/\epsilon)Q(d_G) \le \frac{18C_\beta(1+1/\epsilon)k\eta_S}{\phi_{\mathrm{in}}}\mathrm{cost}_G(T^*).$$

The probability, privacy and running time are those of Theorem 3.5. For Degree-Ordered Hierarchy, combining $\mathrm{cost}_G(T_{\mathrm{degree}}) \le Q(d_G)$ from Theorem 3.4 with (22) gives $\mathrm{cost}_G(T_{\mathrm{degree}}) \le 18k\eta_S \mathrm{cost}_G(T^*)/\phi_{\mathrm{in}}$. □

A.4 Number of Noisy Buckets

Algorithm 1 uses degree budget $\epsilon_d = \epsilon$ in Noisy-Degree Hierarchy and $\epsilon_d = \epsilon/2$ in the two-stage quotient methods introduced in Appendix B.1. The following lemma bounds the number of nonempty buckets.

Lemma A.3 (Number of noisy buckets)**.** *For a nonnegatively weighted graph and $0 < \gamma < 1$, put $L_\gamma = 2\log(n/\gamma)/\epsilon_d$ and $D = \Delta_G(V)$. With probability at least $1-\gamma$, the number of nonempty buckets is*

$$r \le \min\{n,\ 2 + \log_\beta \max\{1, D + L_\gamma\}\}.$$

In particular, $r = O(\log n)$ when $D \le n^{O(1)}$, $\epsilon_d^{-1} \le n^{O(1)}$, $\beta > 1$ is fixed, and $\gamma = n^{-c}$ for fixed $c > 0$.

Proof. The Laplace tail and a union bound give $\Pr[\max_v |Z_v| > L_\gamma] \le ne^{-\epsilon_d L_\gamma/2} = \gamma$. On the complementary event, every noisy degree is at most $D + L_\gamma$. If this is below one, only B_0 can be occupied; otherwise the largest geometric index is at most $\lfloor \log_\beta(D + L_\gamma) \rfloor$. Add the low bucket and use the deterministic bound $r \le n$. □

B Degree Information and Quotient Queries

Noisy-Degree Hierarchy merges degree buckets in a fixed order without using the edges between them. We study two private methods that use this additional edge information to choose how to merge the buckets. Both operate on a quotient graph, whose vertices represent buckets and whose edge weights record the total weight between buckets. We establish their conditional oracle guarantees and show limitations of the fixed interval ordering under the main spectral conditions, and of all degree-information algorithms on a different graph family. Appendix D.3 tests whether these methods improve utility in practice.

B.1 Quotient Methods and Oracle Guarantees

An edge-weighted graph is written $H = (U, F, w_H)$. A vertex-weighted HC instance (H, μ) additionally assigns a weight $\mu : U \to \mathbb{R}_{>0}$ to each vertex. For an HC tree τ on U, define

$$\mathrm{wcost}_{H,\mu}(\tau) := \sum_{\{u,v\} \in F} w_H(u,v) \sum_{x \in \mathrm{leaves}(\tau[u \vee v])} \mu(x), \qquad \mathrm{wopt}_{H,\mu} := \min_\tau \mathrm{wcost}_{H,\mu}(\tau).$$

Fix buckets $\mathcal{B}$ and internal trees T_B as in Section 3.1, and write $r = |\mathcal{B}|$. Define the vertex-weighted quotient by

$$H = G/\mathcal{B}, \quad w_H(B, B') = w_G(B, B') \quad (B \ne B'), \quad \mu(B) = |B|.$$

Internal edges are accounted for by I, rather than by quotient self-loops. Using the notation I, W, and $\mathcal{O}_\mathcal{B}$ from Section 3.1, we have $W = \mathrm{wopt}_{H,\mu}$. Lifting a quotient tree τ by replacing each leaf B with T_B gives

$$\mathrm{cost}_G(\mathrm{lift}(\tau)) = I + \mathrm{wcost}_{H,\mu}(\tau). \tag{31}$$

Both methods use total budget ϵ, allocating $\epsilon_d = \epsilon/2$ to Algorithm 1 and $\epsilon/2$ to the quotient stage. They use the same public-order balanced internal trees as Noisy-Degree Hierarchy. Conditioning on a first-stage output fixes the buckets, masses and public internal trees; it does not fix the private graph.

Quotient Release. For each unordered bucket pair, form $\widetilde{w}(B, B') = \max\{0, w_H(B, B') + Y_{BB'}\}$ with independent $Y_{BB'} \sim \mathrm{Lap}(2/\epsilon)$. Apply an a_r-approximate vertex-weighted HC solver to $(\widetilde{H}, \mu)$, where $a_r \ge 1$, and lift its tree. This gives a modular choice of solver. Weighted recursive sparsest cut admits a polynomial-time choice $a_r = O(\sqrt{\log \max\{r, 2\}})$ using product-demand sparsest cuts (Arora et al., 2009; Kapralov et al., 2025; Laenen et al., 2023).

Quotient Selection. Sample a rooted binary quotient tree with unordered children according to

$$p(\tau) \;\propto\; \exp\Big[-\frac{\epsilon}{4n}\mathrm{wcost}_{H,\mu}(\tau)\Big],$$

and lift it. This is the exponential mechanism (McSherry & Talwar, 2007). Both methods return the lifted tree; I and W are analysis quantities, not additional private releases. There is a unique quotient tree when $r \le 2$. Let $N_r = (2r-3)!!$ for $r \ge 3$, and $N_1 = N_2 = 1$, and define

$$\Delta_r(\delta) = \begin{cases} \dfrac{4n}{\epsilon}[\log N_r + \log(1/\delta)], & r \ge 3, \\ 0, & r \le 2. \end{cases} \tag{32}$$

Proposition B.1 (Conditional quotient oracle guarantees)**.** *Both methods are pure ϵ-edge DP on all nonnegative weighted inputs. Conditional on any first-stage output, let $q = \binom{r}{2}$ and $t = \log(1/\delta)$ for $0 < \delta < 1$. Quotient Release satisfies with conditional probability at least $1 - \delta$,*

$$\mathrm{cost}_G(T_{\mathrm{release}}) \le I + a_r W + \Lambda_r(\delta), \qquad \Lambda_r(\delta) := \frac{2(a_r+1)n}{\epsilon}(q + \sqrt{2qt} + t). \tag{33}$$

Set $\Lambda_1(\delta) = 0$. For fixed δ, this recovers the $O((a_r+1)nr^2/\epsilon)$ additive error. Quotient Selection satisfies with conditional probability at least $1-\delta$,

$$\mathrm{cost}_G(T_{\mathrm{select}}) \le I + W + \Delta_r(\delta), \tag{34}$$

and also satisfies

$$\mathbb{E}[\mathrm{cost}_G(T_{\mathrm{select}}) \mid \mathcal{B}] \le I + W + \frac{4n}{\epsilon}\log N_r. \tag{35}$$

Proof. The first-stage degree vector has sensitivity two. For every fixed release, summing changes over disjoint bucket pairs gives $\|w_H - w_{H'}\|_1 \le \|w_G - w_{G'}\|_1 \le 1$ for adjacent inputs; changes within a bucket do not affect the quotient. The quotient release stage is therefore $\epsilon/2$-DP by the Laplace mechanism. For the selection stage, every quotient-tree cost has sensitivity at most n, so each unnormalized tree weight and the normalizing constant change by a factor of at most $e^{\epsilon/4}$. Thus its conditional law is also $\epsilon/2$-DP. Adaptive composition with the first-stage degree release and lifting give pure ϵ-DP for both complete methods.

For Quotient Release, clipping cannot increase the error from a nonnegative true weight. Thus $D = n\|w_{\widetilde{H}} - w_H\|_1 \le n\sum_{\{B,B'\}}|Y_{BB'}|$, and uniformly over all trees,

$$|\mathrm{wcost}_{\widetilde{H},\mu}(\tau) - \mathrm{wcost}_{H,\mu}(\tau)| \le D.$$

Comparing the released-graph solution with a true optimum gives cost at most $a_rW + (a_r+1)D$ on H. Applying the exponential-sum bound (30) to the q absolute quotient noises, each of mean $2/\epsilon$, and using (31) prove (33). For $r = 1$ there is no quotient query.

There are N_r quotient trees: inserting a leaf into a tree on $r-1$ leaves gives $2r-3$ insertion positions: its $2r-4$ edges and a new root, starting at $N_2 = 1$. The probability that the selected quotient cost exceeds $W + t$ is at most $N_r e^{-\epsilon t/(4n)}$. Substituting (32) and adding I proves (34); for $r \le 2$ the cost is exactly $I + W$. For the expectation, put $\gamma = \epsilon/(4n)$ and let $Z = \sum_\tau e^{-\gamma \mathrm{wcost}_{H,\mu}(\tau)}$ be the normalizing constant. Writing $H(p)$ for the entropy of p, we have $\gamma\mathbb{E}_p[\mathrm{wcost}] = H(p) - \log Z$. Since $H(p) \le \log N_r$ and $\log Z \ge -\gamma W$, adding the internal cost I gives (35). □

From quotient oracles to structural bounds. For each realized degree release, keep its buckets and public internal trees fixed and form T_{ord} by the interval ordering. This is one of the lifts minimized over by the oracle, so, on every realization,

$$I + W \le \mathrm{cost}_G(T_{\mathrm{ord}}).$$

For $d_{\min} \ge 1$, apply Lemma A.2 with the actual degree budget $\epsilon_d = \epsilon/2$. Combining its event with Proposition B.1, for every $\rho, \delta \in (0,1)$, each method satisfies its respective bound with probability at least $1 - \rho - \delta$:

$$\mathrm{cost}_G(T_{\mathrm{select}}) \le F_\beta(\epsilon/2, \rho)Q(d_G) + \Delta_r(\delta), \tag{36}$$

$$\mathrm{cost}_G(T_{\mathrm{release}}) \le a_r F_\beta(\epsilon/2, \rho)Q(d_G) + \Lambda_r(\delta). \tag{37}$$

Here $I + a_rW \le a_r(I+W)$ since $a_r \ge 1$ and $I \ge 0$. A randomized solver's failure probability, if any, is added separately. Substitution of (12) or (22) gives the corresponding spectral or degree-regular guarantee. For $\rho = \delta = n^{-10}/2$, the degree-noise factor is $O_\beta(1 + 1/\epsilon)$ with overall failure at most n^{-10}. The solver factor and the second-stage errors remain; r here is the realized bucket count. Using Lemma A.3 to replace r by a deterministic bound adds its failure probability γ. Unlike these two-stage methods, the main algorithm uses the entire budget for degrees. A better conditional oracle factor therefore need not yield better utility at equal total budget.

Computation. Quotient Release can use a solver polynomial in r. For exact Quotient Selection, fix an order on the buckets and set $Z(\{B\}) = 1$. For a set S of buckets, write $\mu(S) = \sum_{B\in S}\mu(B)$. The recurrence

$$Z(S) = \sum_{\substack{\emptyset \ne A \subsetneq S \\ \min S \in A}} e^{-\epsilon\mu(S)w_H(A, S\setminus A)/(4n)} Z(A)Z(S \setminus A) \tag{38}$$

sums the weights of trees on S. The restriction on $\min S$ counts each unordered root split once. Sampling a split proportionally to its summand and recursing samples p exactly in the arithmetic model, keeping n equal to the original total mass. Enumerating $A \subseteq S \subseteq [r]$ gives $O(3^r \mathrm{poly}(r))$ time. This exponential dependence matters for the practical selector comparison in Appendix D.3.

B.2 Limits of the Fixed Interval Ordering

The next example fixes NOISY-DEGREE HIERARCHY's interval ordering and a shared half-budget degree release.

Proposition B.2 (A three-bucket separation in the spectral regime)**.** *There is a sequence of unweighted graphs G_u with $k = u^2+u$ connected components, $\lambda_k = 0$, $\lambda_{k+1} \geq 1$, and componentwise degree ratio at most four, for which the following holds. Fix $\epsilon = 1$ and $\beta = 4$.* NOISY-DEGREE HIERARCHY *and Quotient Selection share a degree release with $\epsilon_d = \epsilon/2$ and public-order balanced internal trees.* NOISY-DEGREE HIERARCHY *makes no further queries, while Quotient Selection spends another $\epsilon/2$ on its quotient stage. With probability $1 - o(1)$ as $u \to \infty$, there are exactly three buckets and*

$$\mathrm{cost}_{G_u}(T_{\mathrm{noisy}}) = \Omega(\sqrt{k})\mathcal{O}_{\mathcal{B}}, \qquad \mathrm{cost}_{G_u}(T_{\mathrm{select}}) = O(\mathrm{cost}_{G_u}(T^*)). \tag{39}$$

The construction puts small cliques in a large high-degree bucket and bipartite components across two smaller buckets. Joining the latter first avoids the loss of the interval ordering. The instance sequence aligns public labels with the clique blocks.

We use the following identity for an unweighted clique on q vertices:

$$\mathrm{cost}_{K_q}(T) = \frac{q^3 - q}{3} \quad \text{for every rooted binary tree } T. \tag{40}$$

Indeed, a root split into sizes x and y, $x + y = q$, contributes qxy; adding the two inductive expressions proves the identity. It also implies that restricting any larger tree to a clique cannot decrease its cost below this value.

Proof of Proposition B.2. Take $u = 2 \cdot 4^j$ for growing integer j. Set $a = u$, $q = 16u$, and $t = u^2$. Let G be the disjoint union of t copies of K_q and u copies of $K_{a,4a}$. The clique vertices form H, the smaller bipartite sides form A, and the larger sides form B. Their degrees are $q - 1$, $4a$, and a, respectively. Write

$$N = |H| = 16u^3, \quad M = |A| + |B| = 5u^2, \quad n = N + M, \quad E = |E_G(A, B)| = 4u^3. \tag{41}$$

For $\beta = 4$, the three degrees lie in the interiors of three adjacent geometric intervals, each at distance at least $u/2$ from either endpoint. In particular, the lowest degree $a = u = 2 \cdot 4^j$ lies in $[4^j, 4^{j+1}) = [u/2, 2u)$. With $\epsilon = 1$, sufficiently large u has $L = 48 \log n < u/2$. The Laplace tail gives $\max_v |Z_v| \leq L$ with probability at least $1 - n^{-11}$. On this event the noisy partition consists of exactly H, A, B, in that decreasing interval order.

Choose fixed-width public vertex labels so that each clique in H occupies a consecutive block. Since q and t are powers of two, each clique is a subtree of the public-order balanced tree on H. There are no edges within A or B. Therefore (40) gives the exact common internal cost

$$I = \frac{t(q^3 - q)}{3} = \frac{4096u^5 - 16u^3}{3}. \tag{42}$$

The only positive quotient edge is A–B, of weight E. Its minimum LCA mass is M, achieved by joining A, B before H. NOISY-DEGREE HIERARCHY joins H, A first and B last, so

$$\mathcal{O}_{\mathcal{B}} = I + ME, \qquad \mathrm{cost}_G(T_{\mathrm{noisy}}) = I + nE, \qquad \frac{\mathrm{cost}_G(T_{\mathrm{noisy}})}{\mathcal{O}_{\mathcal{B}}} = 1 + \frac{NE}{I + ME} = \Theta(u). \tag{43}$$

Here $ME = 20u^5$ and $NE = 64u^6$. Restricting any optimum to the cliques gives $\mathrm{cost}_G(T^*) \geq I$. Separating all components first, then using the trivial size-times-edge bound on each bipartite component, gives

$$I \leq \mathrm{cost}_G(T^*) \leq I + u\,(5u)(4u^2) = I + 20u^4.$$

Thus $\mathcal{O}_{\mathcal{B}} = O(\mathrm{cost}_G(T^*))$.

Of the three quotient trees, one has cost ME and two have cost nE. Conditional on the stable buckets, the Quotient Selection probability of the good tree is

$$\frac{1}{1 + 2\exp[-NE/(4n)]} = 1 - o(1). \tag{44}$$

Together with the degree event, (44) gives the claimed $1 - o(1)$ success probability.

There are $k = t+u = u^2+u$ components. Cliques have conductance at least $1/2$. For a half-volume set in $K_{a,4a}$, let x, y be the proportions selected from its sides; then $x+y \le 1$ and its conductance is $1-2xy/(x+y) \ge 1/2$. Each component has $\kappa = 1$, outer conductance zero and degree ratio at most four. A clique's positive normalized Laplacian eigenvalues equal $q/(q-1)$; the complete bipartite component has eigenvalues $0, 1, 2$. Hence $\lambda_k = 0$, $\lambda_{k+1} \ge 1$, and the strong spectral hypotheses hold. Since $u = \Theta(\sqrt{k})$, (43) and (44) complete the proof. □

B.3 Limits of Degree Information

Definition B.3 (Degree-information algorithm)**.** An HC algorithm is a *degree-information algorithm* if its output distribution depends on G only through the full labeled degree vector $(d_G(v))_{v \in V}$, for fixed public parameters and vertex labels. It may use arbitrary computation and independent randomness and need not output a bucket-respecting tree. Equivalently, its output laws agree on graphs with the same labeled degree vector.

Any post-processing of a noisy degree release is in this model, as is an algorithm given the exact degree vector. Querying quotient edges or selecting between trees using their private costs does not in general satisfy this restriction.

Proposition B.4 (Degree information does not determine a near-oracle tree)**.** *For every sufficiently large even r there are a finite family $\mathcal{F}_r$ of simple unweighted graphs on the same labeled set of n_r vertices, a fixed r-bucket partition $\mathcal{B}_r$, and a positive scale Γ_r with the following properties. All graphs have the same labeled degree vector and the same pointwise and aggregate certificates, with $h \ge 1/2$, $\kappa \le 2$, $\eta_S = 1$, and $\phi_{out} \le 1/2$. Their common public-order internal trees give*

$$\mathcal{O}_{\mathcal{B}_r}(G) = (1 + o(1))\Gamma_r \qquad (G \in \mathcal{F}_r).$$

For every randomized degree-information algorithm A, there exists $G \in \mathcal{F}_r$ such that

$$\mathbb{E}[\mathrm{cost}_G(A(G))] \ge \frac{r}{4}\Gamma_r, \qquad \Pr\left[\mathrm{cost}_G(A(G)) \ge \frac{r}{8}\Gamma_r\right] \ge \frac{1}{3}. \tag{45}$$

For every $G \in \mathcal{F}_r$, Quotient Selection with $\epsilon = 1$, $\beta = 4$ satisfies

$$\mathbb{E}[\mathrm{cost}_G(T_{\mathrm{select}})] \le (1 + o(1))\Gamma_r, \tag{46}$$

and its cost is at most $(1 + o(1))\Gamma_r$ with probability at least $1 - 2n_r^{-10}$. Here $r = \Theta(\log n_r)$, and all $o(1)$ terms are uniform over the family as $r \to \infty$.

An unknown perfect matching connects degree strata while preserving every labeled degree. A quotient tree can group the matched strata, whereas a degree-information algorithm must use the same output law for every matching. The construction adapts the degree-preserving hidden-matching idea of Agarwal et al. (2022, full version, Section 7.2). The separation follows from the information restriction and holds even without privacy. Its structural conditions are the stated pointwise and degree-regular bounds, with no small-λ_k assumption. This complements Proposition B.2, which fixes the interval ordering under the main spectral conditions; the privacy lower bound is given separately in Theorem 4.1.

Proof of Proposition B.4. Construction and common information. For a sufficiently large even r, take a to be the smallest power of four at least $100r^2$. Put

$$s_i = 2a4^i \ (0 \le i < r), \quad s_* = s_{r-1}, \quad M = s_*^3, \quad n_r = rM, \quad \Gamma_r = raM^2. \tag{47}$$

Each fixed vertex stratum U_i has M vertices, and its fixed internal graph is a union of M/s_i cliques K_{s_i}. The divisibilities hold because all s_i and M are powers of two. Fix public labels so that each clique is a consecutive block inside its stratum. The public-order balanced tree then contains every clique as a subtree.

Choose a perfect matching P on the r strata. For every matched pair, add an a-regular simple bipartite graph between their M-vertex sets; there are no other cross-stratum edges. One such bipartite graph joins left vertex $x \in \mathbb{Z}_M$ to right vertices $x, x+1, \ldots, x+a-1$ modulo M. Permuting its right vertex labels uniformly preserves regularity and makes each left–right vertex

pair an edge with marginal probability a/M. Take $\mathcal{F}_r$ to be the finite family obtained from all perfect matchings and these permutations. We use the distribution with a uniform P and independent uniform permutations conditional on P; the induced graph distribution need not be uniform over distinct graphs.

Every vertex in U_i has the same degree $D_i = s_i - 1 + a$, independent of P and of the permutations. Thus n, m, the labeled degree vector, the public ordering, and all the following certificates are the same across the family. Use each planted clique as a cluster. Then

$$\begin{aligned} k_r &= \sum_i M/s_i, \qquad & h &\geq 1/2, \qquad & \eta_S &= 1, \\ \kappa &= \max_i \left(1 + \frac{a}{s_i - 1}\right) \leq 2, \quad & \phi_{\text{out}} &= \max_i \frac{a}{s_i - 1 + a} \leq 1/2. \end{aligned} \tag{48}$$

A degree-information algorithm may even know the common planted cliques; this does not disclose their unknown cross-stratum pairing.

Stable buckets and the full oracle. If $a = 4^J$, each D_i lies inside $[4^{J+i}, 4^{J+i+1})$ at distance at least a from its endpoints. Since $100r^2 \leq a < 400r^2$ and $\log n_r = \Theta(r)$, we have $a > 48 \log n_r$ for all sufficiently large r. With total $\epsilon = 1$ and degree budget $1/2$, the Laplace tail gives $\max_v |Z_v| \leq 48 \log n_r$ with probability at least $1 - n_r^{-11}$. For $\beta = 4$ this event gives exactly the fixed partition $\mathcal{B}_r = \{U_0, \ldots, U_{r-1}\}$.

The common fixed internal cost and quotient optimum are

$$I = \frac{M}{3} \sum_i (s_i^2 - 1) \leq \frac{16}{45} M s_*^2 = o(\Gamma_r), \tag{49}$$

$$W = (r/2)(aM)(2M) = raM^2 = \Gamma_r. \tag{50}$$

For W, the quotient is precisely P with edge weights aM and equal masses M. Each edge must have LCA mass at least $2M$, and joining all matched pairs first attains these minima simultaneously. Since $M = s_*^3$, (49) is negligible relative to (50). Thus $\mathcal{O}_{\mathcal{B}_r} = (1 + o(1))\Gamma_r$, uniformly over the family.

Quotient Selection: utility and expectation. With $\delta = n_r^{-10}$, (32) gives

$$\frac{\Delta_r(\delta)}{\Gamma_r} = O\left(\frac{r \log r + \log n_r}{aM}\right) = o(1).$$

The degree and selection events together have probability at least $1 - 2n_r^{-10}$. For the expectation, condition on a good first-stage output and use (35). On the bad degree event, any output tree on these unweighted graphs has cost at most $n_r m \leq n_r^3/2$; its expected contribution is at most $n_r^{-7}/2$. Consequently

$$\mathbb{E}[\text{cost}_G(T_{\text{select}})] \leq I + W + 4n_r \log N_r + n_r^{-7}/2 = (1 + o(1))\Gamma_r.$$

Lower bound for degree-information algorithms. Fix such an algorithm A and its public parameters. The output law is the same across $\mathcal{F}_r$. We may therefore sample its tree T independently of the random graph just described. Each pair of vertices from different strata forms a cross edge with probability $a/[(r-1)M]$. For an n_r-leaf binary tree T, let $s_T(u, v)$ be the number of leaves in the subtree rooted at the LCA of u and v. The clique identity gives

$$\sum_{\{u,v\} \subseteq V} s_T(u, v) = \frac{n_r^3 - n_r}{3}.$$

There are $r\binom{M}{2}$ within-stratum pairs, each with LCA size at most n_r. Thus, writing c_{cross} for the cost of the cross-stratum edges only,

$$\begin{aligned} \mathbb{E}_G[c_{\text{cross}}(T)] &\geq \frac{a}{(r-1)M} \left[\frac{n_r^3 - n_r}{3} - r\binom{M}{2} n_r\right] \\ &\geq \frac{ar^2M^2(2r-3)}{6(r-1)} \geq \frac{r^2aM^2}{4} = \frac{r}{4}\Gamma_r. \end{aligned} \tag{51}$$

The omitted lower-order term is positive, and the last step holds for $r \geq 4$. Average also over A's randomness. There exists a fixed $G \in \mathcal{F}_r$ with $\mathbb{E}_A[c_{\text{cross}}(A(G))] \geq r\Gamma_r/4$. For that same G there are exactly $raM/2$ cross edges, hence $c_{\text{cross}}(T) \leq r^2aM^2/2 = r\Gamma_r/2$ for every T. If $p = \Pr_A[c_{\text{cross}}(A(G)) \geq r\Gamma_r/8]$, then

$$r\Gamma_r/4 \leq \mathbb{E}_A[c_{\text{cross}}(A(G))] \leq r\Gamma_r/8 + p(3r\Gamma_r/8),$$

so $p \geq 1/3$. Adding the nonnegative internal cost proves both parts of (45). The selected G can depend on A; the Quotient Selection bounds hold for every member, including this selected graph. □

C Proof of the Pure Edge-DP Lower Bound

We construct the packing by relating the cost of an HC tree to the number of expander components split by a balanced cut. Fix constants $d \geq 3$ and $\phi > 0$ for which there exists an unweighted d-regular ϕ-expander X on some fixed number $s \geq 100$ of vertices, and set $\epsilon_0 := 0.15\ln(2)/d$. For each k, let $n := ks$ and let $\mathcal{G}(n,k)$ be the family of labeled graphs obtained by partitioning the n vertices into k sets of size s and placing an arbitrary relabeling of X on each set. Thus, every graph in $\mathcal{G}(n,k)$ consists of k disjoint copies of X and is $(k,\phi,0)$-clusterable. Observe that each $G \in \mathcal{G}(n,k)$ has $\frac{dn}{2}$ edges. Thus, any $G_1, G_2 \in \mathcal{G}(n,k)$ differ in at most dn edges.

Definition C.1. Let $G \in \mathcal{G}(n,k)$ have expander components $C_1, \ldots, C_k$. A cut $(S, V \setminus S)$ *splits* C_i if $|C_i \cap S| \geq c|C_i|$ and $|C_i \cap (V \setminus S)| \geq c|C_i|$, where throughout this section $c := 1/100$.

Lemma C.2. *For any $G \in \mathcal{G}(n,k)$ and any optimum HC tree T^* of G,*

$$\text{cost}_G(T^*) \leq \frac{dn^2}{2k} = O\left(\frac{n^2}{k}\right). \tag{52}$$

Proof. Let $C_1, \ldots, C_k$ denote the k disjoint d-regular expander components of G. Since there are no edges between distinct components, a hierarchical clustering can first separate these components without incurring any additional cost. Taking an arbitrary HC tree T_i on each C_i and then merging these trees arbitrarily gives

$$\text{cost}_G(T^*) \leq \sum_{i\in[k]} \text{cost}_{G[C_i]}(T_i) \leq \sum_{i\in[k]} |E(C_i)||C_i| = k \cdot \frac{dn}{2k} \cdot \frac{n}{k} = \frac{dn^2}{2k}.$$

□

Lemma C.3. *Let T be an HC tree of G, and let $(A_r, V \setminus A_r)$ be the balanced cut induced by the heavy-child sequence defined below. Then*

$$\text{cost}_G(T) \geq \frac{2n}{3}|E(A_r, V \setminus A_r)|.$$

Proof. Consider a sequence of internal nodes $(A_0, A_1, \ldots, A_r)$ in T defined recursively as follows: $A_0 = V$ is the root node, and for each $i \geq 0$, A_{i+1} is the child of A_i with larger size. Let r be the smallest index such that $|A_r| \leq \frac{2n}{3}$. By the choice of A_r, we have $\frac{n}{3} \leq |A_r| \leq \frac{2n}{3}$ and hence $(A_r, V \setminus A_r)$ forms a balanced cut.

Every edge in $E(A_r, V \setminus A_r)$ has its lowest common ancestor at A_{r-1} or one of its ancestors. Since $|A_{r-1}| > \frac{2n}{3}$, we have

$$\text{cost}_G(T) \geq |E(A_r, V \setminus A_r)| \cdot \frac{2n}{3}.$$

This completes the proof. □

Lemma C.4. *For a graph $G \in \mathcal{G}(n,k)$, let T be an HC tree whose induced balanced cut is $(S, V \setminus S)$, and suppose that $(S, V \setminus S)$ splits at least αk expanders in G, for $0 < \alpha < 1$. Then, $\text{cost}_G(T) \geq \frac{2c\alpha d\phi}{3} \cdot n^2$.*

Proof. Since $(S, V \setminus S)$ splits at least αk expanders in G, we have

$$|E(S, V \setminus S)| \geq \alpha k \cdot \frac{cn}{k} \cdot d \cdot \phi = c\alpha d\phi \cdot n.$$

By Lemma C.3, we have

$$\mathrm{cost}_G(T) \geq c\alpha d\phi \cdot n \cdot \frac{2n}{3} = \frac{2c\alpha d\phi}{3} \cdot n^2.$$

□

For the remainder of the proof, set $\alpha := 1/100$ and $q := \lfloor \alpha k \rfloor$. Write $H_2(x) := -x \log_2 x - (1 - x) \log_2(1 - x)$ for the binary entropy function.

Lemma C.5. *For n sufficiently large, let G_1, G_2 be two graphs drawn independently and uniformly at random from $\mathcal{G}(n, k)$. Recall that $c = \frac{1}{100}$ and $s = n/k \geq 100$. The probability that there exists a balanced cut (A, B) which splits at most q expanders for both G_1 and G_2 is at most $2^{-0.62n}$.*

Proof. Let (A, B) be any balanced cut with $|A| = \beta n$, for $\frac{1}{3} \leq \beta \leq \frac{2}{3}$. Let $\mathcal{E}_1(A, B)$ be the event that (A, B) splits at most q expanders in G_1, and define $\mathcal{E}_2(A, B)$ analogously for G_2. The desired probability can be upper bounded by

$$\sum_{(A,B)} \Pr[\mathcal{E}_1(A, B)] \Pr[\mathcal{E}_2(A, B)].$$

We first analyze $\Pr[\mathcal{E}_1(A, B)]$. In this probability, the cut is fixed and G_1 is drawn uniformly from $\mathcal{G}(n, k)$. The family $\mathcal{G}(n, k)$ is invariant under every permutation of the labeled vertex set. Moreover, every partition of the vertex set into k parts of size s admits the same number of relabelings of X inside its parts. Consequently, the intersections of a fixed set A of size βn with the components of a uniformly random G_1 have the same distribution as the intersections of a uniformly random set A of size βn with the components of any fixed $G_1 \in \mathcal{G}(n, k)$. We may therefore use the equivalent random process in which G_1 is fixed and (A, B) is generated by choosing a uniformly random string $S \in \{0, 1\}^n$ with βn 1s. There are $\binom{n}{\beta n}$ possible strings.

For a string satisfying $\mathcal{E}_1(A, B)$, choose a set of q components containing every component that is split. There are $\binom{k}{q}$ choices for this set, and its string configurations number at most $2^{qs} \leq 2^{\alpha n}$. In each remaining component, either the number of 0s or the number of 1s is at most cs. Hence, the number of configurations in such a component is at most $2 \sum_{i=0}^{\lfloor cs \rfloor} \binom{s}{i} \leq 2^{H_2(c)s+1}$. Thus, the total number of admissible strings is at most

$$\binom{k}{q} 2^{\alpha n} \left(2^{H_2(c)s+1}\right)^{k-q} \leq 2^{H_2(\alpha)k} 2^{\alpha n} 2^{H_2(c)n} 2^k.$$

Using the fact that $2^{H_2(\beta)n - \log_2(n+1)} \leq \binom{n}{\beta n} \leq 2^{H_2(\beta)n}$, we obtain

$$\begin{aligned}
\Pr[\mathcal{E}_1(A, B)] &\leq \frac{2^{H_2(\alpha)k} 2^{\alpha n} 2^{H_2(c)n} 2^k}{2^{H_2(\beta)n - \log_2(n+1)}} \\
&\leq 2^{(\alpha + H_2(c) - H_2(\beta))n + (H_2(\alpha)+1)k + \log_2(n+1)} \\
&\leq 2^{-0.81n},
\end{aligned}$$

where the last inequality follows, for all sufficiently large n, from $H_2(\alpha) \leq 0.081$, $H_2(c) \leq 0.081$, $H_2(\beta) \geq 0.918$, and $s = n/k \geq 100$. Thus,

$$\sum_{(A,B)} \Pr[\mathcal{E}_1(A, B)] \Pr[\mathcal{E}_2(A, B)] \leq 2^n 2^{-0.81n} 2^{-0.81n} = 2^{-0.62n}.$$

□

Lemma C.6. *For n sufficiently large, there exists a family $\mathcal{F} \subseteq \mathcal{G}(n, k)$ of size $M := \lfloor 2^{0.30n} \rfloor$ such that, for any distinct $G, G' \in \mathcal{F}$, no HC tree has cost at most $\frac{2c\alpha d\phi}{3} \cdot n^2$ on both G and G'.*

Proof. Let $\mathcal{F}$ consist of M graphs generated independently and uniformly at random from $\mathcal{G}(n,k)$. For each pair of graphs $G, G' \in \mathcal{F}$, Lemma C.5 implies that, with probability at least $1 - 2^{-0.62n}$, every balanced cut splits at least $q+1 > \alpha k$ expanders in at least one of G and G'.

There are $\binom{M}{2} \le \frac{1}{2}\,2^{0.60n}$ pairs of graphs in $\mathcal{F}$. Applying a union bound over all pairs, with probability at least $1 - 2^{-0.02n-1} > 0$, every balanced cut splits more than αk expanders in at least one graph of every distinct pair $G, G' \in \mathcal{F}$. A repeated graph would admit a balanced cut formed from whole components, contradicting this property; the sampled graphs are therefore distinct. For the balanced cut of any tree, one graph in each pair has at least $q+1 > \alpha k$ split components. The edge count in Lemma C.4 is then strictly greater than $c\alpha d\phi n$, so its cost exceeds $\frac{2c\alpha d\phi}{3}n^2$. This proves the claimed disjointness. □

Proof of Theorem 4.1. Let $\mathcal{F}$ be the packing in Lemma C.6, with $M = \lfloor 2^{0.30n} \rfloor$, and put $\tau = 2c\alpha d\phi n^2/3$. Its low-cost output sets

$$B_G := \{T : \mathrm{cost}_G(T) < \tau\}, \qquad G \in \mathcal{F},$$

are pairwise disjoint. Fix any $G_0 \in \mathcal{F}$. Every G differs from G_0 by at most dn unit-edge additions or deletions, each allowed by the weighted adjacency relation. Thus group privacy gives

$$\sum_{G \in \mathcal{F}} \Pr[A(G) \in B_G] \le e^{dn\epsilon} \sum_{G \in \mathcal{F}} \Pr[A(G_0) \in B_G] \le e^{dn\epsilon}.$$

Hence there is a fixed $G \in \mathcal{F}$ for which

$$\Pr[\mathrm{cost}_G(A(G)) < \tau] \le \frac{e^{dn\epsilon}}{M} \le 2\exp[-(0.30\ln 2 - d\epsilon)n] \le 2e^{-0.15\ln(2)n},$$

where $\epsilon \le \epsilon_0 = 0.15\ln(2)/d$. By (52), $\tau \ge (4c\alpha\phi/3)k\mathrm{cost}_G(T^*)$. Thus $c_1 = 4c\alpha\phi/3$ and $c_2 = 0.15\ln 2$ prove the theorem. The input G may depend on A, ϵ and k, but is fixed before the algorithm's randomness is drawn. Since cost is nonnegative, the same event also implies the expected $\Omega(k)$ lower bound. □

D Supporting Experiments

This appendix gives further details of the utility comparison (Appendix D.1) and running-time measurements (Appendix D.2). Appendix D.3 evaluates Quotient Release and Quotient Selection, both with fixed buckets and at equal total privacy budgets.

D.1 Utility Comparison: Additional Details

For the utility experiments, the five blocks each contain 200 vertices and have within-block edge probabilities of approximately 0.034, 0.053, 0.083, 0.130, and 0.203, respectively. The between-block edge probability is 0.0025. The expected internal degree averages 20 across blocks, and each vertex has expected external degree 2. We fixed this setting after small-scale development experiments and evaluated it on independently generated graphs. Each graph receives an independent random permutation of its public vertex labels, shared by all methods. Email-Eu-core has $n = 1005, m = 16064$, including 19 isolated vertices; Political Blogs has $n = 1490, m = 16715$, including 266 isolated vertices. We use simple undirected, unweighted versions of the real-world graphs, removing self-loops and duplicate edges while retaining all vertices.

Both private methods use total $\epsilon = 1$ under single-edge addition or removal on a fixed public vertex set. Noisy-Degree Hierarchy uses $\beta = 2$ and releases the degree vector, whose sensitivity is two, with Laplace noise of scale $2/\epsilon$. The Imola-DP implementation first releases weights for all unordered vertex pairs with Laplace noise of scale $1/\epsilon$, since this vector has sensitivity one. It then constructs the hierarchy by recursively splitting the released graph using greedy modularity.

We use five independent SBM graphs and five runs of each private method per graph, giving 25 pairs. Each real-world graph has ten pairs. We pair runs by predetermined repetition indices; each method uses its own noise draws. Within each bucket, Noisy-Degree Hierarchy recursively splits the public label order into balanced halves. Its non-private version uses the same $\beta = 2$

intervals and tree construction with exact degrees, giving one deterministic output on each of the seven input graphs. AverageLinkage supplies one deterministic reference cost per graph. Reusing this reference across private runs does not add independent observations.

The exact-degree version also has a theoretical guarantee. On the unweighted experimental graphs, B_0 contains only isolates; when $d_{\min} \geq 1$, it is empty. In either case, every other bucket has distortion $R = \beta$, and interval order agrees with decreasing bucket maxima. The construction in the proof of Lemma 3.1 therefore gives cost at most $2\beta Q(d_G)$, or $4Q(d_G)$ for $\beta = 2$. Under the spectral conditions of Lemma 2.2, this gives the same approximation order as Theorem 3.4. This bucketed construction differs from the vertex-by-vertex DEGREE-ORDERED HIERARCHY in Theorem 3.4.

Table 1 reports the median of the paired private cost ratios, with 95% percentile intervals from 10,000 bootstrap resamples. For synthetic inputs, we first resample graphs and then paired runs within each selected graph. For each real-world graph, we resample paired runs on that fixed input, so the interval describes variation due to mechanism randomness. In the non-private comparison, each synthetic graph contributes one ratio between the exact-degree version and AverageLinkage. We report the median of these five ratios and their observed minimum and maximum. Each real-world graph contributes a single ratio.

We also compare NOISY-DEGREE HIERARCHY with AverageLinkage. The median cost ratios are 1.7772 [1.7672, 1.7866], 1.2210 [1.2198, 1.2230] and 1.1983 [1.1966, 1.2033] for SBM, Email and Political Blogs, respectively. The intervals use the same 95% bootstrap procedure. Table 2 reports the costs of the individual methods. The small cost gap between the noisy and exact-degree versions indicates that NOISY-DEGREE HIERARCHY achieves privacy with only a modest loss of utility on these datasets at $\epsilon = 1$.

Table 2: Dasgupta costs in thousands. Entries show medians, with observed minima and maxima in brackets. The exact-degree version and AverageLinkage each have a single value on each real-world graph.

Input	NOISY-DEGREE HIERARCHY	Imola-DP	Exact degrees	AverageLinkage
SBM	4,290.9 [4,153.8, 4,494.4]	7,202.8 [7,048.8, 7,312.9]	4,051.3 [3,944.2, 4,235.3]	2,414.7 [2,347.9, 2,482.6]
Email	6,211.6 [6,188.1, 6,233.4]	10,316.6 [10,030.0, 10,561.6]	6,175.0	5,087.4
Political Blogs	7,157.4 [7,144.3, 7,204.4]	15,650.0 [15,176.8, 16,229.2]	6,976.0	5,973.1

D.2 RUNNING-TIME MEASUREMENTS

For each $n \in \{240, 480, 960, 1920, 3840\}$, we generate five independent SBM graphs with five equal-sized blocks, giving 25 input graphs in total. Edges are sampled independently with within-block probability $20/(n/5-1)$ and between-block probability $2/(4n/5)$. Thus, the expected internal and external degrees of each vertex remain 20 and 2, respectively, as n grows. All four methods run on the same graphs, with one timed call per method and graph. Each private method uses a fixed noise seed across the graphs. Figure 1 shows the median time at each size, with error bars spanning the minimum and maximum over the five graphs. Table 3 reports the corresponding values.

We run all methods on the same Intel Xeon Platinum 8562Y+ server using Linux, Python 3.11.16, NumPy 2.4.6 and NetworkX 3.6.1, with one worker and numerical libraries restricted to one thread. We time the complete HC function, including its internal checks, with memory-allocation profiling disabled. Timing excludes graph generation and loading, process startup, output validation outside the algorithm, cost evaluation and file I/O.

For NOISY-DEGREE HIERARCHY, the measured time includes computing and releasing degrees, forming geometric buckets, building their public-order balanced trees, and merging them in interval order. For Imola-DP, it includes adding Laplace noise to all vertex-pair weights, clipping negative values, and recursively constructing the tree using greedy modularity. Practical SpecWRSC computes an NJW spectral embedding, runs KMeans with $k = 5$ and ten initializations, forms degree

Table 3: Running times in seconds over five graph instances. Entries show medians, with observed minima and maxima in brackets.

n	NOISY-DEGREE HIERARCHY	Imola-DP	practical SpecWRSC	AverageLinkage
240	0.00185 [0.00183, 0.00187]	1.169 [1.148, 1.254]	0.021 [0.015, 0.023]	0.033 [0.033, 0.033]
480	0.00356 [0.00349, 0.00360]	4.647 [4.213, 4.697]	0.029 [0.029, 0.029]	0.143 [0.142, 0.145]
960	0.00694 [0.00673, 0.00709]	18.770 [18.554, 19.373]	0.051 [0.041, 0.066]	0.656 [0.653, 0.662]
1920	0.01422 [0.01401, 0.01495]	77.117 [76.864, 77.278]	0.105 [0.101, 0.170]	2.876 [2.845, 2.897]
3840	0.03175 [0.03073, 0.06617]	332.183 [329.638, 335.547]	0.232 [0.219, 0.311]	12.669 [12.615, 12.784]

buckets with $\beta = 2$, and lifts quotient cuts to a hierarchy with public-order balanced bucket trees. Its implementation finds product-demand cuts by exhaustive search on small quotients; it does not use the theoretical ARV/SDP solver. AverageLinkage's time includes dense-heap initialization and agglomeration. Both private methods use the budgets specified in Appendix D.1.

D.3 QUOTIENT QUERIES: CONDITIONAL GAINS AND TOTAL BUDGETS

We evaluate whether the bucket-edge information used by Quotient Release and Quotient Selection (Appendix B.1) improves the hierarchy. We first compare their choices of how to merge the same buckets, then compare the complete algorithms at the same total privacy budget.

Protocols and inputs. Protocol A measures the benefit of using bucket-edge information after the buckets are formed. All methods share a degree release with $\epsilon_d = \epsilon/2$, the resulting buckets, and their public internal trees. NOISY-DEGREE HIERARCHY merges these trees in interval order without further queries; each quotient method spends the remaining $\epsilon/2$ on its quotient stage. Protocol B compares complete algorithms at equal total ϵ. NOISY-DEGREE HIERARCHY uses the full budget for degrees, while each quotient method uses its own degree release with half the budget and spends the other half on the quotient stage.

We evaluate both protocols on the full Political Blogs graph, retaining all 1490 vertices and 266 isolated vertices. Each protocol uses ten repetitions on the fixed graph for each $\epsilon \in \{0.5, 1, 2, 4\}$. All methods use $\beta = 2$.

For each paired run, we divide the quotient method's cost by the NOISY-DEGREE HIERARCHY cost. Figure 2 reports the median ratio and a 95% percentile interval for each condition, using 10,000 bootstrap resamples of paired runs on the fixed graph.

Fixed-bucket utility. With fixed buckets, neither method has a cost-ratio interval entirely below one. At $\epsilon = 1$, NOISY-DEGREE HIERARCHY already attains the oracle cost $I + W$ in eight of ten runs: this is the minimum cost obtainable by changing only the merging of the fixed bucket trees.

Equal-budget utility. At equal total budget, neither quotient method achieves lower cost than NOISY-DEGREE HIERARCHY in any of the ten runs at each tested ϵ. On this graph, the small benefit available from changing the bucket merging does not offset the cost of reducing the degree budget.

Quotient Selection does not consistently improve cost over Quotient Release on this graph, despite its stronger oracle guarantee. On Political Blogs, its exact selector takes about 0.63–0.64 seconds, compared with 0.041 seconds for Quotient Release. Our Quotient Release implementation enumerates product-demand cuts on small quotients. Quotient Selection evaluates the exact subset recurrence in floating point, with running time exponential in the number of buckets r. Both selector timings exclude computation of the oracle cost.

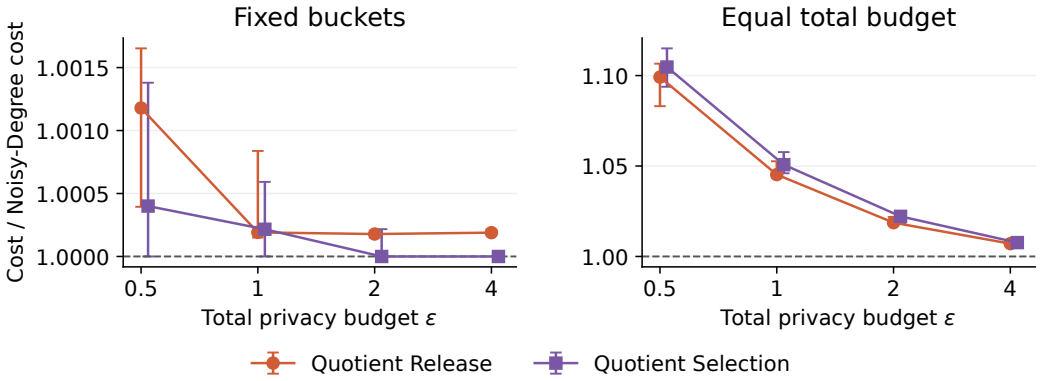


Figure 2: Quotient utility on Political Blogs with fixed buckets (left) and at equal total privacy budget (right). Points show median paired cost ratios relative to NOISY-DEGREE HIERARCHY, with 95% bootstrap intervals. The panels use different vertical scales.

These results favor NOISY-DEGREE HIERARCHY on Political Blogs: additional quotient queries add computation without improving utility at equal total privacy budget.

E COMPARISON WITH GENERAL-GRAPH GUARANTEES

We compare our degree-regular guarantee with the general-graph bound on a family consisting of two disjoint expanders with different sizes and edge weights. Write $\mathrm{OPT} = \mathrm{cost}_G(T^*)$ and fix $0 < \epsilon \leq 1$ and $\delta = n^{-c}$ for constant $c > 0$. By Imola et al. (2023, Theorem 2), their polynomial-time (ϵ, δ)-DP algorithm achieves cost $O(\sqrt{\log n})\mathrm{OPT} + \widetilde{O}(n^{5/2}/\epsilon)$ with constant probability.

Let $n = s^2$ and take disjoint d-regular expanders on s and $n - s$ vertices, with fixed degree $d \geq 3$ and inner conductance at least a constant $h > 0$. Assign weight n to every edge of the smaller component and weight one to every edge of the larger. Then $m = dn/2$, $d_{\min} = d$, $\eta_S = 1$ and $\phi_{\text{out}} = 0$. Uniform weight scaling preserves each component's normalized spectrum, so $\lambda_2 = 0$ and $\lambda_3 \geq h^2/2$; both structural guarantees apply. By Lemma 3.2, a t-vertex component with common edge weight a has optimum $\Theta(adt^2)$. Component optima add, giving

$$\mathrm{OPT} = \Theta(dns^2 + d(n-s)^2) = \Theta(n^2).$$

Substitution into the polynomial-time guarantee gives the ratio $O(\sqrt{\log n}) + \widetilde{O}(\sqrt{n}/\epsilon)$ with constant probability. On the same inputs, Corollary 3.6 with $k = 2$ gives $O(1/\epsilon)$ with probability $1 - n^{-10}$ in $O(n \log n)$ time, for fixed $\beta > 1$.