n-Batch Dynamic Adaptive Spectral Sparsification in $\ddot{O}(n \log^4 n + \epsilon^{-2} n \log^3 n)$ Time

2 November 2026

Abstract

We maintain a weighted graph on n vertices under blocks of at most n edge insertions, deletions and weight changes, each block followed by a request for a $(1 \pm \epsilon)$ spectral sparsifier, against an adaptive adversary. Using a query-driven partition of the edges, a heavy/light split of the edges and a three-step residual estimator of effective resistances – all proved here – and no expander decomposition, a block costs $\ddot{O}(n \log^4 n + \epsilon^{-2} n \log^3 n)$ time, and the output has $O(\epsilon^{-2} n \log n)$ edges. The bound comes from a block lemma: a block whose updates decrease no quadratic form of the Laplacian by more than the factor 10 costs $\ddot{O}(\epsilon^{-2} n \log^3 n)$, and an arbitrary block is simulated by $O(\log n)$ such blocks that halve the deleted weights. Removing the last logarithm reduces to sampling Gaussian vectors with covariance $\mathbf{L}(G - F)^\dagger - \mathbf{L}(G)^\dagger$ for the deleted edge set F; this step is open. Every logarithmic factor is counted; $\ddot{O}(\cdot)$ hides $(\log \log n)^{O(1)}$ factors only.

This manuscript was generated by Richard Peng and Alicia Stepin using GPT6 Astra and the Devin Harness from a manuscript on short-step interior point methods.

1 Main result

The model. A weighted undirected graph $G = (V, E, \mathbf{w})$ on a fixed set V of $n \geq 2^{20}$ vertices, with $\mathbf{w}_e \in [n^{-10}, n^{10}]$, starts empty and changes through *blocks*. A block is a sequence of at most n updates, each an insertion $\textsc{Insert}(e, \mathbf{w}_e)$, a deletion $\textsc{Delete}(e)$ or a weight change $\textsc{Weight}(e, \mathbf{w}'_e)$, followed by one query $\textsc{Sparsify}(\epsilon)$, $n^{-10} \leq \epsilon \leq \frac{1}{2}$ (so $1 + \epsilon^{-2} \leq 2\epsilon^{-2}$, and bounds are written with ϵ^{-2} alone), which must return a graph H on V with $\mathbf{L}(H) \approx_\epsilon \mathbf{L}(G)$ for the current G. The adversary is adaptive: the updates of a block may depend on every graph returned so far. The graph is connected at every query. Logarithms are to the base 2, $\log n$ and $\log m$ are identified ($m \leq n^2$), and $\ddot{O}(\cdot)$ denotes $O(\cdot)$ up to factors of $(\log \log n)^{O(1)}$; no other factor is hidden anywhere. The machine holds records linked by pointers, reals in fixed point and integers of $O(\log n)$ bits, with unit-cost comparison, addition, subtraction, multiplication, integer division and remainder; only the packed distance oracle of Appendix A uses multiplication of packed integers, and it runs on any word RAM with words of $O(\log n)$ bits.

Write G_k for the graph at the k-th query (G_0 is the empty graph) and split the edges of G_k according to the previous query:

- the *retained part* G_k^{ret} has the edges of G_k that are also edges of G_{k-1}, each with the weight $\min\{\mathbf{w}_e^{(k-1)}, \mathbf{w}_e^{(k)}\}$;
- the *new part* G_k^{new} has the edges of G_k that are not edges of G_{k-1}, with their weights, and every edge of G_k^{ret} whose weight increased, with the weight $\mathbf{w}_e^{(k)} - \mathbf{w}_e^{(k-1)}$;

so that $\mathbf{L}(G_k) = \mathbf{L}(G_k^{\text{ret}}) + \mathbf{L}(G_k^{\text{new}})$ and G_k^{new} has at most n edges. The block lemma below needs a block to be *stable*:

Definition 1.1 (stable block)**.** *Block k is* stable *if* $\mathbf{L}(G_k^{\text{ret}}) \succeq \frac{1}{10}\mathbf{L}(G_{k-1})$.

Stability restricts only the deletions and weight decreases of a block: insertions and weight increases are unconstrained. It holds when the deleted and decreased parts of a block have total leverage $\sum_e (\mathbf{w}_e^{(k-1)} - \mathbf{w}_e^{\text{ret}})\mathcal{R}^{\text{eff}}(G_{k-1})_e \le \frac{9}{10}$, and when every weight changes by at most the factor 10, the regime of interior-point methods whose weights drift by a constant factor between refreshes. It fails when a block deletes more than nine tenths of the weight of some cut.

Lemma 1.2 (stable blocks)**.** *If every block is stable, there is a randomized data structure that answers every query* SPARSIFY(ϵ) *of the model with a graph H of at most $10^{12}\epsilon^{-2}n\log n$ edges satisfying $\mathbf{L}(H) \approx_\epsilon \mathbf{L}(G)$, and whose total running time over Q blocks with U updates is*

$$\ddot{O}\Big(U\log^3 n + n\log^3 n \sum_{k\le Q} \epsilon_k^{-2}\Big),$$

with probability $1 - Qn^{-10}$, against an adaptive adversary. A block of n updates and one query with constant ϵ costs $\ddot{O}(n\log^3 n)$.

The three parts of the bound are the maintenance of the partition structure of Lemma 3.2 under the updates ($O(\log^3 n)$ amortized per update), the refresh of the resistance estimate at every query ($\ddot{O}(n\log^3 n)$: the $\log^3 n$ heaviest edges at every vertex, scanned once, and $10^3\log n$ partition queries of $O(n\log^2 n)$ each), and the output itself ($O(\epsilon^{-2}n\log^2 n)$ proposals from the residual route, $O(\log n)$ each).

Arbitrary updates. Stability can be dispensed with at the price of one logarithm: a block whose deletions and decreases remove the weights δ_e, $e \in F$, is simulated by $30\log n + 1$ internal blocks, the i-th leaving the weight $2^{-i}\delta_e$ in place, which are stable with the constant 2, and the last of which removes the remaining $n^{-30}\delta_e$ and is stable because $\sum_{e\in F}\delta_e\mathcal{R}^{\text{eff}}(G_k)_e \le n^{22}$. The structure of Lemma 1.2 run with this schedule gives the main theorem, proved in Section 10.

Theorem 1.3 (main)**.** *Without any hypothesis on the updates, there is a randomized data structure that answers every query* SPARSIFY(ϵ) *with a graph H of at most $10^{12}\epsilon^{-2}n\log n$ edges satisfying $\mathbf{L}(H) \approx_\epsilon \mathbf{L}(G)$, and its total running time over Q blocks with U updates is*

$$\ddot{O}\Big(U\log^4 n + Qn\log^4 n + n\log^3 n \sum_{k\le Q} \epsilon_k^{-2}\Big),$$

with probability $1 - Qn^{-9}$, against an adaptive adversary. A block of n updates and one query with constant ϵ costs $\ddot{O}(n\log^4 n)$.

Where the logarithms come from. Every refresh rebuilds the effective resistance estimate from scratch with fresh randomness, so the adversary learns nothing it can exploit about the randomness of later blocks; the only information carried from one block to the next is a public estimate $\widehat{\mathcal{R}}^{\text{eff}}$ of the effective resistances of the previous graph, which overestimates them within a constant factor (Lemma 7.2). Stability makes $10\widehat{\mathcal{R}}^{\text{eff}}$ an upper bound on the resistances of the retained edges in the new graph (Lemma 8.1), the new edges are kept exactly, and sampling with these intensities gives the sparse references of the new refresh. The estimate is the following: for the

$\theta_{\deg}$ heaviest edges at every vertex it is read off $10^3 \log n$ random projections of a sparse reference, for the remaining light edges it is a proxy at the two endpoints plus a residual whose total energy is $O(n/\theta_{\deg} + n/\log^3 n)$. The light edges with a large residual are found by ℓ_2 queries to the query-driven partition with the residual vectors, whose certificate costs a factor $\log^4 n$ over the energy; this is where $\theta_{\deg} = \log^3 n$ is forced, and the three terms $n\theta_{\deg}$, $n \log^6 n/\theta_{\deg}$ and $n \log^3 n$ of the block cost meet.

Comparison with previous results. The previous bounds, each converted to the cost of n updates followed by one query, are the following.

- Abraham, Durfee, Koutis, Krinninger and Peng [ADK+16] maintain a $(1 \pm \epsilon)$-spectral sparsifier against an oblivious adversary with amortized update time $\mathrm{poly}(\log n, \epsilon^{-1})$; traced through their theorem with $\log \rho = O(\log n)$ for the density $\rho = m/n$, the update time is $O(\epsilon^{-2} \log^9 n)$, so n updates and one query cost $O(\epsilon^{-2} n \log^9 n)$ (the exponent 9 is our reading of the proof; the source states $\mathrm{poly}(\log n, \epsilon^{-1})$). The adversary is oblivious.

- Bernstein, van den Brand, Probst Gutenberg, Nanongkai, Saranurak, Sidford and Sun [BvdBP+22] maintain, against an adaptive adversary, a $\log^{O(1)} n$-approximate spectral sparsifier in $\log^{O(1)} n$ amortized update time, and an $O(k)$-approximate cut sparsifier in $\widetilde{O}(n^{1/k})$ amortized update time; for n updates this gives $n \log^{O(1)} n$, with the exponent unresolved because the source states its bounds with $\widetilde{O}$ and $\log^{O(1)} n$. The query is more limited: the approximation factor is polylogarithmic, not $1 \pm \epsilon$.

- Beimel, Kaplan, Mansour, Nissim, Saranurak and Stemmer [BKM+22] (Lemma 4.11 and the discussion following it) refresh a $(1 + \epsilon)$-spectral sparsifier on top of a dynamic expander decomposition with conductance $\phi = O(1/\log^4 m)$, maintained in $O(\phi^{-2} \log^7 n)$ amortized time per update (their Theorem 4.10, from [BvdBP+22]), in $\widetilde{O}(n + t)$ time per t edge changes of the expanders; per n updates this is $n \log^c n$ with $c \geq 15$ from the decomposition alone and the rest unresolved ($\widetilde{O}$ of the source). The query is more limited: the sparsifier is accurate only if the updates between two refreshes are fixed when the refresh is called (their "blinking adversary"), whereas here the adversary chooses every update after seeing every previous output.

- The leverage score maintenance of van den Brand, Lee, Nanongkai, Peng, Saranurak, Sidford, Song and Wang [vdBLN+20] (Theorem B.1 there) is adaptive but works on a fixed edge set whose weights drift by bounded amounts, with a total time stated with $\widetilde{O}$; it does not support insertions or deletions, so it gives no bound in this regime.

- The batch-dynamic literature processes batches of updates for connectivity, dynamic trees, k-cores and the massively parallel model [AABD19, TDB19, ILMP19, DDK+20, AAB+20, NO21, LSY+22]; none of these maintains a spectral sparsifier, so they give no bound in this regime.

Organization. Section 2 fixes notation and states the tools: the Laplacian solvers, the heavy/light split, the query-driven partition and the packed distance oracle, each proved in place. Section 6 describes the maintained state. Section 7 is the refresh of one block: references, random vectors, public estimate. Section 8 carries the estimate across a block and samples the next references and the output. Section 9 assembles the block procedure and proves Lemma 1.2. Section 10 removes the stability hypothesis by the halving schedule, proves Theorem 1.3, and records what stands between the result and one logarithm less. Appendix A and Appendix B prove the packed distance oracle and the solver statement; the table of symbols is the last appendix.

2 Preliminaries

Graphs and matrices. For $G = (V, E, \mathbf{w})$, $\mathbf{B} \in \mathbb{R}^{E \times V}$ is the signed incidence matrix, $\boldsymbol{\chi}_e = \mathbf{B}^\top \mathbb{1}_e \in \mathbb{R}^V$ the signed indicator of e, $\mathbf{L}(G) = \mathbf{B}^\top \textsc{Diag}(\mathbf{w})\mathbf{B} = \sum_e \mathbf{w}_e \boldsymbol{\chi}_e \boldsymbol{\chi}_e^\top$ the Laplacian, $\mathbf{d}(G)_v$ the weighted degree, $\mathbf{D}(G) = \textsc{Diag}(\mathbf{d}(G))$ and $\mathbf{A}(G) = \mathbf{D}(G) - \mathbf{L}(G)$ the adjacency matrix. $\mathbf{L}(G)^\dagger$ is the pseudoinverse, $\mathbf{\Pi} = \mathbf{I} - \frac{1}{n}\mathbb{1}\mathbb{1}^\top$ the projection onto $\mathbb{1}^\perp$, and $\mathcal{R}^{\text{eff}}(G)_e \coloneqq \boldsymbol{\chi}_e^\top \mathbf{L}(G)^\dagger \boldsymbol{\chi}_e$ the effective resistance, with $\sum_e \mathbf{w}_e \mathcal{R}^{\text{eff}}(G)_e = n - 1$ for connected G. For a vector $\mathbf{x} \in \mathbb{R}^V$ and a set of edges S, $\mathcal{E}_2(\mathbf{x}, S) \coloneqq \sum_{e=uv \in S} \mathbf{w}_e (\mathbf{x}_u - \mathbf{x}_v)^2$ is the energy. For positive semidefinite $\mathbf{A}, \mathbf{B}$ and $\epsilon \in (0, 1)$, $\mathbf{A} \approx_\epsilon \mathbf{B}$ means $(1-\epsilon)\mathbf{B} \preceq \mathbf{A} \preceq (1+\epsilon)\mathbf{B}$, and $a \approx_c b$ for scalars means $b/c \le a \le cb$. If $\mathbf{A} \succeq \frac{1}{c}\mathbf{B}$ and $\mathbf{A} \preceq \mathbf{B}$ then $\ker \mathbf{A} = \ker \mathbf{B}$ and $\mathbf{A}^\dagger \preceq c\mathbf{B}^\dagger$; in particular, by Rayleigh monotonicity, adding edges to a connected graph does not increase any effective resistance.

Solvers. Lemma B.1 (Appendix B) gives the routine LsolverKLPSS of Kyng, Lee, Peng, Sachdeva and Spielman [KLP+16]: on an SDDM matrix $\mathbf{M}$ with n rows, m' nonzero entries and polynomially bounded entries, and an error $\epsilon \in (0, \frac{1}{2}]$, it returns in $O(m' \log n) + O(n \log^2 n)$ time implicit access to a symmetric matrix $\mathbf{Z}(\mathbf{M})$ with $\mathbf{Z}(\mathbf{M}) \approx_\epsilon \mathbf{M}^\dagger$ on the range of $\mathbf{M}$, each application of which takes $O((m' + n \log n) \log(1/\epsilon))$ time. The operator $\mathbf{Z}(\mathbf{M})$ is a fixed matrix once built, so all its applications are consistent; this is what makes the random vectors below Gaussian with a fixed covariance. The source states its failure probability as "with high probability" without an exponent; we write it as n^{-14} throughout, and the exponent is the one quantity of this paper that rests on an unquantified statement of a source (Appendix B).

Sampling.

Lemma 2.1 ([SS11]; Theorem 1.1 of [Tro12])**.** *Let $G = (V, E, \mathbf{w})$ be connected, $\epsilon \in (0, \frac{1}{2}]$, and let $p_e \in (0, 1]$ satisfy $p_e \ge \min\{1,\ 10^3 \epsilon^{-2} \log n \cdot \mathbf{w}_e \mathcal{R}^{\text{eff}}(G)_e\}$ for every edge. Retain every edge independently with probability p_e, with weight $\mathbf{w}_e / p_e$ when retained. The retained graph H satisfies $\mathbf{L}(H) \approx_\epsilon \mathbf{L}(G)$ with probability $1 - n^{-12}$, and has at most $2\sum_e p_e$ edges with probability $1 - \exp(-\frac{1}{3}\sum_e p_e)$ when $\sum_e p_e \ge 10^3 \log n$.*

The first statement is the matrix Chernoff bound of [Tro12] applied to the centered rank-one matrices $(\mathbf{w}_e/p_e)\mathbf{L}(G)^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top\mathbf{L}(G)^{\dagger/2}$, as in [SS11]; the second is the multiplicative Chernoff bound for independent indicators.

Heavy and light edges. Definition 4.1 (Section 4) splits the edges of G into the heavy graph G^{heavy} of the $\theta_{\deg}$ heaviest edges at every vertex and the light graph G^{light}, and sets $\mathbf{M} \coloneqq \mathbf{L}(G^{\text{heavy}}) + \mathbf{D}(G^{\text{light}})$; Lemma 4.2 collects the facts used about this split, in particular $\mathbf{L}(G) \preceq 2\mathbf{M}$, $\boldsymbol{\chi}_e^\top \mathbf{M}^\dagger \boldsymbol{\chi}_e \le 2\mathcal{R}^{\text{eff}}(G)_e$ and $\textsc{Tr}[\mathbf{M}^\dagger \mathbf{A}(G^{\text{light}}) \mathbf{M}^\dagger \mathbf{A}(G^{\text{light}})] \le 32n/\theta_{\deg}$.

The packed distance oracle. Lemma A.1 (Appendix A) returns, for the rows of a matrix $\mathbf{X} \in \mathbb{R}^{n \times 10^3 \log n}$ whose nonzero pairwise distances lie in $[n^{-c}, n^c]$, the squared distance of two rows within the factor 10^3 in $\ddot{O}(1)$ worst-case time after $O(n \log^2 n)$ randomized preprocessing, for all pairs simultaneously with probability $1 - n^{-14}$; its answers are a fixed function of the preprocessing, so adaptive queries are allowed.

3 The query-driven partition

This section builds the data structure QueryDrivenPartition whose interface is used throughout the paper. It maintains a weighted graph G on the fixed vertex set V, $|V| = n$, under edge insertions

and deletions, with at most one edge between any two vertices at any time and all weights in $[n^{-10}, n^{10}]$. The edge set is kept partitioned into clusters, and a query with a vector $\mathbf{x} \in \mathbb{R}^V$ refines the partition until the quadratic form of $\mathbf{x}$ over the edges of G is certified by distances of the vertices to one center per cluster. The structure uses only pointers, comparisons and arithmetic on stored reals, each at unit cost in the model of Section 1. The two tools are the binary logarithmic method of Bentley and Saxe [BS80], which keeps the number of clusters at a vertex logarithmic, and the sweep-cut form of Cheeger's inequality [Chu97], which turns a large quadratic form into a low-conductance split.

Definition 3.1. *Let G be a weighted graph on V with weights in $[n^{-10}, n^{10}]$. A* cluster *C is a nonempty set of edges of G whose weights lie in one interval $[\mathbf{w}_C, 2\mathbf{w}_C)$ with $\mathbf{w}_C$ a power of two; $\mathbf{w}_C$ is the* weight *of C. Its vertex set $V(C)$ consists of the endpoints of its edges, the* degree $\deg_C(v)$ *is the number of edges of C at v, and the* volume $\textsc{Vol}(C) = \sum_{v \in V(C)} \deg_C(v) = 2|C|$*; for $S \subseteq V(C)$, $\textsc{Vol}_C(S) = \sum_{v \in S} \deg_C(v)$ and $\textsc{Cut}_C(S)$ is the number of edges of C with exactly one endpoint in S. A* partition *$\mathcal{C}$ of G is a set of pairwise edge-disjoint clusters whose union is the edge set of G; its* copies *are the pairs (C, v) with $C \in \mathcal{C}$ and $v \in V(C)$. For a vector $\mathbf{x} \in \mathbb{R}^V$ and a cluster C the* weighted median $\mathrm{median}_{\mathbf{x}}(C)$ *is the smallest value x_u, $u \in V(C)$, with $\textsc{Vol}_C(\{v \in V(C) : x_v \le x_u\}) \ge \textsc{Vol}(C)/2$; the vertices with $x_v < \mathrm{median}_{\mathbf{x}}(C)$ and those with $x_v > \mathrm{median}_{\mathbf{x}}(C)$ then each have volume at most $\textsc{Vol}(C)/2$. Each cluster carries a* center $\mathrm{center}_{\mathbf{x}}(C) \in \mathbb{R}$ *for every vector $\mathbf{x}$ queried since the last update, defined by the split recursion: when the query of $\mathbf{x}$ returns, $\mathrm{center}_{\mathbf{x}}(C) = \mathrm{median}_{\mathbf{x}}(C)$ for the clusters present at that moment, and whenever a cluster C is later split into two clusters both children receive the value $\mathrm{center}_{\mathbf{x}}(C)$. The* vertex energy *of a copy with respect to $\mathbf{x}$ is*

$$\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, C, v) = \mathbf{w}_C \left(x_v - \mathrm{center}_{\mathbf{x}}(C)\right)^2, \qquad \mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, \mathcal{C}) = \sum_{C \in \mathcal{C}} \sum_{v \in V(C)} \deg_C(v)\, \mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, C, v),$$

and the cluster energy *of the partition is*

$$\mathcal{E}_2(\mathbf{x}, \mathcal{C}) = \sum_{C \in \mathcal{C}} \mathbf{w}_C \sum_{uv \in C} (x_u - x_v)^2 \le \mathcal{E}_2(\mathbf{x}, G).$$

State of QueryDrivenPartition (vertex set V, weights in $[n^{-10}, n^{10}]$).

- The live edges of G with their weights, one record per edge, with pointers to its two endpoints, to its cluster, to its block, and to its nodes in the lists and trees of the two copies at its endpoints.
- For every integer i with $-10 \log n \le i \le 10 \log n$ the *weight class* i, the edges of weight in $[2^i, 2^{i+1})$, kept as at most one *block* of each *rank* $r \ge 0$; a block of rank r was formed from exactly 2^r insertions and holds the live edges among them, partitioned into clusters.
- For every cluster C: its weight $\mathbf{w}_C$, its block, a doubly linked list of its edges, a doubly linked list of its copies (C, v), and the list of its centers $\text{center}_{\mathbf{x}}(C)$ for the vectors $\mathbf{x}$ queried since the last update, each stored with the number of updates performed so far (a center stored before the last update is treated as absent).
- For every copy (C, v): the degree $\deg_C(v)$, a pointer to v and to C, its node in the global doubly linked list of all copies, a doubly linked list of the edges of C at v, and a balanced binary search tree over these edges whose nodes store subtree sizes.
- For every vertex v: the list of its copies, and the values x_v of the vectors $\mathbf{x}$ queried since the last update, each stored with the number of updates performed so far.
- During a query of $\mathbf{x}$ only: for every cluster C a balanced binary search tree T_C over $V(C)$ keyed by x_v, whose nodes store the sums of $\deg_C(v)$, of $\deg_C(v)\, x_v$ and of $\deg_C(v)\, x_v^2$ over their subtrees; the value $\text{median}_{\mathbf{x}}(C)$; two heaps H_+ and H_- over the clusters; and the running sums $\mathcal{E}_2^{\text{vert}}(\mathbf{x}, \mathcal{C})$ and, for each sign $s \in \{+1, -1\}$, $\sum_C \mathbf{w}_C \sum_{v : s(x_v - \text{median}_{\mathbf{x}}(C)) > 0} \deg_C(v)(x_v - \text{median}_{\mathbf{x}}(C))^2$.

Accessors (each in $O(1)$ per item returned): the list of all copies; for a copy (C, v) the values $\mathbf{w}_C$, $\deg_C(v)$, the pairs $(\text{center}_{\mathbf{x}}(C), \mathcal{E}_2^{\text{vert}}(\mathbf{x}, C, v))$ over the vectors $\mathbf{x}$ queried since the last update, in query order; the list of edges of C at v; and, given an integer $k \in [1, \deg_C(v)]$, the k-th edge of C at v in the order of the tree of (C, v), in $O(\log n)$.

Figure 1: The state of the query-driven partition. Its guarantees are in Lemma 3.2.

Insert$(e = uv, \mathbf{w}_e)$. **Input:** a vertex pair uv not joined by a live edge, a weight $\mathbf{w}_e \in [n^{-10}, n^{10}]$. **Output:** none; G gains the edge.

1. Increase the update counter.
2. Let i be the integer with $2^i \le \mathbf{w}_e < 2^{i+1}$. Form a block of rank 0 in class i holding the single cluster $C = \{e\}$ with $\mathbf{w}_C = 2^i$, $V(C) = \{u, v\}$, degrees 1, and the trees of (C, u) and (C, v) holding e.
3. While class i holds two blocks of the same rank r: merge them into one block of rank $r + 1$ whose live edges form one cluster C with $\mathbf{w}_C = 2^i$; build the lists of C, the copies (C, v) with their degrees, and the tree of each copy over the edges of C at v; discard the clusters of the two old blocks.

Delete(e). **Input:** a live edge $e = uv$. **Output:** none; G loses the edge.

1. Increase the update counter.
2. Let C be the cluster of e. Remove e from the list of C, from the trees of (C, u) and (C, v), and from the live edges; decrease $\deg_C(u)$ and $\deg_C(v)$ by one.
3. Remove a copy whose degree became 0 from C and from the global list of copies; remove C from its block if it has no edge left.

Figure 2: The updates of the query-driven partition. Its guarantees are in Lemma 3.2.

PartitionTwo$(\mathbf{x};\widetilde{\mathcal{E}})$. **Input:** $\mathbf{x}\in\mathbb{R}^V$, and $\widetilde{\mathcal{E}}$ with $\widetilde{\mathcal{E}}\approx_{10}\mathcal{E}_2(\mathbf{x},G)$. **Output:** the set E^{cut} of edges removed from G by this query.

1. Store x_v at every vertex v with the current update counter. Sort V by x_v; for every cluster C build T_C from the sorted order, compute $\mathrm{median}_{\mathbf{x}}(C)$ by descending T_C on the degree sums, and compute the energies of C from the sums at the root of T_C and at the median. For every cluster C and each sign $s\in\{+1,-1\}$ insert into the heap H_s the vertex of C nearest to $\mathrm{median}_{\mathbf{x}}(C)$ among those with $s(x_v-\mathrm{median}_{\mathbf{x}}(C))>0$, if any, keyed by $\mathbf{w}_C(x_v-\mathrm{median}_{\mathbf{x}}(C))^2$. Set $E^{\mathrm{cut}}=\emptyset$.
2. While $\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x},\mathcal{C})>10^8\log^4 n\cdot\widetilde{\mathcal{E}}$ (with $\mathrm{center}_{\mathbf{x}}(C)=\mathrm{median}_{\mathbf{x}}(C)$):
   (a) Let $s\in\{+1,-1\}$ be a sign for which the vertices with $s(x_v-\mathrm{median}_{\mathbf{x}}(C))>0$ carry at least half of $\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x},\mathcal{C})$. Set $\mathrm{Vol}=0$ and $\mathrm{Bd}=0$.
   (b) Repeat: extract the maximum (C,v) of H_s, mark (C,v) as selected, add $\deg_C(v)$ to Vol, and for every edge vu of C at v add 1 to Bd if (C,u) is not selected and subtract 1 otherwise; insert into H_s the next vertex of C on the side s of the median in the order of T_C, if any. Stop as soon as $\mathrm{Bd}\le 10^{-3}\mathrm{Vol}/\log^2 n$.
   (c) For every cluster C with a selected vertex, let S_C be its selected vertices. Add the edges of C with exactly one endpoint in S_C to E^{cut} and remove them from G as in **Delete** (without changing the update counter). Move the copies (C,v), $v\in S_C$, with their trees to a new cluster C' of weight $\mathbf{w}_C$ in the block of C, remove the vertices of S_C from T_C and insert them into a new tree $T_{C'}$, give C' the centers of C for the earlier vectors, recompute the medians, the energies and the entries in H_+ and H_- of C and C', and discard a cluster left without edges. Clear the selected marks.
3. Set $\mathrm{center}_{\mathbf{x}}(C)=\mathrm{median}_{\mathbf{x}}(C)$ for every cluster C, stored with the current update counter, discard the trees T_C, and return E^{cut}.

Figure 3: The query of the query-driven partition. Its guarantees are in Lemma 3.2.

Lemma 3.2. *There is a deterministic data structure* QueryDrivenPartition *(Figures 1 to 3) that maintains a graph G on V with at most one edge between any two vertices and weights in $[n^{-10},n^{10}]$, starting from the empty graph, together with a partition $\mathcal{C}$ of G into clusters in the sense of Definition 3.1, under the operations*

- Insert$(e,\mathbf{w}_e)$ *and* Delete(e)*, the* updates*;*
- PartitionTwo$(\mathbf{x};\widetilde{\mathcal{E}})$ *for $\mathbf{x}\in\mathbb{R}^V$ and $\widetilde{\mathcal{E}}\approx_{10}\mathcal{E}_2(\mathbf{x},G)$, the* queries*, each returning a set E^{cut} of edges that it removes from G.*

Over any sequence of at most n^{10} operations:

- *there are at most $10^3 n\log^2 n$ copies at all times;*
- *the sets E^{cut} returned over the sequence contain in total at most 10^{-1} times the number of updates in the sequence edges;*
- *between two consecutive updates the clusters only split, and for every vector $\mathbf{x}$ queried since the last update every cluster C has a center $\mathrm{center}_{\mathbf{x}}(C)$ as in Definition 3.1 (set by the query*

of $\mathbf{x}$*, inherited by both halves at every later split) with*

$$\sum_{C\in\mathcal{C}}\sum_{v\in V(C)}\deg_C(v)\,\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x},C,v)\le 10^8\log^4 n\cdot\widetilde{\mathcal{E}},$$

where $\widetilde{\mathcal{E}}$ *is the estimate passed with* $\mathbf{x}$*;*

- *the structure returns in* $O(1)$ *per item the list of all copies and, for a copy* (C,v)*, the values* $\mathbf{w}_C$ *and* $\deg_C(v)$*, the list of the edges of* C *at* v*, and the pairs* $(\mathrm{center}_{\mathbf{x}}(C),\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x},C,v))$ *over the vectors* $\mathbf{x}$ *queried since the last update in query order; given an integer* $k\in[1,\deg_C(v)]$ *it returns the* k*-th edge of* C *at* v *in a fixed order in* $O(\log n)$*, so one uniformly random integer* k *yields a uniformly random live edge of* C *at* v *in* $O(\log n)$*;*
- *a sequence of* U *updates and* Q *queries takes* $O(U\log^3 n+Qn\log^2 n)$ *time in total, that is, a query costs amortized* $O(n\log^2 n)$ *and an update amortized* $O(\log^3 n)$*.*

For an edge $e=uv$ in a cluster C of a query vector $\mathbf{x}$, $\mathbf{w}_e<2\mathbf{w}_C$ and $(x_u-x_v)^2\le 2(x_u-\mathrm{center}_{\mathbf{x}}(C))^2+2(x_v-\mathrm{center}_{\mathbf{x}}(C))^2$ give the *envelope*

$$\mathbf{w}_e(x_u-x_v)^2\le 4\left(\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x},C,u)+\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x},C,v)\right). \tag{1}$$

The proof of Lemma 3.2 occupies the rest of this section. The first tool is the sweep-cut form of Cheeger's inequality; we state and prove the version we need, for a multigraph with unit edge weights and a nonnegative vector, following the standard argument [Chu97].

Lemma 3.3. *Let* H *be a multigraph with unit edge weights on a vertex set* W*, with degrees* $d(v)$*, and let* $\mathbf{y}\in\mathbb{R}^W$ *be nonnegative and not identically zero. There is a threshold* $\theta>0$ *such that* $S=\{v\in W:y_v>\theta\}$ *is nonempty and the number of edges of* H *with exactly one endpoint in* S *is at most*

$$\sqrt{\frac{2\sum_{uv\in H}(y_u-y_v)^2}{\sum_{v\in W}d(v)y_v^2}}\cdot\sum_{v\in S}d(v).$$

Proof. Write ρ for the square root in the statement, and for $t\ge 0$ let $S_t=\{v\in W:y_v^2>t\}$, with $c(t)$ edges of H having exactly one endpoint in S_t and $\mathrm{vol}(S_t)=\sum_{v\in S_t}d(v)$. An edge uv with $y_u^2>y_v^2$ has exactly one endpoint in S_t precisely for $t\in[y_v^2,y_u^2)$, and a vertex v lies in S_t precisely for $t\in[0,y_v^2)$, so

$$\int_0^\infty c(t)\,dt=\sum_{uv\in H}\left|y_u^2-y_v^2\right|,\qquad\int_0^\infty\mathrm{vol}(S_t)\,dt=\sum_{v\in W}d(v)y_v^2.$$

By the Cauchy–Schwarz inequality and $(y_u+y_v)^2\le 2y_u^2+2y_v^2$,

$$\sum_{uv\in H}\left|y_u^2-y_v^2\right|=\sum_{uv\in H}|y_u-y_v|(y_u+y_v)\le\sqrt{\sum_{uv\in H}(y_u-y_v)^2}\sqrt{2\sum_{v\in W}d(v)y_v^2}=\rho\sum_{v\in W}d(v)y_v^2.$$

Let $M=\max_v y_v^2>0$. Suppose that $c(t)>\rho\,\mathrm{vol}(S_t)$ for every $t\in[0,M)$. Both c and $\mathrm{vol}(S_t)$ are step functions of t with finitely many steps that vanish for $t\ge M$, so integrating over $[0,M)$ gives $\int_0^\infty c(t)\,dt>\rho\int_0^\infty\mathrm{vol}(S_t)\,dt$, contradicting the two displays. Hence some $t\in[0,M)$ has $c(t)\le\rho\,\mathrm{vol}(S_t)$; the set S_t is nonempty because $t<M$. If $t>0$ take $\theta=\sqrt{t}$; if $t=0$ take θ to be half the smallest positive value of $\mathbf{y}$. In both cases $S=\{v:y_v>\theta\}=S_t$, because $\mathbf{y}\ge 0$. □

The second tool is the guarantee of one sweep, Steps 2(a)–(b) of PartitionTwo.

Lemma 3.4. *Let $\mathcal{C}$ be the current partition and $\mathbf{x} \in \mathbb{R}^V$ a vector with $\mathrm{center}_{\mathbf{x}}(C) = \mathrm{median}_{\mathbf{x}}(C)$ for all $C \in \mathcal{C}$ and $\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, \mathcal{C}) > 10^7 \log^4 n \cdot \mathcal{E}_2(\mathbf{x}, \mathcal{C})$. Then Steps 2(a)–(b) of* PartitionTwo *terminate, and the sets S_C of selected vertices satisfy*

- $s(x_v - \mathrm{median}_{\mathbf{x}}(C)) > 0$ *for every $v \in S_C$, hence $\mathrm{Vol}_C(S_C) \le \mathrm{Vol}(C)/2$ for every C;*
- $\sum_{C \in \mathcal{C}} \mathrm{Cut}_C(S_C) \le 10^{-3} \log^{-2} n \sum_{C \in \mathcal{C}} \mathrm{Vol}_C(S_C)$;
- $\sum_{C \in \mathcal{C}} \mathrm{Vol}_C(S_C) \ge 1$.

If H_s holds, for every cluster, its copy nearest to the median on the side s when the sweep starts, the sweep takes $O(\log n)$ time per selected copy plus $O(1)$ per edge of C at a selected vertex.

Proof. Let H be the multigraph with unit edge weights whose vertices are the copies (C, v) and whose edges are the edges uv of every cluster C, joining (C, u) and (C, v); the degree of (C, v) in H is $\deg_C(v)$, and H is the disjoint union of the clusters. Fix the sign s of Step 2(a) and define $\mathbf{y} \in \mathbb{R}^{V(H)}$ by $y_{(C,v)} = \sqrt{\mathbf{w}_C}\, s(x_v - \mathrm{median}_{\mathbf{x}}(C))$ if this quantity is positive and $y_{(C,v)} = 0$ otherwise. Then $\sum_{(C,v)} \deg_C(v) y_{(C,v)}^2 \ge \mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, \mathcal{C})/2$ by the choice of s, so $\mathbf{y} \ne 0$, and since $|\max(a, 0) - \max(b, 0)| \le |a - b|$,

$$\sum_{uv \in C,\, C \in \mathcal{C}} \big(y_{(C,u)} - y_{(C,v)}\big)^2 \le \sum_{C \in \mathcal{C}} \mathbf{w}_C \sum_{uv \in C} (x_u - x_v)^2 = \mathcal{E}_2(\mathbf{x}, \mathcal{C}).$$

Lemma 3.3 gives $\theta > 0$ such that $S = \{(C, v) : y_{(C,v)} > \theta\}$ is nonempty and has at most

$$\sqrt{\frac{4\,\mathcal{E}_2(\mathbf{x}, \mathcal{C})}{\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, \mathcal{C})}} \sum_{(C,v) \in S} \deg_C(v) < \sqrt{\frac{4}{10^7 \log^4 n}} \sum_{(C,v) \in S} \deg_C(v) < \frac{10^{-3}}{\log^2 n} \sum_{(C,v) \in S} \deg_C(v)$$

edges of H with exactly one endpoint in S.

The sweep extracts copies in decreasing order of $\mathbf{w}_C(x_v - \mathrm{median}_{\mathbf{x}}(C))^2 = y_{(C,v)}^2$ among the copies with $s(x_v - \mathrm{median}_{\mathbf{x}}(C)) > 0$: within a cluster these copies are visited in the order of T_C away from the median, along which the key decreases, and the heap H_s holds the next copy of every cluster, so the global order is decreasing as well. Ties are broken arbitrarily. The quantity Vol is the degree sum of the selected copies, and Bd is the number of edges of H with exactly one selected endpoint, as an edge is counted $+1$ when its first endpoint is selected and -1 when its second one is. Consider the moment when the last copy with $y_{(C,v)}^2 > \theta^2$ has been extracted (it exists since $S \ne \emptyset$, and the heap is not exhausted before it). The selected set is then exactly S, so $\mathrm{Bd} \le 10^{-3} \log^{-2} n \cdot \mathrm{Vol}$ and the sweep stops there if it has not stopped earlier. At the stopping moment the selected copies form the sets S_C; every selected copy satisfies $s(x_v - \mathrm{median}_{\mathbf{x}}(C)) > 0$, the test gives the second item, and at least one copy was extracted, which gives the third item because degrees of copies are at least 1. The vertices $v \in V(C)$ with x_v strictly on one side of $\mathrm{median}_{\mathbf{x}}(C)$ have Vol_C at most $\mathrm{Vol}(C)/2$ by Definition 3.1, which gives the first item. Each extraction and each heap insertion costs $O(\log n)$, locating the next copy of a cluster in T_C costs $O(\log n)$, and the update of Bd visits the edges of C at the selected vertex once. □

Proof of Lemma 3.2. We verify the invariants of the state, then the five items in turn. Throughout, $\log n \ge 20$ and N, D, Q denote the numbers of insertions, deletions and queries of the sequence, with $N + D + Q \le n^{10}$ and $U = N + D$.

Invariants. A block of rank r is formed from exactly 2^r insertions: a rank-0 block from one, and a merge of two blocks of rank $r-1$ from the disjoint union of their insertions. Hence $2^r \le N \le n^{10}$, every rank is at most $10\log n$, each weight class holds at most $10\log n + 1$ blocks, and the number of weight classes is $20\log n + 1$. The clusters of one block have pairwise disjoint vertex sets: this holds after a merge, when the block has one cluster, and a split moves the copies (C, v), $v \in S_C$, to the new cluster C' after removing every edge between S_C and $V(C) \setminus S_C$, so $V(C') = S_C$ and the remaining vertex set of C is $V(C) \setminus S_C$ (a vertex left without edges is removed from both). Every cluster has weights in $[\mathbf{w}_C, 2\mathbf{w}_C)$ with $\mathbf{w}_C = 2^i$ the lower end of its weight class, and the clusters of all blocks partition the live edges, so $\mathcal{C}$ is a partition in the sense of Definition 3.1. Between two updates, DELETE is not called, and the only changes to $\mathcal{C}$ are the splits of Step 2(c), each replacing a cluster C by the two clusters with vertex sets S_C and $V(C) \setminus S_C$ (one of them possibly empty) after the removal of the edges between them; so the clusters only split.

Number of copies. By the invariants a vertex has at most one copy per block, so the number of copies is at most

$$n\,(10\log n + 1)\,(20\log n + 1) \le n \cdot 10.05\log n \cdot 20.05\log n \le 202\,n\log^2 n \le 10^3 n\log^2 n.$$

Potential. Define

$$\Phi = \sum_{C \in \mathcal{C}} \textsc{Vol}(C)\log_2 \textsc{Vol}(C) \ge 0,$$

which is 0 for the empty graph. Since $t \mapsto t\log_2 t$ is nondecreasing on $\{0\} \cup [1, \infty)$ and volumes are even integers, Φ does not increase when edges are removed from clusters (deletions, cut edges) or clusters are discarded. At a merge forming a cluster with $|R|$ live edges, Φ increases by at most $2|R|\log_2(2|R|) \le 2|R| \cdot 2\log n = 4|R|\log n$, because the live graph has at most one edge per vertex pair, so $2|R| \le n(n-1) \le n^2$. At a split of C with $V = \textsc{Vol}(C)$ and $\sigma = \textsc{Vol}_C(S_C) \le V/2$ (Lemma 3.4), the two resulting clusters have volumes at most σ and $V - \sigma$, so Φ decreases by at least

$$V\log_2 V - (V-\sigma)\log_2(V-\sigma) - \sigma\log_2\sigma \ge V\log_2 V - (V-\sigma)\log_2 V - \sigma\log_2\sigma = \sigma\log_2(V/\sigma) \ge \sigma.$$

Every sweep decreases Φ by at least $\sum_C \textsc{Vol}_C(S_C) \ge 1$. The blocks of rank r formed during the sequence have disjoint sets of 2^r insertions each, so at most $N/2^r$ of them are formed, each holding at most 2^r live edges; summing over the ranks $0, \dots, 10\log n$, the total number of live edges over all blocks formed is at most $N(10\log n + 1)$, and the total increase of Φ over the sequence is at most

$$4\log n \cdot N(10\log n + 1) \le 41\,N\log^2 n.$$

Hence the total volume selected by all sweeps of the sequence is at most $41N\log^2 n$, and by Lemma 3.4 the total number of cut edges returned is at most

$$10^{-3}\log^{-2} n \cdot 41N\log^2 n \le 0.041\,N \le 10^{-1}U.$$

Termination and the certificate. Consider a query of $\mathbf{x}$ with estimate $\widetilde{\mathcal{E}}$. At each test of the loop the centers are the current medians, and the cluster energy satisfies $\mathcal{E}_2(\mathbf{x}, \mathcal{C}) \le \mathcal{E}_2(\mathbf{x}, G) \le 10\widetilde{\mathcal{E}}$, where G is the graph at the start of the query, because the sweeps only remove edges and all weights are nonnegative; only the lower bound $\widetilde{\mathcal{E}} \ge \mathcal{E}_2(\mathbf{x}, G)/10$ of the hypothesis $\widetilde{\mathcal{E}} \approx_{10} \mathcal{E}_2(\mathbf{x}, G)$ is used. When the loop body is entered, $\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, \mathcal{C}) > 10^8\log^4 n \cdot \widetilde{\mathcal{E}} \ge 10^7\log^4 n \cdot \mathcal{E}_2(\mathbf{x}, \mathcal{C})$, so Lemma 3.4 applies, the sweep terminates and selects volume at least 1, and the split decreases Φ by at least 1;

as $\Phi \geq 0$, the loop terminates. At exit, $\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, \mathcal{C}) \leq 10^8 \log^4 n \cdot \widetilde{\mathcal{E}}$ with $\mathrm{center}_{\mathbf{x}}(C) = \mathrm{median}_{\mathbf{x}}(C)$, which Step 3 stores; this is the certificate for $\mathbf{x}$ at the end of its query. For the maintained sums: the median of C is found by descending T_C on the subtree degree sums, and for a center c, $\sum_{v \in V(C)} \deg_C(v) \mathbf{w}_C(x_v - c)^2 = \mathbf{w}_C(\sum_v \deg_C(v) x_v^2 - 2c \sum_v \deg_C(v) x_v + c^2 \sum_v \deg_C(v))$ is read from the sums at the root of T_C, and the same expression restricted to the keys strictly above (or below) the median is a range sum in T_C, in $O(\log n)$.

Later splits preserve the certificate. Let $\mathbf{x}$ be a vector queried since the last update, and consider a later split of a cluster C, before the next update, into C' with $V(C') = S_C$ and the rest of C, both inheriting $c = \mathrm{center}_{\mathbf{x}}(C)$. A vertex $v \in V(C)$ belongs to at most one of the two clusters, with degree at most $\deg_C(v)$ there, so

$$\sum_{v \in V(C')} \deg_{C'}(v) \mathbf{w}_C(x_v - c)^2 + \sum_{v \in V(C) \setminus S_C} \deg_C(v) \mathbf{w}_C(x_v - c)^2 \leq \sum_{v \in V(C)} \deg_C(v) \mathbf{w}_C(x_v - c)^2$$

where the degrees on the left are those after the split. Thus $\sum_C \sum_{v \in V(C)} \deg_C(v) \mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, C, v)$ never increases between updates, and the certificate established at the end of the query of $\mathbf{x}$ holds until the next update. The stored values x_v and centers carry the update counter at the time of their query, so the structure recognizes which stored vectors are since the last update, and a split copies only those centers.

Accessors. The list of all copies is a global doubly linked list; $\mathbf{w}_C$ and $\deg_C(v)$ are fields; the edges of C at v form a doubly linked list; the centers of C and the values x_v at v for the vectors queried since the last update are two lists in query order, and the pair $(\mathrm{center}_{\mathbf{x}}(C), \mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}, C, v))$ is computed from their entries and $\mathbf{w}_C$ in $O(1)$ while traversing them in parallel. The tree of (C, v) is a balanced binary search tree whose nodes store subtree sizes, over the edges of C at v in an arbitrary fixed order; it has at most n nodes, so the k-th edge is found by descending on the sizes in $O(\log n)$, and the edge found for a uniformly random $k \in [1, \deg_C(v)]$ is a uniformly random live edge of C at v. The tree is built in linear time at a merge, updated in $O(\log n)$ per edge insertion or removal, and moves with its copy at a split.

Running time. INSERT forms a rank-0 block in $O(1)$. A merge into a block of rank r visits the at most 2^r live edges of the two blocks (through the edge lists of their clusters), groups them by endpoint using a scratch pointer at each vertex to its copy in the cluster being built, and builds the lists and trees of the copies in linear time, so it costs $O(2^r)$; since at most $N/2^r$ blocks of rank r are formed, all merges cost $O(N(10 \log n + 1)) = O(N \log n)$. DELETE costs $O(\log n)$ for the two tree removals. In a query, Step 1 stores n values, sorts V in $O(n \log n)$, appends every vertex in sorted order to the sequence of each of its copies in $O(1)$ per copy, builds each T_C from its sorted sequence with its sums in time linear in $|V(C)|$, computes the median and the energies of each cluster by one traversal of T_C with running sums in time linear in $|V(C)|$, and builds the two heaps H_+, H_- from the first copy of every cluster on each side of its median in time linear in the number of clusters; the total is $O(n \log n)$ plus $O(1)$ per copy, hence $O(n \log^2 n)$ by the copy bound. The heaps persist over the iterations of the loop: the entries of the clusters untouched by a sweep remain valid, and for a cluster touched by the sweep (one with a selected vertex) its stale entries are deleted and the entries of the two resulting clusters inserted, in $O(\log n)$ per touched cluster, of which there are at most as many as selected copies. By Lemma 3.4 a sweep costs $O(\log n)$ per selected copy and $O(1)$ per edge at a selected vertex, that is, $O(\log n)$ per unit of selected volume. The split of Step 2(c) removes each cut edge from two trees and two lists in $O(\log n)$, moves each

selected copy in $O(1)$, moves the edges inside S_C to the list of C' by traversing the lists of the selected copies in $O(1)$ per edge, removes the $|S_C| \le \textsc{Vol}_C(S_C)$ selected vertices from T_C and inserts them into $T_{C'}$ in $O(\log n)$ each, and recomputes the medians, energies and heap entries of C and C' in $O(\log n)$; the number of cut edges is at most the selected volume, so the split also costs $O(\log n)$ per unit of selected volume. Summed over the sequence, the sweeps and splits cost $O(\log n \cdot 41N \log^2 n) = O(N \log^3 n)$ by the potential argument. It remains to account for copying the centers of earlier vectors at a split, $O(1)$ per center. Let Ψ be the number of copies minus twice the number of clusters, which is nonnegative since a cluster has at least one edge, hence at least two copies. A split that produces a nonempty C' increases the number of clusters by one and does not increase the number of copies, so it decreases Ψ by at least 2; removing an edge decreases the number of copies by the number of endpoints left without edges and discards a cluster only when both its endpoints are left without edges in it, so Ψ does not increase. Between two consecutive updates Ψ starts at most $10^3 n \log^2 n$, so at most $500n \log^2 n$ splits with nonempty C' occur, and each copies at most as many centers as there are queries since the last update; if q queries fall between two consecutive updates, their center copying costs $O(q \cdot n \log^2 n)$ in total. Adding up, the sequence costs $O(N \log n + D \log n + N \log^3 n + Qn \log^2 n) = O(U \log^3 n + Qn \log^2 n)$. □

4 Heavy and light edges, and the residual estimator

Throughout this section $G = (V, E, \mathbf{w})$ is a connected weighted graph on $n \ge 2^{20}$ vertices, so that $\log n \ge 20$, with weights in $[n^{-10}, n^{10}]$; $\mathbf{\Pi} \coloneqq \mathbf{I} - \frac{1}{n}\mathbb{1}\mathbb{1}^\top$ is the orthogonal projection onto $\mathbb{1}^\perp$, $\mathbf{L}(G)^\dagger$ is the pseudoinverse of the Laplacian, $\mathbf{L}(G)^{\dagger/2} \coloneqq (\mathbf{L}(G)^\dagger)^{1/2}$, and $\mathcal{R}^{\text{eff}}(G)_e = \boldsymbol{\chi}_e^\top \mathbf{L}(G)^\dagger \boldsymbol{\chi}_e$ is the effective resistance of the edge e. For a symmetric positive semidefinite matrix $\mathbf{N}$, $\mathbf{N}^\dagger$ is its pseudoinverse, $\|\mathbf{N}\|$ its operator norm, $\|\mathbf{N}\|_F$ its Frobenius norm and $\|\mathbf{N}\|_{S_6} \coloneqq \textsc{Tr}[(\mathbf{N}^\top \mathbf{N})^3]^{1/6}$ its Schatten 6-norm, defined for every square matrix. The section splits the edges of G into a heavy part, which is read directly, and a light part, whose contribution to the effective resistances is estimated by a residual; the main statement is Lemma 4.5.

4.1 Heavy and light edges

Definition 4.1 (heavy and light edges)**.** *For $\theta_{\deg} \ge 200$, the* heavy graph G^{heavy} *consists of the $\theta_{\deg}$ heaviest edges at every vertex (all edges at a vertex of degree at most $\theta_{\deg}$; ties broken by a fixed order), the* light graph *is $G^{\text{light}} \coloneqq E \setminus G^{\text{heavy}}$, and*

$$\mathbf{M} \;\coloneqq\; \mathbf{L}(G^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}).$$

An edge is heavy if at least one of its endpoints counts it among its $\theta_{\deg}$ heaviest edges; a light edge is light at both endpoints.

Lemma 4.2 (the heavy and light parts)**.** *Let $G = (V, E, \mathbf{w})$ be connected with the split of Definition 4.1.*

(a) $|E(G^{\text{heavy}})| \le n\theta_{\deg}$.

(b) $\mathbf{w}_e \le \min\{\mathbf{d}(G^{\text{heavy}})_u, \mathbf{d}(G^{\text{heavy}})_v\}/\theta_{\deg}$ *for every light edge $e = uv$.*

(c) $-\mathbf{D}(G^{\text{light}}) \preceq \mathbf{A}(G^{\text{light}}) \preceq \mathbf{D}(G^{\text{light}}) \preceq \mathbf{M}$ *and* $\mathbf{L}(G) \preceq 2\mathbf{M}$.

(d) If $G^{\text{light}} = \emptyset$ then $\mathbf{M} = \mathbf{L}(G)$ and the kernel of $\mathbf{M}$ is spanned by $\mathbb{1}$; if $G^{\text{light}} \ne \emptyset$ then $\mathbf{M}$ is positive definite. In both cases the range of $\mathbf{M}$ contains $\mathbb{1}^\perp$.

(e) $\mathbf{x}^\top \mathbf{M}^\dagger \mathbf{x} \le 2\,\mathbf{x}^\top \mathbf{L}(G)^\dagger \mathbf{x}$ *for every* $\mathbf{x} \perp \mathbb{1}$*; in particular* $\boldsymbol{\chi}_e^\top \mathbf{M}^\dagger \boldsymbol{\chi}_e \le 2\mathcal{R}^{\mathrm{eff}}(G)_e$ *for every edge* e*.*

(f) $\textsc{Tr}[\mathbf{D}(G^{\mathrm{light}})\mathbf{M}^\dagger] \le n$*.*

(g) $\textsc{Tr}\big[\mathbf{M}^\dagger \mathbf{A}(G^{\mathrm{light}})\mathbf{M}^\dagger \mathbf{A}(G^{\mathrm{light}})\big] = \|\mathbf{M}^{\dagger/2}\mathbf{A}(G^{\mathrm{light}})\mathbf{M}^{\dagger/2}\|_F^2 \le 32n/\theta_{\deg}$*.*

Proof. (a) Each vertex contributes at most $\theta_{\deg}$ heavy edges.

(b) A light edge $e = uv$ is not among the $\theta_{\deg}$ heaviest edges at u, so its weight is at most the weight of each of those $\theta_{\deg}$ edges, and $\theta_{\deg}\mathbf{w}_e$ is at most their total weight, which is at most $\mathbf{d}(G^{\mathrm{heavy}})_u$; likewise at v.

(c) The Laplacian $\mathbf{L}(G^{\mathrm{light}}) = \mathbf{D}(G^{\mathrm{light}}) - \mathbf{A}(G^{\mathrm{light}})$ and the signless Laplacian $\mathbf{D}(G^{\mathrm{light}}) + \mathbf{A}(G^{\mathrm{light}})$ are positive semidefinite (the quadratic forms are $\sum_{uv\in G^{\mathrm{light}}} \mathbf{w}_{uv}(\mathbf{x}_u - \mathbf{x}_v)^2$ and $\sum_{uv\in G^{\mathrm{light}}} \mathbf{w}_{uv}(\mathbf{x}_u + \mathbf{x}_v)^2$), so

$$-\mathbf{D}(G^{\mathrm{light}}) \preceq \mathbf{A}(G^{\mathrm{light}}) \preceq \mathbf{D}(G^{\mathrm{light}}) \preceq \mathbf{L}(G^{\mathrm{heavy}}) + \mathbf{D}(G^{\mathrm{light}}) = \mathbf{M},$$
$$\mathbf{L}(G) = \mathbf{L}(G^{\mathrm{heavy}}) + \mathbf{L}(G^{\mathrm{light}}) = \mathbf{M} - \mathbf{A}(G^{\mathrm{light}}) \preceq \mathbf{M} + \mathbf{D}(G^{\mathrm{light}}) \preceq 2\mathbf{M}.$$

(d) If $G^{\mathrm{light}} = \emptyset$ then $G^{\mathrm{heavy}} = G$ and $\mathbf{M} = \mathbf{L}(G)$, whose kernel is spanned by $\mathbb{1}$ since G is connected. Let $G^{\mathrm{light}} \neq \emptyset$ and $\mathbf{M}\mathbf{u} = 0$. Then $\mathbf{u}^\top \mathbf{L}(G^{\mathrm{heavy}})\mathbf{u} = 0$ and $\mathbf{u}^\top \mathbf{D}(G^{\mathrm{light}})\mathbf{u} = 0$, so $\mathbf{u}$ is constant on every connected component of G^{heavy} and vanishes at every vertex of positive light degree. A component of G^{heavy} containing an endpoint of a light edge therefore carries the value 0. A component of G^{heavy} containing no endpoint of a light edge has all its edges in G^{heavy} and is a component of G, hence all of V, which forces $G^{\mathrm{light}} = \emptyset$. So $\mathbf{u} = 0$, and $\mathbf{M}$ is positive definite.

(e) Let $\mathbf{x} \perp \mathbb{1}$. By (d), $\mathbf{x}$ lies in the range of $\mathbf{M}$, so

$$\begin{aligned}\mathbf{x}^\top \mathbf{M}^\dagger \mathbf{x} &= \max_{\mathbf{u}\in\mathbb{R}^V} \big(2\mathbf{x}^\top \mathbf{u} - \mathbf{u}^\top \mathbf{M}\mathbf{u}\big) \\ &\le \max_{\mathbf{u}\in\mathbb{R}^V} \big(2\mathbf{x}^\top \mathbf{u} - \tfrac{1}{2}\mathbf{u}^\top \mathbf{L}(G)\mathbf{u}\big) = 2\,\mathbf{x}^\top \mathbf{L}(G)^\dagger \mathbf{x},\end{aligned}$$

where the inequality is $\mathbf{L}(G) \preceq 2\mathbf{M}$ from (c), and the last maximum is attained at $\mathbf{u} = 2\mathbf{L}(G)^\dagger \mathbf{x}$: the objective is invariant under adding multiples of $\mathbb{1}$ to $\mathbf{u}$ (as $\mathbf{x}^\top \mathbb{1} = 0$) and strictly concave on $\mathbb{1}^\perp$. Taking $\mathbf{x} = \boldsymbol{\chi}_e$ gives the second claim.

(f) If $G^{\mathrm{light}} = \emptyset$ then $\mathbf{D}(G^{\mathrm{light}}) = 0$. Otherwise $\mathbf{M}$ is positive definite, and for a vertex v with $\mathbf{d}(G^{\mathrm{light}})_v > 0$, $\mathbf{M} \succeq \mathbf{D}(G^{\mathrm{light}}) \succeq \mathbf{d}(G^{\mathrm{light}})_v \mathbb{1}_v \mathbb{1}_v^\top$ gives

$$(\mathbf{M}^{-1})_{v,v} = \max_{\mathbf{u}\in\mathbb{R}^V} \big(2\mathbf{u}_v - \mathbf{u}^\top \mathbf{M}\mathbf{u}\big) \le \max_{\mathbf{u}\in\mathbb{R}^V} \big(2\mathbf{u}_v - \mathbf{d}(G^{\mathrm{light}})_v \mathbf{u}_v^2\big) = \frac{1}{\mathbf{d}(G^{\mathrm{light}})_v},$$

so $\textsc{Tr}[\mathbf{D}(G^{\mathrm{light}})\mathbf{M}^{-1}] = \sum_{v:\mathbf{d}(G^{\mathrm{light}})_v>0} \mathbf{d}(G^{\mathrm{light}})_v (\mathbf{M}^{-1})_{v,v} \le n$.

(g) The identity $\textsc{Tr}[\mathbf{M}^\dagger \mathbf{A}(G^{\mathrm{light}})\mathbf{M}^\dagger \mathbf{A}(G^{\mathrm{light}})] = \|\mathbf{M}^{\dagger/2}\mathbf{A}(G^{\mathrm{light}})\mathbf{M}^{\dagger/2}\|_F^2$ is the cyclicity of the trace. If $G^{\mathrm{light}} = \emptyset$ then $\mathbf{A}(G^{\mathrm{light}}) = 0$ and there is nothing to prove, so let $G^{\mathrm{light}} \neq \emptyset$; then $\mathbf{M}$ is positive definite by (d).

A subgraph of the heavy graph. Write $v \to u$ if vu is among the $\theta_{\deg}$ heaviest edges at v ("v chose vu"), and $\max\{\mathbf{w}_f : f \in G^{\mathrm{light}},\ v \in f\}$ for the heaviest light weight at v (zero if v has no light edge). Let H^{heavy} be the graph on V whose edges are the edges $uv \in G^{\mathrm{heavy}}$ with positive weight

$$\begin{aligned}\mathbf{w}(H^{\mathrm{heavy}})_{uv} \;\coloneqq\; \tfrac{1}{2}\big(&\max\{\mathbf{w}_f : f \in G^{\mathrm{light}},\ u \in f\} \cdot [u \to v] \\ &+ \max\{\mathbf{w}_f : f \in G^{\mathrm{light}},\ v \in f\} \cdot [v \to u]\big),\end{aligned}$$

where $[\cdot]$ is the indicator. Then $\mathbf{w}(H^{\text{heavy}})_{uv} \le \mathbf{w}_{uv}$: an edge chosen by u weighs at least every light edge at u, so each of the two terms is at most $\mathbf{w}_{uv}$. Hence $\mathbf{L}(H^{\text{heavy}}) \preceq \mathbf{L}(G^{\text{heavy}})$ and

$$\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}) \;\preceq\; \mathbf{M}.$$

If v has a light edge then v has more than θ_{deg} edges, so v chose exactly θ_{deg} edges, each of weight at least $\frac{1}{2}\max\{\mathbf{w}_f : f \in G^{\text{light}},\ v \in f\}$ in H^{heavy}; therefore

$$\mathbf{d}(H^{\text{heavy}})_v \;\ge\; \tfrac{1}{2}\theta_{\text{deg}} \max\{\mathbf{w}_f : f \in G^{\text{light}},\ v \in f\} \qquad \text{for every } v \text{ with a light edge.} \tag{2}$$

The graph H. Let H be the graph with edge set $E(H^{\text{heavy}}) \cup E(G^{\text{light}})$ (the two sets are disjoint), weights $\mathbf{w}(H^{\text{heavy}})$ on the first part and $\mathbf{w}$ on the second, and vertex set V_H the vertices of positive degree in it. A vertex outside V_H has no light edge and no edge of H^{heavy}, so the corresponding rows and columns of $\mathbf{A}(G^{\text{light}})$, $\mathbf{D}(G^{\text{light}})$ and $\mathbf{L}(H^{\text{heavy}})$ vanish; all matrices below are restricted to $V_H \times V_H$, which changes none of the traces in question. On V_H,

$$\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}) = \mathbf{D}(H) - \mathbf{A}(H^{\text{heavy}}),$$

and this matrix is positive definite: a vector in its kernel is constant on every component of $H^{\text{heavy}}[V_H]$ and zero at every vertex with a light edge, and every component of $H^{\text{heavy}}[V_H]$ contains a vertex with a light edge, because every edge of H^{heavy} has positive weight only if a choosing endpoint has a light edge.

Normalized Frobenius norms. Orient every edge of H: an edge uv of H^{heavy} away from a choosing endpoint u whose heaviest light weight is largest among the choosing endpoints, and a light edge uv away from either endpoint. In both cases the weight $\mathbf{w}(H)_{uv}$ of the oriented edge $u \to v$ is at most $\max\{\mathbf{w}_f : f \in G^{\text{light}},\ u \in f\}$: for a light edge this is the definition of the heaviest light weight, and for an edge of H^{heavy} the weight is the half-sum of at most two terms each at most $\max\{\mathbf{w}_f : f \in G^{\text{light}},\ u \in f\}$. Moreover u has a light edge (the weight is positive), so Equation (2) gives

$$\frac{\mathbf{w}(H)_{uv}}{\mathbf{d}(H)_u} \le \frac{\max\{\mathbf{w}_f : f \in G^{\text{light}},\ u \in f\}}{\mathbf{d}(H^{\text{heavy}})_u} \le \frac{2}{\theta_{\text{deg}}} \qquad \text{for every oriented edge } u \to v \text{ of } H.$$

Hence

$$\|\mathbf{D}(H)^{-1/2}\mathbf{A}(H)\mathbf{D}(H)^{-1/2}\|_F^2 = 2\sum_{uv \in E(H)} \frac{\mathbf{w}(H)_{uv}^2}{\mathbf{d}(H)_u \mathbf{d}(H)_v} \le \frac{4}{\theta_{\text{deg}}} \sum_{u \to v} \frac{\mathbf{w}(H)_{uv}}{\mathbf{d}(H)_v} \le \frac{4n}{\theta_{\text{deg}}},$$

because the total weight of the edges oriented into v is at most $\mathbf{d}(H)_v$ and $|V_H| \le n$. The matrices $\mathbf{A}(H^{\text{heavy}})$ and $\mathbf{A}(G^{\text{light}})$ have disjoint supports and sum to $\mathbf{A}(H)$, so

$$\|\mathbf{D}(H)^{-1/2}\mathbf{A}(H^{\text{heavy}})\mathbf{D}(H)^{-1/2}\|_F^2 + \|\mathbf{D}(H)^{-1/2}\mathbf{A}(G^{\text{light}})\mathbf{D}(H)^{-1/2}\|_F^2 \le \frac{4n}{\theta_{\text{deg}}}. \tag{3}$$

An orthogonal similarity. The matrix $(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))^{-1/2}\mathbf{A}(G^{\text{light}})(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))^{-1/2}$ lies between $-\mathbf{I}$ and $\mathbf{I}$, because $\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}) \pm \mathbf{A}(G^{\text{light}}) = \mathbf{L}(H^{\text{heavy}}) + (\mathbf{D}(G^{\text{light}}) \pm \mathbf{A}(G^{\text{light}})) \succeq 0$. The matrix

$$\mathbf{U} \coloneqq \left(\mathbf{D}(H)^{-1/2}(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))\mathbf{D}(H)^{-1/2}\right)^{-1/2}\mathbf{D}(H)^{-1/2}(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))^{1/2}$$

satisfies $\mathbf{U}\mathbf{U}^\top = \mathbf{I}$, so it is orthogonal, and

$$\begin{aligned}&\mathbf{U}\,(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))^{-1/2}\mathbf{A}(G^{\text{light}})(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))^{-1/2}\,\mathbf{U}^\top \\ &\quad= \big(\mathbf{D}(H)^{-1/2}(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))\mathbf{D}(H)^{-1/2}\big)^{-1/2}\,\mathbf{D}(H)^{-1/2}\mathbf{A}(G^{\text{light}})\mathbf{D}(H)^{-1/2} \\ &\qquad\cdot \big(\mathbf{D}(H)^{-1/2}(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))\mathbf{D}(H)^{-1/2}\big)^{-1/2}.\end{aligned} \tag{4}$$

Both sides have the same Frobenius norm and lie between $-\mathbf{I}$ and $\mathbf{I}$; we write $\mathbf{T}$ for the right-hand side.

Spectral split. Let $\mathbf{P}$ be the spectral projection of

$$\mathbf{D}(H)^{-1/2}(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))\mathbf{D}(H)^{-1/2} = \mathbf{I} - \mathbf{D}(H)^{-1/2}\mathbf{A}(H^{\text{heavy}})\mathbf{D}(H)^{-1/2}$$

onto its eigenvalues below $\frac{1}{2}$. On the range of $\mathbf{P}$ the eigenvalues of the symmetric matrix $\mathbf{D}(H)^{-1/2}\mathbf{A}(H^{\text{heavy}})\mathbf{D}(H)^{-1/2}$ exceed $\frac{1}{2}$, and a symmetric matrix has at most four times its squared Frobenius norm many eigenvalues above $\frac{1}{2}$, so

$$\textsc{Tr}[\mathbf{P}] \le 4\,\|\mathbf{D}(H)^{-1/2}\mathbf{A}(H^{\text{heavy}})\mathbf{D}(H)^{-1/2}\|_F^2.$$

On the range of $\mathbf{I} - \mathbf{P}$ the inverse of $\mathbf{D}(H)^{-1/2}(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))\mathbf{D}(H)^{-1/2}$ has eigenvalues at most 2, and both $\mathbf{P}$ and $\mathbf{I} - \mathbf{P}$ commute with its inverse square root. Write the right-hand side of Equation (4) as the sum of its four compressions by $\mathbf{P}$ and $\mathbf{I} - \mathbf{P}$ on the two sides; these four matrices are orthogonal in the Frobenius inner product, so the squared Frobenius norm of the sum is the sum of their squared Frobenius norms.

- The compression $(\mathbf{I} - \mathbf{P})\mathbf{T}(\mathbf{I} - \mathbf{P})$ is the product of the matrix

  $$(\mathbf{I} - \mathbf{P})\big(\mathbf{D}(H)^{-1/2}(\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}))\mathbf{D}(H)^{-1/2}\big)^{-1/2},$$

  of operator norm at most $\sqrt{2}$, then $\mathbf{D}(H)^{-1/2}\mathbf{A}(G^{\text{light}})\mathbf{D}(H)^{-1/2}$, then the transpose of the first factor; so

  $$\|(\mathbf{I} - \mathbf{P})\mathbf{T}(\mathbf{I} - \mathbf{P})\|_F^2 \le 4\,\|\mathbf{D}(H)^{-1/2}\mathbf{A}(G^{\text{light}})\mathbf{D}(H)^{-1/2}\|_F^2.$$

- The three compressions touching $\mathbf{P}$ have total squared Frobenius norm at most $2\textsc{Tr}[\mathbf{P}]$: $\|\mathbf{P}\mathbf{T}\|_F^2 = \textsc{Tr}[\mathbf{P}\mathbf{T}^2\mathbf{P}] \le \textsc{Tr}[\mathbf{P}]$ since $\mathbf{T}^2 \preceq \mathbf{I}$, this covers the compressions $\mathbf{P}\mathbf{T}\mathbf{P}$ and $\mathbf{P}\mathbf{T}(\mathbf{I} - \mathbf{P})$, and $\|(\mathbf{I} - \mathbf{P})\mathbf{T}\mathbf{P}\|_F = \|\mathbf{P}\mathbf{T}(\mathbf{I} - \mathbf{P})\|_F \le \|\mathbf{P}\mathbf{T}\|_F$.

Therefore, by Equation (3),

$$\begin{aligned}\|\mathbf{T}\|_F^2 &\le 4\,\|\mathbf{D}(H)^{-1/2}\mathbf{A}(G^{\text{light}})\mathbf{D}(H)^{-1/2}\|_F^2 \\ &\quad + 8\,\|\mathbf{D}(H)^{-1/2}\mathbf{A}(H^{\text{heavy}})\mathbf{D}(H)^{-1/2}\|_F^2 \\ &\le \frac{32n}{\theta_{\text{deg}}}.\end{aligned}$$

Monotonicity. The map $\mathbf{N} \mapsto \textsc{Tr}[\mathbf{N}\mathbf{A}(G^{\text{light}})\mathbf{N}\mathbf{A}(G^{\text{light}})]$ is nondecreasing on the positive semidefinite cone: its derivative at $\mathbf{N} \succeq 0$ along $\mathbf{\Delta} \succeq 0$ is $2\textsc{Tr}[\mathbf{\Delta}\,\mathbf{A}(G^{\text{light}})\mathbf{N}\mathbf{A}(G^{\text{light}})] \ge 0$, as $\mathbf{A}(G^{\text{light}})\mathbf{N}\mathbf{A}(G^{\text{light}}) \succeq 0$. For every $t > 0$, $\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}) + t\mathbf{I} \preceq \mathbf{M} + t\mathbf{I}$ gives $(\mathbf{M} + t\mathbf{I})^{-1} \preceq (\mathbf{L}(H^{\text{heavy}}) + \mathbf{D}(G^{\text{light}}) + t\mathbf{I})^{-1}$, on V_H as on V; the latter matrix is block diagonal with respect to V_H and $V \setminus V_H$ and $\mathbf{A}(G^{\text{light}})$ is supported on $V_H \times V_H$, so the trace only involves the V_H block,

which converges to the inverse of the positive definite matrix $\mathbf{L}(H^{\mathrm{heavy}}) + \mathbf{D}(G^{\mathrm{light}})$ on V_H as $t \downarrow 0$. Hence

$$\begin{aligned}\operatorname{Tr}[\mathbf{M}^{-1}\mathbf{A}(G^{\mathrm{light}})\mathbf{M}^{-1}\mathbf{A}(G^{\mathrm{light}})] &\le \operatorname{Tr}\big[(\mathbf{L}(H^{\mathrm{heavy}}) + \mathbf{D}(G^{\mathrm{light}}))^{-1}\mathbf{A}(G^{\mathrm{light}}) \\ &\qquad \cdot (\mathbf{L}(H^{\mathrm{heavy}}) + \mathbf{D}(G^{\mathrm{light}}))^{-1}\mathbf{A}(G^{\mathrm{light}})\big] \\ &= \|\mathbf{T}\|_F^2 \\ &\le \frac{32n}{\theta_{\deg}}.\end{aligned}$$

□

4.2 Two tools

The acceptance tests and the all-pairs estimate of Lemma 4.5 are certified with the following tail bound for Gaussian quadratic forms. The upper tail is Theorem 1 of Hsu, Kakade and Zhang [HKZ12], specialized to a standard Gaussian vector; the lower tail is Lemma 1 of Laurent and Massart [LM00], applied to the eigenvalues of $\mathbf{N}$.

Lemma 4.3 (Theorem 1 of [HKZ12]; Lemma 1 of [LM00])**.** *Let $\mathbf{N} \in \mathbb{R}^{d\times d}$ be symmetric positive semidefinite and let $\mathbf{g} \in \mathbb{R}^d$ have independent standard Gaussian entries. Then $\operatorname{Tr}[\mathbf{N}^2] \le \|\mathbf{N}\|\operatorname{Tr}[\mathbf{N}]$, and for every $t > 0$*

$$\begin{aligned}\Pr\big[\mathbf{g}^\top\mathbf{N}\mathbf{g} > \operatorname{Tr}[\mathbf{N}] + 2\sqrt{\operatorname{Tr}[\mathbf{N}^2]\,t} + 2\|\mathbf{N}\|t\big] &\le e^{-t}, \\ \Pr\big[\mathbf{g}^\top\mathbf{N}\mathbf{g} < \operatorname{Tr}[\mathbf{N}] - 2\sqrt{\operatorname{Tr}[\mathbf{N}^2]\,t}\big] &\le e^{-t}.\end{aligned}$$

Proof. Let $\lambda_1, \dots, \lambda_d \ge 0$ be the eigenvalues of $\mathbf{N}$, so that $\operatorname{Tr}[\mathbf{N}^2] = \sum_i \lambda_i^2 \le \max_i \lambda_i \sum_i \lambda_i = \|\mathbf{N}\|\operatorname{Tr}[\mathbf{N}]$. Theorem 1 of [HKZ12] states, for a matrix $\mathbf{A}'$, the vector $\mathbf{x}$ with $\mathbb{E}[\exp(\boldsymbol{\alpha}^\top\mathbf{x})] \le \exp(\sigma^2\|\boldsymbol{\alpha}\|^2/2)$ for all $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma} \coloneqq \mathbf{A}'^\top\mathbf{A}'$, that $\Pr[\|\mathbf{A}'\mathbf{x}\|^2 > \sigma^2(\operatorname{Tr}[\boldsymbol{\Sigma}] + 2\sqrt{\operatorname{Tr}[\boldsymbol{\Sigma}^2]t} + 2\|\boldsymbol{\Sigma}\|t)] \le e^{-t}$; with $\mathbf{x} = \mathbf{g}$, $\sigma = 1$ and $\mathbf{A}' = \mathbf{N}^{1/2}$ this is the first bound. For the second, rotating $\mathbf{g}$ to the eigenbasis of $\mathbf{N}$ (which preserves its law) writes $\mathbf{g}^\top\mathbf{N}\mathbf{g} = \sum_i \lambda_i \mathbf{g}_i^2$, and Lemma 1 of [LM00] states, for independent standard Gaussians $\mathbf{g}_i$ and nonnegative λ_i, that $\Pr[\sum_i \lambda_i(\mathbf{g}_i^2 - 1) \le -2(\sum_i \lambda_i^2)^{1/2}\sqrt{t}] \le e^{-t}$. □

The residual of Lemma 4.5 is cubic in a random matrix whose sixth moment is bounded by the following lemma; it is proved here because its hypothesis is tailored to the centered retention variables of Equation (5) below.

Lemma 4.4 (sixth moment of a centered rank-one sum)**.** *Let $\mathbf{Q}_i \coloneqq \zeta_i \mathbf{u}_i\mathbf{u}_i^\top$, $i \in I$, with $\mathbf{u}_i \in \mathbb{R}^d$ and independent real random variables ζ_i with $\mathbb{E}[\zeta_i] = 0$ and $\zeta_i \ge -1$ almost surely, such that $\sum_{i\in I} \mathbf{u}_i\mathbf{u}_i^\top \preceq \mathbf{I}$ and, for some $a \le 1$,*

$$\mathbb{E}[|\zeta_i|^p]\,\|\mathbf{u}_i\|^{2p-2} \;\le\; a^{p-1} \qquad \text{for every } i \in I \text{ and } p = 2, \dots, 6.$$

Then $\mathbb{E}\big[\|\sum_{i\in I} \mathbf{Q}_i\|_{S_6}^6\big] = \mathbb{E}\big[\operatorname{Tr}[(\sum_{i\in I} \mathbf{Q}_i)^6]\big] \le 10^2 a^3 d$.

Proof. The equality holds because $\sum_i \mathbf{Q}_i$ is symmetric. Since $(\mathbf{u}_i\mathbf{u}_i^\top)^p = \|\mathbf{u}_i\|^{2p-2}\mathbf{u}_i\mathbf{u}_i^\top$,

$$\sum_i \mathbb{E}[\mathbf{Q}_i^p] = \sum_i \mathbb{E}[\zeta_i^p]\|\mathbf{u}_i\|^{2p-2}\mathbf{u}_i\mathbf{u}_i^\top, \qquad \big|\mathbb{E}[\zeta_i^p]\|\mathbf{u}_i\|^{2p-2}\big| \le a^{p-1},$$

so for $p = 2, \dots, 6$

$$-a^{p-1}\mathbf{I} \preceq -a^{p-1}\sum_i \mathbf{u}_i\mathbf{u}_i^\top \preceq \sum_i \mathbb{E}[\mathbf{Q}_i^p] \preceq a^{p-1}\sum_i \mathbf{u}_i\mathbf{u}_i^\top \preceq a^{p-1}\mathbf{I},$$
$$\Big|\mathrm{Tr}\Big[\sum_i \mathbb{E}[\mathbf{Q}_i^p]\Big]\Big| \le a^{p-1}\mathrm{Tr}\Big[\sum_i \mathbf{u}_i\mathbf{u}_i^\top\Big] \le a^{p-1}d.$$

Four consequences are used repeatedly.

- $\sum_{i,j}(\mathbf{u}_i^\top\mathbf{u}_j)^2 = \mathrm{Tr}[(\sum_i \mathbf{u}_i\mathbf{u}_i^\top)^2] \le \mathrm{Tr}[\sum_i \mathbf{u}_i\mathbf{u}_i^\top] \le d$, as $0 \preceq \sum_i \mathbf{u}_i\mathbf{u}_i^\top \preceq \mathbf{I}$.
- $\mathrm{Tr}[(\sum_i \mathbb{E}[\mathbf{Q}_i^2])^q] \le a^{q-1}\mathrm{Tr}[\sum_i \mathbb{E}[\mathbf{Q}_i^2]] \le a^q d$ for $q = 1, 2, 3$, as $0 \preceq \sum_i \mathbb{E}[\mathbf{Q}_i^2] \preceq a\mathbf{I}$.
- For a fixed symmetric $\mathbf{N}$ with $0 \preceq \mathbf{N} \preceq a'\mathbf{I}$,
$$\sum_i \mathbb{E}[\mathbf{Q}_i\mathbf{N}\mathbf{Q}_i] = \sum_i \mathbb{E}[\zeta_i^2]\,(\mathbf{u}_i^\top\mathbf{N}\mathbf{u}_i)\,\mathbf{u}_i\mathbf{u}_i^\top \preceq a'\sum_i \mathbb{E}[\zeta_i^2]\|\mathbf{u}_i\|^2\mathbf{u}_i\mathbf{u}_i^\top = a'\sum_i \mathbb{E}[\mathbf{Q}_i^2].$$
- For all i, j (not necessarily distinct), $\mathrm{Tr}[\mathbb{E}[\mathbf{Q}_i^2]\,\mathbb{E}[\mathbf{Q}_j^2]] \ge 0$ and $\sum_{i,j}\mathrm{Tr}[\mathbb{E}[\mathbf{Q}_i^2]\,\mathbb{E}[\mathbf{Q}_j^2]] = \mathrm{Tr}[(\sum_i \mathbb{E}[\mathbf{Q}_i^2])^2] \le a^2 d$; for $i \ne j$, $\mathrm{Tr}[\mathbb{E}[\mathbf{Q}_i^2]\,\mathbb{E}[\mathbf{Q}_j^2]] = \mathbb{E}[\|\mathbf{Q}_i\mathbf{Q}_j\|_F^2]$ by independence.

Expansion. $\mathbb{E}[\mathrm{Tr}[(\sum_i \mathbf{Q}_i)^6]] = \sum_{i_1,\dots,i_6 \in I}\mathbb{E}[\mathrm{Tr}[\mathbf{Q}_{i_1}\cdots\mathbf{Q}_{i_6}]]$, and a term vanishes unless every index occurring in it occurs at least twice, because the $\mathbf{Q}_i$ are independent and centered. Group the remaining terms by the partition of the six cyclic positions into blocks of equal indices, with distinct indices on distinct blocks and every block of size at least 2. For each partition we exhibit a nonnegative function of the block indices that bounds the absolute value of the term when the block indices are distinct, and sum it over *all* tuples of block indices; this dominates the sum over distinct tuples. The partitions are, up to rotation, the following $1 + 15 + 10 + 15 = 41$.

- One block, the word $iiiiii$: $\sum_i \mathbb{E}[\mathrm{Tr}[\mathbf{Q}_i^6]] = \mathrm{Tr}[\sum_i \mathbb{E}[\mathbf{Q}_i^6]] \le a^5 d$.
- Blocks of sizes 4 and 2 (15 partitions), the words $iiiijj$, $iiijij$ and $iijiij$ and their rotations. For $iiiijj$,
$$\sum_{i,j}\mathrm{Tr}[\mathbb{E}[\mathbf{Q}_i^4]\,\mathbb{E}[\mathbf{Q}_j^2]] = \mathrm{Tr}\Big[\Big(\sum_i \mathbb{E}[\mathbf{Q}_i^4]\Big)\Big(\sum_j \mathbb{E}[\mathbf{Q}_j^2]\Big)\Big] \le a\,\mathrm{Tr}\Big[\sum_i \mathbb{E}[\mathbf{Q}_i^4]\Big] \le a^4 d,$$
as both matrices are positive semidefinite and the second is at most $a\mathbf{I}$. For $iiijij$ and $iijiij$,
$$\mathrm{Tr}[\mathbf{Q}_i^3\mathbf{Q}_j\mathbf{Q}_i\mathbf{Q}_j] = \mathrm{Tr}[\mathbf{Q}_i^2\mathbf{Q}_j\mathbf{Q}_i^2\mathbf{Q}_j] = \zeta_i^4\zeta_j^2\|\mathbf{u}_i\|^4(\mathbf{u}_i^\top\mathbf{u}_j)^4 \le \zeta_i^4\|\mathbf{u}_i\|^6\cdot\zeta_j^2\|\mathbf{u}_j\|^2\cdot(\mathbf{u}_i^\top\mathbf{u}_j)^2$$
by $(\mathbf{u}_i^\top\mathbf{u}_j)^2 \le \|\mathbf{u}_i\|^2\|\mathbf{u}_j\|^2$; for $i \ne j$ the expectation is at most $a^3\cdot a\cdot(\mathbf{u}_i^\top\mathbf{u}_j)^2$ by independence and the moment hypothesis, and summing over all i, j gives at most $a^4 d$.
- Blocks of sizes 3 and 3 (10 partitions), the words $iiijjj$, $iijijj$ and $ijijij$ and their rotations. For $iiijjj$, with $\mathbb{E}[\mathbf{Q}_i^3] = \mathbb{E}[\zeta_i^3]\|\mathbf{u}_i\|^4\mathbf{u}_i\mathbf{u}_i^\top$ and $|\mathbb{E}[\zeta_i^3]|\|\mathbf{u}_i\|^4 \le a^2$,
$$\sum_{i,j}\big|\mathrm{Tr}[\mathbb{E}[\mathbf{Q}_i^3]\,\mathbb{E}[\mathbf{Q}_j^3]]\big| \le \sum_{i,j} a^2\cdot a^2\cdot(\mathbf{u}_i^\top\mathbf{u}_j)^2 \le a^4 d.$$
For $iijijj$ and $ijijij$,
$$\big|\mathrm{Tr}[\mathbf{Q}_i^2\mathbf{Q}_j\mathbf{Q}_i\mathbf{Q}_j^2]\big| = |\zeta_i|^3|\zeta_j|^3\|\mathbf{u}_i\|^2\|\mathbf{u}_j\|^2(\mathbf{u}_i^\top\mathbf{u}_j)^4 \le |\zeta_i|^3\|\mathbf{u}_i\|^4\cdot|\zeta_j|^3\|\mathbf{u}_j\|^4\cdot(\mathbf{u}_i^\top\mathbf{u}_j)^2,$$
$$\big|\mathrm{Tr}[(\mathbf{Q}_i\mathbf{Q}_j)^3]\big| = |\zeta_i|^3|\zeta_j|^3(\mathbf{u}_i^\top\mathbf{u}_j)^6 \le |\zeta_i|^3\|\mathbf{u}_i\|^4\cdot|\zeta_j|^3\|\mathbf{u}_j\|^4\cdot(\mathbf{u}_i^\top\mathbf{u}_j)^2,$$
with expectations at most $a^2\cdot a^2\cdot(\mathbf{u}_i^\top\mathbf{u}_j)^2$ for $i \ne j$, each summing to at most $a^4 d$.

- Three blocks of size 2 with two equal letters cyclically adjacent (11 partitions): rotate the word to kkW with W a word of length four in i, j, each letter twice. Taking the expectation over ζ_k first (for $k \notin \{i,j\}$) and summing over k leaves $\mathrm{Tr}[(\sum_k \mathbb{E}[\mathbf{Q}_k^2])\, W]$, and the six possible W give:
  - $W \in \{\mathbf{Q}_i^2\mathbf{Q}_j^2, \mathbf{Q}_j^2\mathbf{Q}_i^2\}$: $\sum_{i,j} \mathrm{Tr}[(\sum_k \mathbb{E}[\mathbf{Q}_k^2])\, \mathbb{E}[\mathbf{Q}_i^2]\, \mathbb{E}[\mathbf{Q}_j^2]] = \mathrm{Tr}[(\sum_k \mathbb{E}[\mathbf{Q}_k^2])^3] \le a^3 d$;
  - $W \in \{\mathbf{Q}_i\mathbf{Q}_j^2\mathbf{Q}_i, \mathbf{Q}_j\mathbf{Q}_i^2\mathbf{Q}_j\}$: by the third consequence with $\mathbf{N} = \sum_j \mathbb{E}[\mathbf{Q}_j^2] \preceq a\mathbf{I}$,
  $$\sum_i \mathbb{E}\Big[\mathbf{Q}_i\Big(\sum_j \mathbb{E}[\mathbf{Q}_j^2]\Big)\mathbf{Q}_i\Big] \preceq a \sum_i \mathbb{E}[\mathbf{Q}_i^2],$$
  so the sum is at most $a\,\mathrm{Tr}[(\sum_k \mathbb{E}[\mathbf{Q}_k^2])^2] \le a^3 d$;
  - $W \in \{\mathbf{Q}_i\mathbf{Q}_j\mathbf{Q}_i\mathbf{Q}_j, \mathbf{Q}_j\mathbf{Q}_i\mathbf{Q}_j\mathbf{Q}_i\}$: $|\mathrm{Tr}[(\sum_k \mathbb{E}[\mathbf{Q}_k^2])(\mathbf{Q}_i\mathbf{Q}_j)^2]| \le \|\sum_k \mathbb{E}[\mathbf{Q}_k^2]\| \cdot \|\mathbf{Q}_i\mathbf{Q}_j\|_F^2 \le a\|\mathbf{Q}_i\mathbf{Q}_j\|_F^2$, whose expectation summed over $i \neq j$ is at most $a\sum_{i,j} \mathrm{Tr}[\mathbb{E}[\mathbf{Q}_i^2]\mathbb{E}[\mathbf{Q}_j^2]] \le a^3 d$ by the fourth consequence.

- Three blocks of size 2 with no two equal letters cyclically adjacent (4 partitions): the word $ijkijk$ and the three rotations-with-relabeling of $ijkikj$. For $ijkijk$, with i, j, k distinct,
$$\big|\mathrm{Tr}[(\mathbf{Q}_i\mathbf{Q}_j\mathbf{Q}_k)^2]\big| \le \|\mathbf{Q}_i\mathbf{Q}_j\mathbf{Q}_k\|_F^2 = \mathrm{Tr}[\mathbf{Q}_k\mathbf{Q}_j\mathbf{Q}_i^2\mathbf{Q}_j\mathbf{Q}_k],$$
and taking the expectation over ζ_i, then ζ_j, then ζ_k, with the third consequence applied twice ($\sum_i \mathbb{E}[\mathbf{Q}_i^2] \preceq a\mathbf{I}$, then $\sum_j \mathbb{E}[\mathbf{Q}_j(\sum_i \mathbb{E}[\mathbf{Q}_i^2])\mathbf{Q}_j] \preceq a\sum_j \mathbb{E}[\mathbf{Q}_j^2] \preceq a^2\mathbf{I}$), gives
$$\sum_{i,j,k} \mathbb{E}\big[\mathrm{Tr}[\mathbf{Q}_k\mathbf{Q}_j\mathbf{Q}_i^2\mathbf{Q}_j\mathbf{Q}_k]\big] \le a^2\,\mathrm{Tr}\Big[\sum_k \mathbb{E}[\mathbf{Q}_k^2]\Big] \le a^3 d.$$
For $ijkikj$, a rotation gives $\mathrm{Tr}[\mathbf{Q}_i\mathbf{Q}_j\mathbf{Q}_k\mathbf{Q}_i\mathbf{Q}_k\mathbf{Q}_j] = \mathrm{Tr}[(\mathbf{Q}_j\mathbf{Q}_i\mathbf{Q}_j)(\mathbf{Q}_k\mathbf{Q}_i\mathbf{Q}_k)]$, and $\mathbf{Q}_j\mathbf{Q}_i\mathbf{Q}_j = \zeta_j^2(\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j)\mathbf{u}_j\mathbf{u}_j^\top$ with $\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j = \zeta_i(\mathbf{u}_i^\top\mathbf{u}_j)^2$, so
$$\begin{aligned}\big|\mathrm{Tr}[(\mathbf{Q}_j\mathbf{Q}_i\mathbf{Q}_j)(\mathbf{Q}_k\mathbf{Q}_i\mathbf{Q}_k)]\big| &= \zeta_j^2\zeta_k^2|\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j|\,|\mathbf{u}_k^\top\mathbf{Q}_i\mathbf{u}_k|\,(\mathbf{u}_j^\top\mathbf{u}_k)^2 \\ &\le \zeta_j^2\|\mathbf{u}_j\|^2|\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j| \cdot \zeta_k^2\|\mathbf{u}_k\|^2|\mathbf{u}_k^\top\mathbf{Q}_i\mathbf{u}_k|.\end{aligned}$$
Conditionally on ζ_i the factors indexed by $j \neq k$ are independent, so the sum over distinct $j, k \neq i$ of the conditional expectations is at most
$$\mathbb{E}\Big[\Big(\sum_j \zeta_j^2\|\mathbf{u}_j\|^2|\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j|\Big)^2 \,\Big|\, \zeta_i\Big] \le \sum_j \mathbb{E}[\zeta_j^4]\|\mathbf{u}_j\|^4(\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j)^2 + \Big(\sum_j \mathbb{E}[\zeta_j^2]\|\mathbf{u}_j\|^2|\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j|\Big)^2.$$
With $(\mathbf{u}_j^\top\mathbf{Q}_i\mathbf{u}_j)^2 = \zeta_i^2(\mathbf{u}_i^\top\mathbf{u}_j)^4 \le \zeta_i^2\|\mathbf{u}_i\|^2\|\mathbf{u}_j\|^2(\mathbf{u}_i^\top\mathbf{u}_j)^2$ and $\mathbb{E}[\zeta_j^4]\|\mathbf{u}_j\|^6 \le a^3$, the first sum is at most
$$a^3\zeta_i^2\|\mathbf{u}_i\|^2\,\mathbf{u}_i^\top\Big(\sum_j \mathbf{u}_j\mathbf{u}_j^\top\Big)\mathbf{u}_i \le a^3\zeta_i^2\|\mathbf{u}_i\|^4,$$
and with $\mathbb{E}[\zeta_j^2]\|\mathbf{u}_j\|^2 \le a$ the second is at most $(a|\zeta_i|\sum_j(\mathbf{u}_i^\top\mathbf{u}_j)^2)^2 \le a^2\zeta_i^2\|\mathbf{u}_i\|^4$. Taking the expectation over ζ_i and summing over i, both are at most $a^2\sum_i \mathbb{E}[\zeta_i^2]\|\mathbf{u}_i\|^4 = a^2\mathrm{Tr}[\sum_i \mathbb{E}[\mathbf{Q}_i^2]] \le a^3 d$, so this word contributes at most $2a^3 d$.

Each of the 41 partitions contributes at most $2a^3 d$, as $a \le 1$, and $82a^3 d \le 10^2 a^3 d$. □

4.3 The residual estimator

The sampled references. The estimator works on two sparse random matrices built from G by independent retention of its edges. Each weight is split as $\mathbf{w}_e = \mathbf{w}_e^{\mathrm{s}} + \mathbf{w}_e^{\mathrm{x}}$ into a sampled part $\mathbf{w}_e^{\mathrm{s}} \geq 0$ and an exactly kept part $\mathbf{w}_e^{\mathrm{x}} \geq 0$ (one of them may be zero), and $E_{\mathrm{s}} \coloneqq \{e \in E : \mathbf{w}_e^{\mathrm{s}} > 0\}$. Every $e \in E_{\mathrm{s}}$ has an *intensity* $\varrho_e > 0$ and is *retained* independently with probability $p_e \coloneqq 1 - e^{-\varrho_e}$. The *centered retention variable* ζ_e of e and the *sparse reference* $\widetilde{\mathbf{L}}$ and the *heavy reference* $\widetilde{\mathbf{M}}$ are

$$\begin{aligned}
\zeta_e &\coloneqq \begin{cases} 1/p_e - 1, & e \in E_{\mathrm{s}} \text{ retained}, \\ -1, & e \in E_{\mathrm{s}} \text{ not retained}, \\ 0, & e \notin E_{\mathrm{s}}, \end{cases} \\
\widetilde{\mathbf{L}} &\coloneqq \sum_{e \in E} \big(\mathbf{w}_e^{\mathrm{x}} + (1+\zeta_e)\mathbf{w}_e^{\mathrm{s}}\big)\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}, \\
\widetilde{\mathbf{M}} &\coloneqq \mathbf{D}(G^{\mathrm{light}}) + \sum_{e \in G^{\mathrm{heavy}}} \big(\mathbf{w}_e^{\mathrm{x}} + (1+\zeta_e)\mathbf{w}_e^{\mathrm{s}}\big)\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}, \\
\mathbf{Q}_e &\coloneqq \zeta_e \mathbf{w}_e^{\mathrm{s}} \mathbf{L}(G)^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{L}(G)^{\dagger/2}, \qquad \mathbf{Q}_e^{\mathrm{heavy}} \coloneqq \zeta_e \mathbf{w}_e^{\mathrm{s}} \mathbf{M}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{M}^{\dagger/2}.
\end{aligned} \tag{5}$$

Thus $\mathbb{E}[\zeta_e] = 0$, $\zeta_e \geq -1$, $(1+\zeta_e)\mathbf{w}_e^{\mathrm{s}}$ is $\mathbf{w}_e^{\mathrm{s}}/p_e$ if e is retained and 0 otherwise, $\mathbb{E}[\widetilde{\mathbf{L}}] = \mathbf{L}(G)$, $\mathbb{E}[\widetilde{\mathbf{M}}] = \mathbf{M}$, and

$$\mathbf{L}(G)^{\dagger/2}\widetilde{\mathbf{L}}\mathbf{L}(G)^{\dagger/2} - \boldsymbol{\Pi} = \sum_{e \in E} \mathbf{Q}_e, \qquad \mathbf{M}^{\dagger/2}\widetilde{\mathbf{M}}\mathbf{M}^{\dagger/2} - \mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2} = \sum_{e \in G^{\mathrm{heavy}}} \mathbf{Q}_e^{\mathrm{heavy}}, \tag{6}$$

since $\sum_e \mathbf{w}_e \mathbf{L}(G)^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{L}(G)^{\dagger/2} = \boldsymbol{\Pi}$ and $\mathbf{D}(G^{\mathrm{light}}) + \sum_{e \in G^{\mathrm{heavy}}} \mathbf{w}_e\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top} = \mathbf{M}$. We write $\widetilde{\mathbf{L}} = \widetilde{\mathbf{B}}^{\top}\widetilde{\mathbf{W}}\widetilde{\mathbf{B}}$, where $\widetilde{\mathbf{B}}$ has the row $\boldsymbol{\chi}_e^{\top}$ for every edge e with a positive coefficient in Equation (5) and $\widetilde{\mathbf{W}}$ is the diagonal matrix of these coefficients; $E(\widetilde{\mathbf{L}})$ denotes this set of edges. The intensities are required to dominate the effective resistances:

$$\varrho_e \;\geq\; 10^8 \log n \cdot \mathbf{w}_e^{\mathrm{s}} \mathcal{R}^{\mathrm{eff}}(G)_e \qquad \text{for every } e \in E_{\mathrm{s}}. \tag{7}$$

The references are *accurate* if

$$\widetilde{\mathbf{L}} \approx_{10^{-3}} \mathbf{L}(G) \qquad \text{and} \qquad \widetilde{\mathbf{M}} \approx_{10^{-3}} \mathbf{M}. \tag{8}$$

The attempt. The pseudocode of Attempt is given in Figure 4.

$(\mathbf{x}_j^{\mathrm{rand}}, \mathbf{x}_j^{\mathrm{res}}, \mathbf{x}_j^{\mathrm{prox}})_{j \le 10^3 \log n} = \textsc{Attempt}(\widetilde{\mathbf{L}}, \widetilde{\mathbf{M}}, \mathbf{Z}(\widetilde{\mathbf{L}}), \mathbf{Z}(\widetilde{\mathbf{M}}))$

Input: the references $\widetilde{\mathbf{L}} = \widetilde{\mathbf{B}}^\top \widetilde{\mathbf{W}} \widetilde{\mathbf{B}}$ and $\widetilde{\mathbf{M}}$; the fixed operators $\mathbf{Z}(\widetilde{\mathbf{L}})$, $\mathbf{Z}(\widetilde{\mathbf{M}})$; reads the light degrees $\mathbf{d}(G^{\mathrm{light}})$.

Output: the vectors $\mathbf{x}_j^{\mathrm{rand}}, \mathbf{x}_j^{\mathrm{res}}, \mathbf{x}_j^{\mathrm{prox}} \in \mathbb{R}^V$, $j \le 10^3 \log n$; the verdict *accepted* or *rejected*.

1. Draw $\mathbf{g}_j \in \mathbb{R}^{E(\widetilde{\mathbf{L}})}$, $j \le 10^3 \log n$, with independent standard Gaussian entries.
2. For every $j \le 10^3 \log n$:
   (a) $\mathbf{x}_j^{\mathrm{rand}} \leftarrow \mathbf{Z}(\widetilde{\mathbf{L}}) \widetilde{\mathbf{B}}^\top \widetilde{\mathbf{W}}^{1/2} \mathbf{g}_j$.
   (b) $\mathbf{x}_j^{\mathrm{res}} \leftarrow (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}}) \widetilde{\mathbf{L}})^3 \mathbf{x}_j^{\mathrm{rand}}$.
   (c) $\mathbf{x}_j^{\mathrm{prox}} \leftarrow \mathbf{x}_j^{\mathrm{rand}} - \mathbf{x}_j^{\mathrm{res}}$.
3. The verdict is *accepted* iff

$$\frac{1}{10^3 \log n} \sum_j \mathbf{x}_j^{\mathrm{rand}\top} \widetilde{\mathbf{L}} \mathbf{x}_j^{\mathrm{rand}} \le 1.1n,$$

$$\frac{1}{10^3 \log n} \sum_j \mathbf{x}_j^{\mathrm{res}\top} \widetilde{\mathbf{L}} \mathbf{x}_j^{\mathrm{res}} \le 10^8 (n/\theta_{\deg} + n/\log^3 n),$$

$$\frac{1}{10^3 \log n} \sum_j \mathbf{x}_j^{\mathrm{prox}\top} \mathbf{D}(G^{\mathrm{light}}) \mathbf{x}_j^{\mathrm{prox}} \le 10^3 n.$$

4. Return $(\mathbf{x}_j^{\mathrm{rand}}, \mathbf{x}_j^{\mathrm{res}}, \mathbf{x}_j^{\mathrm{prox}})_{j \le 10^3 \log n}$ and the verdict.

Figure 4: One attempt of a refresh: $10^3 \log n$ random vertex vectors, their cubic residuals and proxies, and the three acceptance tests. Its guarantees are in Lemma 4.5.

The analysis conjugates one application of $\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}}$ by $\mathbf{L}(G)^{1/2}$ and splits it into a deterministic part and a random part:

$$\begin{aligned}
\overline{\mathbf{Y}} &:= \mathbf{\Pi} - \mathbf{L}(G)^{1/2} \mathbf{M}^\dagger \mathbf{L}(G)^{1/2}, \\
\widetilde{\mathbf{Y}} &:= -\mathbf{L}(G)^{1/2} \mathbf{M}^\dagger \mathbf{L}(G)^{1/2} \Big(\sum_{e \in E} \mathbf{Q}_e \Big) \\
&\quad - \mathbf{L}(G)^{1/2} \big(\mathbf{Z}(\widetilde{\mathbf{M}}) - \widetilde{\mathbf{M}}^\dagger \big) \mathbf{L}(G)^{1/2}\, \mathbf{L}(G)^{\dagger/2} \widetilde{\mathbf{L}} \mathbf{L}(G)^{\dagger/2} \\
&\quad + \mathbf{L}(G)^{1/2} \widetilde{\mathbf{M}}^\dagger \mathbf{M}^{1/2} \Big(\sum_{e \in G^{\mathrm{heavy}}} \mathbf{Q}_e^{\mathrm{heavy}} \Big) \mathbf{M}^{\dagger/2} \mathbf{L}(G)^{1/2}\, \mathbf{L}(G)^{\dagger/2} \widetilde{\mathbf{L}} \mathbf{L}(G)^{\dagger/2}.
\end{aligned} \tag{9}$$

Here and below $\mathbf{Z}(\widetilde{\mathbf{M}})$ stands for $\mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2}\mathbf{Z}(\widetilde{\mathbf{M}})\mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2}$ and $\mathbf{Z}(\widetilde{\mathbf{L}})$ for $\mathbf{\Pi}\mathbf{Z}(\widetilde{\mathbf{L}})\mathbf{\Pi}$; the first paragraph of the proof below justifies this.

Lemma 4.5 (the refresh)**.** *Let $G = (V, E, \mathbf{w})$ be connected with the split of Definition 4.1, $\log n \ge 20$, and let $\widetilde{\mathbf{L}}, \widetilde{\mathbf{M}}$ be the references of Equation (5) with intensities satisfying Equation (7). Let $\mathbf{Z}(\widetilde{\mathbf{L}}) \leftarrow \textsc{LsolverKLPSS}(\widetilde{\mathbf{L}}, \frac{1}{40})$ and $\mathbf{Z}(\widetilde{\mathbf{M}}) \leftarrow \textsc{LsolverKLPSS}(\widetilde{\mathbf{M}}, 1/\log n)$, and call the operators* accurate *if they satisfy the conclusion of Lemma B.1: $\mathbf{Z}(\widetilde{\mathbf{L}})$ is symmetric with $\mathbf{Z}(\widetilde{\mathbf{L}}) \approx_{1/40} \widetilde{\mathbf{L}}^\dagger$ on the range of $\widetilde{\mathbf{L}}$ and $\mathbf{Z}(\widetilde{\mathbf{M}})$ is symmetric with $\mathbf{Z}(\widetilde{\mathbf{M}}) \approx_{1/\log n} \widetilde{\mathbf{M}}^\dagger$ on the range of $\widetilde{\mathbf{M}}$. An* attempt *is a run of $\textsc{Attempt}(\widetilde{\mathbf{L}}, \widetilde{\mathbf{M}}, \mathbf{Z}(\widetilde{\mathbf{L}}), \mathbf{Z}(\widetilde{\mathbf{M}}))$ (Figure 4) whose Gaussian vectors are independent of the retention variables and of the randomness of the two solver calls.*

(a) (acceptance) The probability, over the retention variables, that the references are accurate and

$$\Big\|\sum_{e\in E}\mathbf{Q}_e\Big\|_{S_6}^6+\Big\|\sum_{e\in G^{\mathrm{heavy}}}\mathbf{Q}_e^{\mathrm{heavy}}\Big\|_{S_6}^6 \;>\; 10^{-16}\,\frac{n}{\log^3 n} \tag{10}$$

is at most 10^{-2}. If the references and the operators are accurate and Equation (10) fails, an attempt is accepted with probability at least $1-3n^{-15}$ over its Gaussian vectors. If the references are accurate, an accepted attempt satisfies

$$\begin{aligned}\frac{1}{10^3\log n}\sum_j \mathbf{x}_j^{\mathrm{rand}\top}\mathbf{L}(G)\mathbf{x}_j^{\mathrm{rand}} &\le 1.2n,\\ \frac{1}{10^3\log n}\sum_j \mathbf{x}_j^{\mathrm{res}\top}\mathbf{L}(G)\mathbf{x}_j^{\mathrm{res}} &\le 2\cdot 10^8(n/\theta_{\deg}+n/\log^3 n),\\ \frac{1}{10^3\log n}\sum_j \mathbf{x}_j^{\mathrm{prox}\top}\mathbf{D}(G^{\mathrm{light}})\mathbf{x}_j^{\mathrm{prox}} &\le 10^3 n.\end{aligned} \tag{11}$$

Consequently an attempt with fresh retention variables, fresh solver calls and fresh Gaussian vectors is accepted with probability at least $1-10^{-2}-3n^{-15}$ minus the probability that the references or the operators fail to be accurate.

(b) (all pairs) If the references and the operators are accurate, then with probability $1-n^{-14}$ over the Gaussian vectors, for all $u,v\in V$,

$$\frac{1}{10^3\log n}\sum_j\big((\mathbf{x}_j^{\mathrm{rand}})_u-(\mathbf{x}_j^{\mathrm{rand}})_v\big)^2\in\big[\tfrac12,2\big]\cdot(\mathbb{1}_u-\mathbb{1}_v)^\top\mathbf{L}(G)^\dagger(\mathbb{1}_u-\mathbb{1}_v).$$

(c) (cost) If $\widetilde{\mathbf{L}}$ and $\widetilde{\mathbf{M}}$ have $O(n\log n)$ nonzero entries, an attempt takes $\ddot{O}(n\log^2 n)$ time beyond the two calls of LSOLVERKLPSS*, on a machine with a unit-cost primitive returning a standard Gaussian: $O(n\log^2 n)$ Gaussian draws and arithmetic operations, $10^3\log n$ applications of $\mathbf{Z}(\widetilde{\mathbf{L}})$ and $3\cdot 10^3\log n$ applications of $\mathbf{Z}(\widetilde{\mathbf{M}})$.*

Proof. Conventions. By Lemma 4.2(d), $\mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2}$ is the orthogonal projection onto the range of $\mathbf{M}$, which is $\mathbb{R}^V$ if $G^{\mathrm{light}}\neq\emptyset$ and $\mathbb{1}^\perp$ if $G^{\mathrm{light}}=\emptyset$; when the references are accurate, $\widetilde{\mathbf{M}}$ has the same range as $\mathbf{M}$, and $\widetilde{\mathbf{L}}$ has range $\mathbb{1}^\perp$. In Figure 4, $\mathbf{Z}(\widetilde{\mathbf{L}})$ is applied to the vector $\widetilde{\mathbf{B}}^\top\widetilde{\mathbf{W}}^{1/2}\mathbf{g}_j\perp\mathbb{1}$ only, and $\mathbf{Z}(\widetilde{\mathbf{M}})$ to vectors of the form $\widetilde{\mathbf{L}}\mathbf{x}\perp\mathbb{1}$ only; their outputs enter the attempt and the statement only through the quadratic forms of $\widetilde{\mathbf{L}}$ and $\mathbf{L}(G)$, through $\mathbf{D}(G^{\mathrm{light}})$, which vanishes unless the range of $\mathbf{M}$ is $\mathbb{R}^V$, through differences of two coordinates, and through further applications of $\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}}$, which maps $\mathbb{1}$ to $\mathbb{1}$. Hence replacing $\mathbf{Z}(\widetilde{\mathbf{L}})$ by $\mathbf{\Pi}\mathbf{Z}(\widetilde{\mathbf{L}})\mathbf{\Pi}$ and $\mathbf{Z}(\widetilde{\mathbf{M}})$ by $\mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2}\mathbf{Z}(\widetilde{\mathbf{M}})\mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2}$ changes every $\mathbf{x}_j^{\mathrm{rand}}$, $\mathbf{x}_j^{\mathrm{res}}$, $\mathbf{x}_j^{\mathrm{prox}}$ by a multiple of $\mathbb{1}$ and none of the quantities of the statement; we make this replacement, as announced after Equation (9). The accuracy of the operators then reads, as Loewner inequalities on $\mathbb{R}^V$,

$$\tfrac{39}{40}\widetilde{\mathbf{L}}^\dagger\preceq\mathbf{Z}(\widetilde{\mathbf{L}})\preceq\tfrac{41}{40}\widetilde{\mathbf{L}}^\dagger,\qquad\Big(1-\frac{1}{\log n}\Big)\widetilde{\mathbf{M}}^\dagger\preceq\mathbf{Z}(\widetilde{\mathbf{M}})\preceq\Big(1+\frac{1}{\log n}\Big)\widetilde{\mathbf{M}}^\dagger, \tag{12}$$

and $\mathbf{Z}(\widetilde{\mathbf{L}})$, $\mathbf{Z}(\widetilde{\mathbf{M}})$ are supported on the ranges of $\widetilde{\mathbf{L}}$, $\widetilde{\mathbf{M}}$.

The cubic identity. Since $\widetilde{\mathbf{L}}\mathbb{1}=0$, $(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})\mathbb{1}=\mathbb{1}$, so $\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})(\mathbf{I}-\mathbf{\Pi})=0$ and, as $\mathbf{\Pi}=\mathbf{L}(G)^{\dagger/2}\mathbf{L}(G)^{1/2}$,

$$\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})=\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})\mathbf{\Pi}=\big(\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})\mathbf{L}(G)^{\dagger/2}\big)\mathbf{L}(G)^{1/2}.$$

Iterating three times,

$$\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3 = \big(\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})\mathbf{L}(G)^{\dagger/2}\big)^3\mathbf{L}(G)^{1/2}. \tag{13}$$

We show $\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})\mathbf{L}(G)^{\dagger/2} = \overline{\mathbf{Y}}+\widetilde{\mathbf{Y}}$ when the references are accurate. As $\mathbf{\Pi}\widetilde{\mathbf{L}}\mathbf{\Pi} = \widetilde{\mathbf{L}}$, $\widetilde{\mathbf{L}} = \mathbf{L}(G)^{1/2}(\mathbf{L}(G)^{\dagger/2}\widetilde{\mathbf{L}}\mathbf{L}(G)^{\dagger/2})\mathbf{L}(G)^{1/2}$, so

$$\mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})\mathbf{L}(G)^{\dagger/2} = \mathbf{\Pi} - \mathbf{L}(G)^{1/2}\mathbf{Z}(\widetilde{\mathbf{M}})\mathbf{L}(G)^{1/2}\,\mathbf{L}(G)^{\dagger/2}\widetilde{\mathbf{L}}\mathbf{L}(G)^{\dagger/2}.$$

Since $\mathbf{M}$ and $\widetilde{\mathbf{M}}$ have the same range, $\widetilde{\mathbf{M}}^\dagger - \mathbf{M}^\dagger = \widetilde{\mathbf{M}}^\dagger(\mathbf{M}-\widetilde{\mathbf{M}})\mathbf{M}^\dagger$, and by Equation (6) and $\mathbf{M}^{1/2}(\mathbf{M}^{\dagger/2}\widetilde{\mathbf{M}}\mathbf{M}^{\dagger/2} - \mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2})\mathbf{M}^{1/2} = \widetilde{\mathbf{M}}-\mathbf{M}$,

$$\mathbf{Z}(\widetilde{\mathbf{M}}) = \mathbf{M}^\dagger + \big(\mathbf{Z}(\widetilde{\mathbf{M}}) - \widetilde{\mathbf{M}}^\dagger\big) - \widetilde{\mathbf{M}}^\dagger\mathbf{M}^{1/2}\Big(\sum_{e\in G^{\text{heavy}}}\mathbf{Q}_e^{\text{heavy}}\Big)\mathbf{M}^{1/2}\mathbf{M}^\dagger.$$

Substituting this, $\mathbf{L}(G)^{\dagger/2}\widetilde{\mathbf{L}}\mathbf{L}(G)^{\dagger/2} = \mathbf{\Pi} + \sum_e \mathbf{Q}_e$ for the first term, $\mathbf{L}(G)^{1/2}\mathbf{\Pi} = \mathbf{L}(G)^{1/2}$ and $\mathbf{M}^{1/2}\mathbf{M}^\dagger = \mathbf{M}^{\dagger/2}$ gives exactly $\overline{\mathbf{Y}}+\widetilde{\mathbf{Y}}$ of Equation (9). Consequently, by Equation (13), for every j

$$\mathbf{x}_j^{\text{res}\top}\mathbf{L}(G)\mathbf{x}_j^{\text{res}} = \|(\overline{\mathbf{Y}}+\widetilde{\mathbf{Y}})^3\mathbf{L}(G)^{1/2}\mathbf{x}_j^{\text{rand}}\|^2, \qquad \mathbf{L}(G)^{1/2}(\mathbf{I}-\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\mathbf{L}(G)^{\dagger/2} = (\overline{\mathbf{Y}}+\widetilde{\mathbf{Y}})^3, \tag{14}$$

the second because $\overline{\mathbf{Y}}\mathbf{\Pi} = \overline{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}\mathbf{\Pi} = \widetilde{\mathbf{Y}}$.

The deterministic part. For $\mathbf{u}\in\mathbb{R}^V$, $\mathbf{L}(G)^{1/2}\mathbf{u}\perp\mathbb{1}$, so Lemma 4.2(e) gives

$$\mathbf{u}^\top\mathbf{L}(G)^{1/2}\mathbf{M}^\dagger\mathbf{L}(G)^{1/2}\mathbf{u} \le 2\mathbf{u}^\top\mathbf{L}(G)^{1/2}\mathbf{L}(G)^\dagger\mathbf{L}(G)^{1/2}\mathbf{u} = 2\|\mathbf{\Pi}\mathbf{u}\|^2;$$

hence $0\preceq\mathbf{L}(G)^{1/2}\mathbf{M}^\dagger\mathbf{L}(G)^{1/2}\preceq 2\mathbf{\Pi}$, $-\mathbf{\Pi}\preceq\overline{\mathbf{Y}}\preceq\mathbf{\Pi}$ and $\|\overline{\mathbf{Y}}\|\le 1$. With $\mathbf{L}(G) = \mathbf{M}-\mathbf{A}(G^{\text{light}})$ and $r := \operatorname{rank}(\mathbf{M})\in\{n-1,n\}$,

$$\begin{aligned}
\|\overline{\mathbf{Y}}\|_F^2 &= \operatorname{Tr}[\mathbf{\Pi}] - 2\operatorname{Tr}[\mathbf{M}^\dagger\mathbf{L}(G)] + \operatorname{Tr}[\mathbf{M}^\dagger\mathbf{L}(G)\mathbf{M}^\dagger\mathbf{L}(G)]\\
&= (n-1) - 2\big(r - \operatorname{Tr}[\mathbf{M}^\dagger\mathbf{A}(G^{\text{light}})]\big)\\
&\quad + \big(r - 2\operatorname{Tr}[\mathbf{M}^\dagger\mathbf{A}(G^{\text{light}})] + \operatorname{Tr}[\mathbf{M}^\dagger\mathbf{A}(G^{\text{light}})\mathbf{M}^\dagger\mathbf{A}(G^{\text{light}})]\big)\\
&= (n-1-r) + \operatorname{Tr}[\mathbf{M}^\dagger\mathbf{A}(G^{\text{light}})\mathbf{M}^\dagger\mathbf{A}(G^{\text{light}})] \le \frac{32n}{\theta_{\text{deg}}}
\end{aligned}$$

by Lemma 4.2(g), where we used $\operatorname{Tr}[\mathbf{M}^\dagger\mathbf{M}] = r$ and $\mathbf{M}^\dagger\mathbf{M}\mathbf{M}^\dagger\mathbf{A}(G^{\text{light}}) = \mathbf{M}^\dagger\mathbf{A}(G^{\text{light}})$.

Norm bounds. Suppose the references and the operators are accurate. From Equation (6) and Equation (8), $\mathbf{L}(G)^{\dagger/2}\widetilde{\mathbf{L}}\mathbf{L}(G)^{\dagger/2}$ lies between $0.999\mathbf{\Pi}$ and $1.001\mathbf{\Pi}$ and $\mathbf{M}^{\dagger/2}\widetilde{\mathbf{M}}\mathbf{M}^{\dagger/2}$ between

$0.999\mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2}$ and $1.001\mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2}$, so

$$\begin{aligned}
\Big\|\sum_{e\in E}\mathbf{Q}_e\Big\| &\le 10^{-3},\\
\Big\|\sum_{e\in G^{\text{heavy}}}\mathbf{Q}_e^{\text{heavy}}\Big\| &\le 10^{-3},\\
\|\mathbf{L}(G)^{\dagger/2}\widetilde{\mathbf{L}}\mathbf{L}(G)^{\dagger/2}\| &\le 1.001,\\
\|\mathbf{L}(G)^{1/2}\mathbf{M}^{\dagger}\mathbf{L}(G)^{1/2}\| &\le 2,\\
\|\mathbf{M}^{\dagger/2}\mathbf{L}(G)^{1/2}\| = \|\mathbf{L}(G)^{1/2}\mathbf{M}^{\dagger}\mathbf{L}(G)^{1/2}\|^{1/2} &\le 1.5,\\
\|\mathbf{L}(G)^{1/2}\widetilde{\mathbf{M}}^{\dagger/2}\|^2 = \|\widetilde{\mathbf{M}}^{\dagger/2}\mathbf{L}(G)\widetilde{\mathbf{M}}^{\dagger/2}\| &\le 2\|\widetilde{\mathbf{M}}^{\dagger/2}\mathbf{M}\widetilde{\mathbf{M}}^{\dagger/2}\| \le \frac{2}{0.999} \le 2.01,\\
\|\widetilde{\mathbf{M}}^{\dagger/2}\mathbf{M}^{1/2}\|^2 = \|\widetilde{\mathbf{M}}^{\dagger/2}\mathbf{M}\widetilde{\mathbf{M}}^{\dagger/2}\| &\le \frac{1}{0.999} \le 1.002,\\
\|\mathbf{L}(G)^{1/2}\widetilde{\mathbf{M}}^{\dagger}\mathbf{M}^{1/2}\| &\le \|\mathbf{L}(G)^{1/2}\widetilde{\mathbf{M}}^{\dagger/2}\|\cdot\|\widetilde{\mathbf{M}}^{\dagger/2}\mathbf{M}^{1/2}\|\\
&\le \sqrt{2.01\cdot 1.002} \le 1.5,
\end{aligned}$$

using $\mathbf{L}(G)\preceq 2\mathbf{M}$ (Lemma 4.2(c)) and $\mathbf{M}\preceq\widetilde{\mathbf{M}}/0.999$. As $\mathbf{Z}(\widetilde{\mathbf{M}})$ and $\widetilde{\mathbf{M}}^{\dagger}$ are supported on the range of $\widetilde{\mathbf{M}}$, $\mathbf{Z}(\widetilde{\mathbf{M}})-\widetilde{\mathbf{M}}^{\dagger}=\widetilde{\mathbf{M}}^{\dagger/2}(\widetilde{\mathbf{M}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{M}}^{1/2}-\widetilde{\mathbf{M}}^{\dagger/2}\widetilde{\mathbf{M}}\widetilde{\mathbf{M}}^{\dagger/2})\widetilde{\mathbf{M}}^{\dagger/2}$, and the middle factor has norm at most $1/\log n$ by Equation (12); so

$$\|\mathbf{L}(G)^{1/2}(\mathbf{Z}(\widetilde{\mathbf{M}})-\widetilde{\mathbf{M}}^{\dagger})\mathbf{L}(G)^{1/2}\| \le \|\mathbf{L}(G)^{1/2}\widetilde{\mathbf{M}}^{\dagger/2}\|^2\cdot\frac{1}{\log n}\le\frac{2.01}{\log n}.$$

Applying these to the three terms of Equation (9), with $\log n\ge 20$,

$$\|\widetilde{\mathbf{Y}}\| \le 2\cdot 10^{-3}+\frac{2.01\cdot 1.001}{\log n}+1.5\cdot 10^{-3}\cdot 1.5\cdot 1.001 \le 2\cdot 10^{-3}+0.101+2.3\cdot 10^{-3}\le\tfrac{1}{2}.$$

For the Schatten norm, $\|\mathbf{N}\mathbf{N}'\|_{S_6}\le\|\mathbf{N}\|\cdot\|\mathbf{N}'\|_{S_6}$, $\|\mathbf{N}\mathbf{N}'\|_{S_6}\le\|\mathbf{N}\|_{S_6}\cdot\|\mathbf{N}'\|$ and

$$\|\mathbf{L}(G)^{\dagger/2}\widetilde{\mathbf{L}}\mathbf{L}(G)^{\dagger/2}\|_{S_6}\le 1.001\|\mathbf{\Pi}\|_{S_6}\le 1.001\,n^{1/6}$$

give

$$\|\widetilde{\mathbf{Y}}\|_{S_6}\le 2\Big\|\sum_{e\in E}\mathbf{Q}_e\Big\|_{S_6}+\frac{2.02\,n^{1/6}}{\log n}+2.3\Big\|\sum_{e\in G^{\text{heavy}}}\mathbf{Q}_e^{\text{heavy}}\Big\|_{S_6},$$

and $(x+y+z)^6\le 3^5(x^6+y^6+z^6)$ for nonnegative x,y,z gives

$$\|\widetilde{\mathbf{Y}}\|_{S_6}^6\le 243\Big(149\Big(\Big\|\sum_{e\in E}\mathbf{Q}_e\Big\|_{S_6}^6+\Big\|\sum_{e\in G^{\text{heavy}}}\mathbf{Q}_e^{\text{heavy}}\Big\|_{S_6}^6\Big)+\frac{69\,n}{\log^6 n}\Big),\tag{15}$$

as $2^6\le 2.3^6\le 149$ and $2.02^6\le 69$.

The sixth moments.

Claim 4.6. *For every $e\in E_{\text{s}}$ and $p=2,\dots,6$, $\mathbb{E}[|\zeta_e|^p]\,(\mathbf{w}_e^{\text{s}}\mathcal{R}^{\text{eff}}(G)_e)^{p-1}\le(10^{-7}/\log n)^{p-1}$.*

Proof of Claim 4.6. $\mathbb{E}[|\zeta_e|^p]=(1-p_e)+p_e(1/p_e-1)^p$ and $\mathbf{w}_e^{\text{s}}\mathcal{R}^{\text{eff}}(G)_e\le 10^{-8}\varrho_e/\log n$ by Equation (7).

- If $\varrho_e < 1$: $p_e = 1 - e^{-\varrho_e} \geq (1 - e^{-1})\varrho_e \geq 0.63\varrho_e$ by concavity, and $\mathbb{E}[|\zeta_e|^p] \leq 1 + p_e^{1-p} \leq 2p_e^{1-p}$, so

$$\mathbb{E}[|\zeta_e|^p](\mathbf{w}_e^{\mathrm{s}}\mathcal{R}^{\mathrm{eff}}(G)_e)^{p-1} \leq 2\Big(\frac{\mathbf{w}_e^{\mathrm{s}}\mathcal{R}^{\mathrm{eff}}(G)_e}{p_e}\Big)^{p-1} \leq 2\Big(\frac{10^{-8}}{0.63\log n}\Big)^{p-1} \leq \Big(\frac{10^{-7}}{\log n}\Big)^{p-1}.$$

- If $\varrho_e \geq 1$: $p_e \geq 1 - e^{-1} \geq 0.63$ and $1/p_e - 1 = e^{-\varrho_e}/p_e$, so $\mathbb{E}[|\zeta_e|^p] \leq e^{-\varrho_e} + p_e^{1-p}e^{-p\varrho_e} \leq e^{-\varrho_e}(1 + 0.63^{1-p}e^{-(p-1)}) \leq 2e^{-\varrho_e}$, the last because $0.63^{1-p}e^{-(p-1)} = (0.63e)^{1-p} \leq 1$; hence

$$\mathbb{E}[|\zeta_e|^p](\mathbf{w}_e^{\mathrm{s}}\mathcal{R}^{\mathrm{eff}}(G)_e)^{p-1} \leq 2\varrho_e^{p-1}e^{-\varrho_e}\Big(\frac{10^{-8}}{\log n}\Big)^{p-1} \leq 2\Big(\frac{p-1}{e}\Big)^{p-1}\Big(\frac{10^{-8}}{\log n}\Big)^{p-1} \leq \Big(\frac{10^{-7}}{\log n}\Big)^{p-1},$$

as $\max_{\varrho>0} \varrho^{p-1}e^{-\varrho} = ((p-1)/e)^{p-1}$ and $2((p-1)/e)^{p-1} \leq 10^{p-1}$ for $p \leq 6$. □

Lemma 4.4 applies to $\mathbf{Q}_e$, $e \in E_{\mathrm{s}}$ (the other $\mathbf{Q}_e$ vanish), with the vectors $\mathbf{u}_e = (\mathbf{w}_e^{\mathrm{s}})^{1/2}\mathbf{L}(G)^{\dagger/2}\boldsymbol{\chi}_e$, for which $\|\mathbf{u}_e\|^2 = \mathbf{w}_e^{\mathrm{s}}\mathcal{R}^{\mathrm{eff}}(G)_e$ and $\sum_e \mathbf{u}_e\mathbf{u}_e^\top \preceq \sum_e \mathbf{w}_e\mathbf{L}(G)^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top\mathbf{L}(G)^{\dagger/2} = \boldsymbol{\Pi} \preceq \mathbf{I}$, and $a = 10^{-7}/\log n$ by Claim 4.6; and to $\mathbf{Q}_e^{\mathrm{heavy}}$, $e \in G^{\mathrm{heavy}} \cap E_{\mathrm{s}}$, with the vectors $\mathbf{u}_e = (\mathbf{w}_e^{\mathrm{s}})^{1/2}\mathbf{M}^{\dagger/2}\boldsymbol{\chi}_e$, for which $\|\mathbf{u}_e\|^2 = \mathbf{w}_e^{\mathrm{s}}\boldsymbol{\chi}_e^\top\mathbf{M}^\dagger\boldsymbol{\chi}_e \leq 2\mathbf{w}_e^{\mathrm{s}}\mathcal{R}^{\mathrm{eff}}(G)_e$ (Lemma 4.2(e)), $\sum_{e\in G^{\mathrm{heavy}}} \mathbf{u}_e\mathbf{u}_e^\top \preceq \mathbf{M}^{\dagger/2}\mathbf{L}(G^{\mathrm{heavy}})\mathbf{M}^{\dagger/2} \preceq \mathbf{M}^{\dagger/2}\mathbf{M}\mathbf{M}^{\dagger/2} \preceq \mathbf{I}$, and $a = 2 \cdot 10^{-7}/\log n$ (the factor 2^{p-1} from the norm bound). Therefore

$$\mathbb{E}\Big[\Big\|\sum_{e\in E}\mathbf{Q}_e\Big\|_{S_6}^6 + \Big\|\sum_{e\in G^{\mathrm{heavy}}}\mathbf{Q}_e^{\mathrm{heavy}}\Big\|_{S_6}^6\Big] \leq 10^2\big(10^{-21} + 8\cdot 10^{-21}\big)\frac{n}{\log^3 n} = 9\cdot 10^{-19}\,\frac{n}{\log^3 n},$$

the expectation over the retention variables alone, and Markov's inequality bounds the probability of Equation (10) by $9 \cdot 10^{-3} \leq 10^{-2}$. This is the first sentence of (a).

The residual on good references. Suppose the references and the operators are accurate and Equation (10) fails. By Equation (15) and $\log^6 n \geq \log^3 n$, $\|\widetilde{\mathbf{Y}}\|_{S_6}^6 \leq 243(149\cdot 10^{-16}+69)\,n/\log^3 n \leq 1.7 \cdot 10^4\, n/\log^3 n$. The matrix $(\overline{\mathbf{Y}} + \widetilde{\mathbf{Y}})^3$ is the sum of eight words in $\overline{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}$. The word $\widetilde{\mathbf{Y}}^3$ has $\|\widetilde{\mathbf{Y}}^3\|_F \leq \|\widetilde{\mathbf{Y}}\|_{S_6}^3$ (Hölder's inequality for Schatten norms, $\frac{1}{6}+\frac{1}{6}+\frac{1}{6} = \frac{1}{2}$), and each of the other seven contains a factor $\overline{\mathbf{Y}}$ and has Frobenius norm at most $\|\overline{\mathbf{Y}}\|_F \max\{1, \|\widetilde{\mathbf{Y}}\|\}^2 \leq \|\overline{\mathbf{Y}}\|_F$, by $\|\overline{\mathbf{Y}}\| \leq 1$ and $\|\widetilde{\mathbf{Y}}\| \leq \frac{1}{2}$. Hence, by $(7x+y)^2 \leq 98x^2 + 2y^2$,

$$\|(\overline{\mathbf{Y}} + \widetilde{\mathbf{Y}})^3\|_F^2 \leq 98\|\overline{\mathbf{Y}}\|_F^2 + 2\|\widetilde{\mathbf{Y}}\|_{S_6}^6 \leq 98\cdot\frac{32n}{\theta_{\mathrm{deg}}} + 3.4\cdot 10^4\,\frac{n}{\log^3 n} \leq 10^5\Big(\frac{n}{\theta_{\mathrm{deg}}} + \frac{n}{\log^3 n}\Big). \tag{16}$$

The three statistics. Given the references and the operators, each $\mathbf{x}_j^{\mathrm{rand}}$ is a Gaussian vector with covariance $\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{L}})$, and the three statistics of Step 3 of Figure 4 are the averages over $j \leq 10^3 \log n$ of the independent quadratic forms $\mathbf{g}_j^\top\mathbf{N}_1\mathbf{g}_j$, $\mathbf{g}_j^\top\mathbf{N}_2\mathbf{g}_j$, $\mathbf{g}_j^\top\mathbf{N}_3\mathbf{g}_j$ with

$$\begin{aligned}
\mathbf{N}_1 &\coloneqq \widetilde{\mathbf{W}}^{1/2}\widetilde{\mathbf{B}}\mathbf{Z}(\widetilde{\mathbf{L}})\,\widetilde{\mathbf{L}}\,\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{B}}^\top\widetilde{\mathbf{W}}^{1/2},\\
\mathbf{N}_2 &\coloneqq \widetilde{\mathbf{W}}^{1/2}\widetilde{\mathbf{B}}\mathbf{Z}(\widetilde{\mathbf{L}})\big((\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)^\top\widetilde{\mathbf{L}}(\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{B}}^\top\widetilde{\mathbf{W}}^{1/2},\\
\mathbf{N}_3 &\coloneqq \widetilde{\mathbf{W}}^{1/2}\widetilde{\mathbf{B}}\mathbf{Z}(\widetilde{\mathbf{L}})\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)^\top\mathbf{D}(G^{\mathrm{light}})\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{B}}^\top\widetilde{\mathbf{W}}^{1/2}.
\end{aligned} \tag{17}$$

Suppose the references and the operators are accurate. By Equation (12), $\widetilde{\mathbf{L}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}^{1/2}$ lies between $\frac{39}{40}\boldsymbol{\Pi}$ and $\frac{41}{40}\boldsymbol{\Pi}$, so

$$\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{L}}) = \widetilde{\mathbf{L}}^{\dagger/2}\big(\widetilde{\mathbf{L}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}^{1/2}\big)^2\widetilde{\mathbf{L}}^{\dagger/2} \in \Big[\Big(\frac{39}{40}\Big)^2, \Big(\frac{41}{40}\Big)^2\Big]\cdot\widetilde{\mathbf{L}}^\dagger \subseteq [0.94,\, 1.07]\cdot\mathbf{L}(G)^\dagger, \tag{18}$$

using $\widetilde{\mathbf{L}}^\dagger \in [1/1.001, 1/0.999] \cdot \mathbf{L}(G)^\dagger$. The nonzero eigenvalues of $\mathbf{N}_1$ are those of

$$\widetilde{\mathbf{L}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}^{1/2} = (\widetilde{\mathbf{L}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}^{1/2})^2,$$

so

$$\textsc{Tr}[\mathbf{N}_1] \le \left(\frac{41}{40}\right)^2 (n-1) \le 1.06n, \qquad \|\mathbf{N}_1\| \le \left(\frac{41}{40}\right)^2 \le 1.07.$$

For $\mathbf{N}_2$, by Equation (18), $\widetilde{\mathbf{L}} \preceq 1.001\mathbf{L}(G)$ and Equation (14),

$$\begin{aligned}
\textsc{Tr}[\mathbf{N}_2] &= \textsc{Tr}\big[\big((\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)^\top \widetilde{\mathbf{L}}(\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\, \mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{L}})\big] \\
&\le 1.07 \cdot 1.001\, \textsc{Tr}\big[\big((\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)^\top \mathbf{L}(G)(\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\mathbf{L}(G)^\dagger\big] \\
&= 1.08\, \|\mathbf{L}(G)^{1/2}(\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\mathbf{L}(G)^{\dagger/2}\|_F^2 = 1.08\, \|(\overline{\mathbf{Y}} + \widetilde{\mathbf{Y}})^3\|_F^2,
\end{aligned}$$

and likewise, the nonzero eigenvalues of $\mathbf{N}_2$ being those of $\widetilde{\mathbf{L}}^{1/2}(\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\, \mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{L}})\, ((\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3)^\top \widetilde{\mathbf{L}}^{1/2}$,

$$\|\mathbf{N}_2\| \le 1.07 \cdot 1.001\, \|\mathbf{L}(G)^{1/2}(\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\mathbf{L}(G)^{\dagger/2}\|^2 = 1.08\, \|(\overline{\mathbf{Y}} + \widetilde{\mathbf{Y}})^3\|^2 \le 1.08 \cdot (1 + \tfrac{1}{2})^6 \le 13.$$

For $\mathbf{N}_3$ let $G^{\text{light}} \neq \emptyset$ (otherwise $\mathbf{D}(G^{\text{light}}) = 0$ and $\mathbf{N}_3 = 0$), so that $\mathbf{M}$ and $\widetilde{\mathbf{M}}$ are positive definite. Expanding the cube,

$$\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3 = \mathbf{Z}(\widetilde{\mathbf{M}})\big(3\mathbf{I} - 3\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}}) + \widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}})\big)\widetilde{\mathbf{L}}.$$

By Equation (12), $\|\widetilde{\mathbf{M}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{M}}^{1/2}\| \le 1 + 1/\log n \le 1.05$; by $\widetilde{\mathbf{L}} \preceq 1.001\mathbf{L}(G) \preceq 2.002\mathbf{M} \preceq 2.01\widetilde{\mathbf{M}}$, $\|\widetilde{\mathbf{M}}^{-1/2}\widetilde{\mathbf{L}}\widetilde{\mathbf{M}}^{-1/2}\| \le 2.01$; hence

$$\begin{aligned}
&\|\widetilde{\mathbf{M}}^{-1/2}\big(3\mathbf{I} - 3\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}}) + \widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}})\big)\widetilde{\mathbf{M}}^{1/2}\| \\
&\quad = \|3\mathbf{I} - 3(\widetilde{\mathbf{M}}^{-1/2}\widetilde{\mathbf{L}}\widetilde{\mathbf{M}}^{-1/2})(\widetilde{\mathbf{M}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{M}}^{1/2}) + \big((\widetilde{\mathbf{M}}^{-1/2}\widetilde{\mathbf{L}}\widetilde{\mathbf{M}}^{-1/2})(\widetilde{\mathbf{M}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{M}}^{1/2})\big)^2\| \\
&\quad \le 3 + 3 \cdot 2.01 \cdot 1.05 + (2.01 \cdot 1.05)^2 \le 15.2.
\end{aligned}$$

Since $\mathbf{L}(G)^\dagger \preceq 1.001\widetilde{\mathbf{L}}^\dagger$, we have $\widetilde{\mathbf{L}}\mathbf{L}(G)^\dagger\widetilde{\mathbf{L}} \preceq 1.001\widetilde{\mathbf{L}} \preceq 2.02\widetilde{\mathbf{M}}$, and therefore

$$\begin{aligned}
&\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)\mathbf{L}(G)^\dagger\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)^\top \\
&\quad \preceq 2.02\, \mathbf{Z}(\widetilde{\mathbf{M}})\big(3\mathbf{I} - 3\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}}) + \widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{M}})\big)\widetilde{\mathbf{M}}(\cdots)^\top \mathbf{Z}(\widetilde{\mathbf{M}}) \\
&\quad \preceq 2.02 \cdot 15.2^2\, \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{M}}\mathbf{Z}(\widetilde{\mathbf{M}}) \preceq 2.02 \cdot 15.2^2 \cdot 1.05^2\, \widetilde{\mathbf{M}}^{-1} \preceq \frac{2.02 \cdot 15.2^2 \cdot 1.05^2}{0.999}\, \mathbf{M}^{-1} \preceq 520\, \mathbf{M}^{-1},
\end{aligned}$$

where the second step writes $\mathbf{Z}(\widetilde{\mathbf{M}})(\cdots)\widetilde{\mathbf{M}}^{1/2} = \widetilde{\mathbf{M}}^{-1/2}(\widetilde{\mathbf{M}}^{1/2}\mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{M}}^{1/2})(\widetilde{\mathbf{M}}^{-1/2}(\cdots)\widetilde{\mathbf{M}}^{1/2})$ and bounds the last factor by its norm. Consequently, by Equation (18), Lemma 4.2(f) and $\mathbf{D}(G^{\text{light}}) \preceq \mathbf{M}$ (Lemma 4.2(c)),

$$\begin{aligned}
\textsc{Tr}[\mathbf{N}_3] &\le 1.07\, \textsc{Tr}\big[\mathbf{D}(G^{\text{light}})\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)\mathbf{L}(G)^\dagger\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)^\top\big] \\
&\le 1.07 \cdot 520\, \textsc{Tr}[\mathbf{D}(G^{\text{light}})\mathbf{M}^{-1}] \le 560n, \\
\|\mathbf{N}_3\| &\le 1.07\, \|\mathbf{D}(G^{\text{light}})^{1/2}\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)\mathbf{L}(G)^\dagger\big(\mathbf{I} - (\mathbf{I} - \mathbf{Z}(\widetilde{\mathbf{M}})\widetilde{\mathbf{L}})^3\big)^\top \mathbf{D}(G^{\text{light}})^{1/2}\| \\
&\le 1.07 \cdot 520\, \|\mathbf{D}(G^{\text{light}})^{1/2}\mathbf{M}^{-1}\mathbf{D}(G^{\text{light}})^{1/2}\| \le 560.
\end{aligned}$$

The tests. Fix $\mathbf{N} \in \{\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3\}$. The stacked vector $(\mathbf{g}_1, \dots, \mathbf{g}_{10^3 \log n})$ has independent standard Gaussian entries, and $\sum_j \mathbf{g}_j^\top \mathbf{N} \mathbf{g}_j$ is its quadratic form with the block diagonal matrix with $10^3 \log n$ diagonal blocks $\mathbf{N}$, of trace $10^3 \log n \cdot \mathrm{Tr}[\mathbf{N}]$, operator norm $\|\mathbf{N}\|$ and trace of the square $10^3 \log n \cdot \mathrm{Tr}[\mathbf{N}^2] \le 10^3 \log n \cdot \|\mathbf{N}\| \mathrm{Tr}[\mathbf{N}]$. Lemma 4.3 with $t = 15 \ln n$, divided by $10^3 \log n \ge 10^3 \ln n$, gives with probability at least $1 - n^{-15}$

$$\frac{1}{10^3 \log n} \sum_j \mathbf{g}_j^\top \mathbf{N} \mathbf{g}_j \le \mathrm{Tr}[\mathbf{N}] + 2\sqrt{0.015 \, \|\mathbf{N}\| \mathrm{Tr}[\mathbf{N}]} + 0.03 \|\mathbf{N}\|. \tag{19}$$

- For $\mathbf{N}_1$ the right-hand side is at most $1.06n + 2\sqrt{0.015 \cdot 1.07 \cdot 1.06n} + 0.04 \le 1.06n + 0.27\sqrt{n} + 0.04 \le 1.1n$, as $n \ge 2^{20}$.
- For $\mathbf{N}_2$, if moreover Equation (10) fails, $\mathrm{Tr}[\mathbf{N}_2] \le 1.08 \cdot 10^5 (n/\theta_{\deg} + n/\log^3 n)$ by Equation (16), and $2\sqrt{0.015 \cdot 13 \, \mathrm{Tr}[\mathbf{N}_2]} \le 0.04 \mathrm{Tr}[\mathbf{N}_2] + 4.9$ by $2\sqrt{xy} \le x + y$, so the right-hand side is at most $1.04 \cdot 1.08 \cdot 10^5 (n/\theta_{\deg} + n/\log^3 n) + 5.3 \le 10^8 (n/\theta_{\deg} + n/\log^3 n)$, as $n/\log^3 n \ge 1$.
- For $\mathbf{N}_3$ the right-hand side is at most $560n + 2\sqrt{0.015 \cdot 560 \cdot 560n} + 17 \le 560n + 137\sqrt{n} + 17 \le 10^3 n$.

So if the references and the operators are accurate and Equation (10) fails, the three tests pass simultaneously with probability at least $1 - 3n^{-15}$, which is the second sentence of (a). If the references are accurate then $\widetilde{\mathbf{L}} \succeq 0.999 \mathbf{L}(G)$, so on acceptance the first two energies of Equation (11) are at most $1.1n/0.999 \le 1.2n$ and $10^8 (n/\theta_{\deg} + n/\log^3 n)/0.999 \le 2 \cdot 10^8 (n/\theta_{\deg} + n/\log^3 n)$, and the third test is the third energy; this is the third sentence of (a). The last sentence of (a) is the union bound over the failure of the references to be accurate, the failure of the operators to be accurate, the event that the references are accurate and Equation (10) holds (probability at most 10^{-2}), and the failure of the tests (probability at most $3n^{-15}$).

All pairs. Suppose the references and the operators are accurate and fix $u \ne v$. The variables $(\mathbf{x}_j^{\mathrm{rand}})_u - (\mathbf{x}_j^{\mathrm{rand}})_v = (\mathbb{1}_u - \mathbb{1}_v)^\top \mathbf{Z}(\widetilde{\mathbf{L}}) \widetilde{\mathbf{B}}^\top \widetilde{\mathbf{W}}^{1/2} \mathbf{g}_j$, $j \le 10^3 \log n$, are independent centered Gaussians with the common variance $(\mathbb{1}_u - \mathbb{1}_v)^\top \mathbf{Z}(\widetilde{\mathbf{L}}) \widetilde{\mathbf{L}} \mathbf{Z}(\widetilde{\mathbf{L}}) (\mathbb{1}_u - \mathbb{1}_v)$, which lies in $[0.94, 1.06] \cdot (\mathbb{1}_u - \mathbb{1}_v)^\top \mathbf{L}(G)^\dagger (\mathbb{1}_u - \mathbb{1}_v)$ by Equation (18) and $(\mathbb{1}_u - \mathbb{1}_v) \perp \mathbb{1}$. Divided by this variance, $\sum_j ((\mathbf{x}_j^{\mathrm{rand}})_u - (\mathbf{x}_j^{\mathrm{rand}})_v)^2$ has the law of $\mathbf{g}^\top \mathbf{g}$ for a standard Gaussian vector $\mathbf{g} \in \mathbb{R}^{10^3 \log n}$. Lemma 4.3 with $\mathbf{N} = \mathbf{I}$ and $t = 10^3 \log n / 80$ gives

$$\Pr\Big[\Big|\frac{\mathbf{g}^\top \mathbf{g}}{10^3 \log n} - 1\Big| > \frac{1}{4}\Big] \le 2 \exp\Big(-\frac{10^3 \log n}{80}\Big) \le 2n^{-18} \le n^{-16},$$

because $2/\sqrt{80} + 2/80 \le \frac{1}{4}$ and $e^{12.5 \log n} = n^{12.5/\ln 2} \ge n^{18}$. A union bound over the fewer than n^2 pairs gives, with probability $1 - n^{-14}$, for all u, v an average in $[\frac{3}{4} \cdot 0.94, \frac{5}{4} \cdot 1.06] \cdot (\mathbb{1}_u - \mathbb{1}_v)^\top \mathbf{L}(G)^\dagger (\mathbb{1}_u - \mathbb{1}_v) \subseteq [\frac{1}{2}, 2] \cdot (\mathbb{1}_u - \mathbb{1}_v)^\top \mathbf{L}(G)^\dagger (\mathbb{1}_u - \mathbb{1}_v)$, which is (b).

Cost. $E(\widetilde{\mathbf{L}})$ has $O(n \log n)$ edges. Step 1 of Figure 4 draws $10^3 \log n \cdot |E(\widetilde{\mathbf{L}})| = O(n \log^2 n)$ Gaussians. For each j, Step 2(a) forms $\widetilde{\mathbf{B}}^\top \widetilde{\mathbf{W}}^{1/2} \mathbf{g}_j$ in $O(|E(\widetilde{\mathbf{L}})|)$ operations and applies $\mathbf{Z}(\widetilde{\mathbf{L}})$ once, in $\ddot{O}(n \log n \log 40) = \ddot{O}(n \log n)$ time by Lemma B.1; Step 2(b) applies $\widetilde{\mathbf{L}}$ three times, $O(n \log n)$ each, and $\mathbf{Z}(\widetilde{\mathbf{M}})$ three times, $\ddot{O}(n \log n \log \log n) = \ddot{O}(n \log n)$ each by Lemma B.1 (the factor $\log \log n$ is hidden by $\ddot{O}$); Step 2(c) takes $O(n)$. Step 3 evaluates two quadratic forms of $\widetilde{\mathbf{L}}$ per vector, $O(n \log n)$ each, and one of $\mathbf{D}(G^{\mathrm{light}})$, $O(n)$ given $\mathbf{d}(G^{\mathrm{light}})$. Summed over the $10^3 \log n$ vectors this is $\ddot{O}(n \log^2 n)$, which is (c). □

5 References from intensities

The references $\widetilde{\mathbf{L}}$ and $\widetilde{\mathbf{M}}$ of a refresh are drawn by retaining every sampled edge independently, with a probability of the form $1 - e^{-\varrho_e}$ for an intensity $\varrho_e > 0$. The centered retention variable ζ_e of a sampled edge e is that of Equation (5): $\zeta_e = 1/(1 - e^{-\varrho_e}) - 1$ if e is retained and $\zeta_e = -1$ otherwise, so that $\mathbb{E}[\zeta_e] = 0$, $\zeta_e \geq -1$, and the weight of e in the reference is $\mathbf{w}_e(1 + \zeta_e)$. The accuracy of the references is a matrix concentration bound, the master tail bound of Tropp, and the per-edge moments of the retained weights are what the sixth-moment estimate of the residual stream (the proof of Lemma 4.5(a)) uses about the references; both are collected here.

Lemma 5.1 (references from intensities)**.** *Let $G = (V, E, \mathbf{w})$ be connected with the split $G^{\text{heavy}}, G^{\text{light}}$ and $\mathbf{M}$ of Definition 4.1, and let $E = E_{\text{s}} \cup E_{\text{x}}$ be a partition of its edges. Let intensities $\varrho_e > 0$, $e \in E_{\text{s}}$, satisfy*

$$\varrho_e \;\geq\; 10^8 \log n \cdot \mathbf{w}_e \mathcal{R}^{\text{eff}}(G)_e \quad \text{for every } e \in E_{\text{s}}, \qquad \sum_{e \in E_{\text{s}}} \varrho_e \;\leq\; 10^{17} n \log n.$$

Retain every $e \in E_{\text{s}}$ independently with probability $1 - e^{-\varrho_e}$, with weight $\mathbf{w}_e/(1 - e^{-\varrho_e})$, and keep every $e \in E_{\text{x}}$ with its weight; let $\widetilde{\mathbf{L}}$ be the Laplacian of the result and $\widetilde{\mathbf{M}} \coloneqq \mathbf{D}(G^{\text{light}}) + \widetilde{\mathbf{L}}(G^{\text{heavy}})$, where $\widetilde{\mathbf{L}}(G^{\text{heavy}})$ is the part of $\widetilde{\mathbf{L}}$ on the heavy edges. Then:

(a) with probability $1 - 2n^{-14}$, $\widetilde{\mathbf{L}} \approx_{10^{-3}} \mathbf{L}(G)$ and $\widetilde{\mathbf{M}} \approx_{10^{-3}} \mathbf{M}$;

(b) with probability $1 - n^{-14}$, the retained edges number at most $2 \cdot 10^{17} n \log n + |E_{\text{x}}|$;

(c) (per-edge moments) for every $e \in E_{\text{s}}$ and every integer p with $2 \leq p \leq 6$,

$$\mathbb{E}\big[|\zeta_e|^p\big] \big(\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{L}(G)^\dagger \boldsymbol{\chi}_e\big)^{p-1} \leq \Big(\frac{10^{-7}}{\log n}\Big)^{p-1},$$
$$\mathbb{E}\big[|\zeta_e|^p\big] \big(\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{M}^\dagger \boldsymbol{\chi}_e\big)^{p-1} \leq \Big(\frac{2 \cdot 10^{-7}}{\log n}\Big)^{p-1}.$$

Proof. Throughout, $\log n \geq \ln n$ because logarithms are to the base 2. The retained weight of a sampled edge is $\mathbf{w}_e(1 + \zeta_e)$ and the exactly kept edges have their weights on both sides, so

$$\widetilde{\mathbf{L}} - \mathbf{L}(G) = \sum_{e \in E_{\text{s}}} \zeta_e \mathbf{w}_e \boldsymbol{\chi}_e \boldsymbol{\chi}_e^\top, \qquad \widetilde{\mathbf{M}} - \mathbf{M} = \widetilde{\mathbf{L}}(G^{\text{heavy}}) - \mathbf{L}(G^{\text{heavy}}) = \sum_{e \in E_{\text{s}} \cap E(G^{\text{heavy}})} \zeta_e \mathbf{w}_e \boldsymbol{\chi}_e \boldsymbol{\chi}_e^\top. \quad (20)$$

The matrix concentration tool is the master tail bound of Tropp; it is stated here because Claim 5.4 applies it to summands whose norms are not uniformly bounded, which rules out the matrix Chernoff bound with a uniform norm bound.

Claim 5.2 (Theorem 3.6 of [Tro12])**.** *Consider a finite sequence $\{\mathbf{X}_k\}$ of independent, random, self-adjoint matrices. For all $t \in \mathbb{R}$,*

$$\Pr\Big[\lambda_{\max}\Big(\sum_k \mathbf{X}_k\Big) \geq t\Big] \leq \inf_{\theta > 0} \; e^{-\theta t} \cdot \operatorname{Tr}\Big[\exp\Big(\sum_k \log \mathbb{E}\big[e^{\theta \mathbf{X}_k}\big]\Big)\Big].$$

The scalar input to Claim 5.2 is the moment generating function of a centered retention variable.

Claim 5.3. *Let ζ be the centered retention variable of intensity $\varrho > 0$. Then for every real x with $|x| \leq 10^{-3}\varrho$,*

$$\log \mathbb{E}\big[e^{x\zeta}\big] \;\leq\; \frac{0.51\, x^2}{\varrho}.$$

Proof of Claim 5.3. The function $g(x) \coloneqq \mathbb{E}[e^{x\zeta}] = e^{-\varrho}e^{-x} + (1-e^{-\varrho})e^{xe^{-\varrho}/(1-e^{-\varrho})}$ has $g(0) = 1$, $g'(0) = \mathbb{E}[\zeta] = 0$ and

$$g''(y) = e^{-\varrho}e^{-y} + \frac{e^{-2\varrho}}{1-e^{-\varrho}}\, e^{ye^{-\varrho}/(1-e^{-\varrho})}.$$

We show $g''(y) \le 1.002/\varrho$ for every $|y| \le 10^{-3}\varrho$. Since $e^{-\varrho}/(1-e^{-\varrho}) = 1/(e^{\varrho}-1) \le 1/\varrho$, the exponent of the first term is at most $|y| \le 10^{-3}\varrho$ and the exponent of the second term is at most $|y|/\varrho \le 10^{-3}$.

- If $\varrho \le 1$, both exponents are at most 10^{-3}, so

$$g''(y) \le \Big(e^{-\varrho} + \frac{e^{-2\varrho}}{1-e^{-\varrho}}\Big)e^{10^{-3}} = \frac{e^{-\varrho}}{1-e^{-\varrho}}\, e^{10^{-3}} \le \frac{1.0011}{\varrho}.$$

- If $\varrho > 1$, the first term is at most $e^{-\varrho}e^{10^{-3}\varrho} = e^{-0.999\varrho}$ and the second is at most $1.0011\, e^{-2\varrho}/(1-e^{-\varrho})$. The function $\varrho e^{-0.999\varrho}$ is maximized at $\varrho = 1/0.999$ with value at most 0.37, and $\varrho e^{-2\varrho}/(1-e^{-\varrho})$ is decreasing on $\varrho > 1$ with value at most 0.22 at $\varrho = 1$. Hence $g''(y) \le (0.37 + 1.0011 \cdot 0.22)/\varrho \le 1.002/\varrho$.

By Taylor's theorem, $g(x) = 1 + \frac{1}{2}g''(y)x^2$ for some y between 0 and x, so $g(x) \le 1 + 0.501x^2/\varrho$, and $\log g(x) \le g(x) - 1 \le 0.51x^2/\varrho$. □

Both references are handled by one computation.

Claim 5.4. *Let $\mathbf{N}$ be a symmetric positive semidefinite $n \times n$ matrix and $F \subseteq E_{\mathrm{s}}$ a set of sampled edges such that*

$$\sum_{e\in F} \mathbf{w}_e \mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2} \preceq \mathbf{I} \qquad \text{and} \qquad \mathbf{w}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e \le \frac{2\cdot 10^{-8}\varrho_e}{\log n} \quad \text{for every } e \in F.$$

Then with probability at least $1 - n^{-14}$,

$$\Big\|\sum_{e\in F} \zeta_e \mathbf{w}_e \mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}\Big\| \le 10^{-3}.$$

Proof of Claim 5.4. The summands $\zeta_e\mathbf{w}_e\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}$, $e \in F$, are independent symmetric random matrices with mean zero. Put $\theta \coloneqq 5\cdot 10^4 \log n$ in Claim 5.2. For a fixed $e \in F$, the matrix $\theta\zeta_e\mathbf{w}_e\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}$ has rank one with the single nonzero eigenvalue $\theta\zeta_e\mathbf{w}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e$, so its exponential is $\mathbf{I}$ plus $\big(e^{\theta\zeta_e\mathbf{w}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e} - 1\big)$ times the projection $\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}/(\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e)$, and taking the expectation and then the matrix logarithm gives

$$\log \mathbb{E}\big[e^{\pm\theta\zeta_e\mathbf{w}_e\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}}\big] = \log \mathbb{E}\big[e^{\pm\theta\zeta_e\mathbf{w}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e}\big] \cdot \frac{\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}}{\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e}.$$

The hypothesis gives $\theta\mathbf{w}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e \le 5\cdot 10^4 \log n \cdot 2\cdot 10^{-8}\varrho_e/\log n = 10^{-3}\varrho_e$, so Claim 5.3 applies with $x = \pm\theta\mathbf{w}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e$ and yields

$$\begin{aligned}
\log \mathbb{E}\big[e^{\pm\theta\zeta_e\mathbf{w}_e\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}}\big] &\preceq \frac{0.51\,\theta^2\mathbf{w}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger}\boldsymbol{\chi}_e}{\varrho_e} \cdot \mathbf{w}_e\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2} \\
&\preceq 0.51\cdot 25\cdot 10^8 \log^2 n \cdot \frac{2\cdot 10^{-8}}{\log n} \cdot \mathbf{w}_e\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2} \\
&= 25.5 \log n \cdot \mathbf{w}_e\mathbf{N}^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^{\top}\mathbf{N}^{\dagger/2}.
\end{aligned}$$

Summing over $e \in F$ and using the first hypothesis, the sum of the logarithms is at most $25.5 \log n \cdot \mathbf{I}$, whose exponential has trace at most $n\, e^{25.5 \log n}$. Claim 5.2 with $t = 10^{-3}$, for either sign, gives

$$\Pr\Big[\lambda_{\max}\Big(\pm \sum_{e \in F} \zeta_e \mathbf{w}_e \mathbf{N}^{\dagger/2} \boldsymbol{\chi}_e \boldsymbol{\chi}_e^\top \mathbf{N}^{\dagger/2}\Big) \geq 10^{-3}\Big] \leq e^{-50 \log n} \cdot n\, e^{25.5 \log n} = n\, e^{-24.5 \log n} \leq n^{-23.5},$$

using $\log n \geq \ln n$ in the last step. The two signs together fail with probability at most $2n^{-23.5} \leq n^{-14}$. □

Part (a), the Laplacian. Apply Claim 5.4 with $\mathbf{N} = \mathbf{L}(G)$ and $F = E_{\mathrm{s}}$. The first hypothesis holds because $\sum_{e \in E_{\mathrm{s}}} \mathbf{w}_e \mathbf{L}(G)^{\dagger/2} \boldsymbol{\chi}_e \boldsymbol{\chi}_e^\top \mathbf{L}(G)^{\dagger/2} \preceq \mathbf{L}(G)^{\dagger/2} \mathbf{L}(G) \mathbf{L}(G)^{\dagger/2} \preceq \mathbf{I}$, and the second because $\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{L}(G)^\dagger \boldsymbol{\chi}_e = \mathbf{w}_e \mathcal{R}^{\mathrm{eff}}(G)_e \leq 10^{-8} \varrho_e / \log n$. So with probability $1 - n^{-14}$, $\|\mathbf{L}(G)^{\dagger/2}(\widetilde{\mathbf{L}} - \mathbf{L}(G))\mathbf{L}(G)^{\dagger/2}\| \leq 10^{-3}$ by Equation (20). Every $\boldsymbol{\chi}_e$ is orthogonal to $\mathbb{1}$, and G is connected, so the range of $\widetilde{\mathbf{L}} - \mathbf{L}(G)$ lies in the range of $\mathbf{L}(G)$; hence for every $\mathbf{x} \in \mathbb{R}^V$, $\mathbf{x}^\top(\widetilde{\mathbf{L}} - \mathbf{L}(G))\mathbf{x} = (\mathbf{L}(G)^{1/2}\mathbf{x})^\top \mathbf{L}(G)^{\dagger/2}(\widetilde{\mathbf{L}} - \mathbf{L}(G))\mathbf{L}(G)^{\dagger/2}(\mathbf{L}(G)^{1/2}\mathbf{x})$, whose absolute value is at most $10^{-3}\mathbf{x}^\top \mathbf{L}(G)\mathbf{x}$. This is $\widetilde{\mathbf{L}} \approx_{10^{-3}} \mathbf{L}(G)$.

Part (a), the heavy reference. First, $\mathbf{L}(G) \preceq 2\mathbf{M}$: writing $\mathbf{L}(G^{\mathrm{light}}) = \mathbf{D}(G^{\mathrm{light}}) - \mathbf{A}(G^{\mathrm{light}})$,

$$\mathbf{D}(G^{\mathrm{light}}) + \mathbf{A}(G^{\mathrm{light}}) = \sum_{e=uv \in E(G^{\mathrm{light}})} \mathbf{w}_e (\mathbb{1}_u + \mathbb{1}_v)(\mathbb{1}_u + \mathbb{1}_v)^\top \succeq 0,$$

$$\mathbf{L}(G) = \mathbf{L}(G^{\mathrm{heavy}}) + \mathbf{L}(G^{\mathrm{light}}) \preceq \mathbf{L}(G^{\mathrm{heavy}}) + 2\mathbf{D}(G^{\mathrm{light}}) \preceq 2\mathbf{M}.$$

In particular the kernel of $\mathbf{M}$ is contained in the kernel of $\mathbf{L}(G)$, which is the span of $\mathbb{1}$, so every $\boldsymbol{\chi}_e$ lies in the range of $\mathbf{M}$, and for every e, by the variational characterization of the pseudoinverse on the range,

$$\boldsymbol{\chi}_e^\top \mathbf{M}^\dagger \boldsymbol{\chi}_e = \max_{\mathbf{x} \in \mathbb{R}^V} \big(2\boldsymbol{\chi}_e^\top \mathbf{x} - \mathbf{x}^\top \mathbf{M} \mathbf{x}\big) \leq \max_{\mathbf{x} \in \mathbb{R}^V} \big(2\boldsymbol{\chi}_e^\top \mathbf{x} - \tfrac{1}{2}\mathbf{x}^\top \mathbf{L}(G)\mathbf{x}\big) = 2\boldsymbol{\chi}_e^\top \mathbf{L}(G)^\dagger \boldsymbol{\chi}_e = 2\mathcal{R}^{\mathrm{eff}}(G)_e.$$

Apply Claim 5.4 with $\mathbf{N} = \mathbf{M}$ and $F = E_{\mathrm{s}} \cap E(G^{\mathrm{heavy}})$. The first hypothesis holds because $\mathbf{D}(G^{\mathrm{light}}) \succeq 0$ gives $\sum_{e \in E(G^{\mathrm{heavy}})} \mathbf{w}_e \mathbf{M}^{\dagger/2} \boldsymbol{\chi}_e \boldsymbol{\chi}_e^\top \mathbf{M}^{\dagger/2} = \mathbf{M}^{\dagger/2} \mathbf{L}(G^{\mathrm{heavy}}) \mathbf{M}^{\dagger/2} \preceq \mathbf{M}^{\dagger/2} \mathbf{M} \mathbf{M}^{\dagger/2} \preceq \mathbf{I}$, and the second because $\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{M}^\dagger \boldsymbol{\chi}_e \leq 2\mathbf{w}_e \mathcal{R}^{\mathrm{eff}}(G)_e \leq 2 \cdot 10^{-8} \varrho_e / \log n$. So with probability $1 - n^{-14}$, $\|\mathbf{M}^{\dagger/2}(\widetilde{\mathbf{M}} - \mathbf{M})\mathbf{M}^{\dagger/2}\| \leq 10^{-3}$ by Equation (20), and as the range of $\widetilde{\mathbf{M}} - \mathbf{M}$ lies in the span of the $\boldsymbol{\chi}_e$ and hence in the range of $\mathbf{M}$, the same computation as for the Laplacian gives $\widetilde{\mathbf{M}} \approx_{10^{-3}} \mathbf{M}$. The two events hold together with probability at least $1 - 2n^{-14}$.

Part (b), the number of retained edges. The number of retained sampled edges is $\sum_{e \in E_{\mathrm{s}}} (1 + \zeta_e)(1 - e^{-\varrho_e})$, a sum of independent indicators with

$$\begin{aligned} \mathbb{E}\big[e^{\theta(1+\zeta_e)(1-e^{-\varrho_e})}\big] &= 1 + (1 - e^{-\varrho_e})(e^\theta - 1) \\ &\leq \exp\big((e^\theta - 1)(1 - e^{-\varrho_e})\big) \leq \exp\big((e^\theta - 1)\varrho_e\big) \quad \text{for every } \theta > 0. \end{aligned}$$

With $\theta = \ln 2$ and Markov's inequality applied to the exponential of the sum, the sum exceeds $2 \cdot 10^{17} n \log n$ with probability at most

$$\exp\Big(-2\ln 2 \cdot 10^{17} n \log n + \sum_{e \in E_{\mathrm{s}}} \varrho_e\Big) \leq \exp\big(-(2\ln 2 - 1) \cdot 10^{17} n \log n\big) \leq \exp\big(-3 \cdot 10^{16} n \log n\big) \leq n^{-14}.$$

Adding the $|E_{\mathrm{x}}|$ exactly kept edges gives the count. The same computation with any bound $R \geq \sum_{e \in E_{\mathrm{s}}} \varrho_e$ in place of $10^{17} n \log n$ gives at most $2R + |E_{\mathrm{x}}|$ retained edges with probability $1 - e^{-0.38R}$.

Part (c), the per-edge moments. Fix $e \in E_{\mathrm{s}}$ and $2 \le p \le 6$. From the distribution of ζ_e,

$$\mathbb{E}\big[|\zeta_e|^p\big] = e^{-\varrho_e} \cdot 1 + (1 - e^{-\varrho_e})\Big(\frac{e^{-\varrho_e}}{1 - e^{-\varrho_e}}\Big)^p = e^{-\varrho_e} + \frac{e^{-p\varrho_e}}{(1 - e^{-\varrho_e})^{p-1}}.$$

Let $\lambda \in \{1, 2\}$ and let $\mathbf{N} \in \{\mathbf{L}(G), \mathbf{M}\}$ be the matrix with $\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{N}^\dagger \boldsymbol{\chi}_e \le \lambda \cdot 10^{-8} \varrho_e / \log n$ ($\lambda = 1$ for $\mathbf{L}(G)$ by the hypothesis on ϱ_e, $\lambda = 2$ for $\mathbf{M}$ by the inequality $\boldsymbol{\chi}_e^\top \mathbf{M}^\dagger \boldsymbol{\chi}_e \le 2\mathcal{R}^{\mathrm{eff}}(G)_e$ proved in part (a)). We show $\mathbb{E}[|\zeta_e|^p](\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{N}^\dagger \boldsymbol{\chi}_e)^{p-1} \le (\lambda \cdot 10^{-7}/\log n)^{p-1}$.

- If $\varrho_e < 1$: the function $(1 - e^{-\varrho})/\varrho$ is decreasing, so $1 - e^{-\varrho_e} \ge (1 - e^{-1})\varrho_e \ge 0.63\varrho_e$, and $\mathbb{E}[|\zeta_e|^p] \le 1 + (1 - e^{-\varrho_e})^{1-p} \le 2(1 - e^{-\varrho_e})^{1-p}$. Hence, with $\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{N}^\dagger \boldsymbol{\chi}_e \le \lambda \cdot 10^{-8}(1 - e^{-\varrho_e})/(0.63 \log n) \le 1.6\lambda \cdot 10^{-8}(1 - e^{-\varrho_e})/\log n$,

$$\mathbb{E}\big[|\zeta_e|^p\big]\big(\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{N}^\dagger \boldsymbol{\chi}_e\big)^{p-1} \le 2\Big(\frac{1.6\lambda \cdot 10^{-8}}{\log n}\Big)^{p-1} = 2 \cdot 0.16^{p-1}\Big(\frac{\lambda \cdot 10^{-7}}{\log n}\Big)^{p-1} \le \Big(\frac{\lambda \cdot 10^{-7}}{\log n}\Big)^{p-1}.$$

- If $\varrho_e \ge 1$: $1 - e^{-\varrho_e} \ge 1 - e^{-1}$, so $\mathbb{E}[|\zeta_e|^p] \le e^{-\varrho_e}\big(1 + (e^{-1}/(1 - e^{-1}))^{p-1}\big) \le e^{-\varrho_e}(1 + 0.59^{p-1}) \le 2e^{-\varrho_e}$. The function $\varrho^{p-1}e^{-\varrho}$ is maximized at $\varrho = p - 1$, so

$$\begin{aligned}\mathbb{E}\big[|\zeta_e|^p\big]\big(\mathbf{w}_e \boldsymbol{\chi}_e^\top \mathbf{N}^\dagger \boldsymbol{\chi}_e\big)^{p-1} &\le 2\varrho_e^{p-1} e^{-\varrho_e}\Big(\frac{\lambda \cdot 10^{-8}}{\log n}\Big)^{p-1} \\ &\le 2\Big(\frac{p-1}{10e}\Big)^{p-1}\Big(\frac{\lambda \cdot 10^{-7}}{\log n}\Big)^{p-1} \le \Big(\frac{\lambda \cdot 10^{-7}}{\log n}\Big)^{p-1},\end{aligned}$$

  because $(p - 1)/(10e) \le 5/(10e) < \frac{1}{2}$ for $p \le 6$.

This is part (c) with $\lambda = 1$ for $\mathbf{L}(G)$ and $\lambda = 2$ for $\mathbf{M}$. □

6 The maintained state

The structure keeps three kinds of information: the current graph in per-vertex trees, the retained part of the current block in a QueryDrivenPartition, and the *record* of the previous refresh, from which the intensities of the current block are read. Figure 5 lists them; the record is defined in Section 7.

State of the structure during block k, with $\theta_{\deg} = \log^3 n$.

- For every vertex v: a balanced tree T_v of the current edges at v keyed by current weight; the $\theta_{\deg}$ largest keys are the *current heavy* edges at v, every other edge at v is *current light*. An edge is current heavy if it is current heavy at either endpoint.
- For every vertex v: a balanced tree T_v^{light} of the current light edges at v that are edges of G_k^{ret}, augmented with subtree sums of the *active weight*: the weight of the edge in G_k^{ret} if the edge has been light at v since the previous refresh, and 0 otherwise.
- The list Changed of the edges whose heavy/light status at some endpoint changed since the previous refresh, and the list G_k^{new} of the held new edges and weight increases.
- P, a QueryDrivenPartition (Lemma 3.2) holding exactly G_k^{ret} with its weights.
- The record of refresh $k-1$ (Definition 7.1): the sparse reference $\widetilde{\mathbf{L}}_{k-1}$, the heavy list $E(G_{k-1}^{\text{heavy}})$, the vectors $\mathbf{x}_j^{\text{rand}}, \mathbf{x}_j^{\text{res}}, \mathbf{x}_j^{\text{prox}}$, $j \le 10^3 \log n$, stored by coordinate, the proxy masses $\mathbf{x}_v^{\text{vert}}$, and the packed distance oracle D of Lemma A.1 on the rows of $[\mathbf{x}_1^{\text{rand}} \cdots \mathbf{x}_{10^3 \log n}^{\text{rand}}]$.

Updates during block k.

- Insert$(e, \mathbf{w}_e)$: insert e into T_u, T_v and into G_k^{new}; if e becomes current light at an endpoint, insert it into T^{light} there with active weight 0 and append it to Changed.
- Delete(e): remove e from T_u, T_v, from $T_u^{\text{light}}, T_v^{\text{light}}$, from G_k^{new} if present, and P.Delete(e) if present.
- Weight$(e, \mathbf{w}'_e)$ with $\mathbf{w}'_e > \mathbf{w}_e$: update the key in T_u, T_v and append the increase $\mathbf{w}'_e - \mathbf{w}_e$ to G_k^{new}. With $\mathbf{w}'_e < \mathbf{w}_e$: update the key in T_u, T_v, the active weights in $T_u^{\text{light}}, T_v^{\text{light}}$ (if active), and replace the copy in P by P.Delete(e), P.Insert$(e, \mathbf{w}'_e)$.
- After every update, for each endpoint v touched: the edge that left and the edge that entered the $\theta_{\deg}$ largest keys of T_v (at most one each) move between heavy and light at v; move it out of or into T_v^{light}, set its active weight to 0 there, and append it to Changed.

Figure 5: The maintained state and the processing of the updates of a block; the query is Figure 7.

Lemma 6.1 (updates)**.** *An update costs $O(\log n)$ time plus amortized $O(\log^3 n)$ in P, and appends at most 4 edges to* Changed*. At every moment P holds G_k^{ret}, G_k^{new} holds the new edges and increases with $\mathbf{L}(G) = \mathbf{L}(G_k^{\text{ret}}) + \mathbf{L}(G_k^{\text{new}})$, and the active weights of T_v^{light} are positive exactly on the edges that are light at v both at the previous refresh and now, where they equal the weight in G_k^{ret}.*

Proof. An update changes the key of one edge at two vertices. At a vertex v, the set of the $\theta_{\deg}$ largest keys of T_v changes by at most one edge leaving and one entering, found in $O(\log n)$ by rank queries in T_v; so at most two status changes per endpoint, four in all, each an $O(\log n)$ tree operation. The invariants on P and G_k^{new} are immediate from the case list of Figure 5: a deletion removes the retained copy, a decrease replaces it by the smaller weight, an increase or an insertion is held, and the sum $\mathbf{L}(G_k^{\text{ret}}) + \mathbf{L}(G_k^{\text{new}})$ is the current Laplacian in each case. An active weight is set to 0 whenever the status changes and is otherwise the weight of the copy in P, which is the weight in G_k^{ret}. □

7 The refresh of one block

Throughout this section $G = (V, E, \mathbf{w})$ is the connected graph at the query of block k, with the heavy/light split $G^{\text{heavy}}, G^{\text{light}}$ and $\mathbf{M} = \mathbf{L}(G^{\text{heavy}}) + \mathbf{D}(G^{\text{light}})$ of Definition 4.1 for $\theta_{\deg} = \log^3 n$, read off the trees T_v. The refresh takes two sparse references and produces the record of the block. How the references are sampled is the subject of Section 8; here they are given.

The references are those of Equation (5), drawn as in Lemma 5.1; an attempt of the refresh is a run of ATTEMPT (Figure 4), and Lemma 4.5 gives its acceptance probability, the energies Equation (11) of an accepted attempt, its all-pairs guarantee and its cost.

Definition 7.1 (the record of a refresh)**.** *For an accepted attempt, the* proxy masses *are* $\mathbf{x}_v^{\mathrm{vert}} := \frac{1}{10^3 \log n} \sum_j (\mathbf{x}_j^{\mathrm{prox}})_v^2$*, the packed oracle* D *of Lemma A.1 is built on the rows of* $\mathbf{X} := [\mathbf{x}_1^{\mathrm{rand}} \cdots \mathbf{x}_{10^3 \log n}^{\mathrm{rand}}]$*, and the* public resistance estimate *is*

$$\widehat{\mathcal{R}}_{uv}^{\mathrm{eff}} := \begin{cases} D(u,v)/(10^3 \log n), & uv \in G^{\mathrm{heavy}}, \\ \dfrac{1}{10^3 \log n} \displaystyle\sum_j 2\big((\mathbf{x}_j^{\mathrm{res}})_u - (\mathbf{x}_j^{\mathrm{res}})_v\big)^2 + 4\mathbf{x}_u^{\mathrm{vert}} + 4\mathbf{x}_v^{\mathrm{vert}}, & uv \in G^{\mathrm{light}}. \end{cases} \tag{21}$$

The record *of the refresh is* $\widetilde{\mathbf{L}}$*, the list* $E(G^{\mathrm{heavy}})$*, the vectors* $\mathbf{x}_j^{\mathrm{rand}}, \mathbf{x}_j^{\mathrm{res}}, \mathbf{x}_j^{\mathrm{prox}}$*, the masses* $\mathbf{x}^{\mathrm{vert}}$ *and the oracle* D*.*

Lemma 7.2 (the public estimate)**.** *On the events of Lemma 4.5(b) and Lemma A.1:*

(a) $\widehat{\mathcal{R}}_e^{\mathrm{eff}} \geq \frac{1}{2}\mathcal{R}^{\mathrm{eff}}(G)_e$ *for every edge* e*, and the light branch of Equation (21) evaluated at any pair* u, v *is at least* $\frac{1}{2}(\mathbb{1}_u - \mathbb{1}_v)^\top \mathbf{L}(G)^\dagger(\mathbb{1}_u - \mathbb{1}_v)$*;*

(b) $\sum_e \mathbf{w}_e \widehat{\mathcal{R}}_e^{\mathrm{eff}} \leq 10^7 n$*;*

(c) $\widehat{\mathcal{R}}_e^{\mathrm{eff}}$ *is evaluated in* $\ddot{O}(1)$ *time for a heavy edge and* $O(\log n)$ *time for a light edge, from the record.*

Proof. (a) For a heavy edge, $D(u,v) \geq \|\mathbf{X}_u - \mathbf{X}_v\|_2^2 = \sum_j ((\mathbf{x}_j^{\mathrm{rand}})_u - (\mathbf{x}_j^{\mathrm{rand}})_v)^2$ and Lemma 4.5(b) gives the bound. For the light branch, $\mathbf{x}_j^{\mathrm{rand}} = \mathbf{x}_j^{\mathrm{res}} + \mathbf{x}_j^{\mathrm{prox}}$ and $(x+y)^2 \leq 2x^2 + 2y^2$ give

$$((\mathbf{x}_j^{\mathrm{rand}})_u - (\mathbf{x}_j^{\mathrm{rand}})_v)^2 \leq 2((\mathbf{x}_j^{\mathrm{res}})_u - (\mathbf{x}_j^{\mathrm{res}})_v)^2 + 4(\mathbf{x}_j^{\mathrm{prox}})_u^2 + 4(\mathbf{x}_j^{\mathrm{prox}})_v^2,$$

so the light branch dominates the average of Lemma 4.5(b).

(b) For the heavy edges, $D(u,v) \leq 10^3 \|\mathbf{X}_u - \mathbf{X}_v\|_2^2$ and Equation (11) give $\sum_{e \in G^{\mathrm{heavy}}} \mathbf{w}_e \widehat{\mathcal{R}}_e^{\mathrm{eff}} \leq 10^3 \cdot \frac{1}{10^3 \log n} \sum_j \mathbf{x}_j^{\mathrm{rand}\top} \mathbf{L}(G) \mathbf{x}_j^{\mathrm{rand}} \leq 1.2 \cdot 10^3 n$. For the light edges, summing the light branch against $\mathbf{w}_e$ gives $2 \cdot \frac{1}{10^3 \log n} \sum_j \mathcal{E}_2(\mathbf{x}_j^{\mathrm{res}}, G^{\mathrm{light}}) + 4 \sum_v \mathbf{d}(G^{\mathrm{light}})_v \mathbf{x}_v^{\mathrm{vert}}$, which by Equation (11) is at most $4 \cdot 10^8 (n/\theta_{\mathrm{deg}} + n/\log^3 n) + 4 \cdot 10^3 n \leq 4 \cdot 10^6 n + 4 \cdot 10^3 n$ as $\theta_{\mathrm{deg}} \geq 200$. The total is at most $10^7 n$.

(c) The heavy branch is one oracle query; the light branch reads $10^3 \log n$ stored coordinates at each endpoint and the two masses. □

8 From one block to the next

Fix $k \geq 2$, write $G = G_k$, $G^{\mathrm{ret}} = G_k^{\mathrm{ret}}$, $G^{\mathrm{new}} = G_k^{\mathrm{new}}$, and let the record of refresh $k-1$ be that of G_{k-1} with its split $G_{k-1}^{\mathrm{heavy}}, G_{k-1}^{\mathrm{light}}$. All statements are on the events of Lemma 4.5(b), Lemma A.1 and Lemma 5.1 for refresh $k-1$, and block k is stable (Definition 1.1).

Lemma 8.1 (transfer)**.** *If block* k *is stable:*

(a) for every edge e *of* G^{ret}*, with its weight* $\mathbf{w}_e^{\mathrm{ret}}$ *in* G^{ret}*,* $\mathcal{R}^{\mathrm{eff}}(G)_e \leq \mathcal{R}^{\mathrm{eff}}(G^{\mathrm{ret}})_e \leq 10\mathcal{R}^{\mathrm{eff}}(G_{k-1})_e \leq 20\widehat{\mathcal{R}}_e^{\mathrm{eff}}$*, where* $\widehat{\mathcal{R}}^{\mathrm{eff}}$ *is the estimate of refresh* $k-1$*;*

(b) for every $\mathbf{x} \in \mathbb{R}^V$, $\frac{1}{10}\mathcal{E}_2(\mathbf{x}, G_{k-1}) \le \mathcal{E}_2(\mathbf{x}, G^{\mathrm{ret}}) \le \mathcal{E}_2(\mathbf{x}, G_{k-1})$, *and* $\widetilde{\mathcal{E}}(\mathbf{x}) \coloneqq \frac{1}{3}\mathbf{x}^\top \widetilde{\mathbf{L}}_{k-1}\mathbf{x}$ *satisfies* $\widetilde{\mathcal{E}}(\mathbf{x}) \approx_{10} \mathcal{E}_2(\mathbf{x}, G^{\mathrm{ret}})$.

Proof. (a) $\mathbf{L}(G) = \mathbf{L}(G^{\mathrm{ret}}) + \mathbf{L}(G^{\mathrm{new}}) \succeq \mathbf{L}(G^{\mathrm{ret}})$ gives the first inequality (Rayleigh monotonicity; G is connected, and G^{ret} has the kernel of G_{k-1} by Definition 1.1, which is $\mathbb{1}$). $\frac{1}{10}\mathbf{L}(G_{k-1}) \preceq \mathbf{L}(G^{\mathrm{ret}}) \preceq \mathbf{L}(G_{k-1})$ gives $\mathbf{L}(G^{\mathrm{ret}})^\dagger \preceq 10\mathbf{L}(G_{k-1})^\dagger$, the second inequality. The third is Lemma 7.2(a).

(b) The two-sided bound is $\frac{1}{10}\mathbf{L}(G_{k-1}) \preceq \mathbf{L}(G^{\mathrm{ret}}) \preceq \mathbf{L}(G_{k-1})$ evaluated at $\mathbf{x}$. As $\widetilde{\mathbf{L}}_{k-1} \approx_{10^{-3}} \mathbf{L}(G_{k-1})$, $\widetilde{\mathcal{E}}(\mathbf{x}) \in [\frac{0.999}{3}, \frac{1.001}{3}]\mathcal{E}_2(\mathbf{x}, G_{k-1})$, and $\mathcal{E}_2(\mathbf{x}, G^{\mathrm{ret}})/\widetilde{\mathcal{E}}(\mathbf{x}) \in [\frac{0.3}{1.001}, \frac{3}{0.999}] \subseteq [\frac{1}{10}, 10]$. □

The partition queries of the block. At the query of block k, with P holding G^{ret}, the structure computes $\widetilde{\mathcal{E}}_j \coloneqq \widetilde{\mathcal{E}}(\mathbf{x}_j^{\mathrm{res}})$ for the $10^3 \log n$ residual vectors of refresh $k-1$ and runs PartitionTwo$(\mathbf{x}_j^{\mathrm{res}}; \widetilde{\mathcal{E}}_j)$ on P for each j; Lemma 8.1(b) is the hypothesis $\widetilde{\mathcal{E}}_j \approx_{10} \mathcal{E}_2(\mathbf{x}_j^{\mathrm{res}}, G^{\mathrm{ret}})$ of Lemma 3.2. Write $\mathcal{C}$ for the clusters after the last query (all queries are between updates, so later queries only split earlier clusters, the centers and vertex energies of every $\mathbf{x}_j^{\mathrm{res}}$ are inherited, and the certificate of each query holds on the final clusters because a split only divides the degrees), $C(e)$ for the cluster of e in $\mathcal{C}$, and E^{cut} for the union of the cut edges returned, which are held outside P until the end of the block. The certificates give

$$\begin{aligned}\sum_j \sum_{C\in\mathcal{C}} \sum_{v\in V(C)} \deg_C(v)\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}_j^{\mathrm{res}}, C, v) \le Z \;&\coloneqq\; 10^8 \log^4 n \sum_j \widetilde{\mathcal{E}}_j \\ &\le 10^8 \log^4 n \cdot \frac{1.001}{3} \cdot 10^3 \log n \cdot 2 \cdot 10^8 \Big(\frac{n}{\theta_{\deg}} + \frac{n}{\log^3 n}\Big),\end{aligned} \tag{22}$$

the last by Equation (11) for refresh $k-1$ and Lemma 8.1(b).

The three streams. For an edge $e = uv$ of G^{ret} that is in no cut set and whose status at both endpoints is unchanged since refresh $k-1$, write $\mathbf{w}_e$ for its weight in G^{ret} and

$$r_e \;\coloneqq\; \frac{1}{10^3 \log n} \sum_j \big((\mathbf{x}_j^{\mathrm{res}})_u - (\mathbf{x}_j^{\mathrm{res}})_v\big)^2, \qquad \pi_e \;\coloneqq\; \frac{1}{Z} \sum_j \big(\mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}_j^{\mathrm{res}}, C(e), u) + \mathcal{E}_2^{\mathrm{vert}}(\mathbf{x}_j^{\mathrm{res}}, C(e), v)\big),$$

so that $\sum_e \pi_e \le 1$ by Equation (22) and, by the envelope Equation (1) summed over j,

$$\mathbf{w}_e\, r_e \;\le\; \frac{4Z}{10^3 \log n}\, \pi_e. \tag{23}$$

Given a rate $c > 0$, the routine Proposals(c) of Figure 6 proposes edges of G^{ret} through three independent streams and lists the remaining edges directly.

$(E', q) = \textsc{Proposals}(c)$

Input: the state of Figure 5 after the partition queries of the block, with P holding G^{ret} minus E^{cut}; a rate $c > 0$.

Output: a set E' of edges of G^{ret} and, for each $e \in E'$, its proposal probability q_e.

1. (heavy stream) For every $e = uv$ in the stored list $E(G^{\text{heavy}}_{k-1})$ that is still an edge of G^{ret}: propose e independently with probability $\min\{1,\ c\,\mathbf{w}_e\, D(u,v)/(10^3 \log n)\}$.
2. (vertex stream) For every vertex v: run a Poisson process of rate $8c\,\mathbf{x}^{\text{vert}}_v$ per unit of active weight along the active weights of T^{light}_v, and propose every edge hit.
3. (residual stream) Build a tree over the copies (C, v), $C \in \mathcal{C}$, with weights $\deg_C(v) \sum_j \mathcal{E}^{\text{vert}}_2(\mathbf{x}^{\text{res}}_j, C, v)$. Draw Poisson$(8cZ/(10^3 \log n))$ times: with probability $1 - \sum_{C,v} \deg_C(v) \sum_j \mathcal{E}^{\text{vert}}_2(\mathbf{x}^{\text{res}}_j, C, v)/Z$ return nothing, else draw (C, v) from the tree with probability proportional to its weight, then a uniformly random live edge of C at v (Lemma 3.2); propose every edge returned.
4. (direct list) Propose, with $q_e = 1$, every edge of G^{ret} in E^{cut} or in Changed.
5. For every proposed e, compute q_e as 1 minus the product over the streams of Steps 1–3 that cover e of the probability that the stream misses e: $1 - \min\{1, c\mathbf{w}_e D(u,v)/(10^3 \log n)\}$ for the heavy stream, $e^{-8c\mathbf{x}^{\text{vert}}_u \mathbf{w}_e}$ and $e^{-8c\mathbf{x}^{\text{vert}}_v \mathbf{w}_e}$ for the vertex stream at an endpoint where e is active, and $e^{-8cZ\pi_e/(10^3 \log n)}$ for the residual stream. Return E' and q.

Figure 6: The proposals of a block from the record of the previous refresh. Step 1 is the heavy route, Step 2 the vertex route (the proxy terms of Equation (21)), Step 3 the residual route through the certificates of the partition queries, Step 4 the edges those routes do not cover.

Lemma 8.2 (proposals)**.** *On the events named at the start of the section,* $\textsc{Proposals}(c)$ *proposes every edge e of G^{ret} independently with a probability q_e that it returns, with*

$$q_e \;\geq\; \frac{1}{2}\min\{1,\ c\,\mathbf{w}_e \widehat{\mathcal{R}}^{\text{eff}}_e\} \;\geq\; \frac{1}{2}\min\Big\{1,\ \frac{c}{20}\,\mathbf{w}_e \mathcal{R}^{\text{eff}}(G)_e\Big\},$$

the expected number of proposals is at most $10^4 cn + 10^{23} c \log^4 n\,(n/\theta_{\deg} + n/\log^3 n) + |E^{\text{cut}}| + |\text{Changed}|$, and the time is $\ddot{O}(n\theta_{\deg}) + O(n \log^2 n)$ plus $O(\log n)$ per proposal.

Proof. Independence and the value of q_e. The heavy stream is an independent coin per edge. In the vertex stream, the arrivals of a Poisson process along the active weights at v hit disjoint edges as independent Poisson variables of means $8c\mathbf{x}^{\text{vert}}_v \mathbf{w}_e$ (the active weight of an active edge is $\mathbf{w}_e$); the processes at different vertices are independent. In the residual stream, the number of draws returning e is Poisson with mean $8cZ/(10^3 \log n)$ times the probability that one draw returns e, which is $\deg_{C(e)}(u) \sum_j \mathcal{E}^{\text{vert}}_2(\mathbf{x}^{\text{res}}_j, C(e), u)/Z \cdot 1/\deg_{C(e)}(u)$ plus the same at v, that is π_e, so the mean is $8cZ\pi_e/(10^3 \log n)$; thinning a Poisson number of independent draws gives independent counts over the edges. The three streams are independent, so e is proposed independently of the other edges with the probability q_e computed in Step 5, and the edges of Step 4 have $q_e = 1$.

The lower bound. Let $e = uv$ be an edge of G^{ret} not in Step 4; then e is in no cut set and its status at both endpoints is that of refresh $k-1$. If $e \in G^{\text{heavy}}_{k-1}$, the heavy stream alone gives $q_e \geq \min\{1, c\mathbf{w}_e D(u,v)/(10^3 \log n)\} = \min\{1, c\mathbf{w}_e \widehat{\mathcal{R}}^{\text{eff}}_e\}$ by Equation (21). If $e \in G^{\text{light}}_{k-1}$, it is active at both endpoints, so by Equation (23) the three exponents of Step 5 sum to at least $8c\mathbf{w}_e(\mathbf{x}^{\text{vert}}_u + \mathbf{x}^{\text{vert}}_v) + 2c\mathbf{w}_e r_e \geq c\mathbf{w}_e \widehat{\mathcal{R}}^{\text{eff}}_e$ (the light branch of Equation (21) is $2r_e + 4\mathbf{x}^{\text{vert}}_u + 4\mathbf{x}^{\text{vert}}_v$), hence $q_e \geq 1 - e^{-c\mathbf{w}_e \widehat{\mathcal{R}}^{\text{eff}}_e} \geq \frac{1}{2}\min\{1, c\mathbf{w}_e \widehat{\mathcal{R}}^{\text{eff}}_e\}$, as $1 - e^{-x} \geq x/2$ for $x \leq 1$ and $1 - e^{-x} \geq 1 - e^{-1} \geq \frac{1}{2}$ for $x \geq 1$; the heavy case satisfies the same bound. The second inequality of the lemma is Lemma 8.1(a).

The number of proposals. The expected count is the sum of the stream means. Heavy: $\sum_{e\in G_{k-1}^{\text{heavy}}} c\mathbf{w}_e D(u,v)/(10^3\log n) \le c\sum_{e\in G_{k-1}^{\text{heavy}}} \mathbf{w}_e^{(k-1)}\widehat{\mathcal{R}}_e^{\text{eff}} \le 1.2\cdot 10^3 cn$, by $\mathbf{w}_e \le \mathbf{w}_e^{(k-1)}$ and the heavy part of the proof of Lemma 7.2(b). Vertex: $\sum_v 8c\mathbf{x}_v^{\text{vert}}\cdot(\text{active weight at } v) \le 8c\sum_v \mathbf{x}_v^{\text{vert}}\mathbf{d}(G_{k-1}^{\text{light}})_v \le 8\cdot 10^3 cn$, since the active edges at v are light at v in G_{k-1} with weights at most those in G_{k-1}, and $\sum_v \mathbf{d}(G_{k-1}^{\text{light}})_v\mathbf{x}_v^{\text{vert}} = \frac{1}{10^3\log n}\sum_j \mathbf{x}_j^{\text{prox}\top}\mathbf{D}(G_{k-1}^{\text{light}})\mathbf{x}_j^{\text{prox}} \le 10^3 n$ by Equation (11). Residual: $8cZ/(10^3\log n) \le 10^{23}c\log^4 n(n/\theta_{\text{deg}} + n/\log^3 n)$ by Equation (22). Direct: $|E^{\text{cut}}| + |\text{Changed}|$.

Time. Step 1 scans at most $n\theta_{\text{deg}}$ stored edges at $\ddot{O}(1)$ each (Lemma A.1). Step 2 generates each arrival and locates it in T_v^{light} by the subtree sums in $O(\log n)$. Step 3 computes the weight of each of the at most $10^3 n\log^2 n$ copies from $10^3\log n$ stored vertex energies, $O(n\log^3 n)$ in all, builds the tree in $O(n\log^2 n)$, and a draw costs $O(\log n)$ for the copy and $O(\log n)$ for the live edge. Step 4 is a scan of two lists. Step 5 reads, per proposed edge, one oracle value, two masses, and π_e from the stored vertex energies of the two copies, $O(\log n)$ in all (the $10^3\log n$ terms of π_e). □

9 The block procedure and the proof of the block lemma

$H = \textsc{Query}(\epsilon)$ at the end of block k
Input: the state of Figure 5; $n^{-10} \le \epsilon \le \frac{1}{2}$.
Output: a graph H with $\mathbf{L}(H) \approx_\epsilon \mathbf{L}(G_k)$; the record of refresh k replaces that of refresh $k-1$.

1. If $k \ge 2$: for $j \le 10^3\log n$ compute $\widetilde{\mathcal{E}}_j \leftarrow \frac{1}{3}\mathbf{x}_j^{\text{res}\top}\widetilde{\mathbf{L}}_{k-1}\mathbf{x}_j^{\text{res}}$ and run $\textsc{PartitionTwo}(\mathbf{x}_j^{\text{res}}; \widetilde{\mathcal{E}}_j)$ on P; hold the returned E^{cut}.
2. If $k \ge 2$: $(E', q) \leftarrow \textsc{Proposals}(c)$ with $c := 2\cdot 10^{10}\log n + 4\cdot 10^4\epsilon^{-2}\log n$.
3. For every $e \in E'$, with $\mathbf{w}_e$ its weight in G_k^{ret} and $\widehat{\mathcal{R}}_e^{\text{eff}}$ from Equation (21) (heavy branch iff $e \in E(G_{k-1}^{\text{heavy}})$):
$$\varrho_e \leftarrow 10^{10}\log n\cdot \mathbf{w}_e\widehat{\mathcal{R}}_e^{\text{eff}}, \qquad p_e \leftarrow \min\{1,\ 2\cdot 10^4\epsilon^{-2}\log n\cdot \mathbf{w}_e\widehat{\mathcal{R}}_e^{\text{eff}}\};$$
with two independent coins, put e into R with probability $(1-e^{-\varrho_e})/q_e$ and into O with probability p_e/q_e.
4. $\widetilde{\mathbf{L}}_k \leftarrow \sum_{e\in R}\frac{\mathbf{w}_e}{1-e^{-\varrho_e}}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top + \mathbf{L}(G_k^{\text{new}})$; $H \leftarrow \{e \in O \text{ with weight } \mathbf{w}_e/p_e\}\cup G_k^{\text{new}}$, parallel copies merged by adding weights. Read $E(G_k^{\text{heavy}})$ off the trees T_v and put $\widetilde{\mathbf{M}}_k \leftarrow \mathbf{D}(G_k^{\text{light}}) + \widetilde{\mathbf{L}}_k(G_k^{\text{heavy}})$.
5. (refresh) $\mathbf{Z}(\widetilde{\mathbf{L}}_k) \leftarrow \textsc{LsolverKLPSS}(\widetilde{\mathbf{L}}_k, \frac{1}{40})$, $\mathbf{Z}(\widetilde{\mathbf{M}}_k) \leftarrow \textsc{LsolverKLPSS}(\widetilde{\mathbf{M}}_k, 1/\log n)$; run one attempt of Lemma 4.5 (Figure 4); if it is rejected, redraw for every proposed edge the coin of Step 3 that decides its membership in R and repeat Steps 4–5, at most $10^2\log n$ times in all; build the record of Definition 7.1 from the accepted attempt.
6. Reinsert E^{cut} into P; for every edge of G_k^{new}: $P.\textsc{Insert}$ (a new edge) or $P.\textsc{Delete}$ then $P.\textsc{Insert}$ with the current weight (an increase); for every edge of Changed $\cup\, G_k^{\text{new}}$ set its active weight in T^{light} to its current weight where it is current light; empty Changed and G_k^{new}. Return H.

Figure 7: The query of a block. For $k=1$, G_1^{ret} is empty, $R=\emptyset$ and $\widetilde{\mathbf{L}}_1 = \mathbf{L}(G_1^{\text{new}})$.

Lemma 9.1 (the two samples)**.** *On the events named at the start of Section 8, with probability $1-n^{-11}$ over the coins of Steps 2–3 of Figure 7:*

(a) $\widetilde{\mathbf{L}}_k \approx_{10^{-3}} \mathbf{L}(G_k)$, $\widetilde{\mathbf{M}}_k \approx_{10^{-3}} \mathbf{L}(G_k^{\text{heavy}}) + \mathbf{D}(G_k^{\text{light}})$, both with at most $10^{18}n\log n$ nonzero entries, and the intensities ϱ_e satisfy the hypothesis of Lemma 5.1 for G_k;

(b) $\mathbf{L}(H) \approx_\epsilon \mathbf{L}(G_k)$ *and* H *has at most* $10^{12}\epsilon^{-2} n \log n$ *edges.*

Proof. Let e be an edge of G_k^{ret} with weight $\mathbf{w}_e$ there. By Lemma 8.2, $q_e \geq \frac{1}{2}\min\{1, c\mathbf{w}_e \widehat{\mathcal{R}}_e^{\mathrm{eff}}\}$ with $c = 2 \cdot 10^{10} \log n + 4 \cdot 10^4 \epsilon^{-2} \log n$, so $q_e \geq \min\{1, \varrho_e\} \geq 1 - e^{-\varrho_e}$ and $q_e \geq p_e$: the two thinning probabilities of Step 3 are at most 1, and e enters R with probability exactly $1 - e^{-\varrho_e}$ and O with probability exactly p_e, independently over the edges (independent proposals, independent coins).

(a) By Lemma 8.1(a), $\mathcal{R}^{\mathrm{eff}}(G_k)_e \leq 20\widehat{\mathcal{R}}_e^{\mathrm{eff}}$, so $\varrho_e \geq 5 \cdot 10^8 \log n\, \mathbf{w}_e \mathcal{R}^{\mathrm{eff}}(G_k)_e$; and by Lemma 7.2(b) for G_{k-1} and $\mathbf{w}_e \leq \mathbf{w}_e^{(k-1)}$, $\sum_e \varrho_e \leq 10^{10} \log n \cdot 10^7 n = 10^{17} n \log n$. Lemma 5.1 with E_{s} the edges of G_k^{ret}, $E_{\mathrm{x}} = G_k^{\mathrm{new}}$ and $G = G_k$ (a multigraph when an edge has a retained and an increase copy; the lemma is a statement about sums of rank-one matrices and does not use simplicity) gives the approximations with probability $1 - 2n^{-14}$ and at most $2 \cdot 10^{17} n \log n + n$ retained edges with probability $1 - n^{-14}$, two nonzero entries each.

(b) $p_e \geq \min\{1, 10^3 \epsilon^{-2} \log n \cdot \mathbf{w}_e \mathcal{R}^{\mathrm{eff}}(G_k)_e\}$ by Lemma 8.1(a), and the edges of G_k^{new} have $p_e = 1$; Lemma 2.1 on the multigraph $G_k^{\mathrm{ret}} \cup G_k^{\mathrm{new}}$, whose Laplacian is $\mathbf{L}(G_k)$, gives $\mathbf{L}(H) \approx_\epsilon \mathbf{L}(G_k)$ with probability $1 - n^{-12}$ (merging parallel copies does not change the Laplacian) and at most $2(2 \cdot 10^4 \epsilon^{-2} \log n \cdot 10^7 n + n) \leq 10^{12} \epsilon^{-2} n \log n$ edges with probability $1 - \exp(-10^3 \log n/3)$. □

Proof of Lemma 1.2. Correctness and probability. Call block k *good* if the events of Lemma 5.1(a) and Lemma 9.1(a) for the $\widetilde{\mathbf{L}}_k, \widetilde{\mathbf{M}}_k$ of every trial of Step 5, of Lemma B.1 for the two operators of every trial, of Lemma 4.5(b) and Lemma A.1 for the accepted attempt, and of Lemma 9.1(b) for H all hold, and the refresh accepts within $10^2 \log n$ trials. Each trial of Step 5 uses fresh retention coins, fresh solver randomness and fresh Gaussian vectors, so by Lemma 4.5(a), Lemma 5.1(a) and Lemma B.1 it is accepted with probability at least $1 - 10^{-2} - 3n^{-15} - 2n^{-14} - 2n^{-14} \geq 1 - \frac{1}{50}$. Condition on the history up to the query of block k: the graph G_k, the state, and the record of refresh $k-1$ are then fixed, and the randomness of block k – the Gaussian vectors, the coins of Figure 6 and Step 3, the solver and oracle randomness – is fresh and independent of them. If block $k-1$ is good, the hypotheses of Lemma 8.1, Lemma 8.2 and Lemma 9.1 hold for every fixed G_k that is stable (they are properties of the record and of G_{k-1}, not of the adversary's choices), so conditionally on the history block k is good with probability at least $1 - 10^2 \log n\,(n^{-11} + 4n^{-14}) - 2n^{-14} - n^{-12} - n^{-15} - 50^{-10^2 \log n} \geq 1 - n^{-10}$; block 1 is good with the same probability since $G_1^{\mathrm{ret}} = \emptyset$ makes Lemma 9.1(a) trivial. By induction and a union bound, all Q blocks are good with probability $1 - Qn^{-10}$, and on a good block H is as claimed. The adversary sees H, which depends on the record of refresh $k-1$, but the properties of that record used in block k are the deterministic consequences of block $k-1$ being good, and nothing about the fresh randomness of block k is revealed before G_k is chosen; this is the whole adaptivity argument.

Time of the updates. By Lemma 6.1, $O(\log n)$ per update plus amortized $O(\log^3 n)$ per operation of P; the operations of P in block k are the at most n updates, the reinsertion of E^{cut} (amortized 10^{-1} per operation of P, Lemma 3.2), and at most $2n$ operations in Step 6, so $O(n)$ operations per block and $O(U \log^3 n + Qn \log^3 n)$ in all.

Time of a query, with $\theta_{\mathrm{deg}} = \log^3 n$. Step 1: $10^3 \log n$ energies against $\widetilde{\mathbf{L}}_{k-1}$, $O(n \log n)$ nonzeros each, $O(n \log^2 n)$; $10^3 \log n$ partition queries at amortized $O(n \log^2 n)$ each, $O(n \log^3 n)$. Step 2: by Lemma 8.2 with $c = O(\epsilon^{-2} \log n)$, the time is $\ddot{O}(n \log^3 n) + O(n \log^2 n)$ plus $O(\log n)$ per proposal, and the expected number of proposals is

$$O\big(\epsilon^{-2}\, n \log^2 n\big) + |E^{\mathrm{cut}}| + |\mathrm{Changed}| = O\big(\epsilon^{-2}\, n \log^2 n\big),$$

using $\log^4 n(n/\theta_{\mathrm{deg}} + n/\log^3 n) = 2n \log n$ and $|\mathrm{Changed}| \leq 4n$; the Poisson counts exceed twice their means with probability $\exp(-\Omega(n))$, so the proposals cost $O(\epsilon^{-2} n \log^3 n)$. Step 3: $O(\log n)$ per proposed edge (Lemma 7.2(c)), the same bound. Step 4: $O(n \log n)$ for $\widetilde{\mathbf{L}}_k$ and H, $O(n\theta_{\mathrm{deg}}) =$

$O(n\log^3 n)$ for the heavy list and $\widetilde{\mathbf{M}}_k$. Step 5: a trial redraws one coin per proposed edge, $O(\epsilon^{-2}n\log^2 n)$ by the count of Step 2, rebuilds $\widetilde{\mathbf{L}}_k$ and $\widetilde{\mathbf{M}}_k$ in $O(n\log n)$, makes two calls of Lemma B.1, $O(n\log^2 n)$, and runs one attempt, $\ddot{O}(n\log^2 n)$ (Lemma 4.5(c)); there are at most $10^2\log n$ trials on a good block, with expected number at most 2 (each is accepted with probability at least $1-\frac{1}{50}$), so $\ddot{O}(\epsilon^{-2}n\log^2 n)$ in expectation and $\ddot{O}(\epsilon^{-2}n\log^3 n)$ always; the oracle of Lemma A.1 with $c=12$ in $O(n\log^2 n)$ (on the event of Lemma 4.5(b) the nonzero distances between rows of $\mathbf{X}$ lie in $[n^{-6},n^6]$, as every resistance of a connected graph with weights in $[n^{-10},n^{10}]$ lies in $[n^{-11},n^{11}]$); the proxy masses in $O(n\log n)$. Step 6: $O(n)$ operations of P, counted above, and $O(n\log n)$ tree operations. In all a query costs $\ddot{O}(\epsilon^{-2}n\log^3 n)$, which summed over the blocks is the second term of the theorem. □

10 Without the stability hypothesis

Stability (Definition 1.1) enters Lemma 8.1 only: it is what makes the previous estimate an upper bound on the new resistances and the previous energies estimates of the new ones. It can be enforced by the structure itself, at the price of $O(\log n)$ internal refreshes per block.

The halving schedule. The updates of a block are now recorded and applied at its query. Let G_{k-1} be the graph at the previous query and G_k the graph after the block's updates, and write $\delta_e := \mathbf{w}_e^{(k-1)} - \mathbf{w}_e^{(k)} > 0$ for the edges e of G_{k-1} whose weight decreased or that were deleted ($\mathbf{w}_e^{(k)} := 0$), F for the set of these edges, and G_k^{new} for the new part as in Section 1. Put $t := 30\log n$ and, for $0 \le i \le t$,

$$\mathbf{L}(G_{k,i}) := \mathbf{L}(G_k) + 2^{-i}\sum_{e\in F}\delta_e\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top, \qquad \text{so that}$$
$$G_{k,0} = G_{k-1} + G_k^{\mathrm{new}}, \qquad \mathbf{L}(G_{k,t}) = \mathbf{L}(G_k) + n^{-30}\sum_{e\in F}\delta_e\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top.$$

The structure runs $t+1$ internal blocks: block i ($1\le i\le t$) applies $\textsc{Weight}(e,\mathbf{w}_e^{(k)}+2^{-i}\delta_e)$ for every $e\in F$ (block 1 also inserts G_k^{new}) and calls $\textsc{Query}(\frac{1}{2})$, discarding its output; block $t+1$ applies $\textsc{Weight}(e,\mathbf{w}_e^{(k)})$ for $e\in F$ with $\mathbf{w}_e^{(k)}>0$ and $\textsc{Delete}(e)$ for the others, and calls $\textsc{Query}(\epsilon)$, whose output is returned. The intermediate weights lie in $[n^{-40},n^{10}]$; the imported statements are used with this polynomial range, which changes the number of weight classes of Lemma 3.2 from $20\log n+1$ to $50\log n+1$ and nothing else in their proofs, as every bound there is stated for a polynomial range.

Lemma 10.1 (the internal blocks are stable)**.** *With $G_{k,0}$ in the role of G_{k-1}, every internal block satisfies Definition 1.1: for $1\le i\le t$, $\mathbf{L}(G_{k,i})\succeq\frac{1}{2}\mathbf{L}(G_{k,i-1})$, and $\mathbf{L}(G_k)\succeq\frac{1}{2}\mathbf{L}(G_{k,t})$.*

Proof. For $1\le i\le t$, $\mathbf{L}(G_{k,i-1})-\mathbf{L}(G_{k,i}) = 2^{-i}\sum_{e\in F}\delta_e\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top \preceq \mathbf{L}(G_{k,i})$, because $G_{k,i}$ contains every $e\in F$ with weight at least $2^{-i}\delta_e$; so $\mathbf{L}(G_{k,i-1})\preceq 2\mathbf{L}(G_{k,i})$. For the last block, G_k is connected, so for every edge e, $\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top \preceq \mathcal{R}^{\mathrm{eff}}(G_k)_e\mathbf{L}(G_k)$ (for $\mathbf{x}\perp\mathbb{1}$, $(\boldsymbol{\chi}_e^\top\mathbf{x})^2 \le (\boldsymbol{\chi}_e^\top\mathbf{L}(G_k)^\dagger\boldsymbol{\chi}_e)(\mathbf{x}^\top\mathbf{L}(G_k)\mathbf{x})$ by Cauchy–Schwarz in the $\mathbf{L}(G_k)$ inner product), and $\mathcal{R}^{\mathrm{eff}}(G_k)_e\le(n-1)n^{10}$ (a path of at most $n-1$ edges of weight at least n^{-10}), $\delta_e\le n^{10}$, $|F|\le n$; hence

$$\mathbf{L}(G_{k,t})-\mathbf{L}(G_k) = n^{-30}\sum_{e\in F}\delta_e\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top \preceq n^{-30}\cdot n\cdot n^{10}\cdot n^{11}\,\mathbf{L}(G_k)\preceq\mathbf{L}(G_k).$$ □

Proof of Theorem 1.3. The block's updates are fixed when its query arrives, so the internal blocks are a sequence of $t+1 \le 31 \log n$ blocks of the model of Section 1, each with at most $|F|+|G_k^{\text{new}}| \le 2n$ updates and one query, which satisfy Definition 1.1 by Lemma 10.1; the graphs $G_{k,i}$ are connected since G_k is. Lemma 1.2 applies to the whole sequence of internal blocks: each is good with probability $1-n^{-10}$ conditionally on the past, so all $(t+1)Q$ of them are good with probability $1-31Q\log n\cdot n^{-10} \ge 1-Qn^{-9}$, and on a good last block H is as claimed. The time is that of Lemma 1.2 for $(t+1)Q$ blocks with $(t+1)U$ updates, the internal queries at $\epsilon=\frac{1}{2}$ and the last at ϵ_k: $\ddot{O}((t+1)U\log^3 n + (t+1)Qn\log^3 n + n\log^3 n\sum_k \epsilon_k^{-2})$. □

Toward one logarithm less. The following identity shows what a block of arbitrary deletions does to the resistances, and how much of it the same mechanisms could absorb in a single refresh.

Lemma 10.2 (deleted edges)**.** *Let $G=(V,E,\mathbf{w})$ be connected, $F\subseteq E$ with $G-F$ connected, and let $\mathbf{S} \coloneqq \mathbf{W}_F^{1/2}\mathbf{B}_F\mathbf{L}(G)^\dagger\mathbf{B}_F^\top\mathbf{W}_F^{1/2} \in \mathbb{R}^{F\times F}$ be the leverage matrix of F. Then $0 \preceq \mathbf{S} \prec \mathbf{I}$,*

$$\mathbf{L}(G-F)^\dagger - \mathbf{L}(G)^\dagger = \mathbf{L}(G)^\dagger\mathbf{B}_F^\top\mathbf{W}_F^{1/2}(\mathbf{I}-\mathbf{S})^{-1}\mathbf{W}_F^{1/2}\mathbf{B}_F\mathbf{L}(G)^\dagger \;\succeq\; 0,$$
$$\sum_{e\in E\setminus F}\mathbf{w}_e\big(\mathcal{R}^{\text{eff}}(G-F)_e - \mathcal{R}^{\text{eff}}(G)_e\big) = \sum_{e\in F}\mathbf{w}_e\mathcal{R}^{\text{eff}}(G)_e = \textsc{Tr}[\mathbf{S}].$$

Consequently, if $\mathbf{y}_1,\dots,\mathbf{y}_s$ are independent Gaussian vectors with covariance $\mathbf{L}(G-F)^\dagger-\mathbf{L}(G)^\dagger$, $s=10^3\log n$, then with probability $1-n^{-14}$ for every edge e of $G-F$

$$\frac{1}{s}\sum_{i\le s}(\boldsymbol{\chi}_e^\top\mathbf{y}_i)^2 \in \big[\tfrac{1}{2},2\big]\cdot\big(\mathcal{R}^{\text{eff}}(G-F)_e - \mathcal{R}^{\text{eff}}(G)_e\big),$$
$$\sum_{e\in E\setminus F}\mathbf{w}_e\,\frac{1}{s}\sum_{i\le s}(\boldsymbol{\chi}_e^\top\mathbf{y}_i)^2 \le 2\,|F|.$$

Proof. $\mathbf{S}\succeq 0$ is clear, and $\mathbf{S}\preceq\mathbf{I}$ because $\sum_{e\in F}\mathbf{w}_e\mathbf{L}(G)^{\dagger/2}\boldsymbol{\chi}_e\boldsymbol{\chi}_e^\top\mathbf{L}(G)^{\dagger/2}\preceq\boldsymbol{\Pi}$, whose nonzero eigenvalues are those of $\mathbf{S}$; a unit eigenvector $\mathbf{z}$ of $\mathbf{S}$ with eigenvalue 1 would give a vector $\mathbf{L}(G)^{\dagger/2}\mathbf{B}_F^\top\mathbf{W}_F^{1/2}\mathbf{z}\perp\mathbb{1}$ in the kernel of $\mathbf{L}(G)-\mathbf{L}(F)=\mathbf{L}(G-F)$, contradicting connectivity, so $\mathbf{S}\prec\mathbf{I}$. The identity is the Woodbury formula for $\mathbf{L}(G-F)=\mathbf{L}(G)-\mathbf{B}_F^\top\mathbf{W}_F\mathbf{B}_F$ on $\mathbb{1}^\perp$. For the trace identity, $\sum_{e\notin F}\mathbf{w}_e\boldsymbol{\chi}_e^\top\mathbf{L}(G-F)^\dagger\boldsymbol{\chi}_e=\textsc{Tr}[\mathbf{L}(G-F)^\dagger\mathbf{L}(G-F)]=n-1$ and $\sum_{e\notin F}\mathbf{w}_e\boldsymbol{\chi}_e^\top\mathbf{L}(G)^\dagger\boldsymbol{\chi}_e=\textsc{Tr}[\mathbf{L}(G)^\dagger\mathbf{L}(G)]-\sum_{e\in F}\mathbf{w}_e\mathcal{R}^{\text{eff}}(G)_e$; subtracting gives $\sum_{e\in F}\mathbf{w}_e\mathcal{R}^{\text{eff}}(G)_e=\textsc{Tr}[\mathbf{S}]\le|F|$, as every leverage is at most 1. The all-pairs statement is the χ^2 tail with $s=10^3\log n$ degrees of freedom and a union bound over n^2 pairs, exactly as in Lemma 4.5(b), and the energy bound follows from it and the trace identity. □

Thus after a block that deletes F ($|F|\le n$), the retained edges whose leverage rose are exactly the edges with large energy under the vectors $\mathbf{y}_i$, the total energy is at most $2|F|$, and the certificate of Lemma 3.2 with the $\mathbf{y}_i$ as query vectors would find them and sample them with $O(\log^4 n\cdot|F|\cdot\log n)$ proposals per block, the same accounting as the residual stream. What is missing is the vectors themselves: $\mathbf{y}_i=\mathbf{L}(G-F)^\dagger\mathbf{B}_F^\top\mathbf{W}_F^{1/2}(\mathbf{I}-\mathbf{S})^{1/2}\mathbf{g}_i$ (by the identity, $\mathbf{L}(G-F)^\dagger\mathbf{B}_F^\top\mathbf{W}_F^{1/2}=\mathbf{L}(G)^\dagger\mathbf{B}_F^\top\mathbf{W}_F^{1/2}(\mathbf{I}-\mathbf{S})^{-1}$, so the covariance is the difference above) requires solving in $\mathbf{L}(G-F)$, whose only sparse approximation at hand, $\widetilde{\mathbf{L}}_{k-1}$, is a preconditioner of condition number $1/(1-\lambda_{\max}(\mathbf{S}))$, unbounded when a block deletes most of a cut; and $\mathbf{S}$ itself is a dense $|F|\times|F|$ matrix. Drawing the $\mathbf{y}_i$ in $\ddot{O}(n\log^3 n)$ would remove the factor $\log n$ of Theorem 1.3; this is the open item of the verification log.

The old reference does not determine the new resistances. The following example shows that the vectors $\mathbf{y}_i$, and in fact any constant-factor approximation of $\mathbf{L}(G-F)$, cannot be computed from the reference of the previous refresh and the deleted edges alone, however much time is spent: the retained graph $G-F$ itself must be consulted, through the structures of Section 6.

Lemma 10.3 (two graphs with the same reference)**.** *Let $n \geq 2^{20}$ with $n \equiv 2 \pmod 4$. There are connected graphs G_1, G_{11} on the same n vertices with weights in $\{1,2,3\}$, the same weighted degrees, and a common set F of $n-1$ edges of weight 1 with $G_1 - F$ and $G_{11} - F$ connected, such that every $\widetilde{\mathbf{L}}$ with $\widetilde{\mathbf{L}} \approx_{4\cdot 10^{-4}} \mathbf{L}(G_1)$ satisfies $\widetilde{\mathbf{L}} \approx_{10^{-3}} \mathbf{L}(G_{11})$, while no symmetric matrix $\mathbf{X}$ satisfies both $\mathbf{X} \approx_{3/4} \mathbf{L}(G_1 - F)$ and $\mathbf{X} \approx_{3/4} \mathbf{L}(G_{11} - F)$. Hence a procedure whose input is $(\widetilde{\mathbf{L}}, F, \mathbf{w}_F, \mathbf{d}(G))$ outputs a $\frac{3}{4}$-approximation of $\mathbf{L}(G-F)$ with probability at most $\frac{1}{2}$ on one of the two graphs.*

Proof. Let $V = A \cup B$, $A = \{a_1, \ldots, a_{n/2}\}$, $B = \{b_1, \ldots, b_{n/2}\}$, $F \coloneqq \{a_i b_i : i \leq n/2\} \cup \{a_i b_{i+1} : i < n/2\}$, and for $j \in \{1, 11\}$ let $H_j \coloneqq \{a_{2i} b_{2i+3} : i \leq j\}$, j edges outside F with $2j$ distinct endpoints. Fix perfect matchings M_j^A of $A \setminus \{a_{2i} : i \leq j\}$ and M_j^B of $B \setminus \{b_{2i+3} : i \leq j\}$ (sets of even size $n/2 - j$). G_j has every pair inside A and inside B with weight 2, raised to 3 on $M_j^A \cup M_j^B$, the edges of F with weight 1 and the edges of H_j with weight 1. Every vertex of A has weighted degree $2(n/2-1) + 1 + \mathbf{d}(F)_v$, the 1 from its matching edge or from its edge of H_j; likewise in B; so the degrees do not depend on j, and $G_j - F$ is connected through H_j.

Write $Q_A \coloneqq \sum_{\{a,a'\} \subseteq A} (\mathbf{x}_a - \mathbf{x}_{a'})^2$, likewise Q_B; then $2Q_A + 2Q_B \leq \mathbf{x}^\top \mathbf{L}(G_1)\mathbf{x}$ and $Q_A = \frac{n}{2} \sum_{a \in A} (\mathbf{x}_a - \bar{\mathbf{x}}_A)^2$ for the mean $\bar{\mathbf{x}}_A$. $\mathbf{L}(G_{11}) - \mathbf{L}(G_1)$ is the sum of $\boldsymbol{\chi}_h \boldsymbol{\chi}_h^\top$ over the ten edges $h \in H_{11} \setminus H_1$ plus $\mathbf{L}(M_{11}^A) - \mathbf{L}(M_1^A) + \mathbf{L}(M_{11}^B) - \mathbf{L}(M_1^B)$ with unit weights. For $h = a_p b_q$ and every $i \leq n/2$, $(\mathbf{x}_{a_p} - \mathbf{x}_{b_q})^2 \leq 3(\mathbf{x}_{a_p} - \mathbf{x}_{a_i})^2 + 3(\mathbf{x}_{a_i} - \mathbf{x}_{b_i})^2 + 3(\mathbf{x}_{b_i} - \mathbf{x}_{b_q})^2$; averaging over i and using that the edges $a_i b_i$ have weight 1 in G_1,

$$(\mathbf{x}_{a_p} - \mathbf{x}_{b_q})^2 \leq \frac{6}{n}\Big(Q_A + \sum_i (\mathbf{x}_{a_i} - \mathbf{x}_{b_i})^2 + Q_B\Big) \leq \frac{12}{n}\, \mathbf{x}^\top \mathbf{L}(G_1)\mathbf{x}.$$

For a unit perfect matching M of a subset of A, $\mathbf{x}^\top \mathbf{L}(M)\mathbf{x} \leq 2\sum_{a \in A} (\mathbf{x}_a - \bar{\mathbf{x}}_A)^2 = \frac{4}{n} Q_A \leq \frac{2}{n} \mathbf{x}^\top \mathbf{L}(G_1)\mathbf{x}$. Summing, $|\mathbf{x}^\top (\mathbf{L}(G_{11}) - \mathbf{L}(G_1))\mathbf{x}| \leq \frac{128}{n} \mathbf{x}^\top \mathbf{L}(G_1)\mathbf{x} \leq 2 \cdot 10^{-4} \mathbf{x}^\top \mathbf{L}(G_1)\mathbf{x}$ for $n \geq 2^{20}$, so $\mathbf{L}(G_{11}) \approx_{2\cdot 10^{-4}} \mathbf{L}(G_1)$ and $\widetilde{\mathbf{L}} \approx_{4\cdot 10^{-4}} \mathbf{L}(G_1)$ gives $\widetilde{\mathbf{L}} \approx_{10^{-3}} \mathbf{L}(G_{11})$.

For $\mathbf{x} = \mathbb{1}_A - \mathbb{1}_B$ the edges inside A, inside B and in F carry no energy in $G_j - F$, and each edge of H_j carries 4: $\mathbf{x}^\top \mathbf{L}(G_j - F)\mathbf{x} = 4j$. $\mathbf{X} \approx_{3/4} \mathbf{L}(G_1 - F)$ forces $\mathbf{x}^\top \mathbf{X}\mathbf{x} \leq 7$ and $\mathbf{X} \approx_{3/4} \mathbf{L}(G_{11} - F)$ forces $\mathbf{x}^\top \mathbf{X}\mathbf{x} \geq 11$. The input $(\widetilde{\mathbf{L}}, F, \mathbf{w}_F, \mathbf{d}(G))$ is the same for both graphs, so the output has the same law, and the two success events are disjoint. □

The structure of Section 6 holds $G - F$ exactly, in P and in the trees T_v, so Lemma 10.3 does not close the question; it says that the queries to P must carry the information about $G - F$ that the reference lacks – in the example, the energy of $\mathbb{1}_A - \mathbb{1}_B$ in $G - F$ – and the halving schedule obtains it by $O(\log n)$ refreshes. Which $O(1)$ query vectors per block suffice is the open item.

References

[AAB+20] Umut A. Acar, Daniel Anderson, Guy E. Blelloch, Laxman Dhulipala, and Sam Westrick. Parallel batch-dynamic trees via change propagation. In *28th Annual European Symposium on Algorithms (ESA 2020)*, volume 173 of *LIPIcs*, pages 2:1–2:23, 2020.

[AABD19] Umut A. Acar, Daniel Anderson, Guy E. Blelloch, and Laxman Dhulipala. Parallel batch-dynamic graph connectivity. In *Proceedings of the 31st ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 2019)*, pages 381–392, 2019.

[ADK+16] Ittai Abraham, David Durfee, Ioannis Koutis, Sebastian Krinninger, and Richard Peng. On fully dynamic graph sparsifiers. In *57th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2016)*, pages 335–344, 2016. Full version: arXiv:1604.02094v2.

[BKM+22] Amos Beimel, Haim Kaplan, Yishay Mansour, Kobbi Nissim, Thatchaphol Saranurak, and Uri Stemmer. Dynamic algorithms against an adaptive adversary: Generic constructions and lower bounds. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2022)*, pages 1671–1684, 2022. arXiv:2111.03980.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[BS80] Jon Louis Bentley and James B. Saxe. Decomposable searching problems I: Static-to-dynamic transformation. *Journal of Algorithms*, 1(4):301–358, 1980.

[BvdBP+22] Aaron Bernstein, Jan van den Brand, Maximilian Probst Gutenberg, Danupon Nanongkai, Thatchaphol Saranurak, Aaron Sidford, and He Sun. Fully-dynamic graph sparsifiers against an adaptive adversary. In *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022*, volume 229 of *LIPIcs*, pages 20:1–20:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. arXiv:2004.08432.

[Chu97] Fan R. K. Chung. *Spectral Graph Theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[CLM+15] Michael B. Cohen, Yin Tat Lee, Cameron Musco, Christopher Musco, Richard Peng, and Aaron Sidford. Uniform sampling for matrix approximation. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS 2015)*, pages 181–190, 2015.

[DDK+20] Laxman Dhulipala, David Durfee, Janardhan Kulkarni, Richard Peng, Saurabh Sawlani, and Xiaorui Sun. Parallel batch-dynamic graphs: Algorithms and lower bounds. In *Proceedings of the 31st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2020)*, pages 1300–1319, 2020.

[DIIM04] Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the 20th ACM Symposium on Computational Geometry (SCG 2004)*, pages 253–262, 2004.

[FW93] Michael L. Fredman and Dan E. Willard. Surpassing the information theoretic bound with fusion trees. *Journal of Computer and System Sciences*, 47(3):424–436, 1993.

[HKZ12] Daniel Hsu, Sham M. Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17:paper no. 52, 1–6, 2012.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[ILMP19] Giuseppe F. Italiano, Silvio Lattanzi, Vahab S. Mirrokni, and Nikos Parotsidis. Dynamic algorithms for the massively parallel computation model. In *Proceedings of the 31st ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 2019)*, pages 49–58, 2019.

[IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing, STOC 1998*, pages 604–613, 1998.

[KLP+16] Rasmus Kyng, Yin Tat Lee, Richard Peng, Sushant Sachdeva, and Daniel A. Spielman. Sparsified Cholesky and multigrid solvers for connection Laplacians. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 842–850, 2016. Full version: arXiv:1512.01892.

[LM00] Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

[LSY+22] Quanquan C. Liu, Jessica Shi, Shangdi Yu, Laxman Dhulipala, and Julian Shun. Parallel batch-dynamic algorithms for k-core decomposition and related graph problems. In *Proceedings of the 34th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 2022)*, pages 191–204, 2022.

[NO21] Krzysztof Nowicki and Krzysztof Onak. Dynamic graph algorithms with batch updates in the massively parallel computation model. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA 2021)*, pages 2939–2958, 2021.

[SS03] Elias M. Stein and Rami Shakarchi. *Fourier Analysis: An Introduction*, volume 1 of *Princeton Lectures in Analysis*. Princeton University Press, 2003.

[SS11] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.

[TDB19] Thomas Tseng, Laxman Dhulipala, and Guy E. Blelloch. Batch-parallel Euler tour trees. In *Proceedings of the 21st Workshop on Algorithm Engineering and Experiments (ALENEX 2019)*, pages 92–106, 2019.

[Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[vdBLN+20] Jan van den Brand, Yin Tat Lee, Danupon Nanongkai, Richard Peng, Thatchaphol Saranurak, Aaron Sidford, Zhao Song, and Di Wang. Bipartite matching in nearly-linear time on moderately dense graphs. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 919–930. IEEE, 2020.

A The packed distance oracle

This section proves Lemma A.1: given the n rows of the matrix $\mathbf{X} = [\mathbf{x}_1^{\text{rand}} \cdots \mathbf{x}_{10^3 \log n}^{\text{rand}}]$ of a refresh, a data structure built in $O(n \log^2 n)$ time returns the squared distance of any two rows within the factor 10^3 in $\ddot{O}(1)$ time, so that the heavy branch of Equation (21) and Step 1 of Figure 6 cost $\ddot{O}(1)$ per heavy edge instead of the $O(\log n)$ of a direct evaluation. The construction is the random-shift hashing of a Gaussian projection of Indyk and Motwani [IM98] and Datar, Immorlica, Indyk and Mirrokni [DIIM04], read at every dyadic scale and reduced to one parity bit per coordinate and scale; the parity bits of a vertex at a scale are packed into $O(\log \log n)$ integers, and the number of disagreeing bits of two such integers is read off one integer product, the word-parallel use of multiplication of Fredman and Willard [FW93]. No dimension-reduction theorem is used: the Gaussian projection serves only to make every coordinate difference a Gaussian of variance the squared distance, and the concentration is Hoeffding's inequality [Hoe63].

Model of computation. The oracle is a collection of records linked by pointers; every record holds $O(1)$ pointers and integers. An integer held by a record has at most $(4c+7)\log n$ bits, and the operations addition, subtraction, comparison, multiplication, integer division (rounding down) and remainder on such integers cost $O(1)$ each, as does following a pointer. No bitwise operation and no integer-addressed memory is used, so the oracle runs on a pointer machine whose registers support these six operations, and on any word RAM with word length at least $(4c+7)\log n$ and unit-cost multiplication, division and remainder. Real numbers (the entries of $\mathbf{X}$, the Gaussian draws) are held in fixed point, as integer multiples of 2^{-w} for the integer w fixed below, and the Gaussian draws are treated as exact, as in Lemma 4.5; the only rounding the analysis tracks is the rounding of the projected rows in Step 3 of Figure 9. Lemma A.9 at the end of the section states how the same bounds are obtained with bitwise exclusive or and population count, or with exclusive or and a lookup table, in place of multiplication.

Lemma A.1. *Let $\mathbf{X} \in \mathbb{R}^{n \times 10^3 \log n}$ have rows $\mathbf{X}_v$ with all nonzero pairwise distances $\|\mathbf{X}_u - \mathbf{X}_v\|_2$ in $[n^{-c}, n^c]$ for a constant $c \geq 1$. After $O(n \log^2 n)$ randomized preprocessing creating $\ddot{O}(n \log n)$ records, a query (u, v) returns $D(u, v)$ in $\ddot{O}(1)$ worst-case time with $\|\mathbf{X}_u - \mathbf{X}_v\|_2^2 \leq D(u, v) \leq 10^3 \|\mathbf{X}_u - \mathbf{X}_v\|_2^2$ for all pairs simultaneously with probability $1 - n^{-14}$; the answers are a fixed function of the preprocessing, so adaptive queries are allowed.*

The construction. The preprocessing draws $\mathbf{G} \in \mathbb{R}^{10^3 \log n \times 10^3 \log n}$ with independent standard Gaussian entries and computes

$$\mathbf{Y} \;\coloneqq\; \mathbf{X}\mathbf{G} \;\in\; \mathbb{R}^{n \times 10^3 \log n},$$

with rows $\mathbf{Y}_v$, in $n \cdot (10^3 \log n)^2 = O(n \log^2 n)$ operations. Let w be the least integer with $2^w \geq 2^6 n^{c+1}$, let v_0 be a fixed vertex, and let the scales be 2^i for the integers $i \in \{-w+6, \ldots, w\}$, which are $2w - 5 \leq 2(c+1)\log n + 9$ scales (the scales are powers of two, so the offsets 2^6 and 6 are forced to be powers of two). For every coordinate j the preprocessing draws a shift γ_j uniformly from the multiples of 2^{-w} in $[0, 2^{w+1})$ and stores, for every vertex v, the shifted and rounded coordinate

$$\overline{\mathbf{Y}}_{v,j} \;\coloneqq\; 2^{-w} \left\lfloor 2^w (\mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j}) \right\rfloor + \gamma_j + 2^{w+1}, \tag{24}$$

a multiple of 2^{-w}, and for every scale i the bit vector $\mathbf{b}_{v,i} \in \{0,1\}^{10^3 \log n}$ with

$$(\mathbf{b}_{v,i})_j \;\coloneqq\; \left\lfloor \overline{\mathbf{Y}}_{v,j} / 2^i \right\rfloor \bmod 2. \tag{25}$$

A query (u, v) finds by bisection over the scales an index h with $\|\mathbf{b}_{u,h} - \mathbf{b}_{v,h}\|_1 \le 250 \log n$ and $\|\mathbf{b}_{u,h-1} - \mathbf{b}_{v,h-1}\|_1 > 250 \log n$ and returns $D := 2^{2h+1}$; the analysis shows that, with high probability, the bits disagree in more than $250 \log n$ coordinates at every scale $2^i \le \|\mathbf{X}_u - \mathbf{X}_v\|_2$ and in at most $250 \log n$ coordinates at every scale $2^i \ge 10\|\mathbf{X}_u - \mathbf{X}_v\|_2$, which places 2^h in $(\|\mathbf{X}_u - \mathbf{X}_v\|_2,\ 20\|\mathbf{X}_u - \mathbf{X}_v\|_2)$. The threshold $250 \log n$ is not a power of ten: it has to separate the disagreement rates 0.47 and 0.1 per coordinate established in the proof with a margin that survives a union bound over n^2 pairs, and $10^2 \log n$ is too close to the lower rate.

We first record the four facts about one coordinate that the analysis uses: the law of a projected difference, the probability that two parity bits disagree, the mean of that probability under a Gaussian input, and Hoeffding's inequality.

Lemma A.2. *For every pair (u, v) the differences $\mathbf{Y}_{u,j} - \mathbf{Y}_{v,j}$, $j = 1, \dots, 10^3 \log n$, are independent $N(0, \|\mathbf{X}_u - \mathbf{X}_v\|_2^2)$ variables, and the pair $(\mathbf{Y}_{u,j} - \mathbf{Y}_{v_0,j}, \mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j})$ depends on $\mathbf{G}$ only through its j-th column.*

Proof. $\mathbf{Y}_{u,j} - \mathbf{Y}_{v,j} = (\mathbf{X}_u - \mathbf{X}_v)\mathbf{g}_j$ for the j-th column $\mathbf{g}_j$ of $\mathbf{G}$, a linear functional of a standard Gaussian vector with squared norm $\|\mathbf{X}_u - \mathbf{X}_v\|_2^2$, hence $N(0, \|\mathbf{X}_u - \mathbf{X}_v\|_2^2)$; the columns $\mathbf{g}_j$ are independent, and both entries of the pair are functions of $\mathbf{g}_j$. □

Lemma A.3. *Let $T \ge 1$ and $M \ge 1$ be integers, let $a, b \in \mathbb{Z}$, and let g be uniform on the integers of $[0, 2TM)$. Then*

$$\Pr\left[\left\lfloor \frac{a+g}{T} \right\rfloor \not\equiv \left\lfloor \frac{b+g}{T} \right\rfloor \pmod 2\right] = \frac{\textsc{Dist}(b-a,\, 2T\mathbb{Z})}{T}.$$

Proof. The parity of $\lfloor m/T \rfloor$ is $2T$-periodic in $m \in \mathbb{Z}$, equal to 0 for $m \bmod 2T \in [0, T)$ and to 1 for $m \bmod 2T \in [T, 2T)$. Since g is uniform on $2TM$ consecutive integers, $(a + g) \bmod 2T$ is uniform on the residues modulo $2T$, so the probability equals the fraction of residues $m \in \{0, \dots, 2T-1\}$ with $\lfloor m/T \rfloor \not\equiv \lfloor (m+d)/T \rfloor \pmod 2$ for $d := (b - a) \bmod 2T$.

- If $d \le T$: the parities differ exactly for $m \in [T - d, T)$ (where $m + d \in [T, T + d)$) and for $m \in [2T - d, 2T)$ (where $m + d \in [2T, 2T + d)$), that is for $2d$ residues, a fraction d/T.
- If $d > T$: the parity of $\lfloor (m + d)/T \rfloor$ is that of $\lfloor (m - (2T - d))/T \rfloor$, and the substitution $m \mapsto m + (2T - d)$ shows that the count equals the count for the difference $2T - d < T$, that is $2(2T - d)$ residues, a fraction $(2T - d)/T$.

Both fractions equal $\textsc{Dist}(b - a, 2T\mathbb{Z})/T$, as $\textsc{Dist}(b - a, 2T\mathbb{Z}) = \min\{d, 2T - d\}$. □

Lemma A.4. *Let $t > 0$ and let $x \sim N(0, \sigma^2)$. The function $x \mapsto \textsc{Dist}(x, 2t\mathbb{Z})/t$ is $1/t$-Lipschitz with values in $[0, 1]$, and*

$$\begin{aligned} \mathbb{E}\big[\textsc{Dist}(x, 2t\mathbb{Z})/t\big] &\ge 0.49 && \text{if } t \le \sigma, \\ \mathbb{E}\big[\textsc{Dist}(x, 2t\mathbb{Z})/t\big] &\le \frac{\sigma}{t}\sqrt{\frac{2}{\pi}} && \text{for every } \sigma. \end{aligned}$$

Proof. The function $x \mapsto \textsc{Dist}(x, 2t\mathbb{Z})/t$ is continuous, even, $2t$-periodic, equal to $|x|/t$ on $[-t, t]$, with slope $\pm 1/t$ and values in $[0, 1]$. Its Fourier coefficients are, for $k \ge 1$, by integration by parts,

$$\frac{1}{t}\int_{-t}^{t} \frac{|x|}{t} \cos\frac{k\pi x}{t}\, dx = \frac{2}{t}\int_0^t \frac{x}{t}\cos\frac{k\pi x}{t}\, dx = -\frac{2}{k\pi t}\int_0^t \sin\frac{k\pi x}{t}\, dx = -\frac{2\big(1 - (-1)^k\big)}{k^2\pi^2},$$

the sine coefficients vanish by evenness, and the mean is $\frac{1}{t}\int_{-t}^{t} |x|/t\,dx \cdot \frac{1}{2} = \frac{1}{2}$. The coefficients are absolutely summable, so the Fourier series converges uniformly to the function (Corollary 2.3 of Chapter 2 of [SS03]):

$$\frac{\textsc{Dist}(x, 2t\mathbb{Z})}{t} \;=\; \frac{1}{2} - \frac{4}{\pi^2} \sum_{k \geq 1 \text{ odd}} \frac{1}{k^2} \cos \frac{k\pi x}{t} \qquad \text{for all } x \in \mathbb{R}.$$

Taking expectations termwise (the series is uniformly bounded), with the characteristic function of $N(0, \sigma^2)$ at $k\pi/t$,

$$\mathbb{E}\Big[\cos \frac{k\pi x}{t}\Big] = \exp\Big(-\frac{k^2\pi^2\sigma^2}{2t^2}\Big).$$

If $t \leq \sigma$ this is at most $e^{-\pi^2/2} \leq 0.0072$ for every $k \geq 1$, and $\sum_{k \text{ odd}} k^{-2} = \pi^2/8$, so $\mathbb{E}[\textsc{Dist}(x, 2t\mathbb{Z})/t] \geq \frac{1}{2} - \frac{1}{2} \cdot 0.0072 \geq 0.49$. The upper bound is $\textsc{Dist}(x, 2t\mathbb{Z})/t \leq |x|/t$ and $\mathbb{E}[|x|] = \sigma\sqrt{2/\pi}$. □

Lemma A.5 (Theorem 2 of [Hoe63]; Theorem 2.8 of [BLM13])**.** *If $\xi_1, \ldots, \xi_N$ are independent random variables with values in $[0, 1]$, then for every $\epsilon > 0$,*

$$\Pr\Big[\sum_j \xi_j \geq \mathbb{E}\Big[\sum_j \xi_j\Big] + \epsilon N\Big] \leq e^{-2\epsilon^2 N} \quad \textit{and} \quad \Pr\Big[\sum_j \xi_j \leq \mathbb{E}\Big[\sum_j \xi_j\Big] - \epsilon N\Big] \leq e^{-2\epsilon^2 N}.$$

The next two lemmas are the word-parallel Hamming distance: the inner product of two bit strings is one digit of the product of the integer holding the first string in fields of f bits with the integer holding the second string in the same fields in reverse order, because that product is the convolution of the two strings and a field width above $\log k$ keeps every convolution coefficient inside its own field.

Lemma A.6. *Let $k \geq 1$ and $f \geq 1$ be integers with $2^f > k$, and let $\mathbf{u}, \mathbf{v} \in \{0,1\}^k$. Then $0 \leq \sum_{j=1}^k \mathbf{u}_j 2^{(j-1)f} < 2^{kf}$, $0 \leq \sum_{j=1}^k \mathbf{v}_j 2^{(k-j)f} < 2^{kf}$, and*

$$\sum_{j=1}^{k} \mathbf{u}_j \mathbf{v}_j \;=\; \left\lfloor \frac{\big(\sum_{j=1}^k \mathbf{u}_j 2^{(j-1)f}\big)\big(\sum_{j=1}^k \mathbf{v}_j 2^{(k-j)f}\big)}{2^{(k-1)f}} \right\rfloor \bmod 2^f.$$

Proof. The two bounds hold since every bit is at most 1 and $\sum_{j \leq k} 2^{(j-1)f} < 2^{kf}$. Expanding the product and grouping the terms by $d \coloneqq j - l$,

$$\Big(\sum_{j=1}^{k} \mathbf{u}_j 2^{(j-1)f}\Big)\Big(\sum_{l=1}^{k} \mathbf{v}_l 2^{(k-l)f}\Big) = \sum_{j=1}^{k}\sum_{l=1}^{k} \mathbf{u}_j \mathbf{v}_l 2^{(k-1+j-l)f} = \sum_{d=-(k-1)}^{k-1} 2^{(k-1+d)f} \sum_{j-l=d} \mathbf{u}_j \mathbf{v}_l.$$

Each inner sum has $k - |d| \leq k$ terms, each a product of bits, so it lies in $\{0, \ldots, k\} \subseteq \{0, \ldots, 2^f - 1\}$. A nonnegative integer has exactly one expansion in base 2^f with digits in $\{0, \ldots, 2^f - 1\}$, so the inner sums are the base-2^f digits of the product, the sum for d at position $k - 1 + d$. Dividing by $2^{(k-1)f}$ and rounding down removes the digits at positions below $k - 1$, and the remainder modulo 2^f keeps the digit at position $k - 1$, which is $\sum_j \mathbf{u}_j \mathbf{v}_j$. □

Lemma A.7. *Let $k, f, r \geq 1$ be integers with $2^f > k$, and let $\mathbf{u}, \mathbf{v} \in \{0,1\}^{rk}$ be given through the $2r$ integers $\sum_{j=1}^k \mathbf{u}_{(\ell-1)k+j} 2^{(j-1)f}$ and $\sum_{j=1}^k \mathbf{v}_{(\ell-1)k+j} 2^{(k-j)f}$, $\ell = 1, \ldots, r$, and the two integers $\|\mathbf{u}\|_1$ and $\|\mathbf{v}\|_1$. Then*

$$\|\mathbf{u} - \mathbf{v}\|_1 = \|\mathbf{u}\|_1 + \|\mathbf{v}\|_1 - 2\sum_{\ell=1}^{r} \left(\left\lfloor \frac{\big(\sum_{j=1}^k \mathbf{u}_{(\ell-1)k+j} 2^{(j-1)f}\big)\big(\sum_{j=1}^k \mathbf{v}_{(\ell-1)k+j} 2^{(k-j)f}\big)}{2^{(k-1)f}} \right\rfloor \bmod 2^f \right),$$

and the right-hand side is evaluated with at most $4r+3$ *operations of the model on integers below* 2^{2kf}*, given the constants* 2^f *and* $2^{(k-1)f}$*.*

Proof. For bits, $|\mathbf{u}_j-\mathbf{v}_j| = \mathbf{u}_j+\mathbf{v}_j-2\mathbf{u}_j\mathbf{v}_j$; summing over j gives $\|\mathbf{u}-\mathbf{v}\|_1 = \|\mathbf{u}\|_1+\|\mathbf{v}\|_1-2\sum_j \mathbf{u}_j\mathbf{v}_j$, and the inner product splits over the r blocks of k consecutive bits, each given by Lemma A.6. Every block costs one multiplication, one integer division, one remainder and one addition into a running sum, and the remaining operations are one multiplication by 2 and two subtractions. The largest integer handled is a product of two of the given integers, each below 2^{kf} by Lemma A.6. □

The data structure. The bit vectors $\mathbf{b}_{v,i}$ are stored in the fields of Lemma A.7 with the parameters

$$\begin{aligned} f &\coloneqq \text{the least integer with } 2^f > 10^3 \log n, \\ k &\coloneqq \text{the largest integer with } 2^{kf} \le n, \\ r &\coloneqq \text{the least integer with } rk \ge 10^3 \log n, \end{aligned}$$

the bits of index above $10^3 \log n$ being zero, so that $2^f > 10^3 \log n \ge k$ as Lemma A.7 requires, every field integer is below $2^{kf} \le n$, and

$$f \le \log\log n + 11, \qquad k \ge \frac{\log n}{f} - 1, \qquad r \le \frac{10^3 \log n}{k} + 1 \le 2 \cdot 10^3 (\log\log n + 11) + 1 \tag{26}$$

for $\log n \ge 2f$, which holds for $n \ge 2^{40}$; for $2^{20} \le n < 2^{40}$, $k \ge 1$ gives $r \le 4 \cdot 10^4 + 1$. In all cases $f = O(\log\log n)$ and $r = O(\log\log n)$.

The $2w-5$ scales of a vertex are arranged in a binary tree of pointers of the same shape for every vertex, so that the bisection over the scales of a query is a descent of two trees in lockstep without any integer-addressed memory; the shape is defined on intervals of scale indices.

Definition A.8. *For integers* $\ell < h$*,* $\mathrm{Tree}(\ell, h)$ *is the empty tree if* $h = \ell + 1$*, and otherwise the binary tree whose root carries the index* m*, the integer* $(\ell + h)/2$ *rounded down, with left subtree* $\mathrm{Tree}(\ell, m)$ *and right subtree* $\mathrm{Tree}(m, h)$*.*

$\mathrm{Tree}(\ell, h)$ carries the indices $\ell+1, \dots, h-1$ once each, has $h-\ell-1$ nodes, the subtree rooted at its node of index m is $\mathrm{Tree}(\ell', h')$ for an interval with $\ell' < m < h'$, and its depth is at most $\log(h-\ell)+1$, since every level halves the interval length.

Figure 8 lists the state of the oracle, Figure 9 its preprocessing and Figure 10 its query; the proof of Lemma A.1 follows the figures.

Global variables:

- the integers c, w (the least integer with $2^w \geq 2^6 n^{c+1}$), f, k, r as above, and the constants 2^f, $2^{(k-1)f}$.
- one *scale record* per scale index $i \in \{-w+6, \ldots, w\}$, holding i, the integer 2^{i+w} and the number 2^{2i+1}, linked in increasing order of i.
- the shifts $\gamma_1, \ldots, \gamma_{10^3 \log n}$, multiples of 2^{-w} in $[0, 2^{w+1})$, as a list.
- for every vertex v the numbers $\overline{\mathbf{Y}}_{v,1}, \ldots, \overline{\mathbf{Y}}_{v,10^3 \log n}$ as a list.
- for every vertex v and every scale index i a *fingerprint record* holding the integer $\|\mathbf{b}_{v,i}\|_1$ and a list of r *chunk records*, the ℓ-th holding the two integers $F_{v,i,\ell}$ and $F^{\mathrm{rev}}_{v,i,\ell}$.
- for every vertex v a tree T_v of the shape $\mathrm{Tree}(-w+6, w)$ whose node of index i points to the scale record of i and to the fingerprint record of (v, i); the vertex record points to the root of T_v and to the fingerprint record of $(v, -w+6)$.

Invariants:

- $\overline{\mathbf{Y}}_{v,j} = 2^{-w} \lfloor 2^w (\mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j}) \rfloor + \gamma_j + 2^{w+1}$.
- $(\mathbf{b}_{v,i})_j = \lfloor \overline{\mathbf{Y}}_{v,j} / 2^i \rfloor \bmod 2$ for $j \leq 10^3 \log n$ and $(\mathbf{b}_{v,i})_j = 0$ for $10^3 \log n < j \leq rk$.
- $F_{v,i,\ell} = \sum_{j=1}^{k} (\mathbf{b}_{v,i})_{(\ell-1)k+j} 2^{(j-1)f}$ and $F^{\mathrm{rev}}_{v,i,\ell} = \sum_{j=1}^{k} (\mathbf{b}_{v,i})_{(\ell-1)k+j} 2^{(k-j)f}$.

Figure 8: State of the packed distance oracle.

PREPROCESS($\mathbf{X}, c$)
Input: the rows $\mathbf{X}_v$, lists of $10^3 \log n$ numbers at the vertex records, and the constant c;
writes all global variables of Figure 8.
Output: none.

1. Set w to the least integer with $2^w \geq 2^6 n^{c+1}$, f to the least integer with $2^f > 10^3 \log n$, k to the largest integer with $2^{kf} \leq n$, r to the least integer with $rk \geq 10^3 \log n$; compute 2^f, $2^{(k-1)f}$ and the scale records by repeated doubling.
2. Draw $\mathbf{G} \in \mathbb{R}^{10^3 \log n \times 10^3 \log n}$ with independent standard Gaussian entries and compute $\mathbf{Y}_v \leftarrow \mathbf{X}_v \mathbf{G}$ for every v.
3. For $j = 1, \ldots, 10^3 \log n$ draw γ_j uniformly from the multiples of 2^{-w} in $[0, 2^{w+1})$; for every v and j set $\overline{\mathbf{Y}}_{v,j} \leftarrow 2^{-w} \lfloor 2^w (\mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j}) \rfloor + \gamma_j + 2^{w+1}$.
4. For every v and every scale record i: create the fingerprint record of (v, i) with $\|\mathbf{b}_{v,i}\|_1 \leftarrow 0$ and r chunk records with $F \leftarrow 0$, $F^{\mathrm{rev}} \leftarrow 0$; then for $\ell = 1, \ldots, r$ and $j = (\ell-1)k+1, \ldots, \ell k$ with $j \leq 10^3 \log n$:
   (a) set β to the remainder modulo 2 of the quotient of the integer $2^w \overline{\mathbf{Y}}_{v,j}$ by 2^{i+w};
   (b) add β to $\|\mathbf{b}_{v,i}\|_1$, add $\beta\, 2^{(j-(\ell-1)k-1)f}$ to $F_{v,i,\ell}$ and $\beta\, 2^{(\ell k - j)f}$ to $F^{\mathrm{rev}}_{v,i,\ell}$, the two powers of 2^f obtained from those of the previous j by one multiplication and one division by 2^f.
5. For every v build T_v of the shape $\mathrm{Tree}(-w+6, w)$ by the recursion of Definition A.8, pointing every node to its scale record and fingerprint record; store at v the root and the fingerprint record of $(v, -w+6)$.

Figure 9: Preprocessing of the packed distance oracle. Its guarantees are in Lemma A.1.

$D = \textsc{Query}(u, v)$
Input: pointers to the vertex records u and v;
reads f, k, r, 2^f, $2^{(k-1)f}$, the trees T_u, T_v, the scale records and the fingerprint records.
Output: the estimate D of $\|\mathbf{X}_u - \mathbf{X}_v\|_2^2$.

1. If $\|\mathbf{b}_{u,-w+6} - \mathbf{b}_{v,-w+6}\|_1 \le 250 \log n$, return $D \leftarrow 0$.
2. Set $h \leftarrow w$ and let the current nodes be the roots of T_u and T_v.
3. While the current nodes are not empty, with m their common index:
   (a) if $\|\mathbf{b}_{u,m} - \mathbf{b}_{v,m}\|_1 \le 250 \log n$, set $h \leftarrow m$ and move both current nodes to their left children;
   (b) otherwise move both current nodes to their right children.
4. Return $D \leftarrow 2^{2h+1}$, read from the scale record of h.

Figure 10: Query of the packed distance oracle. Every $\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1$ is evaluated by the formula of Lemma A.7 from the fingerprint records of (u, i) and (v, i), their chunk lists traversed in lockstep. Its guarantees are in Lemma A.1.

Proof of Lemma A.1. The good event. Let (i) be the event that $|\mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j}| \le n^{c+1}$ for all v and j, and let (ii) be the event that for every pair (u, v) with $\mathbf{X}_u \ne \mathbf{X}_v$ and every scale index $i \in \{-w+6, \dots, w\}$,

$$\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1 > 250 \log n \qquad \text{if } 2^i \le \|\mathbf{X}_u - \mathbf{X}_v\|_2,$$
$$\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1 \le 250 \log n \qquad \text{if } 2^i \ge 10\|\mathbf{X}_u - \mathbf{X}_v\|_2.$$

By Lemma A.2, $\mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j}$ is $N(0, \|\mathbf{X}_v - \mathbf{X}_{v_0}\|_2^2)$ with $\|\mathbf{X}_v - \mathbf{X}_{v_0}\|_2 \le n^c$, and a Gaussian g of variance at most n^{2c} has, with $\lambda = n^{1-c}$,

$$\Pr[g \ge n^{c+1}] \;\le\; e^{-\lambda n^{c+1}} \mathbb{E}[e^{\lambda g}] \;\le\; e^{-n^2 + n^2/2} \;=\; e^{-n^2/2};$$

a union bound over the $2n \cdot 10^3 \log n$ one-sided events gives $\Pr[\text{not (i)}] \le 2 \cdot 10^3 n \log n \, e^{-n^2/2} \le n^{-40}$.

Event (ii). Fix a pair (u, v) with $\mathbf{X}_u \ne \mathbf{X}_v$ and a scale index i. In units of 2^{-w}, by Equations (24) and (25),

$$(\mathbf{b}_{u,i})_j = \left\lfloor \frac{\lfloor 2^w (\mathbf{Y}_{u,j} - \mathbf{Y}_{v_0,j}) \rfloor + 2^{2w+1} + 2^w \gamma_j}{2^{i+w}} \right\rfloor \bmod 2,$$

and likewise for v, where $2^w \gamma_j$ is uniform on the integers of $[0, 2^{2w+1}) = [0, 2 \cdot 2^{i+w} \cdot 2^{w-i})$ with $w - i \ge 0$, and is independent of $\mathbf{G}$. Conditionally on $\mathbf{G}$, Lemma A.3 with $T = 2^{i+w}$ and $M = 2^{w-i}$ gives that the j-th bits of $\mathbf{b}_{u,i}$ and $\mathbf{b}_{v,i}$ disagree with probability

$$\frac{\textsc{Dist}\big(2^{-w} \lfloor 2^w (\mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j}) \rfloor - 2^{-w} \lfloor 2^w (\mathbf{Y}_{u,j} - \mathbf{Y}_{v_0,j}) \rfloor,\; 2^{i+1}\mathbb{Z}\big)}{2^i}, \tag{27}$$

and the $10^3 \log n$ disagreement indicators are independent, since the j-th depends on $(\mathbf{G}, \gamma)$ only through the j-th column of $\mathbf{G}$ and γ_j (Lemma A.2). The argument of $\textsc{Dist}$ in Equation (27) differs from $\mathbf{Y}_{v,j} - \mathbf{Y}_{u,j}$ by less than $2^{-w} \le 2^i/2^6$, as $i \ge -w+6$, so by the Lipschitz bound of Lemma A.4 the expectation of Equation (27) over $\mathbf{G}$ differs from $\mathbb{E}[\textsc{Dist}(x, 2^{i+1}\mathbb{Z})/2^i]$ for $x \sim N(0, \|\mathbf{X}_u - \mathbf{X}_v\|_2^2)$ (Lemma A.2) by at most $2^{-6} \le 0.016$. Hence the probability that the j-th bits disagree is at least $0.49 - 0.016 \ge 0.47$ when $2^i \le \|\mathbf{X}_u - \mathbf{X}_v\|_2$, and at most $\frac{1}{10}\sqrt{2/\pi} + 0.016 \le 0.1$ when $2^i \ge 10\|\mathbf{X}_u - \mathbf{X}_v\|_2$, by Lemma A.4. Lemma A.5 with $N = 10^3 \log n$ gives, in the first case,

$$\Pr\big[\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1 \le 250 \log n\big] \le \Pr\Big[\sum_j \xi_j \le \mathbb{E}\Big[\sum_j \xi_j\Big] - 0.22 \cdot 10^3 \log n\Big] \le e^{-2 \cdot 0.22^2 \cdot 10^3 \log n} \le n^{-90},$$

and in the second case

$$\Pr\big[\|\mathbf{b}_{u,i}-\mathbf{b}_{v,i}\|_1 > 250\log n\big] \le \Pr\Big[\sum_j \xi_j \ge \mathbb{E}\Big[\sum_j \xi_j\Big] + 0.15\cdot 10^3 \log n\Big] \le e^{-2\cdot 0.15^2\cdot 10^3 \log n} \le n^{-45},$$

where ξ_j is the indicator that the j-th bits disagree and $e^{-\log n} \le n^{-1}$. A union bound over the at most n^2 pairs and the $2w-5 \le 2(c+1)\log n + 9 \le n^3$ scales (as c is a constant and $n \ge 2^{20}$) gives $\Pr[\text{not (ii)}] \le n^5 \cdot n^{-45} = n^{-40}$. Both events hold with probability at least $1-2n^{-40} \ge 1-n^{-14}$; the rest of the proof is on these events.

Correctness. The fingerprint records hold the integers of Lemma A.7 for $\mathbf{u} = \mathbf{b}_{u,i}$ and $\mathbf{v} = \mathbf{b}_{v,i}$ (invariants of Figure 8), so every evaluation of $\|\mathbf{b}_{u,i}-\mathbf{b}_{v,i}\|_1$ in Figure 10 is exact. If $\mathbf{X}_u = \mathbf{X}_v$ then $\mathbf{Y}_u = \mathbf{Y}_v$, so $\mathbf{b}_{u,i} = \mathbf{b}_{v,i}$ for every i, Step 1 returns $D = 0 = \|\mathbf{X}_u - \mathbf{X}_v\|_2^2$. Otherwise $n^{-c} \le \|\mathbf{X}_u - \mathbf{X}_v\|_2 \le n^c$, so $2^{-w+6} \le 2^6/(2^6 n^{c+1}) < n^{-c} \le \|\mathbf{X}_u - \mathbf{X}_v\|_2$ and $2^w \ge 2^6 n^{c+1} \ge 10\|\mathbf{X}_u - \mathbf{X}_v\|_2$; by (ii), $\|\mathbf{b}_{u,-w+6} - \mathbf{b}_{v,-w+6}\|_1 > 250\log n$, so Step 1 does not return, and $\|\mathbf{b}_{u,w} - \mathbf{b}_{v,w}\|_1 \le 250\log n$. The descent of Step 3 maintains the invariant that the current nodes of T_u and T_v are the roots of the subtrees $\mathrm{Tree}(\ell, h)$ of both trees for the same interval (ℓ, h), with $\|\mathbf{b}_{u,\ell} - \mathbf{b}_{v,\ell}\|_1 > 250\log n$ and $\|\mathbf{b}_{u,h} - \mathbf{b}_{v,h}\|_1 \le 250\log n$: it holds at the start with $(\ell, h) = (-w+6, w)$ by the two facts just established, the two trees have the same shape so the current nodes carry the same index m and their left (right) children are the roots of $\mathrm{Tree}(\ell, m)$ ($\mathrm{Tree}(m, h)$) in both trees, and Step 3(a) or 3(b) moves to the child whose interval keeps the invariant. When the current nodes are empty, $h = \ell + 1$ with $\|\mathbf{b}_{u,\ell} - \mathbf{b}_{v,\ell}\|_1 > 250\log n$ and $\|\mathbf{b}_{u,h} - \mathbf{b}_{v,h}\|_1 \le 250\log n$; by (ii) the first forces $2^\ell < 10\|\mathbf{X}_u - \mathbf{X}_v\|_2$ and the second forces $2^h > \|\mathbf{X}_u - \mathbf{X}_v\|_2$, so $\|\mathbf{X}_u - \mathbf{X}_v\|_2 < 2^h = 2\cdot 2^\ell < 20\|\mathbf{X}_u - \mathbf{X}_v\|_2$ and

$$\|\mathbf{X}_u - \mathbf{X}_v\|_2^2 \;<\; 2\|\mathbf{X}_u - \mathbf{X}_v\|_2^2 \;<\; D = 2^{2h+1} \;<\; 2\cdot 20^2\|\mathbf{X}_u - \mathbf{X}_v\|_2^2 \;\le\; 10^3\|\mathbf{X}_u - \mathbf{X}_v\|_2^2.$$

The query reads only stored records and draws no randomness, so $D(u, v)$ is a function of the preprocessing, and the bounds hold for all pairs simultaneously on the events (i) and (ii).

Sizes of the integers. On (i), $|2^{-w}\lfloor 2^w(\mathbf{Y}_{v,j} - \mathbf{Y}_{v_0,j})\rfloor| \le n^{c+1} + 2^{-w} < 2^w$, so $0 < \overline{\mathbf{Y}}_{v,j} < 2^{w+3}$ and the integer $2^w\overline{\mathbf{Y}}_{v,j}$ is below 2^{2w+3}; the scale records hold $2^{i+w} \le 2^{2w}$ and $2^{2i+1} = 2^{-2w}\cdot 2^{2(i+w)+1}$, whose numerator is below 2^{4w+1}; and the integers of Lemma A.7 are below $2^{2kf} \le n^2$. Since $2^w < 2^7 n^{c+1}$ and $n \ge 2^{20}$, all integers are below $2^{29} n^{4c+4} \le n^{4c+7}$, within the $(4c+7)\log n$ bits of the model.

Time and space. Step 1 of Figure 9 takes $O(w + kf) = O(\log n)$ operations and creates $2w-5$ scale records; Step 2 takes $(10^3 \log n)^2$ draws and $n(10^3\log n)^2$ operations; Step 3 takes $10^3 \log n$ draws and $O(n\log n)$ operations. Step 4 takes, per vertex and scale, $r+1$ record creations and $O(1)$ operations per $j \le 10^3\log n$, that is $O(n(2w-5)(10^3\log n + r)) = O(n\log^2 n)$ operations and $n(2w-5)(r+1)$ records in all, which is $O(n\log n\cdot\log\log n) = \ddot{O}(n\log n)$ by Equation (26). Step 5 creates $n(2w-7)$ nodes in $O(n\log n)$ operations. The total is $O(n\log^2 n)$ operations and $\ddot{O}(n\log n)$ records. A query evaluates $\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1$ once in Step 1 and once per level of the descent, at most $\log(2w-5)+1 = O(\log\log n)$ times, each evaluation costing at most $4r+3 = O(\log\log n)$ operations by Lemma A.7 and Equation (26), plus the comparison with $250\log n$; the moves between nodes cost $O(1)$ each. The total is $O((\log\log n)^2) = \ddot{O}(1)$ operations, in the worst case. □

The multiplication of Lemma A.7 is the only place where the instruction set matters; the following lemma records the two replacements available on a word RAM, so that Lemma A.1 holds with the same bounds under either of the common instruction sets.

Lemma A.9. *Let the state of Figure 8 be built by Figure 9.*

(a) If the bitwise exclusive or $\oplus$ and the population count Popcount *(the number of ones in the binary expansion) of integers below n cost $O(1)$, then*

$$\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1 = \sum_{\ell=1}^{r} \textsc{Popcount}(F_{u,i,\ell} \oplus F_{v,i,\ell}),$$

evaluated in $3r$ operations; the integers F^{rev} and $\|\mathbf{b}_{v,i}\|_1$ are then not needed.

(b) If the memory is integer-addressed and $\oplus$ costs $O(1)$ but Popcount *is not available, a table of the population counts of the integers below 2^s, s the least integer with $2s \geq kf$, has at most $2\sqrt{n}$ entries, is built in $O(\sqrt{n})$ operations, and gives*

$$\textsc{Popcount}(x) = \textsc{Popcount}(\lfloor x/2^s \rfloor) + \textsc{Popcount}(x \bmod 2^s)$$

for $x < 2^{kf}$ by two lookups, one division, one remainder and one addition.

In both cases the preprocessing remains $O(n \log^2 n)$ operations and the query $O((\log\log n)^2) = \ddot{O}(1)$ operations, and Lemma A.1 holds with the same bounds.

Proof. (a) The integer $F_{v,i,\ell}$ has the bit $(\mathbf{b}_{v,i})_{(\ell-1)k+j}$ at binary position $(j-1)f$ and zeros elsewhere (invariants of Figure 8); the exclusive or acts position by position, so $F_{u,i,\ell} \oplus F_{v,i,\ell}$ has at position $(j-1)f$ the bit $|(\mathbf{b}_{u,i})_{(\ell-1)k+j} - (\mathbf{b}_{v,i})_{(\ell-1)k+j}|$ and zeros elsewhere, and its population count is the number of disagreements in the ℓ-th block; summing over the r blocks gives $\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1$ with one exclusive or, one population count and one addition per block. The integers are below $2^{kf} \leq n$.

(b) The table is filled in increasing order by $\textsc{Popcount}(0) = 0$ and $\textsc{Popcount}(x) = \textsc{Popcount}(\lfloor x/2 \rfloor) + (x \bmod 2)$, three operations per entry; $2^s < 2 \cdot 2^{kf/2} \leq 2\sqrt{n}$ as $2(s-1) < kf$ and $2^{kf} \leq n$. The identity splits the binary expansion of $x < 2^{kf} \leq 2^{2s}$ into its s low bits and its at most s high bits, both below 2^s.

In both cases every evaluation of $\|\mathbf{b}_{u,i} - \mathbf{b}_{v,i}\|_1$ costs $O(r) = O(\log\log n)$ operations, as in Lemma A.7, and the correctness argument of Lemma A.1 is unchanged, as it uses only that the evaluations are exact; the table adds $O(\sqrt{n})$ to the preprocessing. □

B The fixed solver

The random vertex vectors of a refresh are $\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{B}}^\top\widetilde{\mathbf{W}}^{1/2}\mathbf{g}$ for a Gaussian $\mathbf{g}$; for them to be Gaussian with covariance $\mathbf{Z}(\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}\mathbf{Z}(\widetilde{\mathbf{L}})$, the solve must be one fixed linear operator, applied consistently to every right-hand side, and for the trace identities of the refresh this operator must be symmetric. The solver of Kyng, Lee, Peng, Sachdeva and Spielman [KLP+16] builds such an operator; we state what this paper uses and derive it from the statements of the arXiv version (arXiv:1512.01892), quoted verbatim below with their numbers.

Lemma B.1 (from Theorem 1.2, Claim 2.4, Lemma 3.3 and Lemmas I.3 and I.5 of [KLP+16])**.** *There is a routine* LsolverKLPSS *that, on an SDDM matrix $\mathbf{M}$ with n rows, m' nonzero entries and polynomially bounded entries, and an error $\epsilon \in (0, \frac{1}{2}]$, returns $\mathbf{Z}(\mathbf{M}) \leftarrow \textsc{LsolverKLPSS}(\mathbf{M}, \epsilon)$, implicit access to a symmetric matrix $\mathbf{Z}(\mathbf{M})$, such that with probability $1 - n^{-c}$, for the absolute constant $c > 0$ of the phrase "with high probability" in Lemma I.5 of [KLP+16]:*

(a) $\mathbf{Z}(\mathbf{M}) \approx_\epsilon \mathbf{M}^\dagger$ *on the range of* $\mathbf{M}$*, that is,* $\mathbf{x}^\top\mathbf{Z}(\mathbf{M})\mathbf{x} \approx_\epsilon \mathbf{x}^\top\mathbf{M}^\dagger\mathbf{x}$ *for every* $\mathbf{x}$ *in the range of* $\mathbf{M}$*;*

(b) the construction takes $O(m' \log n) + O(n \log^2 n)$ *time;*

(c) one application of $\mathbf{Z}(\mathbf{M})$ *to a vector takes* $O\big((m' + n \log n)\log(1/\epsilon)\big)$ *time.*

The constants in (b) and (c) are $2^{O(k \log^2 k)}$ *for the phase length k of the recursive construction of [KLP+16], which we fix at $k = 28$; see the accounting after the proof.*

The source statements. In [KLP+16] a bDD matrix is a Hermitian block matrix $\mathbf{M}$ with $r \times r$ blocks satisfying $\mathbf{M}_{[i,i]} \succeq \sum_{j \neq i} \|\mathbf{M}_{[i,j]}\| \cdot \mathbf{I}_r$ for every block row i (its Definition 3.4 with $\alpha = 0$, "for $r = 1$, Laplacian matrices are 0-bDD"), and the approximation relation is multiplicative in $e^{\pm\epsilon}$:

> "We say that $\mathbf{A}$ is an ϵ-approximation of $\mathbf{B}$, written $\mathbf{A} \approx_\epsilon \mathbf{B}$, if $e^\epsilon\mathbf{B} \succeq \mathbf{A} \succeq e^{-\epsilon}\mathbf{B}$. This relation is symmetric." (Section 2 of [KLP+16].)

The statements used are the following.

> "**Theorem 1.2** (Sparsified Multigrid). There is an algorithm that, when given a bDD matrix $\mathbf{M}$ with n block rows and m nonzero blocks, produces a solver for $\mathbf{M}$ in $O(m \log n + n \log^{2+o(1)} n)$ work and $O(n^{o(1)})$ depth so that the solver finds ϵ-approximate solutions to systems of equations in $\mathbf{M}$ in $O((m + n \log^{1+o(1)} n)\log(1/\epsilon))$ work and $O(\log^2 n \log\log n \log(1/\epsilon))$ depth."
>
> "**Claim 2.4.** Suppose that all nonzero eigenvalues of $\mathbf{M}$ are at least μ and $\mathbf{Z} \approx_\epsilon (\mathbf{M} + \epsilon\mu\mathbf{I})^{-1}$ for some $0 < \epsilon < 1/2$. Given any $\mathbf{b}$ such that $\mathbf{M}\mathbf{x} = \mathbf{b}$ for some $\mathbf{x}$, we have $\|\mathbf{x} - \mathbf{Z}\mathbf{b}\|_\mathbf{M}^2 \leq 6\epsilon\|\mathbf{x}\|_\mathbf{M}^2$."
>
> "**Lemma 3.3.** Consider an ϵ-SCC for $\mathbf{M}^{(0)}$ where $\mathbf{M}^{(i)}$ and $\mathbf{Z}^{(i)}$ can be applied to a vector in work $W_{\mathbf{M}^{(i)}}$, $W_{\mathbf{Z}^{(i)}}$ and depth $D_{\mathbf{M}^{(i)}}$, $D_{\mathbf{Z}^{(i)}}$ respectively. The algorithm $\textsc{ApplyChain}\big((\mathbf{M}^{(1)}, \mathbf{Z}^{(1)}), \ldots, (\mathbf{M}^{(d)}, \mathbf{Z}^{(d)}); F_1, \ldots, F_d; \mathbf{b}\big)$ corresponds to a linear operator $\mathbf{W}$ acting on $\mathbf{b}$ such that 1. $\mathbf{W}^{-1} \approx_{\sum_{i=1}^d 2\epsilon_i} \mathbf{M}^{(0)}$, and 2. for any vector $\mathbf{b}$, $\textsc{ApplyChain}(\ldots; \mathbf{b})$ runs in $O(\sum_{i=1}^d (D_{\mathbf{M}^{(i)}} + D_{\mathbf{Z}^{(i)}}))$ depth and $O(\sum_{i=1}^d (W_{\mathbf{M}^{(i)}} + W_{\mathbf{Z}^{(i)}}))$ work."
>
> "**Lemma 3.5.** Given a bDD matrix $\mathbf{M}$ with n block-rows, and an $\alpha \geq 0$, bDDSubset computes a subset F of size least $n/(8(1 + \alpha))$ such that $\mathbf{M}_{[F,F]}$ is α-bDD. It runs in

runs in $O(m)$ expected work and $O(\log n)$ expected depth, where m is the number of nonzero blocks in $\mathbf{M}$."

"**Lemma I.3.** For any bDD matrix $\mathbf{M}^{(0)} = \xi\mathbf{I} + \mathbf{B}^{(0)}(\mathbf{B}^{(0)})^*$ for some $\xi > 0$ that has n blocks, the algorithm RecursiveConstruct($\mathbf{M}^{(0)}$) returns a Schur Complement Chain $(\mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \dots; F_1, F_2, \dots)$ whose corresponding the linear operator $\mathbf{W}$ satisfies $\mathbf{W} \approx_{O(1)} (\mathbf{M}^{(0)})^{-1}$. Also, we can evaluate $\mathbf{Wb}$ in $O(\log^2 n \log\log n)$ depth and $2^{O(k\log^2 k)} n \log n$ work for any vector $\mathbf{b}$."

"**Lemma I.5.** With high probability, RecursiveConstruct($\mathbf{M}^{(0)}$) takes $2^{O(\log n/k)}$ depth and $m \log n + 2^{O(k\log^2 k)} n \log^2 n$ work."

Two displayed formulas of the proofs are also used. The proof of Theorem 1.3 in Appendix C of [KLP+16] writes the operator of ApplyChain as

"$(\mathbf{W}^{(1)})^{-1} = \mathbf{U}^\top \operatorname{diag}\big((\mathbf{Z}^{(1)})^{-1}, \dots, (\mathbf{Z}^{(d-1)})^{-1}, \mathbf{I}_{|C_{d-1}|}\big)\mathbf{U}$"

for an upper triangular $\mathbf{U}$ (the displayed factorization on page 26 of arXiv:1512.01892, with $\mathbf{U}$ assembled from the identity and the blocks $\mathbf{Z}^{(i)}\mathbf{M}_{[F_i,C_i]}$), and equation (7) of Appendix E defines the Jacobi operators as

"$\mathbf{Z}^{(k)} \coloneqq \sum_{i=0}^{k} \mathbf{X}_{[F,F]}^{-1}\big(-\mathbf{L}_{[F,F]}\mathbf{X}_{[F,F]}^{-1}\big)^i$"

for the splitting $\mathbf{M}_{[F,F]} = \mathbf{X}_{[F,F]} + \mathbf{L}_{[F,F]}$ of Lemma 3.6 there, with $\mathbf{X}_{[F,F]}$ block diagonal. In the source the proof of Theorem 1.2 chooses $k = \log\log\log n$, so that its $2^{O(k\log^2 k)}$ is $\log^{o(1)} n$ and its depth is $n^{o(1)}$; this is the only place where k is fixed, and only the depth bound needs k to grow.

Proof. Let $\mathbf{M}$ be an SDDM matrix: symmetric, $\mathbf{M}_{u,v} \le 0$ for $u \ne v$, and $\mathbf{M}_{u,u} \ge \sum_{v\ne u} |\mathbf{M}_{u,v}|$ for every u. Then

$$\mathbf{M} = \sum_{u<v} (-\mathbf{M}_{u,v})(\mathbb{1}_u - \mathbb{1}_v)(\mathbb{1}_u - \mathbb{1}_v)^\top + \sum_u \Big(\mathbf{M}_{u,u} - \sum_{v\ne u} |\mathbf{M}_{u,v}|\Big)\mathbb{1}_u\mathbb{1}_u^\top \succeq 0, \tag{28}$$

and $\mathbf{M}$ is a bDD matrix with $r = 1$.

The smallest nonzero eigenvalue. Let $\mathbf{x}$ be a unit vector orthogonal to the kernel of $\mathbf{M}$, and let H be the graph on V with an edge uv for every nonzero $\mathbf{M}_{u,v}$, $u \ne v$. The quadratic form Equation (28) splits over the connected components of H.

- If a component C contains a vertex v with $\mathbf{M}_{v,v} > \sum_{u\ne v} |\mathbf{M}_{v,u}|$, then for every $u \in C$ a path from u to v in H has at most $n-1$ edges, so by the Cauchy–Schwarz inequality $\mathbf{x}_u^2 \le 2(\mathbf{x}_u - \mathbf{x}_v)^2 + 2\mathbf{x}_v^2 \le 2n \sum_{ab \text{ on the path}} (\mathbf{x}_a - \mathbf{x}_b)^2 + 2\mathbf{x}_v^2$, and summing over $u \in C$, $\sum_{u\in C} \mathbf{x}_u^2 \le 2n^2 \cdot \mathbf{x}_C^\top \mathbf{M}\mathbf{x}_C / \min\{|\mathbf{M}_{a,b}|, \mathbf{M}_{v,v} - \sum_{u\ne v}|\mathbf{M}_{v,u}|\}$, the minimum over the nonzero off-diagonal entries and nonzero excesses $\mathbf{M}_{v,v} - \sum_{u\ne v}|\mathbf{M}_{v,u}|$ of $\mathbf{M}$.
- Otherwise $\mathbf{M}$ restricted to C is the Laplacian of a connected graph, $\mathbb{1}_C$ is in the kernel of $\mathbf{M}$, so $\sum_{u\in C} \mathbf{x}_u = 0$, $\max_{u\in C} \mathbf{x}_u - \min_{u\in C} \mathbf{x}_u \ge \max_{u\in C} |\mathbf{x}_u| \ge (\sum_{u\in C} \mathbf{x}_u^2)^{1/2}/\sqrt{n}$, and along a path of at most $n-1$ edges between the maximizer and the minimizer the Cauchy–Schwarz inequality gives $\mathbf{x}_C^\top \mathbf{M}\mathbf{x}_C \ge \min_{a\ne b,\ \mathbf{M}_{a,b}\ne 0} |\mathbf{M}_{a,b}| \cdot \sum_{u\in C} \mathbf{x}_u^2/n^2$.

Summing over the components, every nonzero eigenvalue of $\mathbf{M}$ is at least n^{-3} times the least nonzero off-diagonal entry or excess of $\mathbf{M}$, and for polynomially bounded entries this is $n^{-O(1)}$; write μ for this lower bound, a quantity computed from $\mathbf{M}$ in $O(m' + n)$ time.

The shift. Put $\xi := \epsilon\mu/10$. By Equation (28), $\mathbf{M}+\xi\mathbf{I} = \xi\mathbf{I} + \mathbf{B}^{(0)}(\mathbf{B}^{(0)})^\top$ with $\mathbf{B}^{(0)}$ the matrix whose columns are $(-\mathbf{M}_{u,v})^{1/2}(\mathbb{1}_u - \mathbb{1}_v)$ over the edges of H and $(\mathbf{M}_{u,u} - \sum_{v\neq u}|\mathbf{M}_{u,v}|)^{1/2}\mathbb{1}_u$ over the vertices with positive excess, which is the input form of Lemmas I.3 and I.5. Run RecursiveConstruct$(\mathbf{M}+\xi\mathbf{I})$ with phase length $k = 28$; by Lemma I.5 this takes $O(m'\log n) + O(n\log^2 n)$ time, and by Lemma I.3 it returns a chain whose operator $\mathbf{W}$ satisfies, for an absolute constant $a > 0$,

$$e^{-a}(\mathbf{M}+\xi\mathbf{I})^{-1} \preceq \mathbf{W} \preceq e^{a}(\mathbf{M}+\xi\mathbf{I})^{-1}, \tag{29}$$

and one application of $\mathbf{W}$ takes $O(n\log n)$ time, where the constants hidden in the two $O(\cdot)$ are $2^{O(k\log^2 k)}$ with $k = 28$.

$\mathbf{W}$ *is a fixed symmetric matrix.* By Lemma 3.3, ApplyChain is a linear operator $\mathbf{W}$ in $\mathbf{b}$; the chain is fixed once built, so every application uses the same $\mathbf{W}$. In the displayed factorization $(\mathbf{W}^{(1)})^{-1} = \mathbf{U}^\top \operatorname{diag}((\mathbf{Z}^{(1)})^{-1},\dots,(\mathbf{Z}^{(d-1)})^{-1},\mathbf{I})\mathbf{U}$ of Appendix C, each $\mathbf{Z}^{(i)}$ is the Jacobi operator of equation (7), $\sum_{i=0}^{k}\mathbf{X}_{[F,F]}^{-1}(-\mathbf{L}_{[F,F]}\mathbf{X}_{[F,F]}^{-1})^i = \mathbf{X}_{[F,F]}^{-1/2}\sum_{i=0}^{k}(-\mathbf{X}_{[F,F]}^{-1/2}\mathbf{L}_{[F,F]}\mathbf{X}_{[F,F]}^{-1/2})^i\mathbf{X}_{[F,F]}^{-1/2}$, a symmetric matrix since $\mathbf{X}_{[F,F]}$ is a positive diagonal matrix and $\mathbf{L}_{[F,F]}$ is symmetric (for $r = 1$ all matrices are real). Hence $(\mathbf{W}^{(1)})^{-1}$ and $\mathbf{W} = \mathbf{W}^{(1)}$ are symmetric.

From constant to ϵ accuracy. Let T be the least integer with $(1-e^{-2a})^T \le \epsilon/10$, so $T \le e^{2a}\ln(10/\epsilon) + 1 = O(\log(1/\epsilon))$, and define

$$\mathbf{Z}(\mathbf{M}) := \sum_{i=0}^{T-1} e^{-a}\big(\mathbf{I} - e^{-a}\mathbf{W}(\mathbf{M}+\xi\mathbf{I})\big)^i\mathbf{W}.$$

Since $\mathbf{W}$ is symmetric positive definite by Equation (29), $(\mathbf{I}-e^{-a}\mathbf{W}(\mathbf{M}+\xi\mathbf{I}))^i\mathbf{W} = \mathbf{W}^{1/2}(\mathbf{I} - e^{-a}\mathbf{W}^{1/2}(\mathbf{M}+\xi\mathbf{I})\mathbf{W}^{1/2})^i\mathbf{W}^{1/2}$, so

$$\mathbf{Z}(\mathbf{M}) = \mathbf{W}^{1/2}\, f\big(\mathbf{W}^{1/2}(\mathbf{M}+\xi\mathbf{I})\mathbf{W}^{1/2}\big)\,\mathbf{W}^{1/2}, \qquad f(\lambda) := \sum_{i=0}^{T-1} e^{-a}(1-e^{-a}\lambda)^i = \frac{1-(1-e^{-a}\lambda)^T}{\lambda},$$

a symmetric matrix. By Equation (29), the eigenvalues λ of $\mathbf{W}^{1/2}(\mathbf{M}+\xi\mathbf{I})\mathbf{W}^{1/2}$ lie in $[e^{-a}, e^{a}]$, so $0 \le 1-e^{-a}\lambda \le 1-e^{-2a}$ and $(1-\epsilon/10)/\lambda \le f(\lambda) \le 1/\lambda$; therefore

$$(1-\epsilon/10)\,\mathbf{W}^{1/2}\big(\mathbf{W}^{1/2}(\mathbf{M}+\xi\mathbf{I})\mathbf{W}^{1/2}\big)^{-1}\mathbf{W}^{1/2} \preceq \mathbf{Z}(\mathbf{M}) \preceq \mathbf{W}^{1/2}\big(\mathbf{W}^{1/2}(\mathbf{M}+\xi\mathbf{I})\mathbf{W}^{1/2}\big)^{-1}\mathbf{W}^{1/2},$$

and both outer matrices equal $(\mathbf{M}+\xi\mathbf{I})^{-1}$; that is, $\mathbf{Z}(\mathbf{M}) \approx_{\epsilon/10} (\mathbf{M}+\xi\mathbf{I})^{-1}$. For $\mathbf{x}$ in the range of $\mathbf{M}$, expanding $\mathbf{x}$ in the eigenvectors of $\mathbf{M}$ with eigenvalues $\lambda_i \ge \mu \ge 10\xi/\epsilon$,

$$\mathbf{x}^\top(\mathbf{M}+\xi\mathbf{I})^{-1}\mathbf{x} = \sum_i \frac{(\mathbf{x}^\top\mathbf{v}_i)^2}{\lambda_i+\xi} \in \Big[\frac{1}{1+\epsilon/10}, 1\Big]\cdot\mathbf{x}^\top\mathbf{M}^\dagger\mathbf{x},$$

so $\mathbf{x}^\top\mathbf{Z}(\mathbf{M})\mathbf{x} \in [(1-\epsilon/10)^2, 1]\cdot\mathbf{x}^\top\mathbf{M}^\dagger\mathbf{x}$, which is part (a). Each of the T terms is accumulated with one application of $\mathbf{W}$ ($O(n\log n)$ time) and one multiplication by $\mathbf{M}+\xi\mathbf{I}$ ($O(m'+n)$ time), which is part (c); part (b) is the construction time of Lemma I.5 plus the $O(m'+n)$ time to compute μ and ξ. The probability is that of Lemma I.5, whose phrase "with high probability" the source does not quantify. □

Exponent accounting for Lemma B.1. Every logarithmic factor of Lemma B.1 is traced in Table 1; the source's $\log^{o(1)} n$ factors come from its choice $k = \log\log\log n$ and are replaced here by the constant $2^{O(k\log^2 k)}$, $k = 28$, as explained in the table.

factor	lemma or step producing it	logarithm of	exp.	explicit or derived
build, $m' \log n$	Lemma I.5, the term $m \log n$ (the recurrence $W_{\text{dense}}(m) \le W_{\text{dense}}(m/2) + m \log n + \dots$ in its proof), inherited from the $m \log n$ work of SPARSIFY (Lemma H.1)	n	1	explicit in Lemmas I.5, H.1
build, $n \log^2 n$	Lemma I.5, the term $2^{O(k \log^2 k)} n \log^2 n$; in Lemma I.4 one $\log n$ is the $O(Kn \log n/\epsilon^2)$ density of the output of SPARSIFY (Lemma H.1), the other the $O(\log n)$ linear solves per sparsification ("$w_{\text{eval}} \log n$" in Lemma H.1)	n, n	2	explicit in Lemmas I.4, I.5
apply, $n \log n$	Lemma I.3, the term $2^{O(k \log^2 k)} n \log n$; the $\log n$ is the density $n^{(i)} \log n^{(i)}$ of the chain matrices (Lemma I.2, item 2), summed geometrically over the levels	n	1	explicit in Lemmas I.2, I.3
apply, $\log(1/\epsilon)$	the $T = O(\log(1/\epsilon))$ steps of the iteration in the proof above, from $\mathbf{W} \approx_{O(1)} (\mathbf{M}+\xi\mathbf{I})^{-1}$ (Lemma I.3) to ϵ; the source's Theorem 1.2 states the same factor	$1/\epsilon$	1	derived (our iteration); explicit in Theorem 1.2
apply, $m' \log(1/\epsilon)$	one multiplication by $\mathbf{M} + \xi\mathbf{I}$ per step of the iteration; Theorem 1.2 states "$(m + n \log^{1+o(1)} n) \log(1/\epsilon)$"	$1/\epsilon$	1	explicit in Theorem 1.2
constant $2^{O(k \log^2 k)}$	Lemma I.2 (densities $2^{3ck \log^2(jk)}$), Lemmas I.3, I.5; the source fixes $k = \log \log \log n$ in the proof of Theorem 1.2 for the depth $2^{O(\log n/k)} = n^{o(1)}$, which turns the constant into $\log^{o(1)} n$; the work sums of Lemmas I.3 and I.5 converge for every k with $2\beta^k < 1$, where $\beta = 1 - 1/40$ is the vertex reduction of Lemma 3.5 with $\alpha = 4$, so $k = 28$ makes the constant absolute	none	0	derived from the proofs of Lemmas I.3, I.5; not stated in the source
no $\log n$ from ϵ or ξ	the construction does not depend on ϵ or on the shift ξ (Lemma I.5 has no such term); the $\log(1/\xi) = O(\log n)$ bits of the shift only enter the word size	—	0	explicit
probability n^{-c}	Lemma I.5 "with high probability", inherited from Lemma H.1 and Theorem H.2 (Theorem 3 of [CLM+15]); the exponent is set by the $O(\log n)$ rows of the random projection in the proof of Lemma H.1	n	c	exponent not stated in the source

Table 1: Exponent accounting for Lemma B.1; "explicit" means the factor is written in the quoted statement, "derived" that it was traced by us through the proof named.

C Table of symbols

The symbols of the paper in order of first appearance.

symbol	meaning (where defined)
$\ddot{O}(\cdot)$	$O(\cdot)$ up to $(\log\log n)^{O(1)}$ factors (Section 1)
$G=(V,E,\mathbf{w})$, n	the current graph on the fixed vertex set V, $\lvert V\rvert = n \geq 2^{20}$, weights $\mathbf{w}_e \in [n^{-10}, n^{10}]$ (Section 1)
$\mathbf{L}(G)$, $\mathbf{L}(G)^\dagger$	Laplacian and its pseudoinverse (Section 2)
ϵ, ϵ_k	accuracy of a query, $n^{-10} \leq \epsilon \leq \frac{1}{2}$; of the k-th query (Section 1)
$\mathbf{A} \approx_\epsilon \mathbf{B}$	$(1-\epsilon)\mathbf{B} \preceq \mathbf{A} \preceq (1+\epsilon)\mathbf{B}$ (Section 2)
H	the sparsifier returned by a query (Section 1)
G_k, G_k^{ret}, G_k^{new}	the graph at the k-th query; its retained part (edges of G_{k-1} at the smaller of the two weights) and new part (new edges and increases) (Section 1)
$\mathcal{R}^{\text{eff}}(G)_e$	effective resistance $\boldsymbol{\chi}_e^\top \mathbf{L}(G)^\dagger \boldsymbol{\chi}_e$ (Section 1)
U, Q	total numbers of updates and queries (Lemma 1.2)
F, δ_e	the edges deleted or decreased in a block and the amounts (Section 1)
$\widehat{\mathcal{R}}^{\text{eff}}$	the public resistance estimate of a refresh (Definition 7.1)
θ_{deg}	heavy threshold, $\theta_{\text{deg}} = \log^3 n$ (Definition 4.1)
$\boldsymbol{\chi}_e$, $\mathbb{1}$, $\mathbb{1}_v$	$\mathbb{1}_u - \mathbb{1}_v$ for $e = uv$, the all-ones vector, the v-th unit vector (Section 2)
$\mathbf{d}(G)$, $\mathbf{D}(G)$, $\mathbf{A}(G)$	weighted degrees, their diagonal matrix, the adjacency matrix (Section 2)
$\mathbf{\Pi}$	$\mathbf{I} - \frac{1}{n}\mathbb{1}\mathbb{1}^\top$ (Section 2)
$\mathbf{M}$	$\mathbf{L}(G^{\text{heavy}}) + \mathbf{D}(G^{\text{light}})$ (Definition 4.1)
$\mathbf{Z}(\mathbf{M})$	the fixed solver operator for $\mathbf{M}$ (Lemma B.1)
G^{heavy}, G^{light}	heavy and light graphs (Definition 4.1)
Tr, $\lVert\cdot\rVert$, $\lVert\cdot\rVert_F$, $\lVert\cdot\rVert_{S_6}$	trace, operator, Frobenius and Schatten-6 norms (Section 4)
$\mathbf{X}$	the matrix $[\mathbf{x}_1^{\text{rand}} \cdots \mathbf{x}_{10^3 \log n}^{\text{rand}}]$ (Definition 7.1)
Vol, Cut	volume of a vertex set and the edges across a cut (Definition 3.1)
$\mathcal{C}$, C, (C, v)	the partition, a cluster, a copy (Definition 3.1)
$\text{center}_{\mathbf{x}}(C)$	the center of a cluster for the query vector $\mathbf{x}$ (Definition 3.1)
$\mathcal{E}_2(\mathbf{x}, G)$, $\mathcal{E}_2^{\text{vert}}(\mathbf{x}, C, v)$	energy $\mathbf{x}^\top \mathbf{L}(G)\mathbf{x}$ and its vertex share in a cluster (Definition 3.1)
$\widetilde{\mathcal{E}}$	the energy estimate given to PartitionTwo (Lemma 3.2)
E^{cut}	edges cut by a query and returned to the caller (Lemma 3.2)
$\mathbb{E}$	expectation (Section 4)
ϱ_e, p_e	intensity of an edge and its retention probability $1 - e^{-\varrho_e}$ (Equation (5))
ζ_e	centered retention variable (Equation (5))
$\widetilde{\mathbf{L}}$, $\widetilde{\mathbf{M}}$	sparse and heavy references (Equation (5))
$\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{W}}$	incidence and weight matrices of $\widetilde{\mathbf{L}} = \widetilde{\mathbf{B}}^\top \widetilde{\mathbf{W}} \widetilde{\mathbf{B}}$ (Section 4)
$\mathbf{Q}_e$, $\mathbf{Q}_e^{\text{heavy}}$	the deviation of one edge in the two references (Equation (5))
$\mathbf{x}_j^{\text{rand}}$, $\mathbf{x}_j^{\text{res}}$, $\mathbf{x}_j^{\text{prox}}$	random, residual and proxy vectors of an attempt (Figure 4)
$\overline{\mathbf{Y}}$, $\widetilde{\mathbf{Y}}$	deterministic and random parts of one residual step (Equation (9))
T_v, T_v^{light}	per-vertex trees of the current edges and of the active light edges (Figure 5)
Changed	edges whose heavy/light status changed in the block (Figure 5)
P	the QueryDrivenPartition holding G_k^{ret} (Figure 5)
$\mathbf{x}_v^{\text{vert}}$	proxy mass at v (Definition 7.1)
$D(u, v)$	the packed distance oracle (Lemma A.1)
q_e	proposal probability of an edge (Lemma 8.2)
R, O	the edges sampled for the reference and for the output (Figure 7)
$\mathbf{S}$	$\mathbf{W}_F^{1/2} \mathbf{B}_F \mathbf{L}(G)^\dagger \mathbf{B}_F^\top \mathbf{W}_F^{1/2}$ (Lemma 10.2)
$\mathbb{Z}$, Dist	the integers; the squared-distance table of the packed oracle (Appendix A)
$\mathbf{b}$, $\mathbf{v}$	bit vectors and word vectors of the packed oracle (Appendix A)