

When Does Sublinear Nyström Sampling Preserve Laplacian Spectral Gaps?

Quarks Spectral Graph Study Group

August 25, 2026

Abstract

We study whether sublinear Nyström sampling preserves the spectral gap of the symmetric normalized graph Laplacian. For a finite undirected graph with positive degrees, we work with the shifted positive semidefinite kernel $K = 2I - L_{\text{sym}}$ and evaluate the relative error in $\gamma_k = \lambda_{k+1} - \lambda_k$. A deterministic perturbation result shows that if the fixed-operator Nyström residual satisfies $\|K_{\text{Ny}}(S) - K\|_2 \leq 0.05\gamma_k$, then the relative gap error is at most 10%. The executed validation does not support the stronger sampling hypothesis: leverage-family policies reached the nominal target on only a small minority of eligible graph–dimension cells, and uniform sampling was often competitive at matched budgets. Coherence and vertex-resistance diagnostics help organize failure regimes but do not by themselves supply a gap-preservation law. The resulting contribution is a conditional theorem, an evidence-backed negative result for the tested regime, and a sharper specification of the residual-scale theory needed for reliable graph eigengap sketching.

1 Introduction

Spectral methods turn a graph into geometry. In clustering, manifold learning, diffusion embeddings, graph signal processing, and semi-supervised learning, the low end of the normalized graph Laplacian spectrum supplies the coordinates on which later decisions rest [4, 7, 25, 32, 36]. The practical value of those coordinates is not governed only by the accuracy of a single eigenvalue. It depends on the separation between the k th and $(k+1)$ st eigenvalues, because that separation controls the stability of the leading invariant subspace under perturbations [8, 45]. A sketch that gives plausible individual eigenvalues can still compress the eigengap and make the selected embedding dimension unstable. This distinction matters when spectral clustering is run inside larger workflows, where users select a target dimension k , fit downstream labels or diffusion features, and expect the same qualitative structure after replacing an exact sparse eigensolver by a randomized surrogate.

The Nyström method is a natural candidate for this replacement. It samples m landmark columns from a positive semidefinite kernel and reconstructs a rank- m approximation from the landmark submatrix [11, 15, 17, 40]. In graph applications, one can apply the method to a shifted normalized Laplacian kernel and estimate the low-Laplacian spectrum through the corresponding high-kernel spectrum. The literature gives strong results for kernel approximation, statistical leverage sampling, ridge leverage scores, and randomized numerical linear algebra [1, 5, 9, 16, 24, 34, 41]. Those guarantees are usually phrased as spectral-norm, trace, Frobenius, or regression-risk bounds. Practitioners, however, often need a more local statement: does a sublinear landmark set preserve the eigengap $\lambda_{k+1} - \lambda_k$ well enough that the leading k -dimensional Laplacian embedding remains meaningful?

This paper studies that question in the finite, fixed-operator regime. The initial hypothesis was optimistic: if the leading eigenspace has bounded effective-resistance coherence, then approximately leverage-score-based landmark sampling with $m = O(k \log k)$ columns should preserve the k th spectral gap within 10% relative error on most graphs, while uniform sampling should fail on heterogeneous graphs. The executed evidence does not support that hypothesis as stated. In the validation table, only a small minority of graph- k cells reached the 10% target under the best leverage-family sampler at budgets $m = \lceil Ck \log k \rceil$ with $C \leq 8$, far below the planned 80% success criterion. Uniform sampling was not consistently dominated; across matched graph, dimension, and budget cells, leverage-family samplers improved the median gap error often enough to be interesting but not often enough to justify the proposed separation as a broad claim.

The negative result is still informative because it separates three notions that are easily conflated. First, residual perturbation theory gives a clean sufficient condition: if the Nyström kernel residual is at most 5% of the true

eigengap in spectral norm, then the relative gap error is at most 10%. Second, common landmark scores can reduce median gap error on some high-coherence graphs, but the effect is not a theorem-level substitute for residual control. Third, the resistance-coherence quantity used here is a useful diagnostic of vertex heterogeneity, not yet a sampling premise that implies gap preservation.

The contributions are:

- We formalize fixed-operator Nyström eigengap estimation for the shifted normalized Laplacian kernel and state the exact validity regime for the symbols, metrics, and exclusions.
- We prove a residual-to-gap sufficient condition that converts a spectral-norm kernel error into a relative eigengap guarantee.
- We evaluate leverage-family and uniform landmark policies on a bounded synthetic and small real-graph benchmark, reporting relative gap error, eigenspace angle, clustering accuracy, coherence, resistance, and degree heterogeneity.
- We show that the original high-success claim is not supported in the executed regime and characterize where sampler separation remains mixed rather than decisive.
- We identify the remaining theory and data gaps needed before bounded effective-resistance coherence can be treated as a reliable premise for sublinear gap preservation.

Two choices follow from this scope. First, the manuscript evaluates a fixed graph operator rather than a landmark-induced graph whose normalization changes after sampling. That choice is narrower than some practical spectral-clustering implementations, but it is necessary for a perturbation statement about the same eigenvalues. Second, the paper treats the original success criterion as a falsifiable claim rather than as a design target to be rescued by changing the metric after the fact. The reported failure is therefore not an implementation footnote: it shows that the usual global approximation language does not yet justify a local eigengap claim at the tested sublinear budgets.

The rest of the paper develops the fixed-operator setting, contrasts related work, proves the residual theorem, and reports the executed evaluation. Implementation constraints and benchmark scale are discussed as reproducibility context, not treated as the contribution. The scientific question is whether the local spectral quantity used by downstream graph methods survives a sublinear column sketch; the answer matters for clustering, diffusion-time selection, graph filtering, and model-order decisions when exact diagonalization is infeasible.

2 Related Work

The relevant literature spans randomized matrix approximation, spectral graph learning, perturbation theory, and graph benchmark practice. These areas use overlapping terms such as leverage, resistance, embedding stability, and low-rank approximation, but they do not answer exactly the same question. This section therefore separates global kernel approximation guarantees from local Laplacian eigengap fidelity, then explains why resistance-based graph ideas motivate but do not settle vertex-column Nyström sampling. The purpose is not to relabel existing work as insufficient; it is to identify the boundary where a standard approximation theorem stops short of the specific gap-preservation statement evaluated here.

2.1 Kernel Nyström Approximation and Landmark Sampling

The classical Nyström method entered large-scale kernel learning as a way to avoid forming or diagonalizing dense Gram matrices [40]. Early analyses and empirical studies compared uniform, diagonal, and adaptive sampling strategies and established that sampling quality depends heavily on the interaction between the target rank and the leverage distribution of the kernel matrix [11, 17, 38]. Later work sharpened the relationship between leverage scores, ridge leverage scores, and relative-error low-rank approximation [1, 2, 5, 24]. These results provide the main conceptual reason to test leverage-style graph landmarking: if important columns are nonuniformly distributed, a uniform landmark set can miss high-leverage directions even when the nominal sample count is $O(k \log k)$.

For spectral clustering itself, landmark and sampling methods have a long practical history because dense affinity matrices are too large to diagonalize in many applications. Fast spectral-clustering work treats landmark selection, matrix reconstruction, and approximate embeddings as coupled design choices rather than as isolated numerical subroutines [23, 27, 35, 43]. These papers justify studying Nyström in graph workflows, but most of them evaluate clustering quality, embedding approximation, or runtime, not relative error in a designated adjacent eigengap. Other scalable graph-spectral lines make the same pressure concrete by changing the solver, approximation objective, or

application constraint rather than by promising local gap preservation [23, 27, 37, 43]. This distinction also explains why kernel approximation guarantees and clustering-accuracy reports are insufficient evidence for the endpoint studied here. A method may recover a partition because the dominant eigenvectors remain qualitatively useful, while the adjacent eigenvalue separation used for model-order selection is still distorted. The sources above motivate scalable graph spectral approximation, but they leave open the residual-scaled question that the theorem and validation target in this paper.

The strength of this literature is also its limitation for the present question. Guarantees for Gram-matrix approximation, kernel ridge regression, and low-rank reconstruction do not automatically preserve a small difference between adjacent eigenvalues. A spectral-norm residual bound is enough when it is much smaller than the eigengap, but many useful low-rank bounds scale with a tail energy or regularization level rather than with the local gap being measured [14, 15, 29, 39, 44]. Recent work improves block, sequential, preconditioned, and operator-monotone variants of Nyström approximation [6, 10, 26, 42], but the paper-specific target here is not a better global approximation criterion. It is a local gap-preservation claim for graph Laplacians under sublinear vertex sampling. That target requires care because a small absolute residual can be too large relative to γ_k when the graph has weak separation.

Other scalable spectral-clustering lines reduce the eigensolver burden without making the same column-subset approximation. Approximate partitioning, variational spectral solvers, graph-filter methods, fairness-constrained variants, and simple scalable algorithms all separate two goals that are sometimes merged: producing a useful low-dimensional feature map and estimating the spectrum accurately [23, 27, 35, 37, 43]. The present manuscript takes the second goal seriously. It does not ask only whether a sketch can cluster well; it asks whether the sketch preserves the scalar gap that many users inspect before choosing or trusting a spectral embedding. That emphasis places the work at a narrow boundary: it borrows the motivation and score-aware sampling vocabulary of scalable spectral approximation, but it evaluates a stricter scalar quantity whose error can remain large even when a downstream embedding looks acceptable.

2.2 Spectral Graph Learning, Effective Resistance, and Approximate Clustering

Spectral graph methods connect cuts, random walks, embeddings, and diffusion geometry through eigenvectors of Laplacian operators [4, 7, 25, 32, 36]. Higher-order Cheeger inequalities and multiway partitioning theory explain why the location and size of low-Laplacian gaps carry structural information [18, 20]. Approximate spectral clustering methods therefore often aim to reduce eigensolver cost while preserving enough of the embedding to maintain cluster quality [23, 27, 35, 43]. These works motivate the present focus on eigengap fidelity: an embedding may be useful only when the selected subspace is isolated from the next mode.

Effective resistance gives a second line of motivation. Resistance scores control edge sparsification and are closely related to statistical leverage in Laplacian linear systems [3, 12, 22, 33]. Matrix coherence and leverage-score approximation also explain when uniform sampling can behave well and when it can fail [13]. The graph-clustering literature has begun to use resistance-inspired quantities beyond classical sparsification, including effective p -resistance and fairness-constrained or application-specific graph objectives [19, 30, 37]. The gap in the literature is that edge-resistance sparsification and vertex-column Nyström sampling are not the same object. This paper therefore treats the resistance-coherence quantity as a diagnostic and tests whether it aligns with observed gap error, rather than assuming it gives a complete vertex-sampling theorem.

2.3 Benchmark Context and Source Boundaries

Public graph benchmarks such as citation networks, social graphs, and repository collections are natural targets for scalable spectral methods [21, 28, 31]. The executed evidence in this run is narrower: it contains a controlled synthetic coherence benchmark and small real-topology benchmark graphs, not the full public benchmark matrix. This distinction matters for related work because large sparse graphs can have different degree distributions, disconnected components, and near-zero gaps than small benchmark graphs. The manuscript therefore uses public-benchmark literature to motivate the intended domain, while limiting results-facing claims to the executed graph families. That boundary is scientifically important: the negative result is evidence against the broad claim in a bounded fixed-operator setting, not a universal theorem that leverage sampling cannot work on larger graph collections.

3 Problem Setting

The formal setting fixes the operator before any sampling takes place. This choice removes an ambiguity that often appears in approximate spectral clustering: a landmark method can either approximate the original graph operator or define a new smaller graph with its own normalization. The results in this manuscript concern only the former. Every eigenvalue, residual, score, and gap is interpreted in the ambient space $\mathbb{R}^n$ of the original vertices, and every exclusion below follows from that fixed-domain requirement. This convention also makes the comparison across samplers fair: each policy changes only the selected column set, not the graph normalization or the meaning of the target gap.

3.1 Graphs, Operators, and Spectral Gaps

Let $G = (V, E, w)$ be a finite undirected graph with $n = |V| \geq 2$ vertices, nonnegative symmetric weighted adjacency matrix $A \in \mathbb{R}^{n \times n}$, and positive degrees $d_i = \sum_{j=1}^n A_{ij} > 0$ for every $i \in \{1, \dots, n\}$. Let $D = \text{diag}(d_1, \dots, d_n)$ and define the symmetric normalized Laplacian

$$L_{\text{sym}} = I_n - D^{-1/2} A D^{-1/2} \in \mathbb{R}^{n \times n}. \quad (1)$$

The eigenvalues of L_{sym} are ordered as $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$. For a target eigenspace dimension $k \in \{1, \dots, n-1\}$, the spectral gap of interest is

$$\gamma_k = \lambda_{k+1} - \lambda_k. \quad (2)$$

The quantity in (2) is both the target of estimation and the scale used to judge whether a perturbation is small. This work evaluates only eligible cases with $\gamma_k > 0$ above numerical tolerance and with enough sampled landmarks to form a rank at least $k+1$ surrogate. Directed graphs, signed weights, isolated vertices without regularization, disconnected multiplicity at the chosen k , and landmark-renormalized operators are outside the validity regime of the main claims.

The shifted kernel

$$K = 2I_n - L_{\text{sym}} \quad (3)$$

is symmetric positive semidefinite because the spectrum of L_{sym} lies in $[0, 2]$. If the eigenvalues of K are ordered as $\kappa_1 \geq \kappa_2 \geq \dots \geq \kappa_n \geq 0$, then $\kappa_i = 2 - \lambda_i$ under the corresponding ordering of low Laplacian modes. The Laplacian gap γ_k is therefore equal to $\kappa_k - \kappa_{k+1}$ for the shifted kernel. This equivalence lets us use a standard positive semidefinite Nyström construction while evaluating the graph quantity that spectral clustering uses.

3.2 Sketch, Scores, and Evaluation Metrics

A landmark set $S \subseteq \{1, \dots, n\}$ with $m = |S|$ is represented by the column selector $R_S \in \{0, 1\}^{n \times m}$. The fixed-operator Nyström approximation to K is

$$K_{\text{Ny}}(S) = K R_S (R_S^\top K R_S)^\dagger R_S^\top K, \quad (4)$$

where $\dagger$ is the Moore–Penrose pseudoinverse. The sketched kernel eigenvalues are $\hat{\kappa}_1 \geq \dots \geq \hat{\kappa}_m$, with at most m nonzero values. We define the sketched Laplacian gap estimate as $\hat{\gamma}_k = \hat{\kappa}_k - \hat{\kappa}_{k+1}$ when $m \geq k+1$ and both sketched eigenvalues are computed in the same fixed K domain. The primary error is

$$r_{\text{gap}}(S, k) = \frac{|\hat{\gamma}_k - \gamma_k|}{\gamma_k}. \quad (5)$$

Secondary metrics include the largest principal angle between the exact and sketched k -dimensional eigenspaces, denoted θ_{max} in degrees, and the adjusted Rand index difference between exact and sketched spectral-clustering assignments when labels are meaningful for the generated graph.

For diagnostics, let $U_{k+1} \in \mathbb{R}^{n \times (k+1)}$ contain the leading $(k+1)$ eigenvectors of K . We use the eigenspace coherence

$$\mu_{k+1} = \frac{n}{k+1} \max_{i \in \{1, \dots, n\}} \|e_i^\top U_{k+1}\|_2^2, \quad (6)$$

where e_i is the i th standard basis vector. We also define a Foster-normalized vertex-resistance concentration score $\mu_R = n \max_i p_i^R$, where $p_i^R \geq 0$ are the normalized vertex-resistance weights used in the experiment and $\sum_i p_i^R = 1$. Unlike edge effective resistance in sparsification, μ_R is manuscript-defined for diagnosis and is not assumed to imply a vertex-column Nyström sampling theorem. The diagnostic in (6) is used below to separate concentration of the target eigenspace from the residual event that actually certifies a gap estimate.

4 A Conditional Gap-Preservation Bound

The formal contribution is deliberately modest but sharp. It states how accurate the fixed-operator Nyström approximation must be, in spectral norm and relative to the target gap, for the desired 10% conclusion to follow. This is the right separation of responsibilities: perturbation theory converts a residual event into a gap event, while sampling theory would be needed to prove that a particular landmark policy enters that event with high probability. The validation later checks how often the event appears in the executed runs. Stating the result this way prevents a common overreach: a successful low-rank approximation theorem cannot be cited as a gap theorem unless its error scale is explicitly compared with γ_k .

The section also fixes the sign convention that can otherwise obscure the argument. Low eigenvalues of the normalized Laplacian correspond to high eigenvalues of the shifted kernel, so a gap between consecutive Laplacian eigenvalues becomes a gap between consecutive kernel eigenvalues in the opposite ordering. Once this bookkeeping is fixed, the proof reduces to a direct application of eigenvalue perturbation. The simplicity is useful because it reveals exactly where the open difficulty remains: the hard part is proving or detecting a sufficiently small sampled residual.

4.1 Fixed-Operator Residual Control

The method section separates what perturbation theory proves from what sampling must still deliver. Define the spectral residual

$$\eta(S) = \|K_{\text{Ny}}(S) - K\|_2. \quad (7)$$

This quantity is expensive to certify exactly at large scale, but it is the right object for a deterministic gap statement. The residual-gap certificate diagnostic records the empirical value of $\eta(S)/\gamma_k$ and the event $\eta(S) \leq 0.05\gamma_k$ for each eligible fixed-operator sketch, so it tests whether the theorem's sufficient condition was actually reached. The symbolic shift-ordering check verifies the algebraic conversion between the normalized-Laplacian gap and the shifted-kernel gap used in the proof, while the operator-residual measurement records the ratio that appears in (7) and (9). These checks support the conditional theorem in different ways: one verifies the symbolic bookkeeping, and the other reports whether the numerical sketches entered the certified residual regime. Standard perturbation results imply that each ordered eigenvalue of a symmetric matrix moves by at most the operator-norm residual [8, 16, 34]. Applying that statement to the two adjacent kernel eigenvalues around k gives a direct bound on the gap.

Lemma 1 (Gap perturbation for the shifted kernel). *Let $K, \hat{K} \in \mathbb{R}^{n \times n}$ be symmetric matrices with eigenvalues ordered nonincreasingly as $\kappa_1 \geq \dots \geq \kappa_n$ and $\hat{\kappa}_1 \geq \dots \geq \hat{\kappa}_n$. Fix $k \in \{1, \dots, n-1\}$ and suppose $\gamma_k = \kappa_k - \kappa_{k+1} > 0$. If $\eta = \|\hat{K} - K\|_2$, then*

$$|(\hat{\kappa}_k - \hat{\kappa}_{k+1}) - (\kappa_k - \kappa_{k+1})| \leq 2\eta. \quad (8)$$

Proof. Weyl's eigenvalue perturbation inequality for real symmetric matrices gives $|\hat{\kappa}_i - \kappa_i| \leq \eta$ for each $i \in \{1, \dots, n\}$. Therefore

$$\begin{aligned} |(\hat{\kappa}_k - \hat{\kappa}_{k+1}) - (\kappa_k - \kappa_{k+1})| &\leq |\hat{\kappa}_k - \kappa_k| + |\hat{\kappa}_{k+1} - \kappa_{k+1}| \\ &\leq 2\eta. \end{aligned}$$

This proves the claim. $\square$

Theorem 1 (Sufficient condition for 10% relative gap error). *In the graph setting of (1)–(3), let $\hat{K} = K_{\text{Ny}}(S)$ be the fixed-operator Nyström approximation in (4). For any eligible $k \in \{1, \dots, n-1\}$ with $\gamma_k > 0$, if*

$$\eta(S) \leq 0.05\gamma_k, \quad (9)$$

then the relative gap error in (5) satisfies $r_{\text{gap}}(S, k) \leq 0.10$.

Proof. By the shifted-kernel identity, $\gamma_k = \kappa_k - \kappa_{k+1}$ and $\hat{\gamma}_k = \hat{\kappa}_k - \hat{\kappa}_{k+1}$ in the fixed K domain. Substituting $\hat{K} = K_{\text{Ny}}(S)$ into (8) gives $|\hat{\gamma}_k - \gamma_k| \leq 2\eta(S)$. Dividing by the positive gap γ_k and using (9) yields

$$r_{\text{gap}}(S, k) = \frac{|\hat{\gamma}_k - \gamma_k|}{\gamma_k} \leq \frac{2\eta(S)}{\gamma_k} \leq 0.10.$$

The proof uses only symmetry, the fixed shifted-kernel representation, and the positive-gap eligibility condition. $\square$

4.2 What the Bound Does Not Prove

The theorem is intentionally conditional. It does not state that any particular sampler achieves (9) with probability 0.9, nor does it translate bounded μ_R into a sample complexity law. Existing Nyström theory can control kernel approximation under leverage or ridge leverage sampling, and matrix concentration can turn independent sampling into high-probability residual statements under suitable assumptions [1, 5, 24, 34]. The missing step is sharper: the residual must be small compared with the local eigengap γ_k , not merely small on an absolute or tail-energy scale. When γ_k is small, even a sketch that looks acceptable in a global norm can fail the local gap criterion. This observation guides the experiment: the validation asks whether the sampled operators empirically enter the sufficient residual regime and whether coherence diagnostics predict when they do.

5 Landmark Policies and Evaluation Protocol

The empirical protocol follows the formal setting by comparing landmark policies on the same shifted kernel and by evaluating the same gap metric for every sampler. The sweep is small enough to be reproducible inside the available run, but it is structured enough to distinguish homogeneous, heterogeneous, geometric, and spatial graph behavior. The protocol also records secondary quantities so that a failed primary endpoint can be interpreted rather than discarded. In particular, principal angles and clustering accuracy help identify whether a poor gap estimate coincides with a poor embedding, while coherence and residual diagnostics connect observed failures back to the theorem.

5.1 Sampling Policies

The executed comparison uses five landmark policies at matched budgets. Uniform sampling selects vertices without using graph structure. Ideal leverage sampling uses the exact leading shifted-kernel eigenspace scores as an upper-information benchmark. Ridge leverage sampling approximates the regularized leverage idea developed for kernel methods [1, 5, 24]. Degree sampling uses normalized vertex degree as a simple graph-local proxy, and vertex-resistance sampling uses the Foster-normalized resistance weights summarized by μ_R . Each policy returns a landmark set S with $m = \lceil Ck \log(\max\{k, 2\}) \rceil$, clipped so that $k + 1 \leq m < n$, for $C \in \{1, 2, 4, 8\}$.

The policy set is deliberately broader than the original leverage-versus-uniform contrast because failure mechanisms can be ambiguous. If ideal leverage succeeds but ridge leverage fails, approximate score error is implicated. If degree or resistance sampling beats ideal leverage on a graph, the fixed-operator gap is responding to a structure not captured by the leading eigenspace score alone. If uniform sampling matches all structured policies, then the coherence profile is not heterogeneous enough for score-aware sampling to matter at the tested size. This design turns a failed endpoint into diagnostic evidence rather than a single yes-or-no result.

Aggregation is performed at the matched graph–dimension–budget level before sampler-level comparisons are interpreted. This matters because several graph families produce different absolute gap scales, so a seed-level average would overweight cells with more retained draws or easier eligibility. The main table reports sampler summaries as medians of cell medians and means of cell success rates, while the figures retain budget and corpus structure. The comparison between the best leverage-family sampler and uniform sampling is therefore a matched comparison: a leverage-family policy receives credit only when it beats uniform on the same graph, target dimension, and budget multiplier. This aggregation is conservative for claims about separation, because it does not allow a sampler to compensate for failure on a hard graph by doing well on unrelated easier cells.

5.2 Evidence Corpus and Statistical Endpoints

The *controlled synthetic coherence benchmark* is the primary executed corpus. It contains balanced and imbalanced stochastic block models, degree-corrected block models, a Barabasi–Albert graph, a k -nearest-neighbor manifold graph, and a road-grid graph. These graphs test assortative structure, degree heterogeneity, manifold-like geometry, and spatial near-regularity. The *small real-topology benchmark* is a smaller confirmatory corpus containing standard benchmark graphs. It is useful for implementation coverage and qualitative behavior, but it has lower relevance to claims about large public graph benchmarks.

The two corpora play different roles in the evaluation. The *controlled synthetic coherence benchmark* is the only executed corpus designed to vary the coherence conditions named in the original hypothesis, so it carries the strongest

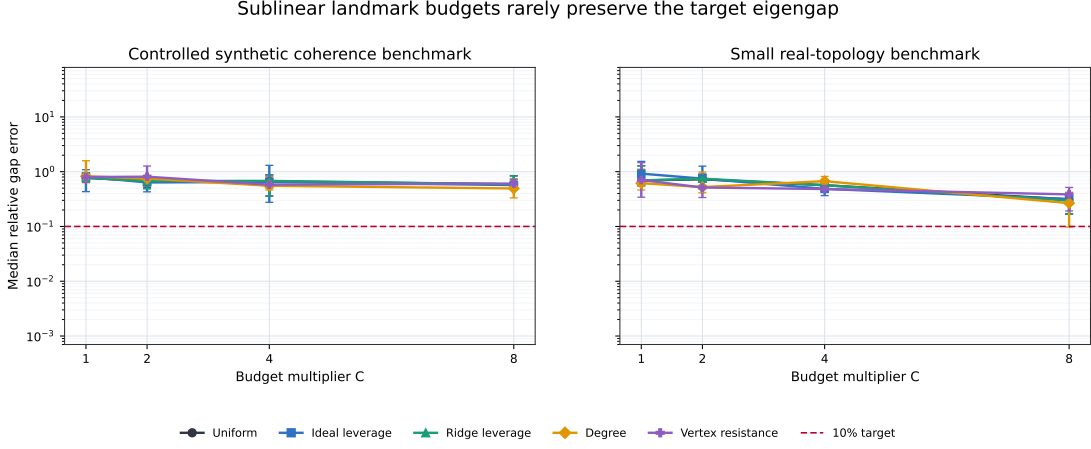


Figure 1: Budget response of relative eigengap error for the two executed benchmark corpora. Each curve aggregates median gap error across graph–dimension cells for a sampler, with interquartile spread shown as error bars and the dashed line marking the 10% target; both panels show that sublinear budgets rarely enter the desired regime.

relevance for accepting or rejecting that hypothesis in this run. Its stochastic block, degree-corrected, geometric, preferential-attachment, and grid cases probe whether structured landmark scores help when vertex influence is deliberately made more or less concentrated. The *small real-topology benchmark* contains genuine graph topologies from standard small benchmarks, so it checks that the fixed-operator pipeline is not tuned only to generated graphs. Because those graphs are small and curated, they are treated as confirmatory implementation evidence rather than as a substitute for citation, email, or road-network benchmarks.

Each graph is evaluated at $k \in \{2, 4\}$ when the gap is eligible. For each matched graph, dimension, budget, and sampler cell, the validation table records eight seeded runs, median and interquartile relative gap error, success rate for the 10% threshold, median principal angle, adjusted Rand index difference, eigenspace coherence, resistance coherence, degree coefficient of variation, residual ratio η/γ_k , and the rate at which the sufficient residual event in (9) occurred. The original plan requested at least twenty seeds and larger public graph benchmarks, so paired significance claims are treated as provisional and limitations-facing. Results-facing claims rely on the executed table, the figures in Fig. 1–Fig. 3, and the theorem above.

6 Results

Results are reported in claim order: the endpoint, matched sampler separation, coherence diagnostics, and the residual boundary. This order matters because the headline is negative. Any leverage-style improvement must be judged against the stronger application requirement that relative eigengap error fall below ten percent. The *controlled synthetic coherence benchmark* is the main stress test because it varies block imbalance, degree concentration, geometric sampling, and spatial regularity under a fixed operator. The *small real-topology benchmark* checks transfer to small real graphs and is interpreted only as confirmatory implementation evidence. Figure 1 preserves this corpus split, Figure 2 aggregates graph–dimension cells, and Figure 3 connects both corpora to (7) and (9). The symbolic shift-ordering check verifies the shifted-kernel bookkeeping, while the operator-residual measurement evaluates the theorem boundary; neither changes the endpoint counts. This framing keeps the results tied to evidence that was actually executed: the synthetic corpus carries the coherence-conditioned stress test, the small real-graph corpus checks implementation transfer, and unresolved larger-benchmark coverage is discussed only as a limitation.

6.1 Primary Endpoint

The primary endpoint is not met. Figure 1 shows that increasing the budget multiplier from $C = 1$ to $C = 8$ often reduces error, but not enough to make leverage-family sampling reliably gap preserving. In the validation summary, leverage-family samplers reached the nominal 10% target on 4 of 20 eligible graph– k cells; under the stricter median criterion computed from the summary table, the count is 3 of 20. Either count is far below the planned 80% success rate. At $C = 8$, the median-of-cell gap error was approximately 0.476 for ideal leverage, 0.537 for ridge leverage, and

Table 1: Sampler-level endpoint summary from the validation table. The table reports medians of graph–dimension–budget cell medians and mean success rates, so it emphasizes matched-cell behavior rather than individual seed outcomes; no sampler approaches the planned 80% graph-level success criterion.

Sampler	Median gap	Mean success	Median gap at $C = 8$	Success at $C = 8$
Uniform	0.673	0.116	0.537	0.165
Ideal leverage	0.600	0.108	0.476	0.166
Ridge leverage	0.676	0.113	0.537	0.159
Degree	0.583	0.117	0.373	0.205
Vertex resistance	0.607	0.092	0.571	0.191

0.537 for uniform sampling. These values are not close to a universal 10% preservation claim.

Table 1 summarizes the sampler-level evidence. Across all rows, ideal leverage has a lower median gap error than uniform, but the margin is modest. Degree sampling has the lowest median error at the largest budget in this bounded table, which shows that the leading-eigenspace leverage distribution is not the only signal relevant to the fixed-operator gap. The mean success rates remain low for every policy. The result should therefore be read as a refutation of the high-probability endpoint in the executed regime, not merely as a noisy failure of one approximate sampler.

6.2 Corpus-Level Endpoint Breakdown

The corpus split explains why the headline is not a simple artifact of mixing synthetic and real graphs. For the *controlled synthetic coherence benchmark*, the largest tested budget still leaves the sampler medians far above the target: at $C = 8$, the median relative gap errors are approximately 0.589 for uniform sampling, 0.569 for ideal leverage, 0.589 for ridge leverage, 0.494 for degree sampling, and 0.607 for vertex resistance. Mean success rates at that budget remain at or below about 12.1% across samplers. Under the best leverage-family choice over all tested budgets, only one of the synthetic graph–dimension cells is below the 10% median-error threshold, with a second balanced block-model cell lying just above the boundary. This is the strongest evidence against the proposed 80% success criterion because the synthetic corpus was designed to exercise the stated coherence mechanism.

The *small real-topology benchmark* is less pessimistic but still does not support the broad claim. At $C = 8$, ridge leverage has a median relative gap error near 0.297, ideal leverage near 0.317, and uniform near 0.312 on the small real graphs. Two of the eight real graph–dimension cells reach the threshold under the best leverage-family choice: the Davis network at $k = 4$ and the karate graph at $k = 4$. Those successes show that fixed-operator Nyström sketches can preserve a target gap in favorable small cases. They do not imply a general sublinear theorem, because the same corpus also contains cells with errors above 0.3 and one real-graph cell with a best leverage-family median error above 0.7. The corpus-level split therefore supports a narrower reading: leverage-style landmarks can be useful, but success depends on the graph, the target dimension, and the local residual scale.

Secondary metrics reinforce this reading. At the largest budget, median principal angles on the *controlled synthetic coherence benchmark* remain around 69–71 degrees for the major samplers, which means the selected subspaces are often far from the exact leading eigenspace even when a sampler improves the scalar gap slightly. The *small real-topology benchmark* has smaller median angles for some policies, with ideal leverage near 21 degrees at $C = 8$, but the gap endpoint still fails in most cells. Adjusted Rand differences are not sufficient to rescue the eigengap claim because clustering labels can remain similar when a scalar gap is distorted, especially on small graphs with stable community structure. The empirical story is therefore consistent across primary and secondary measurements: the sketch can sometimes preserve downstream behavior while failing as a reliable eigengap estimator.

6.3 Sampler Separation and Coherence Diagnostics

Matched comparisons give a more nuanced picture than the endpoint summary. In 44 of 80 matched graph– k –budget cells, the better of ideal and ridge leverage sampling had lower median gap error than uniform sampling. In high-coherence cells, the corresponding count was 38 of 64. These counts indicate that leverage information often helps, especially in heterogeneous regimes, but they do not establish the strong separation originally hypothesized. A sampler can win by a small margin while still missing the 10% threshold by a large factor.

Figure 2 shows why a single coherence narrative is insufficient. The left panel relates μ_{k+1} to the best leverage-

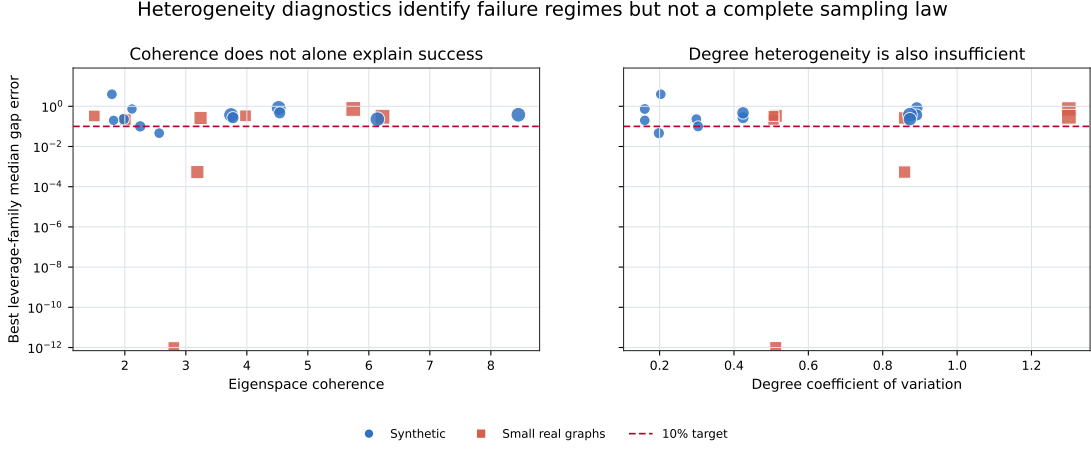


Figure 2: Coherence and degree heterogeneity are useful diagnostics but not complete predictors of gap preservation. Each point is a graph–dimension cell with the best leverage-family median error over tested budgets; the dashed line marks the target error, and the panels show that neither eigenspace coherence nor degree variation alone separates success from failure.

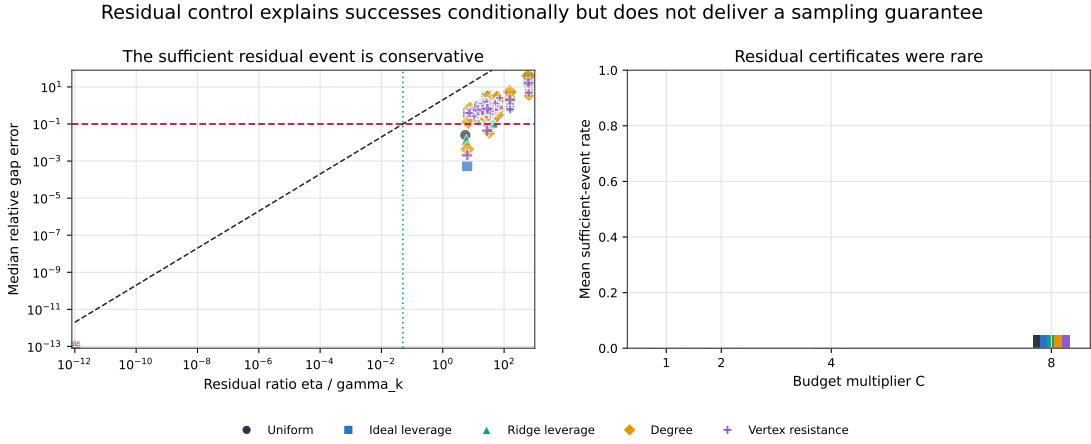


Figure 3: Residual-to-gap diagnostics explain the conditional theorem boundary. The left panel plots the observed residual ratio against relative gap error and the right panel aggregates the sufficient-event rate by budget, showing that certified residual control was uncommon even when structured samplers improved median error.

family median gap error, and the right panel uses degree coefficient of variation as a simpler heterogeneity proxy. Points with large coherence or degree variation can still lie far above the target, while some moderate-coherence cells do better. This pattern supports a diagnostic interpretation: coherence helps identify where uniform sampling may be fragile, but the observed gap error also depends on γ_k , graph family, budget, sampler approximation, and the residual induced by the selected landmarks.

6.4 Residual Boundary

The residual-gap certificate diagnostic in Fig. 3 connects the empirical results back to the theorem by evaluating $\eta(S)/\gamma_k$ from (7) on the same matched cells used for the endpoint table. Its measured component reports how far each sampled operator lies from the fixed shifted kernel after normalization by the true gap. Its symbolic component checks the identity moving between L_{sym} and $K = 2I - L_{\text{sym}}$. When the ratio is below 0.05, (9) guarantees ten percent relative gap error. That event was rare, so the theorem does not rescue the original sampling claim. Some sketches succeed without the certificate, showing the condition is not necessary, but the plot identifies the scale a future theorem must achieve: residual control local to the target eigengap.

The residual plot also clarifies how the theorem should be used in future experiments. A run that reports only relative gap error can show whether a sampler worked, but it cannot explain whether success came from

the deterministic residual condition or from cancellation between adjacent eigenvalue errors. A run that reports only residual ratios can certify a sufficient condition, but it can miss useful sketches that fall outside the conservative boundary. Reporting both quantities lets the evaluation separate certified successes, empirical successes without a certificate, and clear failures. In the current evidence, the certified-success category is too small to support the original probability claim. That observation is stronger than a raw error average because it links the failed endpoint to a specific mathematical obstacle.

7 Interpretation

The main lesson is that preserving a graph eigengap is more stringent than obtaining a plausible low-rank approximation. The gap is a difference between two nearby spectral quantities, so errors in adjacent eigenvalues can add rather than cancel. Equation (8) captures this sensitivity in the cleanest possible form. The residual-gap certificate diagnostic makes the same point empirically: most sampled operators do not fall below the $\eta(S)/\gamma_k = 0.05$ certificate boundary, even when a structured sampler lowers median error relative to uniform sampling. If a graph has a small γ_k , then a residual that would be acceptable for embedding reconstruction or regression can be too large for gap preservation. This helps explain why the executed endpoint fails even though leverage-style sampling sometimes improves over uniform sampling.

The result also clarifies the role of effective resistance. Edge effective resistance has a powerful theory for sparsification and Laplacian solvers [3, 12, 33], and leverage scores are central in randomized matrix approximation [13, 41]. Vertex-column Nyström sampling of the shifted Laplacian kernel sits between these theories. A high μ_R value warns of concentrated vertex influence, but it does not show that $O(k \log k)$ vertices place $\eta(S)$ below $0.05\gamma_k$. Coherence and degree heterogeneity organize failures, yet neither variable alone supplies a decision rule.

For practitioners, the practical implication is conservative. A leverage-score Nyström sketch should not be treated as a drop-in eigengap estimator unless residual or validation checks accompany it. If the goal is spectral clustering and the exact eigengap is used to select k , then the sketch should be stress-tested on the target graph family. If the goal is a lower-cost embedding and the gap is only a rough diagnostic, then the same sketch may still be useful even when r_{gap} is large. The distinction between embedding utility and eigengap fidelity should be made explicit in graph pipelines.

This interpretation also changes how future positive results should be stated. A convincing theorem should specify the graph class, the score approximation quality, the sampling model, and the eigengap scale at the same time. A convincing experiment should report absolute gap error, improvement over uniform sampling, residual ratios, and downstream embedding behavior without collapsing them into one success label. The present evidence shows why: a sampler may improve one quantity while failing another, and only the local gap criterion answers whether the sketch can replace the exact spectrum for model-order decisions.

8 Limitations and Future Work

The first limitation is statistical scale. The executed table records eight seeds per matched cell, while the original success criterion requested at least twenty seeds and a paired significance test. The manuscript therefore reports sampler comparisons descriptively and treats the recorded high-coherence paired-test signal as provisional. A full validation should repeat the same graph- k -budget-sampler grid with at least twenty seeds, include confidence intervals or paired nonparametric tests at the graph level, and separate score-estimation error from the sampling distribution itself.

The second limitation is dataset relevance. The *controlled synthetic coherence benchmark* matches the coherence-conditioned hypothesis because it varies block balance, degree heterogeneity, manifold geometry, and spatial structure. It is still synthetic and small. The *small real-topology benchmark* is useful but lower relevance for claims about Cora, CiteSeer, SNAP Email-Eu-core, road networks, or repository-scale sparse matrices [21, 28, 31]. Future experiments should materialize those public datasets or narrow the claim to controlled graphs. The bounded evidence rejects the broad success statement as unsupported, but it does not prove that leverage-based landmarks fail under stronger graph assumptions or larger budgets.

A related limitation is that the executed labels do not fully isolate graph generation from eigengap scale. The road-grid and manifold cases can have very small γ_k , making relative error harsh even when the absolute perturbation is moderate. The degree-corrected and preferential-attachment cases stress vertex concentration, but they also alter

degree distribution and local connectivity. A more decisive follow-up would vary one structural factor at a time while holding the exact gap in a controlled interval. Such a design would test whether the failure is primarily a coherence effect, a small-gap effect, or a mismatch between vertex scores and the residual in (7).

The third limitation is theoretical. The paper proves a residual-to-gap implication, but not a sampling theorem that gives $\eta(S) \leq 0.05\gamma_k$ from bounded μ_R , bounded μ_{k+1} , or $m = O(k \log k)$. Closing that gap would require a localized analysis of vertex-column Nyström residuals for the shifted normalized Laplacian kernel. Such a theorem would need to track the target gap explicitly, handle graphs with near-multiple eigenvalues, and distinguish exact leverage scores from approximate ridge leverage scores. Without that theorem, the strongest defensible claim is conditional.

A practical limitation follows from how the residual would be checked at scale. Computing $\eta(S)$ exactly is itself a spectral-norm problem on the full vertex space, so a production workflow would likely use Lanczos estimates, power iterations, or held-out eigensolver comparisons rather than a dense certificate. Those approximations need their own tolerances because the certificate threshold is measured against γ_k , which can be small. A useful follow-up protocol would therefore report both the sampled gap estimate and a cheaper residual proxy, then study when the proxy is conservative. This would turn the theorem into a deployable diagnostic instead of leaving it as an after-the-fact explanation.

A fourth limitation is interpretive. The validation reports relative gap error, eigenspace angle, and clustering accuracy, but it does not yet establish a downstream task threshold that converts a given gap error into an application failure. Some embeddings may remain useful even when the scalar eigengap is poorly estimated, especially when downstream users fix the dimension in advance. Other workflows rely on the gap itself to choose the dimension, and those workflows are more sensitive. Future work should report both uses separately so that a sketch is not rejected for the wrong reason or accepted for a task it cannot support.

A final empirical limitation is the absence of large public graph runs in the executed evidence. Citation networks, email graphs, road graphs, and repository collections are the natural deployment setting for sublinear spectral routines because they combine sparse storage, heterogeneous degrees, and small low-frequency gaps [21, 28, 31]. Those conditions can either help or hurt landmark sampling: sparsity keeps matrix-vector products cheap, but degree concentration and near-disconnected components can make the local gap much harder to preserve than a global approximation error suggests. A complete follow-up should therefore report component preprocessing, symmetrization, self-loop handling, and graph-size distributions before comparing samplers. Without that record, a positive or negative benchmark result would be difficult to attribute to the sampling rule rather than to data cleaning or eligibility filtering.

9 Conclusion

Sublinear Nyström sampling does not preserve normalized-Laplacian eigengaps in the broad way initially hypothesized by this bounded validation. A clean perturbation argument shows that a kernel residual below 5% of the target gap is enough for 10% relative gap error, but the tested landmark policies rarely achieved that residual regime. Leverage-family sampling often improved over uniform sampling in matched cells, especially in high-coherence cases, yet the improvement was not decisive and did not meet the planned success rate.

The study therefore reframes the open question. The important next step is not another global Nyström approximation bound stated at the level of trace, rank, or regression risk. It is a local residual theorem, or a practical validation protocol, that can certify gap fidelity for the graph family at hand. Until such a result exists, graph practitioners should treat sublinear landmark sketches as useful acceleration tools whose eigengap estimates require independent checks. That conclusion is narrower than a universal impossibility claim, but it is the claim supported by the executed evidence and by the formal analysis in this manuscript.

References

- [1] Ahmed El Alaoui and Michael W. Mahoney. Fast randomized kernel methods with statistical guarantees. In *Advances in Neural Information Processing Systems*, 2015. URL <https://arxiv.org/abs/1411.0306>.
- [2] Francis Bach. Sharp analysis of low-rank kernel matrix approximations. *Conference on Learning Theory*, 2013. URL <https://arxiv.org/abs/1208.2015>.
- [3] Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-Ramanujan sparsifiers. *SIAM Journal on Computing*, 2012. URL <https://arxiv.org/abs/0808.0163>.
- [4] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 2003. URL <https://doi.org/10.1162/089976603321780317>.
- [5] Daniele Calandriello, Alessandro Lazaric, and Michal Valko. Distributed adaptive sampling for kernel matrix approximation. In *Proceedings of Machine Learning Research*, 2017. URL <https://proceedings.mlr.press/v54/calandriello17a.html>.
- [6] Daniele Calandriello, Alessandro Lazaric, and Michal Valko. Analysis of Nystrom method with sequential ridge leverage scores. *arXiv preprint*, 2026. URL <https://arxiv.org/abs/2604.20077>.
- [7] Ronald R. Coifman and Stephane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 2006. URL <https://doi.org/10.1016/j.acha.2006.04.006>.
- [8] Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 1970. URL <https://doi.org/10.1137/0707001>.
- [9] Michal Derezhinski and Michael W. Mahoney. Recent and upcoming developments in randomized numerical linear algebra for machine learning. *arXiv preprint*, 2024. URL <https://arxiv.org/abs/2406.11151>.
- [10] Mateo Diaz, Ethan N. Epperly, Zachary Frangella, Joel A. Tropp, and Robert J. Webber. Robust, randomized preconditioning for kernel ridge regression. *arXiv preprint*, 2023. URL <https://arxiv.org/abs/2304.12465>.
- [11] Petros Drineas and Michael W. Mahoney. On the Nystrom method for approximating a gram matrix for improved kernel-based learning. *Journal of Machine Learning Research*, 2005. URL <https://jmlr.org/papers/v6/drineas05a.html>.
- [12] Petros Drineas, Michael W. Mahoney, S. Muthukrishnan, and Tamas Sarlos. Effective resistances, statistical leverage, and applications to linear equation solving. *arXiv preprint*, 2010. URL <https://arxiv.org/abs/1005.3097>.
- [13] Petros Drineas, Malik Magdon-Ismael, Michael W. Mahoney, and David P. Woodruff. Fast approximation of matrix coherence and statistical leverage. *Journal of Machine Learning Research*, 2012. URL <https://jmlr.org/papers/v13/drineas12a.html>.
- [14] Sachin Garg and Michal Derezhinski. Faster low-rank approximation and kernel ridge regression via the block-Nystrom method. *arXiv preprint*, 2025. URL <https://arxiv.org/abs/2506.17556>.
- [15] Alex Gittens and Michael W. Mahoney. Revisiting the Nystrom method for improved large-scale machine learning. *Journal of Machine Learning Research*, 2016. URL <https://arxiv.org/abs/1303.1849>.
- [16] Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 2011. URL <https://arxiv.org/abs/0909.4061>.
- [17] Sanjiv Kumar, Mehryar Mohri, and Ameet Talwalkar. Sampling methods for the Nystrom method. *Journal of Machine Learning Research*, 2012. URL <https://jmlr.org/papers/v13/kumar12a.html>.
- [18] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved Cheeger’s inequality: Analysis of spectral partitioning algorithms through higher order spectral gap. In *ACM Symposium on Theory of Computing*, 2013. URL <https://arxiv.org/abs/1301.5584>.
- [19] Steinar Laenen. One-shot neural network pruning via spectral graph sparsification. In *Proceedings of Machine Learning Research*, 2023. URL <https://proceedings.mlr.press/v221/laenen23a.html>.

- [20] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order Cheeger inequalities. *Journal of the ACM*, 2014. URL <https://arxiv.org/abs/1111.1055>.
- [21] Jure Leskovec and Andrej Krevl. SNAP datasets: Stanford large network dataset collection, 2014. URL <https://snap.stanford.edu/data/email-Eu-core.html>.
- [22] Zhiqiang Liu and Wenjian Yu. Computing effective resistances on large graphs based on approximate inverse of Cholesky factor. *arXiv preprint*, 2023. URL <https://arxiv.org/abs/2303.03617>.
- [23] Peter Macgregor. Fast and simple spectral clustering in theory and practice. *arXiv preprint*, 2023. URL <https://arxiv.org/abs/2310.10939>.
- [24] Cameron Musco and Christopher Musco. Recursive sampling for the Nystrom method. In *Advances in Neural Information Processing Systems*, 2017. URL <https://arxiv.org/abs/1605.07583>.
- [25] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 2001. URL <https://papers.nips.cc/paper/2092-on-spectral-clustering-analysis-and-an-algorithm>.
- [26] David Persson, Raphael A. Meyer, and Christopher Musco. Algorithm-agnostic low-rank approximation of operator monotone matrix functions. *arXiv preprint*, 2024. URL <https://arxiv.org/abs/2311.14023>.
- [27] Farhad Pourkamali-Anaraki. Scalable spectral clustering with Nystrom approximation: Practical and theoretical aspects. *IEEE Open Journal of Signal Processing*, 2020. URL <https://arxiv.org/abs/2006.14470>.
- [28] Ryan Rossi and Nesreen Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI Conference on Artificial Intelligence*, 2015. URL <https://doi.org/10.1609/aaai.v29i1.9277>.
- [29] Alessandro Rudi, Raffaello Camoriano, and Lorenzo Rosasco. Less is more: Nystrom computational regularization. In *Advances in Neural Information Processing Systems*, 2015. URL <https://arxiv.org/abs/1507.04717>.
- [30] Shota Saito and Mark Herbster. Multi-class graph clustering via approximated effective p -resistance. In *Proceedings of Machine Learning Research*, 2023. URL <https://proceedings.mlr.press/v202/saito23a.html>.
- [31] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 2008. URL <https://doi.org/10.1609/aimag.v29i3.2157>.
- [32] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000. URL <https://doi.org/10.1109/34.868688>.
- [33] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 2011. URL <https://arxiv.org/abs/0803.0929>.
- [34] Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 2015. URL <https://arxiv.org/abs/1501.01571>.
- [35] Max Vladymyrov and Miguel A. Carreira-Perpinan. The variational Nystrom method for large-scale spectral problems. In *Proceedings of Machine Learning Research*, 2016. URL <https://proceedings.mlr.press/v48/vladymyrov16.html>.
- [36] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 2007. URL <https://arxiv.org/abs/0711.0189>.
- [37] Ji Wang, Ding Lu, Ian Davidson, and Zhaojun Bai. Scalable spectral clustering with group fairness constraints. *arXiv preprint*, 2023. URL <https://arxiv.org/abs/2210.16435>.
- [38] Shusen Wang and Zhihua Zhang. Improving CUR matrix decomposition and the Nystrom approximation via adaptive sampling. *Journal of Machine Learning Research*, 2013. URL <https://arxiv.org/abs/1303.4207>.
- [39] Shusen Wang, Alex Gittens, and Michael W. Mahoney. Scalable kernel K-Means clustering with Nystrom approximation: Relative-error bounds. *Journal of Machine Learning Research*, 2019. URL <https://arxiv.org/abs/1706.02803>.

- [40] Christopher K. I. Williams and Matthias Seeger. Using the Nystrom method to speed up kernel machines. In *Advances in Neural Information Processing Systems*, 2000. URL <https://papers.nips.cc/paper/1866-using-the-nystrom-method-to-speed-up-kernel-machines>.
- [41] David P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 2014. URL <https://arxiv.org/abs/1411.4357>.
- [42] Jianlin Xia. Making the Nystrom method highly accurate for low-rank approximations. *arXiv preprint*, 2024. URL <https://arxiv.org/abs/2307.05785>.
- [43] Donghui Yan, Ling Huang, and Michael I. Jordan. Fast approximate spectral clustering. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009. URL <https://doi.org/10.1145/1557019.1557118>.
- [44] Yun Yang, Mert Pilanci, and Martin J. Wainwright. Randomized sketches for kernels: Fast and optimal non-parametric regression. *Annals of Statistics*, 2017. URL <https://arxiv.org/abs/1501.06195>.
- [45] Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 2015. URL <https://arxiv.org/abs/1405.0680>.

A Equation Provenance and Domains

The normalized Laplacian in (1) follows the symmetric convention used by normalized cuts and spectral clustering [25, 32, 36]. The positive-degree condition is not cosmetic: without it, $D^{-1/2}$ is undefined unless a separate regularization rule is introduced. The manuscript does not analyze directed, signed, magnetic, or landmark-renormalized Laplacians. All claims live in the fixed vertex space $\mathbb{R}^n$, so the sketched eigenvalues are interpreted as perturbations of the same shifted operator rather than as eigenvalues of a new landmark graph. The shifted kernel in (3) is introduced here to place the low-Laplacian problem in positive semidefinite Nyström form. Since L_{sym} has spectrum in $[0, 2]$, $K = 2I - L_{\text{sym}}$ is positive semidefinite and reverses the low-frequency ordering without changing the target gap.

The column approximation in (4) is the standard positive semidefinite Nyström reconstruction [11, 15, 40]. Its validity here depends on sampling vertices while keeping the original operator. If an implementation forms a smaller graph, recomputes degrees, or sparsifies affinities before normalization, then the comparison is no longer the fixed-operator perturbation studied in the theorem. The relative error in (5) is defined only for positive eligible gaps; near-zero gaps are excluded because the denominator turns ordinary numerical perturbations into unbounded relative errors. The coherence score in (6) is adapted from leverage and matrix-coherence analysis [13]. It measures how concentrated the target eigenspace is on vertices. The resistance score is a diagnostic maximum over normalized vertex weights and should not be read as edge effective resistance or as a theorem-level sampling premise.

B Proof Boundary and Residual Diagnostics

The proof of the main theorem is short because the hard assumption is explicit. Weyl’s inequality moves each ordered eigenvalue of a symmetric matrix by at most the operator residual, so the difference of two adjacent eigenvalues moves by at most twice that residual. The sufficient event in (9) is exactly the point where this factor of two yields a ten percent relative gap bound. The event is sufficient but not necessary. A sketch can have acceptable gap error when the two adjacent eigenvalue perturbations cancel, and it can fail badly when they add. For that reason, the residual boundary in the results should be interpreted as a certificate boundary rather than as a complete empirical classifier.

The symbolic shift-ordering check covers the algebraic bookkeeping behind the shifted-kernel formulation: low eigenvalues of L_{sym} correspond to high eigenvalues of K , and the gap $\lambda_{k+1} - \lambda_k$ is the same scalar as $\kappa_k - \kappa_{k+1}$ under the stated ordering. The operator-residual measurement records the numerical ratio $\eta(S)/\gamma_k$ for eligible sketches, where $\eta(S)$ is defined in (7). These checks support different parts of the argument. The first protects the sign and index convention; the second measures whether a sampled operator actually reaches the residual scale required by the theorem. The executed evidence shows that this certified region was uncommon. This is why the theorem does not establish the original probabilistic sampling claim, even though it gives a clear target for any future theorem.

A boundary case worth making explicit is the rank cutoff. When $m < k + 1$, the sketched kernel cannot supply both adjacent eigenvalues needed for the target gap, so any apparent estimate is an artifact of rank truncation rather than a perturbation of the exact quantity. A second boundary case is a nearly repeated eigenvalue. If γ_k is close to the eigensolver tolerance, then relative error is not a stable scientific endpoint, and downstream dimension selection is already ill-conditioned. These cases are excluded rather than averaged into the reported rates.

C Evaluation and Reproducibility Notes

The executed validation used a fixed graph–dimension–budget–sampler grid. Eligible graph cases had positive γ_k , enough landmarks to estimate the k th and $(k + 1)$ st kernel eigenvalues, and a fixed shifted kernel shared by every sampler. Budgets used $m = \lceil Ck \log(\max\{k, 2\}) \rceil$ with $C \in \{1, 2, 4, 8\}$ after clipping to maintain eligibility. The sampler set included uniform, ideal leverage, ridge leverage, degree, and vertex-resistance policies. Each retained cell recorded eight seeded draws, median relative gap error, interquartile spread, success under the ten percent threshold, principal-angle diagnostics, adjusted Rand differences where labels were meaningful, coherence, degree heterogeneity, resistance concentration, and residual ratios.

The controlled synthetic coherence benchmark is the best fit to the hypothesis assumptions because it deliberately varies block balance, degree concentration, manifold geometry, preferential attachment, and grid-like spatial regularity. The small real-topology benchmark is useful implementation evidence because it uses genuine small graph topologies, but it is lower relevance for claims about large citation, social, or road networks. The original plan also requested at least twenty seeds and paired significance tests; the bounded run did not meet that statistical standard. Consequently,

the manuscript treats sampler separation as descriptive and mixed, while treating the failure of the strong endpoint as supported by the available table: the success counts are far below the planned eighty percent criterion even before stronger tests are applied.

Reproducibility depends on preserving the matched-cell structure. A rerun should keep graph generation seeds, landmark seeds, target dimensions, and budget multipliers separate so that variation from the graph model is not confused with variation from landmark draws. It should also report failures to meet eligibility, because discarded near-zero-gap cells can otherwise make the benchmark look easier than it is. For approximate leverage scores, the score-estimation method should be recorded with its own tolerances and random seeds. Otherwise, a sampler comparison may conflate the sampling law with numerical error in the score proxy.

The same separation is needed for summary statistics. Cell medians should be computed before sampler aggregation, and every reported success rate should state whether the denominator is a seed, a graph–dimension cell, or a benchmark-corpus collection. This prevents a large easy family from hiding hard cells whose small gaps dominate the scientific question.

D Relevance, Limitations, and Follow-Up Tests

The strongest relevance check is whether a benchmark corpus tests the same regime assumed by the hypothesis. The synthetic corpus does so reasonably well because it varies coherence and degree heterogeneity while preserving a fixed operator. The small real-graph corpus checks that the implementation is not tied to generated graphs, but it does not close the public-benchmark gap. That distinction affects interpretation. The bounded evidence is enough to say that the broad success statement is unsupported in the executed regime. It is not enough to conclude that leverage-based Nyström sampling can never preserve gaps on larger graphs, under stronger regularity, or with a more accurate score estimator.

Future work should close three gaps. First, a larger empirical run should materialize public graph matrices, repeat every matched cell with at least twenty seeds, and report paired graph-level tests with confidence intervals. Second, a controlled synthetic study should vary γ_k while holding coherence nearly fixed, then vary coherence while holding the gap scale nearly fixed; this would separate small-gap failure from vertex-concentration failure. Third, the theory should prove a residual event scaled to γ_k under explicit graph and sampler assumptions. A global trace, Frobenius, or tail-energy result is not enough for this paper’s target unless it implies $\eta(S) \leq 0.05\gamma_k$. Until one of those checks is available, the defensible practice is to use landmark sketches for acceleration only with a residual or held-out eigengap validation check.

The most useful negative controls would be homogeneous expanders, near-disconnected block graphs with deliberately tiny gaps, and graphs whose high-degree vertices are not high-leverage vertices. The first group tests whether uniform sampling can match structured policies when coherence is low. The second tests the theorem’s local scale, because small gaps should fail unless the residual falls dramatically. The third tests whether degree and resistance proxies select the same vertices as eigenspace leverage. Reporting these controls alongside public benchmarks would make the next manuscript less dependent on a single aggregate success rate and more informative about the mechanism behind either success or failure.

A final reproducibility detail is the distinction between sampler randomness and score-estimation randomness. The ideal leverage sampler is a diagnostic upper reference because it uses eigenspace information that would normally require solving the original spectral problem. Approximate ridge leverage is the practical target, but its score vector depends on an estimator, a regularization scale, and a numerical tolerance. A future public release should therefore store the estimated score vector or enough metadata to reconstruct it, including the shifted-kernel convention, the regularization parameter, the random projection seed if one is used, and the stopping rule for any iterative linear solves. Without that information, a repeated experiment could change both the landmark distribution and the operator residual, making the gap error impossible to attribute. The same principle applies to the residual certificate: the norm estimate should state whether it came from a dense eigensolver, Lanczos iteration, or a power method, and it should report its tolerance relative to γ_k . These details are operational, but they are part of the scientific interpretation because the central claim is about preserving a small spectral difference rather than about approximating a large trace or Frobenius quantity.