KRYLOV SUBSPACE-PRESERVING GRAPH COARSENING FOR SCALABLE GRAPH LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph coarsening merges nodes into supernodes so that a graph neural network (GNN) can be trained on the smaller graph and deployed on the original. Spectral GNNs apply polynomial filters of degree at most K, whose output combines the node features and their first K propagations with coefficients learned only after coarsening. All such outputs lie in the span of these propagated features, the block Krylov subspace. Existing coarsening methods target the Laplacian spectrum, the topology, the node attributes or the output of one fixed convolution, and none of these criteria accounts for the full range of combinations of the propagated features. We therefore propose to preserve the Krylov subspace itself. We quantify how coarsening changes this subspace by the worst-case error of a filter learned on the coarse graph and deployed on the original. This error is bounded above and below by a k-means objective on the propagated features, which we minimize by multilevel Ward merging. Many existing methods also densify the coarse graph, limiting the training speedup, and several are expensive for graphs with millions of nodes. To avoid densification, we prune edges after each level under a budget that bounds the average degree. We evaluate on nine homophilic and heterophilic node-classification benchmarks against state-of-the-art coarsening baselines, with both polynomial-filter and nonlinear GNNs. At 90% node reduction, a GCN trained on our coarse graphs is significantly more accurate than every baseline on seven of the nine datasets, with a median gain of 2.4 percentage points over the best baseline. On CPU, our coarsening method is $3\times$ faster than the most competitive baseline, and on one A100 GPU, using our coarsening method followed by training is $5\times$ faster than training on the full graph.

1 INTRODUCTION

Graph neural networks (GNNs) are the dominant paradigm for learning on graph-structured data (Kipf & Welling, 2017; Hamilton et al., 2017; Velickovic et al., 2018), but the cost of a message-passing layer scales with the number of edges and nodes, so training on large graphs is expensive (Hu et al., 2020; Zeng et al., 2020). *Graph coarsening* addresses this cost with a preprocessing step that merges nodes into supernodes; a GNN is then trained on the reduced graph and deployed on the original, and the cost of coarsening is amortized over epochs, hyperparameter settings and architectures (Loukas, 2019; Huang et al., 2021; Hashemi et al., 2024). For any coarsening method to be useful, it must reduce training cost while retaining the information the GNN needs.

The extent to which the coarsening method retains this information depends on how the GNN's output varies with the graph. Spectral GNNs make this dependence explicit by defining graph convolution as a filter in the spectral domain of the graph Laplacian (Bruna et al., 2014). Application of this filter directly requires the eigendecomposition of the Laplacian, so the filter is instead parameterized as a polynomial of degree at most K in a propagation operator (Hammond et al., 2011; Defferrard et al., 2016). There are various ways to use this polynomial: GCN applies a first-order polynomial in each layer, with pointwise nonlinearities between layers (Kipf & Welling, 2017), whereas several later models apply a single polynomial of degree at most K, expanded in a chosen basis with fixed or learned coefficients (Wu et al., 2019; Klicpera et al., 2019a; Chien et al., 2021; He et al., 2021). Even a linear model of this form, JacobiConv (Wang & Zhang, 2022), reaches node-classification accuracy competitive with nonlinear GNNs, making the linear-filter assumption natural and not restrictive in practice. Irrespective of the basis and the coefficients, a filter of degree

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

at most K outputs a linear combination of its input and the first K propagations of that input, so, for a fixed input, every such output lies in their span, i.e., the block Krylov subspace (Luan et al., 2019; Huang et al., 2024). This subspace depends not only on the graph structure but also on the input features: in the spectral domain, the filter rescales each frequency component of the features, so a frequency that the features do not contain is absent from every filter output (Wang & Zhang, 2022).

Coarsening changes both the graph structure and the features, so the coarse graph has a block Krylov subspace of its own, spanned by the coarse features and their propagations. The filter coefficients are learned from these coarse propagations and then applied to the propagations on the original graph; thus this transfer depends on what the coarsening method preserves. The coarsening process, irrespective of the method, cannot preserve any variation within a supernode because the merged nodes share a single representation on the coarse graph, even when their propagated features on the original graph differ. Some existing coarsening methods choose which nodes to merge so as to preserve spectral or topological properties of the graph (Loukas, 2019; Hermsdorff & Gunderson, 2019; Cohen & Talmon, 2025; Wu et al., 2026), and spectral approximation can also yield propagation guarantees under additional assumptions (Joly & Keriven, 2024). Others preserve node attributes together with graph structure (Kumar et al., 2023; Kataria et al., 2024; Bai et al., 2026) or the output of a single graph convolution (Dickens et al., 2024). Since the partition is fixed before the filter coefficients are learned, none of these criteria accounts for the full range of combinations of the propagated features. This motivates preserving the input features and their first K propagations, and hence preserving the block Krylov subspace they span.

Preserving this subspace is just one requirement of a useful coarsening method; the other is reducing the training cost, which depends not only on the node count of the coarse graph but also on its edge count. The edge count, however, need not fall in proportion to the reduction in node count by the coarsening method. A supernode inherits the union of its members' neighborhoods, so a merge removes only the edges between the merged nodes and coalesces their edges to shared neighbors; as a result, the coarse graph may become denser. This densification is more pronounced in methods that allow merges of non-adjacent nodes: at 90% node reduction, UGC (Kataria et al., 2024) raises the average degree by 7–10$\times$ on ogbn-arxiv, ogbn-products and Reddit2, and ACM (Dickens et al., 2024) by 3–5$\times$ (Figure 4a). These extra edges limit the training speedup: UGC and ACM reach only 9–35% and 26–45%, respectively, of the speedup obtained at the same node reduction with the edges sparsified to preserve the original average degree (Figure 4b; Appendix B). Beyond training, the coarsening process itself has a computation cost, which is often significant for the strongest baselines: the variation methods require the low-frequency Laplacian eigenvectors (Loukas, 2019), FGC (Kumar et al., 2023) and INGC (Cohen & Talmon, 2025) are quadratic and SGBGC (Xia et al., 2025) superlinear in the node count, and ACM requires a nearest-neighbor search over node embeddings (Section 4).

In this paper, we propose a graph coarsening method that chooses the partition to preserve the block Krylov subspace of the node features for polynomial filters of degree at most K. We show that the worst-case transfer error of these filters is bounded above and below by the per-hop projection residuals of the propagated features on the partition (Section 3). Minimizing these residuals reduces to a k-means problem on the propagated features, so the cost of a candidate merge is its Ward increment. The partition is then built by batched Ward agglomeration: each level merges vertex-disjoint pairs, favoring adjacent ones and admitting a non-adjacent merge only when its cost is at most a fixed fraction of the cheapest adjacent one. Since merging alone need not reduce the edge count in proportion to the node count, each level is followed by an edge selection under a budget that bounds the average degree at each level. Our main contributions can be summarized as follows:

- We identify the block Krylov subspace of the node features as the quantity a coarsening must preserve for polynomial filters of degree at most K, independently of the basis and the coefficients that are later learned (Section 3).
- We bound the worst-case transfer error of polynomial filters of degree at most K above and below by per-hop projection residuals, reducing the objective to a k-means cost (Section 3.1).
- We approximately minimize the transfer error by batched Ward agglomeration with cost-gated non-adjacent merges, interleaved with an edge pruning step that bounds the average degree at each level (Sections 3.2 and 3.3).

- On nine benchmarks at 90% node reduction, the proposed method is significantly more accurate than every state-of-the-art coarsening method on seven datasets (Welch's t-test, Holm-corrected, Appendix F), while being $3\times$ faster than the most competitive baseline (Section 4).

2 Technical Background

2.1 Graph notation

Let $G = (V, E)$ be an undirected graph with $n = |V|$ nodes, adjacency matrix A, diagonal degree matrix D, features $X \in \mathbb{R}^{n\times d}$, and normalized Laplacian $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$. The model propagates with a symmetric operator S built from A and satisfying $\|S\|_2 \leq 1$; for a GCN it is $S = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ with $\tilde{A} = A + I$ and $\tilde{D}$ the degree matrix of $\tilde{A}$.

2.2 Graph coarsening

Definition 1 (Graph coarsening)**.** A *coarsening* of G into $k < n$ supernodes is a surjective map $\pi : V \to [k]$, where $[k] = \{1, \dots, k\}$. Let $C_c = \pi^{-1}(c)$ denote the nodes assigned to supernode c, and let $P \in \{0, 1\}^{n\times k}$ be the partition matrix with $P_{ic} = 1$ if and only if $i \in C_c$. The diagonal matrix $H = P^\top P = \operatorname{diag}(|C_1|, \dots, |C_k|)$ records the supernode sizes. The associated weighted quotient graph G_c has adjacency $A_c = P^\top AP$ and features $X_c = H^{-1}P^\top X$.

Let $R = H^{-1}P^\top$ denote the restriction operator, which averages node features over each supernode, so that $X_c = RX$, and let $R^+ = P$ denote its pseudoinverse, which lifts a coarse signal to V by assigning to each node the value of its supernode. The product $\Pi = R^+R$ is the orthogonal projector onto the subspace of signals that are constant on each supernode, and $\Pi^\perp = I - \Pi$ is the projector onto its orthogonal complement, the component of a signal that averaging removes. For the analysis, we adopt the coarse propagation operator $S_c := RSR^+$ of Joly & Keriven (2024), which lifts a coarse signal to V, applies S, and averages the result over each supernode. Appendix A.1 summarizes the merge criteria and computational cost of existing coarsening methods.

2.3 Polynomial filters and the Krylov subspace

Let S be a fixed propagation operator and $x \in \mathbb{R}^n$ a graph signal. A polynomial filter of degree at most K is the map $x \mapsto p(S)x$ with $p(t) = \sum_{\ell=0}^{K} w_\ell t^\ell$ and $w \in \mathbb{R}^{K+1}$. The set of all such outputs is the Krylov subspace (Huang et al., 2024)

$$\mathcal{K}_{K+1}(S, x) \;=\; \operatorname{span}\{x, Sx, \dots, S^K x\} \;=\; \{\, p(S)x : \deg p \leq K \,\}. \tag{1}$$

Since any two bases of the polynomials of degree at most K are related by an invertible change of coefficients, a spectral GNN with unconstrained coefficients realizes the same set of filters in every polynomial basis, although the basis affects the conditioning of the optimization problem (Wang & Zhang, 2022); models with fixed or constrained coefficients realize only a subset of Equation (1). Table 3 in Appendix A.2 lists the basis, operator and coefficients of representative spectral GNNs. For a feature matrix $X \in \mathbb{R}^{n\times d}$, the block Krylov subspace is

$$\mathcal{K}_{K+1}(S, X) \;=\; \operatorname{range}\big[\, X,\; SX,\; \dots,\; S^K X \,\big], \tag{2}$$

a subspace of $\mathbb{R}^n$ of dimension at most $(K+1)d$. For any coefficient matrices $W_0, \dots, W_K \in \mathbb{R}^{d\times d}$, every column of $\sum_{\ell=0}^{K} S^\ell X W_\ell$ lies in $\mathcal{K}_{K+1}(S, X)$ (Luan et al., 2019).

3 Methodology

Given G, S, X, and a target size k, we seek a coarsening π that preserves the block Krylov subspace $\mathcal{K}_{K+1}(S, X)$ while limiting the number of retained edges. Over a bounded coefficient class, the squared worst-case transfer error of a polynomial filter of degree at most K admits two-sided bounds in terms of an objective that depends only on the propagated features $X, SX, \dots, S^K X$ and the partition, and decomposes over supernodes. These features are computed once on the original graph and used to score merges in a multilevel scheme, with edge selection after each level enforcing the edge budget (Figure 1). Sections 3.1–3.3 develop the criterion, the merge algorithm and the edge selection; Algorithm 1 in Appendix C gives the complete algorithm and Appendix D the proofs.

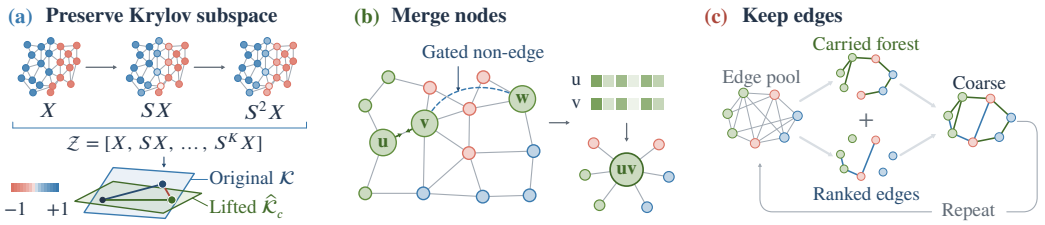


Figure 1: Overview of the proposed coarsening. *(a)* The propagated features $X, SX, \ldots, S^K X$ span the Krylov subspace $\mathcal{K}$, which the lifted coarse subspace $\hat{\mathcal{K}}_c$ should preserve. *(b)* Multilevel merging by this criterion, with gated non-adjacent merges. *(c)* After each level, edges are ranked and pruned to a degree budget while a spanning forest keeps the graph connected.

3.1 Merge Criterion

A polynomial filter of degree at most K with coefficients $W = (W_0, \ldots, W_K)$, $W_\ell \in \mathbb{R}^{d\times d}$, can be applied on G or, with the same coefficients, on G_c, where it propagates $X_c = RX$ with S_c and its output is lifted to the original nodes by R^+. Let $\Gamma_\ell(\pi) := S^\ell - R^+ S_c^\ell R$ for $0 \le \ell \le K$. We define the *transfer error* of W under π as the Frobenius norm of the difference between its output on G and its lifted output on G_c,

$$\mathcal{T}_\pi(W) \;:=\; \left\| \sum_{\ell=0}^{K} \Gamma_\ell(\pi)\, X W_\ell \right\|_F . \tag{3}$$

The coefficients are unknown when the partition is chosen, so we consider the worst-case transfer error over a bounded coefficient class. The coefficient bound prevents arbitrary rescaling of the filter from making this error unbounded.

Definition 2. Assume $\|X\|_F = 1$ and let $\mathcal{W} := \{W : \sum_{\ell=0}^{K} \|W_\ell\|_2 \le 1\}$, with $\|\cdot\|_2$ the spectral norm. The *Krylov approximation constant* of the coarsening π is its worst-case transfer error over $\mathcal{W}$,

$$\mathcal{E}^\star(\pi) \;:=\; \sup_{W\in\mathcal{W}} \mathcal{T}_\pi(W). \tag{4}$$

Proposition 1. *Let $\|S\|_2 \le 1$ and $\|X\|_F = 1$. For every coarsening π, $\mathcal{E}^\star(\pi) = \max_{0\le\ell\le K} \|\Gamma_\ell(\pi)\, X\|_F \le 2$. For any index $\ell^\star$ attaining the maximum, the supremum in Equation (4) is attained by $W_{\ell^\star} = I_d$ and $W_\ell = 0$ for $\ell \ne \ell^\star$.*

Evaluating $\mathcal{E}^\star(\pi)$ requires the coarse operator S_c, which changes with the partition, so recomputing it after every candidate merge is expensive. We instead use an objective that requires no propagation on the coarse graph. Let $\mathcal{Z} := [\, X, SX, \ldots, S^K X \,]$ be the *trajectory*, computed once on G. For each order $\ell \ge 0$, we define the *projection residual* $\delta_\ell(\pi) := \left\|\Pi^\perp S^\ell X\right\|_F$, the error of approximating $S^\ell X$ by its supernode averages $\Pi S^\ell X$. We define the objective as

$$\mathcal{E}(\pi) \;:=\; \sum_{\ell=0}^{K} \delta_\ell(\pi)^2 \;=\; \left\|\Pi^\perp \mathcal{Z}\right\|_F^2 , \tag{5}$$

the squared residual of the entire trajectory. Because $\mathcal{E}$ is a sum over supernodes, a candidate merge changes only the terms of the two supernodes involved, and this change is evaluated without any propagation on the coarse graph. Theorem 1 shows that $\mathcal{E}$ bounds the squared worst-case transfer error from both sides.

Theorem 1. *Let $\|S\|_2 \le 1$ and $1 \le k < n$.*

(i) *For every coarsening π of G into k supernodes,*

$$\mathcal{E}(\pi)/(K+1) \;\le\; \max_{0\le\ell\le K} \delta_\ell(\pi)^2 \;\le\; \mathcal{E}^\star(\pi)^2 \;\le\; \mathcal{E}(\pi). \tag{6}$$

(ii) *For every such coarsening π, every $\ell \ge 0$, and every $Z \in \mathbb{R}^{k\times d}$,*

$$\left\|S^\ell X - R^+ Z\right\|_F^2 \;=\; \delta_\ell(\pi)^2 + \left\|H^{1/2}\left(RS^\ell X - Z\right)\right\|_F^2 \;\ge\; \delta_\ell(\pi)^2. \tag{7}$$

In particular, the bound holds with $Z = (S_c')^\ell RX$ for every coarse operator $S_c' \in \mathbb{R}^{k\times k}$.

(iii) *Let $\pi^\sharp$ and $\pi^\star$ minimize $\mathcal{E}$ and $\mathcal{E}^\star$, respectively, over the coarsenings of G into k supernodes. Then $\mathcal{E}^\star(\pi^\sharp) \le \sqrt{K+1}\,\mathcal{E}^\star(\pi^\star)$.*

Part (i) bounds the squared worst-case transfer error between $\mathcal{E}(\pi)/(K+1)$ and $\mathcal{E}(\pi)$. Part (ii) shows that no coarse propagation operator can reduce the order-ℓ error below $\delta_\ell(\pi)$, and identifies the excess exactly. A small $\mathcal{E}$ is therefore necessary for uniformly small transfer error under any coarse operator, and sufficient under $S_c = RSR^+$. Part (iii) gives the corresponding approximation guarantee for a global minimizer of $\mathcal{E}$. We compare the sets of attainable outputs on the two graphs, allowing different coefficients within $\mathcal{W}$,

$$\mathcal{O}_G \;:=\; \Big\{\sum_{\ell=0}^{K} S^\ell X W_\ell : W \in \mathcal{W}\Big\}, \qquad \mathcal{O}_c^\uparrow \;:=\; \Big\{R^+ \sum_{\ell=0}^{K} S_c^\ell RX W_\ell : W \in \mathcal{W}\Big\}. \tag{8}$$

The coarse outputs are lifted to the original nodes so that both sets lie in $\mathbb{R}^{n\times d}$. Let $d_{\mathrm{H}}(\pi)$ denote their Hausdorff distance under the Frobenius norm (Definition 3 in Appendix A.3).

Corollary 1. *Let $\|S\|_2 \le 1$. Then $\max_{0\le\ell\le K}\delta_\ell(\pi) \le d_{\mathrm{H}}(\pi) \le \mathcal{E}^\star(\pi)$, hence $\mathcal{E}(\pi)/(K+1) \le d_{\mathrm{H}}(\pi)^2 \le \mathcal{E}(\pi)$; and $\mathcal{E}(\pi) = 0$ if and only if $\mathcal{O}_G = \mathcal{O}_c^\uparrow$.*

Corollary 1 shows that the objective $\mathcal{E}(\pi)$, the squared worst-case transfer error $\mathcal{E}^\star(\pi)^2$, and the squared Hausdorff distance $d_{\mathrm{H}}(\pi)^2$ are comparable within a factor of $K+1$. All three vanish exactly when the bounded sets of attainable outputs coincide. Corollary 2 in Appendix C.3 shows in addition that $\mathcal{E}^\star(\pi)$ bounds the differences in feature-weighted spectral moments through degree $2K+1$.

The objective $\mathcal{E}(\pi)$ weights each left singular direction of $\mathcal{Z}$ by its squared singular value. When these values differ substantially, a few directions can dominate the objective. To reduce this imbalance, we rescale the trajectory with a ridge $\lambda = \lambda_{\mathrm{rel}}\,\sigma_1^2$, where σ_1 is the largest singular value of $\mathcal{Z}$ and $\lambda_{\mathrm{rel}} \ge 0$ is a scale-free parameter that sets the ridge as a fraction of σ_1^2. With $\mathcal{Z} = Q\Sigma V^\top$ a compact SVD, we define $\mathcal{Z}_\lambda := Q\Sigma(\Sigma^2+\lambda I)^{-1/2}$ and use the objective

$$\mathcal{E}_\lambda(\pi) \;:=\; \big\|\Pi^\perp \mathcal{Z}_\lambda\big\|_F^2 . \tag{9}$$

For every partition π and every $\lambda_{\mathrm{rel}} \ge 0$, $\mathcal{E}^\star(\pi)^2 \le \mathcal{E}(\pi) \le \sigma_1^2(1+\lambda_{\mathrm{rel}})\,\mathcal{E}_\lambda(\pi)$, so $\mathcal{E}_\lambda$ ranks partitions by a π-independent upper bound on the squared transfer error, and $\mathcal{E}_\lambda(\pi) = 0$ exactly when $\mathcal{E}(\pi) = 0$ (derivation in Appendix C.2).

3.2 Merge algorithm

Given the target size k, we seek a coarsening of G into k supernodes (Definition 1) that minimizes the objective. Since Π replaces each row of $\mathcal{Z}_\lambda$ by the mean of its supernode, computing this minimizer is the k-means problem on the rows z_i of $\mathcal{Z}_\lambda$ (Huang et al., 2021). Starting from singleton supernodes, we approximate it by multilevel merging: with $\bar z_C$ the centroid of supernode C, merging C and C' increases the objective by the Ward increment $\Delta\mathcal{E}_\lambda(C,C')$ (Ward, 1963),

$$\pi_\lambda^\star \;\in\; \operatorname*{arg\,min}_{\pi:\,V\to[k]} \mathcal{E}_\lambda(\pi), \qquad \Delta\mathcal{E}_\lambda(C,C') \;=\; \tfrac{|C|\,|C'|}{|C|+|C'|}\,\|\bar z_C - \bar z_{C'}\|_2^2 . \tag{10}$$

At each level, the candidate pairs are the adjacent pairs and at most b non-adjacent pairs per supernode, selected by a neighbor-of-neighbor search with graph neighbors and uniformly sampled supernodes (Algorithm 2), which avoids comparing all $O(n^2)$ pairs. Adjacent merges absorb an edge and keep the merged supernode connected, whereas a non-adjacent merge may have a lower increment. Let $N_E(C)$ denote the neighbors of supernode C at level t, and let $\Delta_C^E = \min_{C'\in N_E(C)}\Delta\mathcal{E}_\lambda(C,C')$ be the cheapest adjacent merge at C, taken as $+\infty$ if $N_E(C)$ is empty. For an *escape ratio* $\zeta \in [0,1]$, a non-adjacent candidate for C is admitted when $\Delta\mathcal{E}_\lambda(C,C') \le \zeta\,\Delta_C^E$. At $\zeta = 1$, a non-adjacent candidate is admitted whenever its increment does not exceed that of the cheapest adjacent merge at C; at $\zeta = 0$, only adjacent pairs are candidates. Merging without regard to adjacency absorbs few edges (Appendix B). For example, UGC's supernodes are close to independent sets of G and leave on average 96% of the edges in place across nine datasets at 90% node reduction.

Each supernode then selects, among its admitted candidates, the pair with the smallest increment, and a pair is merged when both supernodes select each other, as in reciprocal nearest-neighbor clustering (Murtagh, 1983). Because the merged pairs share no supernode and $\mathcal{Z}_\lambda$ does not change during merging, the increments of the pairs merged at a level sum exactly to the resulting increase of $\mathcal{E}_\lambda$, so merging in batches costs nothing relative to merging one pair at a time. The selection is repeated a fixed number of times over the supernodes not yet paired, and levels continue until k supernodes remain.

3.3 Per-level edge selection

After each merging level, we sparsify the contracted graph so that edge reduction accompanies node reduction; sparsification during coarsening can also reduce the cost of the multilevel hierarchy (Gottesbüren et al., 2025). Let $\mathcal{Q}_t$ denote the graph at level t, with $\mathcal{Q}_0 = G$. The edges of the contracted graph are ranked by $w^c_{uv}\gamma_{uv}$, where w^c_{uv} counts the retained edges of G between supernodes C_u and C_v, and $\gamma_{uv} = \min\{a_{uv}/\left|C_u\right|, a_{vu}/\left|C_v\right|\} \in (0, 1]$, where a_{uv} is the number of nodes of C_u adjacent in G to C_v. The score favors pairs joined by many edges spread across both supernodes. A spanning forest of the contracted graph is retained unconditionally, and the highest-scoring remaining edges are added up to a budget, so pruning does not split any connected component of the contracted graph.

Let m_t be the number of nodes of $\mathcal{Q}_t$ and ρ_t the number of its edges beyond a spanning forest. Merging at level t contracts $\mathcal{Q}_t$ to a graph on m_{t+1} nodes with $\widehat{\rho}_{t+1}$ edges beyond a spanning forest. With $r_t = m_{t+1}/m_t$, $(x)_+ = \max\{x, 0\}$ and a budget parameter $\nu \geq -1$, the number of edges retained beyond the spanning forest is

$$s_{t+1}(\nu) \;=\; \min\Big\{\widehat{\rho}_{t+1},\; \big\lfloor \rho_t\, r_t^{1+\nu} + (-\nu)_+\,(\widehat{\rho}_{t+1} - \rho_t)_+ \big\rfloor\Big\}, \tag{11}$$

and the pruned graph is $\mathcal{Q}_{t+1}$. At $\nu = -1$, all edges of the contracted graph are retained; as ν increases, the budget shrinks, and for large ν only the spanning forest remains. For $\nu \geq 0$ the second term vanishes.

The normalized propagation operator $\widetilde{S}_c$ of the final retained graph $\mathcal{Q}_T$, where T is the number of levels, differs from $S_c = RSR^+$, so the upper bound of Theorem 1(i) does not apply to it. Let $\mathcal{Z}_c := [\, RX, \widetilde{S}_c RX, \ldots, \widetilde{S}_c^K RX \,]$. Applying Theorem 1(ii) with $Z = \widetilde{S}_c^\ell RX$ and summing over $0 \leq \ell \leq K$,

$$\left\|\mathcal{Z} - R^+\mathcal{Z}_c\right\|_F^2 \;=\; \mathcal{E}(\pi) + \mathcal{D}(\pi, \widetilde{S}_c), \qquad \mathcal{D}(\pi, \widetilde{S}_c) := \left\|H^{1/2}(R\mathcal{Z} - \mathcal{Z}_c)\right\|_F^2.$$

The first term is the objective $\mathcal{E}(\pi)$ in Equation (5) and depends only on the partition. The second is the difference between propagation on $\mathcal{Q}_T$ and the supernode averages of propagation on G, weighted by supernode size. At a fixed partition, edge selection changes only this term, so pruning trades propagation cost against $\mathcal{D}$; Section 4.4 evaluates this trade-off.

With K, d and b treated as constants, the complete procedure, from forming $\mathcal{Z}$ to the final edge selection, requires $O(T(n+|E|)\log n)$ time and $O(n+|E|)$ memory; a detailed analysis is presented in Appendix C.1.

4 Experiments

4.1 Setup

We evaluate our method on nine node-classification benchmarks (Table 7): five homophilic graphs (ogbn-arxiv, ogbn-products, the paper–paper citation subgraph of ogbn-mag, reddit2 and igb-small) and four heterophilic graphs (arxiv-year, snap-patents, pokec and twitch-gamers). All graphs are symmetrized and we use the official splits. We compare against the following baselines: Var-Neigh and Var-Edges (Loukas, 2019), A-ConvMatch (Dickens et al., 2024), UGC (Kataria et al., 2024), SGBGC (Xia et al., 2025), STPGC (Wu et al., 2026) and Random Matching (Karypis & Kumar, 1998); FGC (Kumar et al., 2023) and INGC (Cohen & Talmon, 2025) are excluded because their cost is quadratic in the number of nodes. Each graph is coarsened at reduction ratios of 80, 90, 95 and 99%, and each supernode is labeled with the majority label of its training nodes (Huang

Table 1: GCN test accuracy (%) on the coarsened graph at 90% reduction ratio, evaluated on the original graph. Best in **bold**, second best underlined, ranked among the methods with a result; *mean rank* averages these ranks. ‡ marks a coarsening whose achieved ratio missed the $90 \pm 3\%$ tolerance, OOT one that exceeded the six-hour time limit, and OOM one that ran out of memory; Appendix E reports every backbone at all four ratios.

Dataset	Full graph	A-ConvMatch	Random	SGBGC	STPGC	UGC	Var-Edges	Var-Neigh	Ours
igb-small	$72.31_{\pm.01}$	$70.10_{\pm.09}$	$70.49_{\pm.10}$	$70.78_{\pm.06}$	$68.61_{\pm.49}$	$64.53_{\pm.91}$	$70.23_{\pm.03}$	$69.53_{\pm.14}$	**71.82**$_{\pm.03}$
ogbn-arxiv	$71.90_{\pm.17}$	$64.86_{\pm.57}$	$59.70_{\pm1.18}$	$64.00_{\pm.41}$	$64.41_{\pm.36}$	$33.41_{\pm1.38}$	$66.01_{\pm.58}$	$66.04_{\pm.51}$	**68.93**$_{\pm.11}$
ogbn-mag	$35.64_{\pm.12}$	$31.97_{\pm.39}$	$14.70_{\pm.57}$	$25.92_{\pm.54}$	$29.76_{\pm.34}$	$13.44_{\pm.96}$	$28.46_{\pm.32}$	$27.35_{\pm.31}$	**35.11**$_{\pm.19}$
ogbn-products	$76.67_{\pm.68}$	$72.99_{\pm.30}$	$35.31_{\pm1.70}$	$66.53_{\pm.34}$	$60.26_{\pm.20}$	$65.24_{\pm.30}$	OOT	OOT	**76.43**$_{\pm.20}$
reddit2	$95.34_{\pm.01}$	$93.80_{\pm.06}$	$91.90_{\pm.26}$	$91.54_{\pm.17}$	$74.73_{\pm.93}$	$25.39_{\pm2.45}$	**93.88**$_{\pm.12}$	$91.52_{\pm.56}$	$93.55_{\pm.05}$
arxiv-year	$45.97_{\pm.13}$	$40.59_{\pm.29}$	$38.46_{\pm.40}$	$34.69_{\pm.77}$	$36.85_{\pm.35}$	$32.54_{\pm1.41}$	$37.42_{\pm.68}$	$38.80_{\pm.33}$	**40.62**$_{\pm.18}$
pokec	$72.50_{\pm.44}$	$51.06_{\pm.26}$	‡	$52.06_{\pm.34}$	$51.77_{\pm.52}$	$51.79_{\pm.15}$	$51.90_{\pm.21}$	$50.86_{\pm.12}$	**58.47**$_{\pm1.03}$
snap-patents	$41.63_{\pm.04}$	OOM	$24.54_{\pm2.28}$	$29.82_{\pm.12}$	$32.32_{\pm.22}$	$26.21_{\pm.58}$	‡	‡	**34.71**$_{\pm.22}$
twitch-gamers	$63.68_{\pm.10}$	$60.73_{\pm.49}$	$58.26_{\pm.84}$	$59.66_{\pm.70}$	$57.38_{\pm1.33}$	$58.65_{\pm.74}$	‡	$59.08_{\pm.46}$	**61.54**$_{\pm.13}$
Mean rank	—	3.1	5.2	4.1	5.2	6.3	3.3	4.7	**1.2**

et al., 2021). We train five backbones on each coarsened graph (hyperparameters in Appendix E.2): two polynomial-filter GNNs (Wang & Zhang, 2022), SGC (Wu et al., 2019) and GPR-GNN (Chien et al., 2021), and three nonlinear GNNs, GCN (Kipf & Welling, 2017), GraphSAGE (Hamilton et al., 2017) and SIGN (Frasca et al., 2020). All models are evaluated on the original graph and we report the mean and standard deviation over five training seeds. All experiments are run on an AMD EPYC 7543P server with 16 CPU cores, 128 GB of RAM and one NVIDIA A100 GPU with 80 GB of memory.

The default coarsening parameters of our method are degree $K = 3$, ridge $\lambda_{\text{rel}} = 10^{-4}$, escape ratio $\zeta = 0.5$ and budget parameter $\nu = 0.3$; the remaining settings are listed in Appendix E.3. Nothing was tuned per dataset.

4.2 Evaluation on node-classification

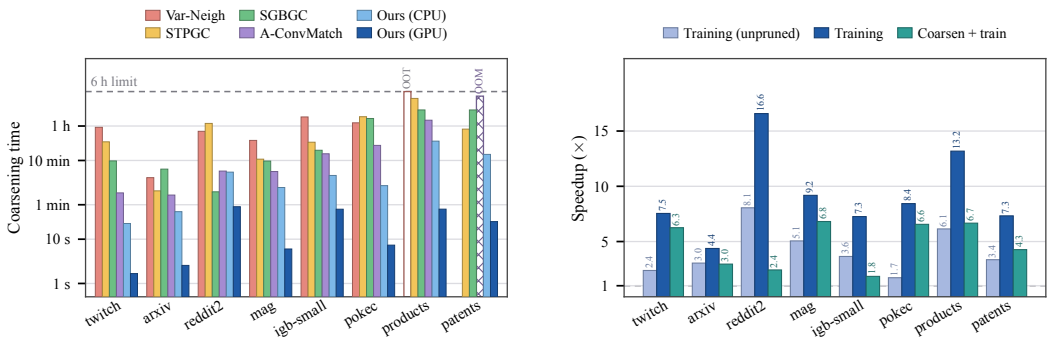
At reduction ratio 90 (Table 1), our method achieves significantly higher GCN accuracy on seven of the nine datasets, with a mean rank of 1.2 compared with 3.1 for the strongest baseline, A-ConvMatch. Relative to the best baseline on each dataset, the median improvement is 2.39 percentage points (pp), significant under a Wilcoxon signed-rank test across datasets ($p = 0.012$). The advantage is not specific to GCN: at ratio 90, our coarsening remains the most accurate on seven of the nine datasets with the nonlinear backbones GraphSAGE and SIGN, and on six with the polynomial-filter backbones SGC and GPR-GNN (Table 10).

The advantage persists under more aggressive coarsening as well. At 95% and 99% node reduction, a GCN trained on our coarse graphs remains the most accurate on seven of the nine datasets (Tables 11 and 12), and with each of the other four backbones our coarsening method places in the top two on at least seven of the nine datasets at 95%, increasing to eight at 99%.

Appendix F tests the ratio-90 differences at three levels with Holm correction throughout. Under Welch's t-test our method is significantly more accurate in 95% of valid dataset–baseline pairs; a Wilcoxon signed-rank test over datasets separates it from six of the seven baselines; and a Friedman test with post-hoc comparisons ranks it significantly above every baseline ($p \leq 0.043$).

4.3 Computational cost

Coarsening cost. Figure 2a reports the time each coarsening method needs to reach ratio 90 under a common CPU protocol, together with that of our method on one A100 GPU; Table 8 lists the runtime values. Our method is $3\times$ (median, throughout this section) faster than A-ConvMatch which is the most competitve baseline with second best performance in 5 out of 9 datasets. The speedup advantage becomes more notable for larger graphs such as ogbn-products, for which our method is $7\times$ faster than all the competitive baselines. Random matching and UGC, non-competitive in Table 1, are omitted; their runtimes are in Table 8.



(a) (b)

Figure 2: Computational cost at 90% reduction ratio. *(a)* Coarsening time on CPU for each method (logarithmic axis), together with that of our method on a single GPU. *(b)* Speedup of GCN training on the coarsened graph relative to training on the full graph, with and without edge pruning, and end to end, that is, including the time spent on coarsening (on GPU) and pruning.

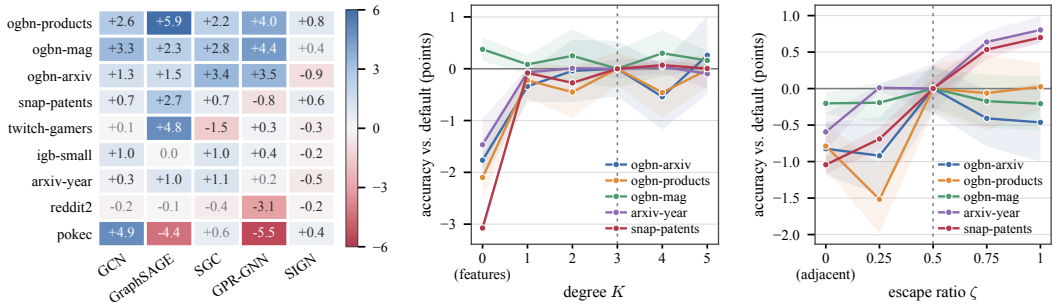
Training cost. Figure 2b reports the reduction in training time obtained with the coarse graph on eight datasets; arxiv-year is omitted because it uses the same graph as ogbn-arxiv, with different labels. Using the coarse graph output from our method, training is $8\times$ faster than training on full graph, and only $3.5\times$ faster without edge pruning, consistent with the densification of the coarse graph discussed in Section 1. In terms of memory usage, we see a $10\times$ reduction, which is in line with the tenfold reduction in nodes. To make a fair comparison for speedup in training using a coarse graph as compared to the full graph, we attribute the coarsening runtime cost to a single training run (the least favorable accounting, as one coarsening can be used multiple times across seeds and backbones), and we observe that coarsening and training together are $5\times$ faster than full-graph training.

4.4 Analysis and Ablations

Objective and Krylov approximation constant. Table 15 in Appendix H evaluates the Krylov approximation constant $\mathcal{E}^\star(\pi)$ of Definition 2 and the objective $\mathcal{E}(\pi)$ of Equation (5) on the partition produced by every method at 90% reduction ratio. Ours attains the smallest $\mathcal{E}^\star(\pi)$ on all nine datasets and the smallest $\mathcal{E}(\pi)$ on six; A-ConvMatch reaches a lower objective on ogbn-arxiv, arxiv-year and ogbn-mag. We also empirically measure the tightness of the two-sided bound of Theorem 1(i) for $K = 1$ to 5 (Figure 8).

Basis and operator. We empirically evaluate our method across different polynomial bases and propagation operators (Appendix G), at ratio 90 with a GCN on ogbn-arxiv, ogbn-products, pokec and snap-patents. The basis experiment compares six polynomial bases (Appendix G.1). As every basis spans the same Krylov subspace (Appendix A.2), the objective at $\lambda_{\text{rel}} = 0$ is invariant to the change of basis; for $\lambda_{\text{rel}} > 0$ the ridge weights are set by the singular values of the trajectory, which are not basis-invariant, so the merge can select different partitions. Empirically (Figure 6, Appendix G.1), the objective values $\mathcal{E}(\pi)$ across the six bases agree to a coefficient of variation below 0.4% at $\lambda_{\text{rel}} = 0$ on every graph, and for $\lambda_{\text{rel}} > 0$ accuracy differs across bases only on pokec. The operator experiment replaces $\hat{A}$ by the self-loop-weighted adjacency, the heat kernel and personalized PageRank (Table 14), operators that may change the Krylov subspace. Accuracy stays within seed noise on three graphs (Figure 7, Appendix G.2); on pokec the self-loop weights stay close to the default $\hat{A}$ while every heat and PPR variant loses about four points.

Effect of pruning. We evaluate our method without the edge pruning step of Section 3.3 across the five backbones, keeping every edge of the contracted graph at every level, and compare against the default coarsening in Figure 3a. Across the 45 dataset–backbone pairs, pruning changes accuracy by less than 1 pp in 23, improves it by 1 pp or more in 18, and lowers it by more than 1 pp in 4. On



(a) (b) (c)

Figure 3: Ablations at 90% reduction ratio. *(a)* Difference in accuracy between the pruned and the unpruned coarsening, for the five backbones; differences within one training-seed standard deviation are shown in Grey. *(b)* GCN accuracy as a function of the polynomial degree K, relative to the default $K = 3$. *(c)* GCN accuracy as a function of the escape ratio ζ, relative to the default $\zeta = 0.5$. In (b) and (c), each line is one dataset and the band spans one training-seed standard deviation.

the three OGB graphs, four of the five backbones are more accurate on the pruned coarsening, by 1–6 pp (+5.9 for GraphSAGE on ogbn-products, +4.4 for GPR-GNN on ogbn-mag). The four cases in which pruning lowers accuracy by more than 1 pp are GraphSAGE and GPR-GNN on pokec (-4.4 and -5.5 pp, the latter within seed noise), GPR-GNN on Reddit2 (-3.1 pp) and SGC on twitch-gamers (-1.5 pp). The budget parameter ν controls the final number of edges in the coarsened graph (Figure 5a): a larger ν removes more edges at every level. On ogbn-products, accuracy keeps improving as the graph gets sparser, by 3.7 pp from $\nu = 0$ to $\nu = 1$; snap-patents moves the other way and loses 3.0 pp over the same range; and ogbn-arxiv, arxiv-year and ogbn-mag are largely indifferent, staying within 1.6 pp across all of ν. Edge pruning therefore yields a smaller graph and faster training without a consistent loss in accuracy.

Hyperparameter sensitivity. We vary one hyperparameter at a time around the default configuration of Appendix E, at ratio 90 with a GCN and three training seeds, on ogbn-arxiv, ogbn-mag, ogbn-products, arxiv-year and snap-patents. Figure 3(b,c) show the polynomial degree K and the escape ratio ζ. With $K = 0$ the merge reduces to Ward clustering of the raw features and loses 1.5–3.1 pp on four of the five datasets; one propagation step recovers most of this, and all $K \geq 1$ lie within 0.5 pp of the default $K = 3$. Setting $\zeta = 0$, which restricts the candidates to adjacent pairs, costs 1.7 pp on snap-patents, 1.4 on arxiv-year and 0.8 on ogbn-products, and nothing measurable on ogbn-arxiv and ogbn-mag. The ridge λ_{rel} is swept in Figure 5c: accuracy stays within 0.6 pp of the default for every $\lambda_{\text{rel}} \leq 10^{-2}$, and only ogbn-arxiv changes at larger values, losing 1.1–2.6 pp. The remaining parameters are swept in Appendix E.4.

5 Conclusion

We propose a graph coarsening method that preserves the block Krylov subspace spanned by the node features and their first K propagations, so that the coarsened graph accounts for the outputs of polynomial filters before their coefficients are learned. A projection-residual objective bounds the worst-case transfer error from both sides; its ridge-weighted form is a k-means objective, so merges are scored by Ward increments, while a per-level edge budget controls densification. On nine benchmarks at 90% node reduction, our method attains a mean rank of 1.2 with GCN, improving on the best baseline by a median of 2.39 pp; coarsening is $3\times$ faster than the strongest baseline on CPU, and coarsening followed by training is $5\times$ faster than training on the full graph. Two directions remain for future work. First, the guarantee covers node merging alone, and edge pruning is applied after it without a bound of its own; a joint analysis of the two errors would connect the objective to the sparse propagation operator actually used in training. The second is to extend the transfer analysis to nonlinear message passing, where learned transformations are interleaved with propagation.

AI USE STATEMENT

In this work, we used large language models (LLMs) to edit the text of the paper for readability, to assist with LaTeX formatting, and to identify and summarize relevant literature. We did not use LLMs to develop the method or its theoretical framework, to formulate or prove the mathematical claims, to design the experiments, or to interpret the results; generating synthetic data, translation, dataset cleaning and qualitative data analysis are not applicable to this work. Additionally, we used LLMs as a coding assistant for the implementation and for the scripts that render the figures and tables. We reviewed all AI-assisted output: every edited passage was checked against the intended meaning, every reference suggested by an LLM was verified against the original source, and LLM-assisted code was read, tested and verified for correctness by the authors. We take full responsibility for the final content of this work, including any text or code prepared with the aid of generative AI.

REFERENCES

Xu Bai, Bin Lu, Kun Zhang, Shengbo Chen, Xinbing Wang, Chenghu Zhou, and Meng Jin. Rethinking efficient graph coarsening via a non-selfishness principle. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=WceHgf9JOZ.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In Yoshua Bengio and Yann LeCun (eds.), *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014. URL http://arxiv.org/abs/1312.6203.

Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=n6jl7fLxrP.

Ido Cohen and Ronen Talmon. Spectral graph coarsening using inner product preservation and the grassmann manifold. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 117415–117446. Curran Associates, Inc., 2025. doi: 10.52202/085713-3915. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/aa4bee5e2720d7e8f073d762ee8c7fd4-Paper-Conference.pdf.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 3837–3845, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/04df4d434d481c5bb723be1b6df1ee65-Abstract.html.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1):1–30, 2006. URL http://jmlr.org/papers/v7/demsar06a.html.

Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Weighted graph cuts without eigenvectors A multilevel approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(11):1944–1957, 2007. doi: 10.1109/TPAMI.2007.1115. URL https://doi.org/10.1109/TPAMI.2007.1115.

Charles Dickens, Edward W. Huang, Aishwarya Reganti, Jiong Zhu, Karthik Subbian, and Danai Koutra. Graph coarsening via convolution matching for scalable graph neural network training. In Tat-Seng Chua, Chong-Wah Ngo, Roy Ka-Wei Lee, Ravi Kumar, and Hady W. Lauw (eds.), *Companion Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, Singapore, May 13-17, 2024*, pp. 1502–1510. ACM, 2024. doi: 10.1145/3589335.3651920. URL https://doi.org/10.1145/3589335.3651920.

Florian Dörfler and Francesco Bullo. Kron reduction of graphs with applications to electrical networks. *IEEE Trans. Circuits Syst. I Regul. Pap.*, 60-I(1):150–163, 2013. doi: 10.1109/TCSI.2012.2215780. URL https://doi.org/10.1109/TCSI.2012.2215780.

Fabrizio Frasca, Emanuele Rossi, Davide Eynard, Benjamin Chamberlain, Michael Bronstein, and Federico Monti. Sign: Scalable inception graph neural networks. In *ICML 2020 Workshop on Graph Representation Learning and Beyond*, 2020.

Lars Gottesbüren, Nikolai Maas, Dominik Rosch, Peter Sanders, and Daniel Seemaier. Linear-time multilevel graph partitioning via edge sparsification. In *33rd Annual European Symposium on Algorithms, ESA 2025, September 15-17, 2025, Warsaw, Poland*, volume 351 of *LIPIcs*, pp. 32:1–32:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025. doi: 10.4230/LIPICS.ESA.2025.32. URL https://doi.org/10.4230/LIPIcs.ESA.2025.32.

Yuhe Guo and Zhewei Wei. Graph neural networks with learnable and optimal polynomial bases. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12077–12097. PMLR, 2023. URL https://proceedings.mlr.press/v202/guo23i.html.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 1024–1034, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html.

David K. Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011. doi: 10.1016/j.acha.2010.04.005.

Mohammad Hashemi, Shengbo Gong, Juntong Ni, Wenqi Fan, B. Aditya Prakash, and Wei Jin. A comprehensive survey on graph reduction: Sparsification, coarsening, and condensation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024, Jeju, South Korea, August 3-9, 2024*, pp. 8058–8066. ijcai.org, 2024. URL https://www.ijcai.org/proceedings/2024/891.

Mingguo He, Zhewei Wei, Zengfeng Huang, and Hongteng Xu. Bernnet: Learning arbitrary graph spectral filters via bernstein approximation. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 14239–14251, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/76f1cfd7754a6e4fc3281bcccb3d0902-Abstract.html.

Mingguo He, Zhewei Wei, and Ji-Rong Wen. Convolutional neural networks on graphs with chebyshev approximation, revisited. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/2f9b3ee2bcea04b327c09d7e3145bd1e-Abstract-Conference.html.

Gecia Bravo Hermsdorff and Lee M. Gunderson. A unifying framework for spectrum-preserving graph sparsification and coarsening. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 7734–7745, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/cd474f6341aeffd65f93084d0dae3453-Abstract.html.

N.J. Higham. *Functions of Matrices: Theory and Computation*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, 2008. ISBN 9780898717778. URL https://books.google.co.in/books?id=S6gpNn1JmbgC.

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 42(1):80–86, February 2000. ISSN 0040-1706. doi: 10.2307/1271436. URL https://doi.org/10.2307/1271436.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/fb60d411a5c5b72b2e7d3527cfc84fd0-Abstract.html.

Keke Huang, Wencai Cao, Hoang Ta, Xiaokui Xiao, and Pietro Liò. Optimizing polynomial graph filters: A novel adaptive krylov subspace approach. In Tat-Seng Chua, Chong-Wah Ngo, Ravi Kumar, Hady W. Lauw, and Roy Ka-Wei Lee (eds.), *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*, pp. 1057–1068. ACM, 2024. doi: 10.1145/3589334.3645705. URL https://doi.org/10.1145/3589334.3645705.

Zengfeng Huang, Shengzhong Zhang, Chong Xi, Tang Liu, and Min Zhou. Scaling up graph neural networks via graph coarsening. In Feida Zhu, Beng Chin Ooi, and Chunyan Miao (eds.), *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, pp. 675–684. ACM, 2021. doi: 10.1145/3447548.3467256. URL https://doi.org/10.1145/3447548.3467256.

Antonin Joly and Nicolas Keriven. Graph coarsening with message-passing guarantees. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/d041d6bb47c01a4ce327a1773703e9a0-Abstract-Conference.html.

George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.*, 20(1):359–392, 1998. doi: 10.1137/S1064827595287997. URL https://doi.org/10.1137/S1064827595287997.

Mohit Kataria, Sandeep Kumar, and Jayadeva. UGC: universal graph coarsening. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/733209a1f12071a7ec979e8ffaeb1d99-Abstract-Conference.html.

Arpandeep Khatua, Vikram Sharma Mailthody, Bhagyashree Taleka, Tengfei Ma, Xiang Song, and Wen-Mei Hwu. IGB: addressing the gaps in labeling, features, heterogeneity, and size of public graph datasets for deep learning research. In Ambuj K. Singh, Yizhou Sun, Leman Akoglu, Dimitrios Gunopulos, Xifeng Yan, Ravi Kumar, Fatma Ozcan, and Jieping Ye (eds.), *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, pp. 4284–4295. ACM, 2023. doi: 10.1145/3580305.3599843. URL https://doi.org/10.1145/3580305.3599843.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Johannes Klicpera, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019a. URL https://openreview.net/forum?id=H1gL-2A9Ym.

Johannes Klicpera, Stefan Weißenberger, and Stephan Günnemann. Diffusion improves graph learning. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 13333–13345, 2019b. URL https://proceedings.neurips.cc/paper/2019/hash/23c894276a2c5a16470e6a31f4618d73-Abstract.html.

Manoj Kumar, Anurag Sharma, Shashwat Saxena, and Sandeep Kumar. Featured graph coarsening with similarity guarantees. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17953–17975. PMLR, 2023. URL https://proceedings.mlr.press/v202/kumar23a.html.

Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser-Nam Lim. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 20887–20902, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/ae816a80e4c1c56caa2eb4e1819cbb2f-Abstract.html.

Andreas Loukas. Graph reduction with spectral and cut guarantees. *J. Mach. Learn. Res.*, 20: 116:1–116:42, 2019. URL https://jmlr.org/papers/v20/18-680.html.

Sitao Luan, Mingde Zhao, Xiao-Wen Chang, and Doina Precup. Break the ceiling: Stronger multi-scale deep graph convolutional networks. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 10943–10953, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/ccdf3864e2fa9089f9eca4fc7a48ea0a-Abstract.html.

Fionn Murtagh. A survey of recent advances in hierarchical clustering algorithms. *Comput. J.*, 26 (4):354–359, 1983. doi: 10.1093/COMJNL/26.4.354. URL https://doi.org/10.1093/comjnl/26.4.354.

Oleg Platonov, Denis Kuznedelev, Artem Babenko, and Liudmila Prokhorenkova. Characterizing graph datasets for node classification: Homophily-heterophily dichotomy and beyond. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/01b681025fdbda8e935a66cc5bb6e9de-Abstract-Conference.html.

Yousef Saad. *Iterative methods for sparse linear systems.* SIAM, 2003. ISBN 978-0-89871-534-7.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Xiyuan Wang and Muhan Zhang. How powerful are spectral graph neural networks. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 23341–23362. PMLR, 2022. URL https://proceedings.mlr.press/v162/wang22am.html.

Joe H. Ward, Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963. doi: 10.1080/01621459.1963.10500845. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845.

Felix Wu, Amauri H. Souza Jr., Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6861–6871. PMLR, 2019. URL http://proceedings.mlr.press/v97/wu19e.html.

Xiang Wu, Rong-Hua Li, Xunkai Li, Kangfei Zhao, Hongchao Qin, and Guoren Wang. Scalable topology-preserving graph coarsening: Concepts and algorithms. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=WtINGEMjTB.

Shuyin Xia, Xinjun Ma, Zhiyuan Liu, Cheng Liu, Sen Zhao, and Guoyin Wang. Graph coarsening via supervised granular-ball for scalable graph neural network training. In Toby Walsh, Julie Shah, and Zico Kolter (eds.), *Thirty-Ninth AAAI Conference on Artificial Intelligence, Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence, Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2025, Philadelphia, PA, USA, February 25 - March 4, 2025*, pp. 12872–12880. AAAI Press, 2025. doi: 10.1609/AAAI.V39I12.33404. URL https://doi.org/10.1609/aaai.v39i12.33404.

Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor K. Prasanna. Graphsaint: Graph sampling based inductive learning method. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=BJe8pkHFwS.

APPENDIX

A EXTENDED BACKGROUND

A.1 COARSENING METHODS

Table 2 summarizes existing coarsening methods, including the baselines of Section 4.1; the merge criterion of each is as follows, in order of publication.

- **Random matching** (Karypis & Kumar, 1998) contracts a random maximal matching at every level. It uses no information about the graph beyond adjacency and serves as the criterion-free reference.
- **Heavy-edge matching** (Dhillon et al., 2007) contracts a maximal matching in which every node is paired with the neighbor joined to it by the heaviest edge, so that the edge weight absorbed at each level is large.
- **Kron reduction** (Dörfler & Bullo, 2013) keeps a chosen subset of nodes and replaces the Laplacian by its Schur complement on the removed ones, which preserves effective resistances between kept nodes.
- **Var-Neighborhoods (VN)** and **Var-Edges (VE)** (Loukas, 2019) assign each candidate contraction set (an edge or a one-hop neighborhood) a *local variation* cost, which measures how much contracting that set would alter the leading r-dimensional eigenspace of the Laplacian, and greedily contract the sets of smallest cost.
- **FGC** (Kumar et al., 2023) optimizes the coarse adjacency and coarse features jointly, minimizing a smoothness objective under a similarity constraint between original and coarse features; it operates on a dense $n \times n$ matrix.
- **A-ConvMatch (ACM)** (Dickens et al., 2024) merges the node pair whose contraction least changes the output of one GCN convolution, using an $O(1)$ upper bound on that change instead of recomputing the convolution.
- **UGC** (Kataria et al., 2024) hashes the concatenation of each node's features and adjacency row with locality-sensitive hashing; nodes that fall into the same bucket form a supernode, without a merge criterion on the graph itself. The weight that balances the two parts of the concatenation is a heterophily factor computed from the training labels.
- **SGBGC** (Xia et al., 2025) grows granular balls around high-degree labeled nodes and splits any ball whose training-label purity falls below a threshold, so supernodes are label-pure by construction.
- **INGC** (Cohen & Talmon, 2025) treats the node features as functions on the graph and chooses the coarsening that best preserves their inner products, by optimization on the Grassmann manifold of coarsening matrices.
- **STPGC** (Wu et al., 2026) collapses a node onto a neighbor whose closed neighbourhood contains its own, a strong collapse from algebraic topology that preserves the homotopy type of the graph, restricted to nodes below a degree threshold θ; among the admissible neighbors, those sharing the node's training label are preferred.
- **NOPE** (Bai et al., 2026) replaces pairwise similarity matching, in which every node searches for its own best partner, by a *non-selfishness* principle that scores the collective interference of a node's neighborhood; this reaches linear memory and near-linear time in n. The variant NOPE* assumes local isotropy to cut the per-node evaluation from $O(d_{\max}d)$ to $O(d)$, removing the bottleneck at high-degree nodes.

A.2 POLYNOMIAL FILTERS AND KRYLOV SUBSPACES

The merge criterion of Section 3.1 preserves the Krylov subspace $\mathcal{K}_{K+1}(S, x)$ of Equation (1). Table 3 lists the spectral GNNs of Section 2.3 by polynomial basis, propagation operator and coefficient scheme; we show below that the subspace depends on the operator alone.

Table 2: Graph coarsening methods. *Node data* is the information beyond the graph structure that the merge criterion requires; *Conn.* indicates whether every supernode is constrained to induce a connected subgraph. Complexities are stated for an input graph with n nodes, m edges, maximum degree $d_{\max}$ and d-dimensional node features, coarsened to k nodes. c is the number of levels of a multilevel scheme, r the number of eigenpairs preserved by the local-variation methods, and θ the degree threshold of STPGC.

Method	Node data	Conn.	Time	Memory
Random matching (Karypis & Kumar, 1998)	none	✓	$O(c(n+m))$	$O(n+m)$
Heavy-edge matching (Dhillon et al., 2007)	none	✓	$O(cm \log m)$	$O(n+m)$
Var-Neighborhoods (Loukas, 2019)	none	✓	$\tilde{O}(crm + r^2 n + cr^3)$	$O(m+nr)$
Var-Edges (Loukas, 2019)	none	✓	$\tilde{O}(crm + r^2 n + cr^3)$	$O(m+nr)$
Kron reduction (Dörfler & Bullo, 2013)	none	n/a	$O(n^3)$	$O(n^2)$
UGC (Kataria et al., 2024)	features, **labels**	✗	$O(nd+m)$	$O(m+nd)$
FGC (Kumar et al., 2023)	features	✗	$O(n^2(k+d))$	$O(n^2)$
INGC (Cohen & Talmon, 2025)	features	✗	$O(n^2(k+d))$	$O(m+n(k+d))$
A-ConvMatch (Dickens et al., 2024)	features	✗	$O(c(m+nd))$	$O(m+nd)$
SGBGC (Xia et al., 2025)	**labels**	✗	$O(n^{3/2}+m\sqrt{n})$	$O(m+nd)$
STPGC (Wu et al., 2026)	**labels**	n/a	$O(n\theta d_{\max}^2)$	$O(n+m)$
NOPE / NOPE* (Bai et al., 2026)	features	✗	$\tilde{O}(nd_{\max}d)$ / $\tilde{O}(nd)$	$O(n+m)$

Table 3: Polynomial basis, propagation operator S and coefficient scheme of spectral GNNs of the form $\sum_{\ell=0}^{K} w_\ell S^\ell x$. $\mathcal{L}$ denotes the normalized Laplacian, $\tilde{\mathcal{L}}$ the normalized Laplacian with self-loops, and $\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ the renormalized adjacency of GCN.

Method	Basis	Operator S	Coefficients
ChebNet (Defferrard et al., 2016)	Chebyshev	$2\mathcal{L}/\lambda_{\max} - I$	learned
GCN (Kipf & Welling, 2017)	monomial, $K{=}1$	$\hat{A}$	fixed
APPNP (Klicpera et al., 2019a)	monomial	$\hat{A}$	fixed, personalized PageRank
SGC (Wu et al., 2019)	monomial	$\hat{A}$	fixed, single power S^K
GPR-GNN (Chien et al., 2021)	monomial	$I - \tilde{\mathcal{L}}$	learned, signed
BernNet (He et al., 2021)	Bernstein	$I - \mathcal{L}/2$	learned, non-negative
JacobiConv (Wang & Zhang, 2022)	Jacobi	$I - \mathcal{L}$	learned, per-channel
ChebNetII (He et al., 2022)	Chebyshev	$2\mathcal{L}/\lambda_{\max} - I$	learned at Chebyshev nodes
OptBasisGNN (Guo & Wei, 2023)	learned, orthonormal	learned	learned
AdaptKry (Huang et al., 2024)	monomial	tunable P_τ	learned

Basis and coefficients. Let $p_0, \ldots, p_K$ be polynomials of degree at most K, $p_\ell(S) = \sum_{j=0}^{K} \Psi_{\ell j} S^j$, with $\Psi \in \mathbb{R}^{(K+1)\times(K+1)}$. Then

$$[\, p_0(S)x,\ \ldots,\ p_K(S)x \,] \;=\; [\, x,\ Sx,\ \ldots,\ S^K x \,]\, \Psi^\top. \tag{12}$$

Hence $\operatorname{span}\{p_\ell(S)x\}_{\ell \le K} = \mathcal{K}_{K+1}(S, x)$ whenever Ψ is invertible, i.e. whenever $\{p_\ell\}$ is a basis of the degree-K polynomials (Huang et al., 2024, Prop. 1). This holds for the Chebyshev, Bernstein and Jacobi bases of Table 3 (Defferrard et al., 2016; He et al., 2021; Wang & Zhang, 2022). For $p_\ell(S) = (aS + bI)^\ell$ with $a \neq 0$, Ψ is triangular with nonzero diagonal, so $\mathcal{K}_{K+1}(aS + bI, x) = \mathcal{K}_{K+1}(S, x)$; the spectral rescaling of ChebNet is of this form. The coefficients w, learned or fixed, select an element of $\mathcal{K}_{K+1}(S, x)$ by Equation (1).

Propagation operator. A non-affine change of operator $S \mapsto S'$ can move $\mathcal{K}_{K+1}(S, x)$ through two mechanisms. If S' has different eigenvectors, as when the self-loop weight in $\hat{A}$ is changed (Kipf & Welling, 2017), the subspaces differ already for an eigenvector x of S: $\mathcal{K}_{K+1}(S, x) = \operatorname{span}\{x\}$, while $\dim \mathcal{K}_2(S', x) = 2$. If $S' = g(S)$ shares the eigenvectors of S but reweights its eigenvalues,

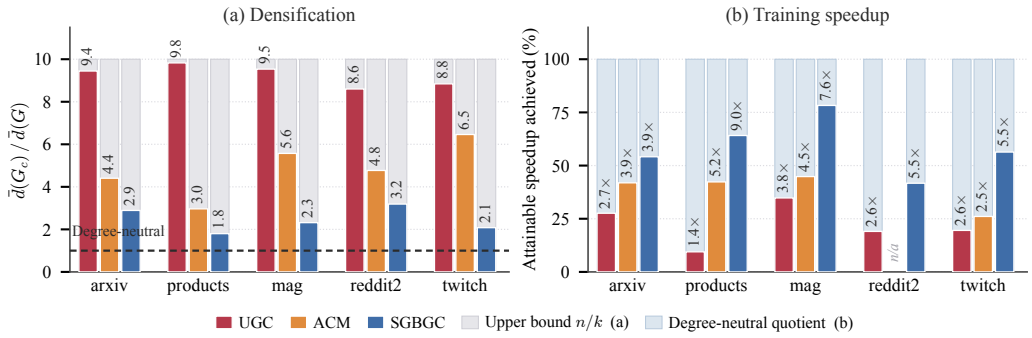


Figure 4: *(a)* Degree multiplier at 90% node reduction, against the ceiling $n/k = 10$ that no contraction can exceed. *(b)* Measured GCN speedup over the full graph, as a share of what a degree-matched control reaches; the absolute speedup is annotated.

as for the heat kernel $e^{-\tau(I-S)}$ and personalized PageRank $\alpha(I-(1-\alpha)S)^{-1}$ (Klicpera et al., 2019b), then $g(S)$ is a polynomial in S (Higham, 2008), so $\mathcal{K}_{K+1}(g(S), x)$ lies in the smallest S-invariant subspace containing x (Saad, 2003), but need not coincide with $\mathcal{K}_{K+1}(S, x)$ unless g is affine. Table 4 collects the cases; Section 4.4 tests both operator changes.

Table 4: Effect of the three design axes of Table 3 on the Krylov subspace $\mathcal{K}_{K+1}(S, x)$. A change of basis or of coefficients leaves the subspace unchanged (Equations (1) and (12)); of the operator changes, only the affine one does.

Axis	**Change**	$\mathcal{K}_{K+1}(S, x)$	**Condition**
Basis	$\{S^\ell\} \mapsto \{p_\ell(S)\}$	unchanged	Ψ invertible
Coefficients	$w \mapsto w'$	unchanged	none
Operator	$S \mapsto aS + bI$	unchanged	$a \neq 0$
Operator	$S \mapsto g(S)$, g non-affine	may change	eigenvectors of S retained
Operator	$S \mapsto S'$	changes	x an eigenvector of S but not of S'

A.3 Additional concepts

Definition 3 (Hausdorff distance)**.** Let $(\mathcal{M}, \mathrm{dist})$ be a metric space and $\mathcal{A}, \mathcal{B} \subseteq \mathcal{M}$ nonempty and compact. The directed distance from $\mathcal{A}$ to $\mathcal{B}$ is $\vec{\mathrm{dist}}(\mathcal{A}, \mathcal{B}) := \sup_{\xi \in \mathcal{A}} \inf_{\omega \in \mathcal{B}} \mathrm{dist}(\xi, \omega)$, and the Hausdorff distance is

$$d_{\mathrm{H}}(\mathcal{A}, \mathcal{B}) \;:=\; \max\{\vec{\mathrm{dist}}(\mathcal{A}, \mathcal{B}),\, \vec{\mathrm{dist}}(\mathcal{B}, \mathcal{A})\} \;=\; \inf\{\varepsilon \geq 0 :\; \mathcal{A} \subseteq \mathcal{B}_\varepsilon,\; \mathcal{B} \subseteq \mathcal{A}_\varepsilon\},$$

where $\mathcal{A}_\varepsilon := \{\xi \in \mathcal{M} : \inf_{\omega \in \mathcal{A}} \mathrm{dist}(\xi, \omega) \leq \varepsilon\}$ is the closed ε-neighborhood of $\mathcal{A}$.

d_{H} is a metric on the nonempty compact subsets of $\mathcal{M}$ (Burago et al., 2001, Def. 7.3.1); in particular $d_{\mathrm{H}}(\mathcal{A}, \mathcal{B}) = 0$ if and only if $\mathcal{A} = \mathcal{B}$. In Section 3.1, $\mathcal{M} = \mathbb{R}^{n \times d}$ with the Frobenius distance, and $\mathcal{A} = \mathcal{O}_G$, $\mathcal{B} = \mathcal{O}_c^{\uparrow}$ are the images of the compact coefficient set $\mathcal{W}$ under linear maps, hence compact; $d_{\mathrm{H}}(\pi) := d_{\mathrm{H}}(\mathcal{O}_G, \mathcal{O}_c^{\uparrow})$.

B Edge reduction under coarsening

Under a coarsening π an edge of G is *absorbed* if both endpoints fall in the same supernode; otherwise it survives as part of a coarse edge, and edges of G joining the same pair of supernodes *coalesce* into one. With E_{int} the absorbed edges and w_{uv}^c the number of edges of G between supernodes C_u

Table 5: Ratio of the average degree of the coarsened graph to that of the input, $\bar{d}(G_c)/\bar{d}(G)$, for each coarsener and dataset at four node-reduction targets. A value of 1 indicates that edges were reduced in the same proportion as nodes. ‡ marks a coarsening that did not reach the target reduction; dashes mark runs that ran out of memory or time.

target	VN	VE	UGC	ACM	SGBGC	target	VN	VE	UGC	ACM	SGBGC
ogbn-arxiv ($\bar{d} = 13.7$)						*ogbn-products* ($\bar{d} = 50.5$)					
80%	1.6	1.9	4.9	3.2	2.2	80%	—	—	4.9	2.0	—
90%	1.4	1.5	9.4	4.4	2.9	90%	—	—	9.8	3.0	1.8
95%	**0.3**	**0.3**	18.3	5.9	3.0	95%	—	—	18.9	3.8	1.8
99%	**0.4**	‡	30.1	8.2	4.3	99%	—	—	38.8	7.3	1.2
arxiv-year ($\bar{d} = 13.7$)						*pokec* ($\bar{d} = 27.3$)					
80%	1.6	1.9	4.9	3.2	1.9	80%	2.2	3.2	4.9	4.7	3.8
90%	1.4	1.5	9.9	4.4	2.4	90%	4.2	**0.1**	9.8	9.3	5.9
95%	**0.3**	**0.3**	18.1	5.9	3.5	95%	**0.2**	‡	19.5	16.0	8.5
99%	**0.4**	‡	29.7	8.2	6.6	99%	**0.1**	‡	67.0	38.1	24.8
reddit2 ($\bar{d} = 99.7$)						*snap-patents* ($\bar{d} = 9.6$)					
80%	1.8	2.6	4.9	3.1	2.2	80%	‡	‡	4.9	—	—
90%	2.4	2.1	8.6	3.7	3.2	90%	‡	‡	9.7	—	3.3
95%	2.4	‡	11.1	4.2	4.1	95%	‡	‡	19.8	—	2.7
99%	**0.8**	‡	8.4	—	5.0	99%	‡	‡	85.0	—	‡
igb-small ($\bar{d} = 12.1$)						*twitch-gamers* ($\bar{d} = 80.9$)					
80%	3.4	3.3	5.0	4.2	2.6	80%	2.4	1.8	4.8	4.2	2.5
90%	5.6	3.9	9.9	7.7	3.9	90%	**0.3**	‡	8.8	6.5	2.1
95%	7.5	‡	19.3	12.6	5.7	95%	**0.2**	‡	12.7	8.8	2.4
99%	2.0	‡	64.2	23.7	8.5	99%	‡	‡	9.4	7.9	3.2
ogbn-mag ($\bar{d} = 14.7$)											
80%	2.1	2.2	5.0	3.3	1.9						
90%	2.8	2.3	9.5	5.6	2.3						
95%	3.3	1.0	19.5	8.0	3.1						
99%	**0.7**	‡	47.9	12.8	6.0						

Table 6: Normalized component excess $\varphi = \Phi(\pi)/(n-k)$ for the two coarseners with the largest values, by dataset and node-reduction target. $\varphi = 0$ corresponds to a connected partition and $\varphi = 1$ to cells that are all independent sets; dashes mark unavailable results.

	A-ConvMatch				UGC			
dataset	80%	90%	95%	99%	80%	90%	95%	99%
ogbn-arxiv	0.653	0.580	0.512	0.377	1.000	0.999	0.998	0.981
arxiv-year	0.653	0.580	0.512	0.377	1.000	0.999	0.998	0.980
IGB-small	0.863	0.835	—	—	1.000	1.000	1.000	0.997
ogbn-mag	0.754	0.723	0.679	0.542	1.000	1.000	0.999	0.992
ogbn-products	0.572	0.522	0.475	—	0.999	0.999	0.999	0.987
Pokec	0.984	0.982	0.969	—	1.000	1.000	1.000	0.998
Reddit2	0.678	—	—	—	0.997	0.991	0.978	0.835
snap-patents	—	—	—	—	0.999	0.999	0.999	0.999
Twitch-gamers	0.978	0.958	0.934	0.836	0.998	0.995	0.987	0.869

and C_v,

$$|E| - |E_c| \;=\; |E_{\text{int}}| \;+\; \sum_{\{u,v\}\in E_c} \big(w^c_{uv} - 1\big). \tag{13}$$

A merge criterion acts on the first term only: contracting an adjacent pair absorbs one edge, contracting a non-adjacent pair absorbs none, and the coalesced count is set by the neighbors the two supernodes share. Table 5 reports $\bar{d}(G_c)/\bar{d}(G)$ for every coarsener, dataset and target; values above 1 mean the quotient is denser than the input.

For a supernode C let $\kappa(C)$ denote the number of connected components of the induced subgraph $G[C]$. A partition is *connected* if $\kappa(C_u) = 1$ for every supernode, the condition imposed by Loukas (2019). The departure from this condition is measured by the *component excess*

$$\Phi(\pi) \;=\; \sum_{u=1}^{k} \big(\kappa(C_u) - 1\big), \qquad 0 \le \Phi(\pi) \le n - k,$$

where the lower bound is attained exactly by connected partitions and the upper bound exactly when every superndoe is an independent set of G. Table 6 reports the normalized value $\varphi = \Phi(\pi)/(n-k)$.

Since a connected graph on p vertices has at least $p-1$ edges, $|E_{\text{int}}| \ge \sum_u (|C_u| - \kappa(C_u)) = (n-k) - \Phi(\pi)$, and by Equation (13)

$$|E_c| \;\le\; |E| - (n-k) + \Phi(\pi). \tag{14}$$

A connected partition thus absorbs at least $n-k$ edges, whereas degree neutrality, $|E_c| \le |E|\ k/n$, requires removing $|E|\,(1-k/n)$; absorption covers at most a fraction $2/\bar{d}(G)$ of this, between 2% and 15% on the graphs of Table 7. The remaining reduction must come from coalescence, which no merge criterion controls, or from pruning (Section 3.3). UGC's supernodes are close to independent sets (Table 6); on ogbn-arxiv at target 90 they absorb 100 edges in total and 97% of the edges remain in the quotient.

C Algorithm and additional details

Algorithm 1 states the multilevel construction of Section 3, and Algorithm 2 the candidate family it uses at each level.

C.1 Time and memory complexity

Let $p = (K+1)d$ be the width of $\mathcal{Z}$, let $q \le p$ be the number of columns of $\mathcal{Z}_\lambda$ retained after truncation (Appendix C.2), and let $n_t \le n$ and $m_t \le |E|$ be the numbers of supernodes and retained edges at level t. We count array operations and charge sorting a keys $O(a \log n)$; the propagation operator S is sparse with $O(n + |E|)$ nonzeros. The procedure has two one-time stages and two per-level stages.

1. *Trajectory.* $\mathcal{Z} = [X, SX, \dots, S^K X]$ requires K sparse products with an $n \times d$ matrix, $O(K(n+|E|)d)$ time and $O(np)$ memory.
2. *Ridge scaling.* Since $\mathcal{Z}_\lambda = \mathcal{Z}V(\Sigma^2 + \lambda I)^{-1/2}$, only the $p \times p$ Gram matrix $\mathcal{Z}^\top\mathcal{Z} = V\Sigma^2 V^\top$ and its eigendecomposition are needed, $O(np^2 + p^3)$ time and $O(p^2)$ memory, followed by the projection of $\mathcal{Z}$ onto the retained q directions in $O(npq)$.
3. *Merging (per level).* A Ward increment costs $O(q)$. Scoring the m_t adjacent pairs, the $O(b)$ sampled candidates per supernode and their $O(b^2)$ neighbor-of-neighbor refinements (Algorithm 2) evaluates $O(m_t + b^2 n_t)$ increments; a fixed number of mutual-best matching passes over the resulting $O(m_t + bn_t)$ pairs and the contraction of the edge list each sort at most as many keys. Updating the assignment of the original nodes costs $O(n)$. The level therefore costs $O((q + \log n)(m_t + b^2 n_t) + n)$ time and $O(n_t(q + b^2) + m_t)$ memory.
4. *Edge selection (per level).* The spanning forest of the contracted graph is extracted from the forest of the previous level in $O(n_t)$ time by breadth-first search. The coverage γ_{uv} counts distinct members of C_u and C_v joined by an edge of G and is recomputed from the edges of G in $O(|E| \log n)$ whenever the budget Equation (11) binds; ranking the residual edges sorts m_t keys. The level therefore costs $O((|E| + m_t)\log n)$ time and $O(n + |E|)$ memory.

Summing over T levels and using $n_t \le n$, $m_t \le |E|$ and $q \le p$, the total time is

$$O\big(K(n+|E|)d \;+\; np^2 + p^3 \;+\; T\,(p + \log n)(|E| + b^2 n)\big), \tag{15}$$

where the level term absorbs the assignment updates and the coverage recomputation. The peak memory is $O(n(p+b^2) + p^2 + |E|)$: the trajectory, the Gram matrix, the candidate pools and the edge list, none of which is retained across levels. A level at most halves the number of supernodes, so $T \ge \log_2(n/k)$, but T is not bounded above in terms of n/k; at 90% node reduction no evaluated dataset required more than 30 levels (Reddit2), and twitch-gamers required 5. For fixed K, d and b, the cost is therefore $O(T(n+|E|)\log n)$, near-linear in the size of G for every level.

Algorithm 1 Krylov subspace-preserving coarsening

Input: graph $G = (V, E)$, features $X \neq 0$, operator S, degree $K \geq 0$, target $1 \leq k \leq n$
ridge $\lambda_{\text{rel}} \geq 0$, escape ratio $\zeta \in [0, 1]$, budget parameter $\nu \geq -1$, pairing rounds $J \geq 1$
Output: coarsening $\pi : V \to [k]$ and retained graph $\mathcal{Q}$ on its k supernodes

$\mathcal{Z} \leftarrow [\, X, SX, \ldots, S^K X \,] = Q\Sigma V^\top$; $\mathcal{Z}_\lambda \leftarrow Q\Sigma\big(\Sigma^2 + \lambda_{\text{rel}}\sigma_1^2 I\big)^{-1/2}$ ▷ Equation (9)
$m_0 \leftarrow n$; $\pi_0 \leftarrow$ a bijection $V \to [n]$ ▷ every node is its own supernode
$\mathcal{Q}_0 \leftarrow G$; $\mathcal{F}_0 \leftarrow$ a spanning forest of G; $t \leftarrow 0$
while $m_t > k$ **do**
  Merge phase
  $\mathcal{P}_t(\zeta) \leftarrow$ candidate pairs of $\mathcal{Q}_t$ ▷ Algorithm 2
  $\Omega_t \leftarrow \varnothing$; $\mathcal{C}_t^{\text{avail}} \leftarrow \mathcal{C}_t = \{\pi_t^{-1}(c) : c \in [m_t]\}$ ▷ all supernodes available
  for $j = 1, \ldots, J$ **do**
    $\text{best}(C) \leftarrow$ none for every $C \in \mathcal{C}_t^{\text{avail}}$ ▷ no choice yet
    for each $C \in \mathcal{C}_t^{\text{avail}}$ that has an available candidate **do**
      $\text{best}(C) \leftarrow \arg\min\{\Delta\mathcal{E}_\lambda(C, C') : C' \in \mathcal{C}_t^{\text{avail}},\ \{C, C'\} \in \mathcal{P}_t(\zeta)\}$
    end for
    $\mathcal{N} \leftarrow \big\{\{C, C'\} \subseteq \mathcal{C}_t^{\text{avail}} : \text{best}(C) = C' \text{ and } \text{best}(C') = C\big\}$ ▷ mutual choices
    for each $\{C, C'\} \in \mathcal{N}$ in increasing order of $\Delta\mathcal{E}_\lambda(C, C')$ **do**
      if $|\Omega_t| < m_t - k$ **then** ▷ ensures $m_{t+1} = m_t - |\Omega_t| \geq k$
        $\Omega_t \leftarrow \Omega_t \cup \big\{\{C, C'\}\big\}$; $\mathcal{C}_t^{\text{avail}} \leftarrow \mathcal{C}_t^{\text{avail}} \setminus \{C, C'\}$
      end if
    end for
  end for
  $m_{t+1} \leftarrow m_t - |\Omega_t|$
  $(\pi_{t+1}, \widehat{\mathcal{Q}}_{t+1}) \leftarrow (\pi_t, \mathcal{Q}_t)$ with every pair of Ω_t merged
  Sparsify phase
  $\mathcal{F}_{t+1} \leftarrow$ a spanning forest of $\mathcal{F}_t$ after the merge
  $\mathcal{R}_{t+1} \leftarrow$ top $s_{t+1}(\nu)$ edges of $\widehat{\mathcal{Q}}_{t+1} \setminus \mathcal{F}_{t+1}$ by score $w_{uv}^c \gamma_{uv}$
  $\mathcal{Q}_{t+1} \leftarrow \mathcal{F}_{t+1} \cup \mathcal{R}_{t+1}$; $t \leftarrow t + 1$
end while
return π_t, $\mathcal{Q}_t$

C.2 Ridge scaling

Equation (9) replaces the trajectory $\mathcal{Z} = Q\Sigma V^\top$ by $\mathcal{Z}_\lambda = Q\Sigma(\Sigma^2 + \lambda_{\text{rel}}\sigma_1^2 I)^{-1/2} = Q\Phi$, where $\Phi = \text{diag}(\phi_1, \phi_2, \ldots)$ with

$$\phi_i^2 = \frac{\sigma_i^2}{\sigma_i^2 + \lambda_{\text{rel}}\sigma_1^2} = \frac{\rho_i^2}{\rho_i^2 + \lambda_{\text{rel}}}, \qquad \rho_i = \frac{\sigma_i}{\sigma_1}, \qquad d_{\text{eff}}(\lambda) \;:=\; \textstyle\sum_i \phi_i^2 \;=\; \|\mathcal{Z}_\lambda\|_F^2 . \tag{16}$$

Here $d_{\text{eff}}(\lambda)$ is the *effective dimension* of the ridge-scaled trajectory: the ridge shrinks each singular value σ_i of $\mathcal{Z}$ to $\phi_i \in (0, 1]$ while leaving the left singular vectors Q unchanged, so $d_{\text{eff}}(\lambda)$ is the sum of the squared singular values of $\mathcal{Z}_\lambda$, the effective degrees of freedom of ridge regression (Hoerl & Kennard, 2000). It decreases from $\text{rank}(\mathcal{Z})$ at $\lambda_{\text{rel}} = 0$ towards 0 as $\lambda_{\text{rel}} \to \infty$, and is not necessarily an integer. Since the columns q_i of Q are orthonormal, both objectives are weighted sums of the same per-direction residuals,

$$\mathcal{E}(\pi) = \sum_i \sigma_i^2 \, \big\|\Pi^\perp q_i\big\|_2^2 , \qquad \mathcal{E}_\lambda(\pi) = \sum_i \phi_i^2 \, \big\|\Pi^\perp q_i\big\|_2^2 ,$$

and differ only in the weights: $\mathcal{E}$ weights direction i by its energy σ_i^2, $\mathcal{E}_\lambda$ by $\phi_i^2 \in (0, 1]$, which is close to 1 for $\rho_i \gg \sqrt{\lambda_{\text{rel}}}$ and close to 0 for $\rho_i \ll \sqrt{\lambda_{\text{rel}}}$. Thus $\sqrt{\lambda_{\text{rel}}}$ acts as a soft cutoff on the normalized spectrum and d_{eff} counts the directions that survive it.

Relation to the transfer error. Termwise, $\sigma_i^2 = \phi_i^2(\sigma_i^2 + \lambda_{\text{rel}}\sigma_i^2) \leq \phi_i^2\sigma_1^2(1 + \lambda_{\text{rel}})$, and summing over i gives, for every partition π and every $\lambda_{\text{rel}} \geq 0$, the relation stated in Section 3.1,

$$\mathcal{E}^\star(\pi)^2 \;\leq\; \mathcal{E}(\pi) \;\leq\; \sigma_1^2(1 + \lambda_{\text{rel}})\, \mathcal{E}_\lambda(\pi), \tag{17}$$

Algorithm 2 Candidate family $\mathcal{P}_t(\zeta)$ for one level

Input: retained graph $\mathcal{Q}_t$ on $\mathcal{C}_t$, Ward increment $\Delta\mathcal{E}_\lambda$, width $b \geq 1$, escape ratio $\zeta \in [0,1]$, search depth $h \geq 0$
Output: candidate family $\mathcal{P}_t(\zeta)$, a set of unordered pairs of supernodes

if $\zeta = 0$ **then**
  return $E(\mathcal{Q}_t)$ ▷ adjacent pairs only
end if
for each $C \in \mathcal{C}_t$ **do**
  $N_E(C) \leftarrow \{C' : \{C, C'\} \in E(\mathcal{Q}_t)\}$ ▷ graph neighbors
  $\Delta_C^E \leftarrow \min_{C' \in N_E(C)} \Delta\mathcal{E}_\lambda(C, C')$ ▷ $+\infty$ if $N_E(C) = \varnothing$
  $\mathcal{S}_C \leftarrow b$ random graph neighbors and b random supernodes
  $N^{(0)}(C) \leftarrow$ the b cheapest supernodes of $\mathcal{S}_C \setminus \{C\}$ by $\Delta\mathcal{E}_\lambda(C, \cdot)$
end for
for $j = 1, \ldots, h$ **do**
  for each $C \in \mathcal{C}_t$ **do**
    $\mathcal{S}_C \leftarrow N^{(j-1)}(C) \cup \bigcup_{C' \in N^{(j-1)}(C)} N^{(j-1)}(C')$ ▷ neighbors of neighbors
    $N^{(j)}(C) \leftarrow$ the b cheapest supernodes of $\mathcal{S}_C \setminus \{C\}$ by $\Delta\mathcal{E}_\lambda(C, \cdot)$
  end for
end for
return $E(\mathcal{Q}_t) \cup \left\{\{C, C'\} : C' \in N^{(h)}(C) \setminus N_E(C),\ \Delta\mathcal{E}_\lambda(C, C') \leq \zeta\, \Delta_C^E\right\}$

where the first inequality is Equation (6); the second is an equality when every σ_i with $\Pi^\perp q_i \neq 0$ equals σ_1. Ranking partitions by $\mathcal{E}_\lambda$ therefore ranks them by an upper bound on the squared transfer error whose prefactor does not depend on π. Moreover, Φ is invertible on the rank of $\mathcal{Z}$, so $\Pi^\perp \mathcal{Z}_\lambda = 0$ exactly when $\Pi^\perp \mathcal{Z} = 0$: the ridge changes how residuals are weighted, not which partitions are exact.

Truncating the trajectory. Because the weights ϕ_i^2 decrease with i, the trailing columns of $\mathcal{Z}_\lambda$ contribute little to $\mathcal{E}_\lambda$, and we can drop them to reduce the width of the trajectory the merge works with. Let $\mathcal{E}_{\lambda,q}(\pi)$ denote the objective computed from the leading q columns of $\mathcal{Z}_\lambda$ only, and let

$$\tau_q \;:=\; \sum_{i>q} \phi_i^2$$

be the total weight of the discarded directions. Each discarded term $\phi_i^2 \left\|\Pi^\perp q_i\right\|_2^2$ lies between 0 and ϕ_i^2, because $\left\|\Pi^\perp q_i\right\|_2 \leq \|q_i\|_2 = 1$, so for *every* partition

$$\mathcal{E}_{\lambda,q}(\pi) \;\leq\; \mathcal{E}_\lambda(\pi) \;\leq\; \mathcal{E}_{\lambda,q}(\pi) + \tau_q. \tag{18}$$

The truncated objective therefore approximates the full one to within τ_q uniformly over partitions, and τ_q depends on the spectrum of $\mathcal{Z}$ alone, so the approximation error is known before the merge starts rather than estimated after it. In practice we fix a tolerance on τ_q (the default is 0.05, Appendix E.3) and take q as the smallest width that meets it. Since $\phi_i^2 \leq 1$, meeting a tolerance τ_q requires $q \geq d_{\text{eff}}(\lambda) - \tau_q$, so d_{eff} is a lower bound on the width that any tolerance allows.

C.3 Additional theoretical results

Theorem 1 and Corollary 1 compare filter outputs. When S is symmetric, the Krylov approximation constant also controls the feature-weighted spectral moments $X^\top S^j X$. These determine the output energy $\left\|\sum_\ell S^\ell X W_\ell\right\|_F^2$ of every polynomial filter of degree at most K through the moments of degree at most $2K$; Corollary 2 controls one more, up to degree $2K+1$, and shows that this range is sharp.

Corollary 2. *Let π be a coarsening of G with coarse operator $S_c = RSR^+$, and assume $S = S^\top$, $\|S\|_2 \leq 1$ and $\|X\|_F = 1$. For every integer $j \geq 0$, define the feature-weighted spectral moments of order j on the original and on the coarsened graph,*

$$M_j \;:=\; X^\top S^j X \;\in\; \mathbb{R}^{d\times d}, \qquad M_j^c \;:=\; (RX)^\top H\, S_c^j\, RX \;\in\; \mathbb{R}^{d\times d}. \tag{19}$$

Then the following hold.

(i) *For* $0 \le j \le K$, $\left\|M_j - M_j^c\right\|_F \;\le\; \mathcal{E}^\star(\pi)$.

(ii) *For* $K < j \le 2K+1$, $\left\|M_j - M_j^c\right\|_F \;\le\; 2\,\mathcal{E}^\star(\pi)$.

(iii) *The degree range in* (ii) *is sharp: for every* $K \ge 0$ *and* $d \ge 1$ *there exist* S, X *and* π *satisfying the assumptions above with* $\mathcal{E}^\star(\pi) = 0$ *and* $M_{2K+2} \ne M^c_{2K+2}$.

Parts (i) and (ii) follow from the bilinear form of the moments, which needs symmetry of S; the counterexample of (iii) is a path graph with a single merged pair, constructed in the proof (Appendix D).

D PROOFS

Notation. Throughout, $\Pi = R^+R = PH^{-1}P^\top$ denotes the orthogonal projector onto $\mathrm{range}(P)$, so that $\Pi = \Pi^\top = \Pi^2$, and $\Pi^\perp = I - \Pi$ its complement. Since $RR^+ = H^{-1}P^\top P = I_k$, R is a left inverse of R^+. The proofs below use the following identity repeatedly.

Lemma 1. *Let* $S_c = RSR^+$. *For every integer* $\ell \ge 0$,

$$R^+ S_c^\ell R \;=\; (\Pi S)^\ell \Pi. \tag{20}$$

Proof. For $\ell = 0$, both sides equal $R^+R = \Pi$. Suppose the identity holds for some $\ell \ge 0$. Using $S_c = RSR^+$ and $R^+R = \Pi$, we obtain

$$\begin{aligned} R^+ S_c^{\ell+1} R &= R^+(RSR^+)S_c^\ell R = \Pi S\big(R^+ S_c^\ell R\big) \\ &= \Pi S(\Pi S)^\ell \Pi = (\Pi S)^{\ell+1}\Pi. \end{aligned}$$

The claim follows by induction. □

Proposition 1. *Let* $\|S\|_2 \le 1$ *and* $\|X\|_F = 1$. *For every coarsening* π,

$$\mathcal{E}^\star(\pi) \;=\; \max_{0\le \ell \le K} \|\Gamma_\ell(\pi)\, X\|_F \;\le\; 2.$$

For any index $\ell^\star$ *attaining the maximum, the supremum in Equation* (4) *is attained by* $W_{\ell^\star} = I_d$ *and* $W_\ell = 0$ *for* $\ell \ne \ell^\star$.

Proof. For every $W \in \mathcal{W}$, the nonnegative weights $\|W_\ell\|_2$ satisfy $\sum_{\ell=0}^K \|W_\ell\|_2 \le 1$. By the triangle inequality and $\|AB\|_F \le \|A\|_F \|B\|_2$,

$$\begin{aligned} \mathcal{T}_\pi(W) &= \left\| \sum_{\ell=0}^K \Gamma_\ell(\pi) X W_\ell \right\|_F \\ &\le \sum_{\ell=0}^K \|\Gamma_\ell(\pi) X\|_F \; \|W_\ell\|_2 \\ &\le \Big(\max_{0\le j\le K} \|\Gamma_j(\pi) X\|_F \Big) \sum_{\ell=0}^K \|W_\ell\|_2 \\ &\le \max_{0\le j\le K} \|\Gamma_j(\pi) X\|_F \,. \end{aligned}$$

For any maximizing index $\ell^\star$, the choice $W_{\ell^\star} = I_d$ and $W_\ell = 0$ for $\ell \ne \ell^\star$ belongs to $\mathcal{W}$ and satisfies $\mathcal{T}_\pi(W) = \|\Gamma_{\ell^\star}(\pi) X\|_F$. Hence the upper bound is attained, which proves the claimed identity.

Finally, $\|\Pi\|_2 \le 1$ because Π is an orthogonal projector, and $\|S\|_2 \le 1$ by hypothesis. Repeated application of $\|AB\|_2 \le \|A\|_2 \|B\|_2$ (submultiplicativity) gives, for $0 \le \ell \le K$,

$$\begin{aligned} \left\|S^\ell\right\|_2 &\le \|S\|_2^\ell \le 1, \\ \left\|(\Pi S)^\ell \Pi\right\|_2 &\le \|\Pi S\|_2^\ell \, \|\Pi\|_2 \le \|\Pi\|_2^{\ell+1} \, \|S\|_2^\ell \le 1. \end{aligned}$$

Using $\|AX\|_F \le \|A\|_2 \|X\|_F$ and $\|X\|_F = 1$, these bounds imply

$$\|S^\ell X\|_F \le \|S^\ell\|_2 \|X\|_F \le 1,$$
$$\|(\Pi S)^\ell \Pi X\|_F \le \|(\Pi S)^\ell \Pi\|_2 \|X\|_F \le 1.$$

Applying Lemma 1 and the triangle inequality,

$$\begin{aligned}\|\Gamma_\ell(\pi)X\|_F &= \|S^\ell X - (\Pi S)^\ell \Pi X\|_F \\ &\le \|S^\ell X\|_F + \|(\Pi S)^\ell \Pi X\|_F \le 1 + 1 = 2.\end{aligned}$$

Taking the maximum over ℓ gives $\mathcal{E}^\star(\pi) \le 2$. □

Theorem 1. *Let $\|S\|_2 \le 1$ and $1 \le k < n$.*

(i) *For every coarsening π of G into k supernodes,*

$$\mathcal{E}(\pi)/(K+1) \;\le\; \max_{0\le\ell\le K} \delta_\ell(\pi)^2 \;\le\; \mathcal{E}^\star(\pi)^2 \;\le\; \mathcal{E}(\pi). \tag{6}$$

(ii) *For every such coarsening π, every $\ell \ge 0$, and every $Z \in \mathbb{R}^{k\times d}$,*

$$\|S^\ell X - R^+ Z\|_F^2 \;=\; \delta_\ell(\pi)^2 + \left\|H^{1/2}\big(RS^\ell X - Z\big)\right\|_F^2 \;\ge\; \delta_\ell(\pi)^2. \tag{7}$$

In particular, the bound holds with $Z = (S'_c)^\ell RX$ for every coarse operator $S'_c \in \mathbb{R}^{k\times k}$.

(iii) *Let $\pi^\sharp$ and $\pi^\star$ minimize $\mathcal{E}$ and $\mathcal{E}^\star$, respectively, over the coarsenings of G into k supernodes. Then $\mathcal{E}^\star(\pi^\sharp) \le \sqrt{K+1}\,\mathcal{E}^\star(\pi^\star)$.*

Proof. Let π be a coarsening of G into k supernodes. We first prove (ii), which also gives the lower bound in (i).

Part (ii). For every $\ell \ge 0$ and $Z \in \mathbb{R}^{k\times d}$, splitting $S^\ell X$ using $I = \Pi^\perp + \Pi$ and then substituting $\Pi = R^+R$ gives

$$\begin{aligned}S^\ell X - R^+ Z &= \Pi^\perp S^\ell X + \Pi S^\ell X - R^+ Z \\ &= \Pi^\perp S^\ell X + R^+ RS^\ell X - R^+ Z \\ &= \Pi^\perp S^\ell X + R^+\big(RS^\ell X - Z\big).\end{aligned}$$

The columns of the two summands lie in the orthogonal subspaces $\mathrm{range}(\Pi^\perp)$ and $\mathrm{range}(R^+) = \mathrm{range}(\Pi)$. Expanding the squared Frobenius norm, with $\langle A, B\rangle_F = \mathrm{tr}(A^\top B)$, gives

$$\begin{aligned}\|S^\ell X - R^+ Z\|_F^2 &= \|\Pi^\perp S^\ell X\|_F^2 + \|R^+\big(RS^\ell X - Z\big)\|_F^2 \\ &\quad + 2\underbrace{\langle \Pi^\perp S^\ell X, R^+\big(RS^\ell X - Z\big)\rangle_F}_{=0} \\ &= \delta_\ell(\pi)^2 + \|R^+\big(RS^\ell X - Z\big)\|_F^2,\end{aligned}$$

where the cross term vanishes by orthogonality and $\delta_\ell(\pi) = \|\Pi^\perp S^\ell X\|_F$ by definition. Since H is diagonal with positive entries, $H = (H^{1/2})^\top H^{1/2}$. Using $\|B\|_F^2 = \mathrm{tr}(B^\top B)$, $R^+ = P$, and $P^\top P = H$, we obtain

$$\begin{aligned}\|R^+\big(RS^\ell X - Z\big)\|_F^2 &= \|P\big(RS^\ell X - Z\big)\|_F^2 \\ &= \mathrm{tr}\Big(\big(RS^\ell X - Z\big)^\top P^\top P\big(RS^\ell X - Z\big)\Big) \\ &= \mathrm{tr}\Big(\big(RS^\ell X - Z\big)^\top H\big(RS^\ell X - Z\big)\Big) \\ &= \mathrm{tr}\Big(\big[H^{1/2}\big(RS^\ell X - Z\big)\big]^\top \big[H^{1/2}\big(RS^\ell X - Z\big)\big]\Big) \\ &= \left\|H^{1/2}\big(RS^\ell X - Z\big)\right\|_F^2.\end{aligned}$$

Together with the squared-norm expansion above, this proves Equation (7), whose second term is nonnegative. Choosing $Z = (S'_c)^\ell RX$ gives the stated special case for every coarse operator S'_c.

Part (i). For $0 \le \ell \le K$, let

$$e_\ell := S^\ell X - R^+ S_c^\ell RX = S^\ell X - (\Pi S)^\ell \Pi X,$$

where the equality follows from Lemma 1. By Proposition 1, $\mathcal{E}^\star(\pi) = \max_{0\le\ell\le K} \|e_\ell\|_F$, and part (ii), with $Z = S_c^\ell RX$, gives $\delta_\ell(\pi)^2 \le \|e_\ell\|_F^2$. Since $\mathcal{E}(\pi)/(K+1)$ is the average of the $K+1$ residual squares,

$$\mathcal{E}(\pi)/(K+1) \;\le\; \max_{0\le\ell\le K} \delta_\ell(\pi)^2 \;\le\; \max_{0\le\ell\le K} \|e_\ell\|_F^2 \;=\; \mathcal{E}^\star(\pi)^2.$$

This proves the first two inequalities in Equation (6).

For the upper bound, note that $e_0 = \Pi^\perp X$. For $1 \le \ell \le K$, the definition at order $\ell - 1$ is

$$e_{\ell-1} = S^{\ell-1}X - (\Pi S)^{\ell-1}\Pi X.$$

Rearranging gives $(\Pi S)^{\ell-1}\Pi X = S^{\ell-1}X - e_{\ell-1}$. Substituting this into the expression for e_ℓ, we obtain

$$\begin{aligned} e_\ell &= S^\ell X - \Pi S(\Pi S)^{\ell-1}\Pi X \\ &= S^\ell X - \Pi S\big(S^{\ell-1}X - e_{\ell-1}\big) \\ &= \Pi^\perp S^\ell X + \Pi S\, e_{\ell-1}. \end{aligned} \tag{21}$$

Again, the columns of the two summands lie in the orthogonal ranges of $\Pi^\perp$ and Π. Moreover, $\|\Pi S\|_2 \le \|\Pi\|_2 \|S\|_2 \le 1$. Hence

$$\begin{aligned} \|e_\ell\|_F^2 &= \delta_\ell(\pi)^2 + \|\Pi S\, e_{\ell-1}\|_F^2 \\ &\le \delta_\ell(\pi)^2 + \|\Pi S\|_2^2 \, \|e_{\ell-1}\|_F^2 \\ &\le \delta_\ell(\pi)^2 + \|e_{\ell-1}\|_F^2 . \end{aligned}$$

Induction from $\|e_0\|_F^2 = \delta_0(\pi)^2$ gives

$$\|e_\ell\|_F^2 \;\le\; \sum_{m=0}^{\ell} \delta_m(\pi)^2 \;\le\; \sum_{m=0}^{K} \delta_m(\pi)^2 \;=\; \mathcal{E}(\pi) \qquad (0 \le \ell \le K).$$

Taking the maximum over ℓ proves $\mathcal{E}^\star(\pi)^2 \le \mathcal{E}(\pi)$.

Part (iii). Recall that $\pi^\sharp$ minimizes $\mathcal{E}$, whereas $\pi^\star$ minimizes $\mathcal{E}^\star$, over the same set of coarsenings of G into k supernodes. In particular, $\pi^\star$ is admissible in the minimization of $\mathcal{E}$. Thus

$$\begin{aligned} \mathcal{E}^\star(\pi^\sharp)^2 &\le \mathcal{E}(\pi^\sharp) && \text{by the upper bound in (i),} \\ &\le \mathcal{E}(\pi^\star) && \text{by minimality of } \pi^\sharp \text{ for } \mathcal{E}, \\ &\le (K+1)\,\mathcal{E}^\star(\pi^\star)^2 && \text{by the lower bound in (i).} \end{aligned}$$

The last step multiplies $\mathcal{E}(\pi^\star)/(K+1) \le \mathcal{E}^\star(\pi^\star)^2$ by $K+1$. Taking square roots gives

$$\mathcal{E}^\star(\pi^\sharp) \;\le\; \sqrt{K+1}\,\mathcal{E}^\star(\pi^\star),$$

which proves (iii). □

Part (ii) uses only the orthogonal-projector identities; it requires neither $\|S\|_2 \le 1$ nor a specific coarse operator. The upper bound in (i) uses $S_c = RSR^+$ and $\|S\|_2 \le 1$, but requires neither symmetry of S nor invariance of $\mathrm{range}(P)$ under S.

Corollary 1. *Let* $\|S\|_2 \le 1$. *Then* $\max_{0\le\ell\le K} \delta_\ell(\pi) \le d_{\mathrm{H}}(\pi) \le \mathcal{E}^\star(\pi)$, *hence* $\mathcal{E}(\pi)/(K+1) \le d_{\mathrm{H}}(\pi)^2 \le \mathcal{E}(\pi)$; *and* $\mathcal{E}(\pi) = 0$ *if and only if* $\mathcal{O}_G = \mathcal{O}_c^\uparrow$.

Proof. For every $W \in \mathcal{W}$, the fine and lifted coarse outputs with the same coefficients satisfy

$$\left\| \sum_{\ell=0}^{K} S^\ell X W_\ell - R^+ \sum_{\ell=0}^{K} S_c^\ell R X W_\ell \right\|_F = \mathcal{T}_\pi(W) \le \mathcal{E}^\star(\pi).$$

In each directed distance, let W specify the first output and let $\widetilde{W} \in \mathcal{W}$ vary over the coefficients of the second output. Choosing $\widetilde{W} = W$ in each infimum gives

$$\begin{aligned} \vec{\mathrm{dist}}(\mathcal{O}_G, \mathcal{O}_c^\uparrow) &= \sup_{W \in \mathcal{W}} \inf_{\widetilde{W} \in \mathcal{W}} \left\| \sum_{\ell=0}^{K} \big(S^\ell X W_\ell - R^+ S_c^\ell R X \widetilde{W}_\ell \big) \right\|_F \\ &\le \sup_{W \in \mathcal{W}} \mathcal{T}_\pi(W) = \mathcal{E}^\star(\pi), \\ \vec{\mathrm{dist}}(\mathcal{O}_c^\uparrow, \mathcal{O}_G) &= \sup_{W \in \mathcal{W}} \inf_{\widetilde{W} \in \mathcal{W}} \left\| \sum_{\ell=0}^{K} \big(R^+ S_c^\ell R X W_\ell - S^\ell X \widetilde{W}_\ell \big) \right\|_F \\ &\le \sup_{W \in \mathcal{W}} \mathcal{T}_\pi(W) = \mathcal{E}^\star(\pi). \end{aligned}$$

Taking the maximum of the two directed distances yields $d_{\mathrm{H}}(\pi) \le \mathcal{E}^\star(\pi)$ by Definition 3.

For each $0 \le \ell \le K$, choosing $W_\ell = I_d$ and all other blocks zero is admissible and shows that $S^\ell X \in \mathcal{O}_G$. Let $Y \in \mathcal{O}_c^\uparrow \subseteq \mathbb{R}^{n \times d}$ be an arbitrary lifted coarse output. By the definition of $\mathcal{O}_c^\uparrow$, there is some $\widetilde{W} \in \mathcal{W}$ such that

$$Y = R^+ \sum_{j=0}^{K} S_c^j R X \widetilde{W}_j.$$

The sum before multiplication by R^+ lies in $\mathbb{R}^{k \times d}$, so Theorem 1(ii) gives $\left\| S^\ell X - Y \right\|_F \ge \delta_\ell(\pi)$. Taking the infimum over all lifted coarse outputs preserves this lower bound:

$$\inf_{Y \in \mathcal{O}_c^\uparrow} \left\| S^\ell X - Y \right\|_F \;\ge\; \delta_\ell(\pi).$$

This infimum is the distance from $S^\ell X$ to $\mathcal{O}_c^\uparrow$. The directed distance $\vec{\mathrm{dist}}(\mathcal{O}_G, \mathcal{O}_c^\uparrow)$ is the supremum of these distances over all fine outputs in $\mathcal{O}_G$. Since $S^\ell X \in \mathcal{O}_G$, we have

$$\vec{\mathrm{dist}}(\mathcal{O}_G, \mathcal{O}_c^\uparrow) \;\ge\; \inf_{Y \in \mathcal{O}_c^\uparrow} \left\| S^\ell X - Y \right\|_F \;\ge\; \delta_\ell(\pi).$$

Finally, the Hausdorff distance is the maximum of the two directed distances, so $d_{\mathrm{H}}(\pi) \ge \vec{\mathrm{dist}}(\mathcal{O}_G, \mathcal{O}_c^\uparrow) \ge \delta_\ell(\pi)$. Taking the maximum over $0 \le \ell \le K$ proves the lower bound.

Combining these bounds with Theorem 1(i), we obtain

$$\mathcal{E}(\pi)/(K+1) \;\le\; \max_{0 \le \ell \le K} \delta_\ell(\pi)^2 \;\le\; d_{\mathrm{H}}(\pi)^2 \;\le\; \mathcal{E}^\star(\pi)^2 \;\le\; \mathcal{E}(\pi).$$

In particular, $\mathcal{E}(\pi) = 0$ if and only if $d_{\mathrm{H}}(\pi) = 0$. The two output sets are nonempty and compact, being linear images of the nonempty compact coefficient set $\mathcal{W}$. Hence $d_{\mathrm{H}}(\pi) = 0$ if and only if $\mathcal{O}_G = \mathcal{O}_c^\uparrow$, which proves the final claim. □

Corollary 2. *Let π be a coarsening of G with coarse operator $S_c = RSR^+$, and assume $S = S^\top$, $\|S\|_2 \le 1$ and $\|X\|_F = 1$. For every integer $j \ge 0$, let $M_j := X^\top S^j X$ and $M_j^c := (RX)^\top H S_c^j R X$ be the moments of Equation* (19)*. Then the following hold.*

(i) *For $0 \le j \le K$,* $\left\| M_j - M_j^c \right\|_F \;\le\; \mathcal{E}^\star(\pi)$.

(ii) *For $K < j \le 2K+1$,* $\left\| M_j - M_j^c \right\|_F \;\le\; 2\,\mathcal{E}^\star(\pi)$.

(iii) *The degree range in* (ii) *is sharp: for every $K \ge 0$ and $d \ge 1$ there exist S, X and π satisfying the assumptions above with $\mathcal{E}^\star(\pi) = 0$ and $M_{2K+2} \ne M_{2K+2}^c$.*

Proof. For each integer $\ell \geq 0$, let $\widehat{X}_\ell := R^+ S_c^\ell R X \in \mathbb{R}^{n\times d}$ denote the lifted coarse features. By Lemma 1, $\widehat{X}_\ell = (\Pi S)^\ell \Pi X$. Proposition 1 and the contraction bounds in its proof give

$$\left\|S^\ell X\right\|_F, \ \left\|\widehat{X}_\ell\right\|_F \leq 1 \quad (\ell \geq 0), \qquad \left\|S^\ell X - \widehat{X}_\ell\right\|_F \leq \mathcal{E}^\star(\pi) \quad (0 \leq \ell \leq K).$$

For $0 \leq j \leq K$, the identity $R^\top H = P = R^+$ gives $M_j^c = X^\top R^\top H S_c^j R X = X^\top \widehat{X}_j$. Hence

$$\begin{aligned}\left\|M_j - M_j^c\right\|_F &= \left\|X^\top\big(S^j X - \widehat{X}_j\big)\right\|_F \\ &\leq \|X\|_2 \left\|S^j X - \widehat{X}_j\right\|_F \leq \mathcal{E}^\star(\pi).\end{aligned}$$

The last step uses $\|X\|_2 \leq \|X\|_F = 1$.

For the remaining degrees, symmetry gives a bilinear representation of the moments. Since $S = S^\top$ and $S_c = H^{-1}P^\top S P$, we have

$$S_c^\top H = H S_c, \qquad P^\top S^b P = H S_c^b \quad (b \in \{0,1\}).$$

The first identity implies $(S_c^a)^\top H = H S_c^a$. Thus, for every integer $a \geq 0$ and $b \in \{0,1\}$,

$$\begin{aligned}\widehat{X}_a^\top S^b \widehat{X}_a &= (RX)^\top (S_c^a)^\top P^\top S^b P S_c^a R X \\ &= (RX)^\top H S_c^{2a+b} R X = M_{2a+b}^c.\end{aligned}$$

Likewise, symmetry of S gives $M_{2a+b} = (S^a X)^\top S^b (S^a X)$.

Now let $K < j \leq 2K+1$ and write $j = 2a+b$, where $a = \lfloor j/2 \rfloor \leq K$ and $b \in \{0,1\}$. Adding and subtracting $\widehat{X}_a^\top S^b (S^a X)$ gives

$$\begin{aligned}M_j - M_j^c &= \big(S^a X - \widehat{X}_a\big)^\top S^b (S^a X) \\ &\quad + \widehat{X}_a^\top S^b \big(S^a X - \widehat{X}_a\big).\end{aligned}$$

Using the triangle inequality, $\left\|A^\top B\right\|_F \leq \|A\|_F \|B\|_F$, and $\left\|S^b\right\|_2 \leq 1$, we conclude that

$$\begin{aligned}\left\|M_j - M_j^c\right\|_F &\leq \left\|S^a X - \widehat{X}_a\right\|_F \left(\|S^a X\|_F + \left\|\widehat{X}_a\right\|_F\right) \\ &\leq \mathcal{E}^\star(\pi)(1+1) = 2\,\mathcal{E}^\star(\pi).\end{aligned}$$

To prove sharpness, fix $K \geq 0$ and $d \geq 1$. Let S be the self-loop GCN operator of the path $v_1, \ldots, v_{K+3}$. Merge its last two vertices and leave every other vertex as a singleton supernode. Let the first column of X be the indicator of v_1 and the remaining columns be zero. These choices satisfy $S = S^\top$, $\|S\|_2 = 1$, and $\|X\|_F = 1$. For $0 \leq \ell \leq K$, the propagated features vanish on the merged pair and are therefore constant on every supernode. Hence $\mathcal{E}(\pi) = 0$ and $\mathcal{E}^\star(\pi) = 0$ by Theorem 1(i). In particular, $\widehat{X}_K = S^K X$, and therefore

$$\widehat{X}_{K+1} = \Pi S \widehat{X}_K = \Pi S^{K+1} X.$$

The moment identities above, with $a = K+1$ and $b = 0$, now give

$$\begin{aligned}M_{2K+2} - M_{2K+2}^c &= (S^{K+1}X)^\top (I - \Pi) S^{K+1} X \\ &= (\Pi^\perp S^{K+1} X)^\top (\Pi^\perp S^{K+1} X).\end{aligned}$$

At order $K+1$, the first feature is positive at v_{K+2} and zero at v_{K+3}. Thus $\Pi^\perp S^{K+1} X \neq 0$, so the displayed Gram matrix is nonzero. This proves $M_{2K+2} \neq M_{2K+2}^c$. □

E Experiments: datasets, hyperparameters, full tables

E.1 Datasets

Table 7 lists the nine graphs. Directed sources are symmetrized once at load and the same graph serves coarsening, training and evaluation, since the spectral baselines require a symmetric Laplacian; edge counts therefore count each undirected edge once. For ogbn-mag we use the paper–paper citation subgraph. Splits are those released with each benchmark and are never resampled.

Table 7: Datasets: n nodes, m undirected edges, d features, C classes, average degree $\bar{d} = 2m/n$, edge homophily h_{edge} and adjusted homophily h_{adj} (Platonov et al., 2023), negative under genuine heterophily; blocks split at $h_{\mathrm{adj}} = 0.25$. Splits are the official train/validation/test splits (%). Homophily is computed over labeled nodes.

Dataset	n	m	d	C	$\bar{d}$	h_{edge}	h_{adj}	Split	Source
Homophilic									
ogbn-products	2,449,029	61,859,076	100	47	50.5	0.81	0.79	8/2/90	Hu et al. (2020)
Reddit2	232,965	11,606,919	602	41	99.6	0.78	0.77	66/10/24	Zeng et al. (2020)
ogbn-arxiv	169,343	1,157,799	128	40	13.7	0.65	0.59	54/18/29	Hu et al. (2020)
IGB-small	1,000,000	6,034,065	1024	19	12.1	0.57	0.46	60/20/20	Khatua et al. (2023)
ogbn-mag	736,389	5,396,336	128	349	14.7	0.30	0.28	85/9/6	Hu et al. (2020)
Heterophilic									
twitch-gamers	168,114	6,797,557	7	2	80.9	0.55	0.09	50/25/25	Lim et al. (2021)
arxiv-year	169,343	1,157,799	128	5	13.7	0.22	0.01	50/25/25	Lim et al. (2021)
snap-patents	2,923,922	13,972,546	269	5	9.6	0.22	-0.00	50/25/25	Lim et al. (2021)
pokec	1,632,803	22,301,964	65	2	27.3	0.42	-0.15	50/25/25	Lim et al. (2021)

E.2 Training configuration

Every backbone is trained with one configuration for every dataset, coarsener and ratio; nothing was tuned per method. GCN, GraphSAGE and SIGN share a common configuration, and SGC and GPR-GNN use the settings of their original papers (Wu et al., 2019; Chien et al., 2021). Models are trained on the coarsened graph and evaluated on the original graph.

GCN and GraphSAGE have three layers, SGC and SIGN three propagation hops, and GPR-GNN ten. The hidden width is 256 with BatchNorm after each hidden layer for GCN, GraphSAGE and SIGN; SGC has no hidden layer, and GPR-GNN uses no normalization. Dropout is 0.5 except for SGC, which uses none. All models are trained with Adam without weight decay at learning rate 0.01 (0.2 for SGC), for 300 epochs (GCN, GraphSAGE, SIGN), 500 (SGC) or 1000 (GPR-GNN), minimizing the cross-entropy of the log-softmax outputs. Validation accuracy on the original graph is computed every five epochs and the epoch with the best value is selected. Each cell uses one coarsening and five training seeds (0–4) for GCN, GraphSAGE and SIGN, three (0–2) for SGC and GPR-GNN.

GraphSAGE uses mean aggregation. SIGN passes hops 0–3 each through a linear layer, BatchNorm and a ReLU, concatenates them and applies a linear classifier. GPR-GNN applies a two-layer MLP to the features and then propagates its output with learned coefficients over hops 0–10, initialized as personalized PageRank with $\alpha = 0.1$.

E.3 Default configuration of the coarsener

Our coarsener runs with one configuration for every dataset, ratio and backbone; nothing was tuned per dataset. The propagation operator is the renormalized adjacency $\hat{A} = \tilde{D}^{-1/2}(A + I)\tilde{D}^{-1/2}$ of Section 2.1, the trajectory has degree $K = 3$ (Equation (1)) in the monomial basis (Equation (12)), and the relative ridge is $\lambda_{\mathrm{rel}} = 10^{-4}$ (Equation (9)) with discarded ridge energy $\tau_q = 0.05$ (Equation (18)). The candidate search of Algorithm 2 uses width $b = 8$, depth $h = 2$ and escape ratio $\zeta = 0.5$ (Section 3.2); each level runs $J = 5$ pairing rounds (Algorithm 1); and the edge budget is set by $\nu = 0.3$ (Equation (11)).

E.4 Sensitivity to the budget, search width, ridge and pairing rounds

Figure 5 extends the sweep of Section 4.4, under the same protocol, to the budget parameter ν, the search width b, the relative ridge λ_{rel} and the number of pairing rounds J. The budget parameter is the only one of the four that moves accuracy beyond seed noise: raising ν from 0 to 1 lowers the average degree of the coarsened graph from 9–46 to 2.5–6.6, with the effect on each dataset reported in Section 4.4. The search width leaves accuracy within 1.2 points of the default from $b = 4$ to 16, while coarsening time grows 4–12$\times$ (ogbn-products: 47 to 580 s). For the ridge, in

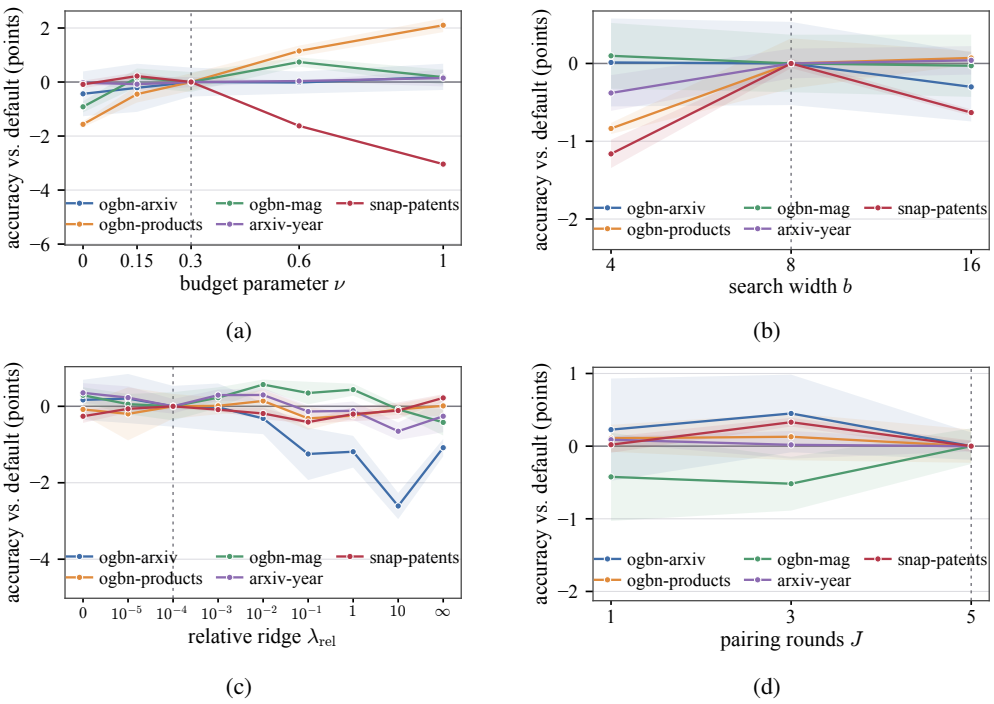


Figure 5: Sensitivity of GCN accuracy at 90% reduction ratio to the remaining parameters of our method: *(a)* the budget parameter ν, *(b)* the search width b, *(c)* the relative ridge λ_{rel}, where ∞ denotes the unscaled objective $\mathcal{E}$, and *(d)* the number of pairing rounds J. Each line is one dataset and reports the difference in accuracy from the default setting; shaded bands show one standard deviation over training seeds. The protocol is that of Figure 3(b,c).

Table 8: Coarsening wall-clock time at ratio 90, in minutes except for the last column. Every column but the last follows the controlled CPU protocol of Figure 2a: each run alone, with a six-hour limit. *Ours (GPU)* is the coarsening time in seconds recorded on one A100. OOT: exceeded the six-hour limit; OOM: terminated for memory; ‡: the method did not reach the $90 \pm 3\%$ tolerance on the achieved ratio (Table 1).

Dataset	A-ConvMatch	Random	SGBGC	STPGC	UGC	Var-Edges	Var-Neigh	Ours (CPU)	Ours (GPU, s)
igb-small	14.2	0.58	17.0	25.8	1.2	128.6	96.2	4.6	47.6
ogbn-arxiv	1.7	0.29	6.4	2.1	0.27	7.7	4.1	0.70	2.6
ogbn-mag	5.6	0.40	9.7	10.7	0.68	43.7	28.4	2.5	6.0
ogbn-products	81.2	6.8	138.5	252.1	4.2	‡	OOT	27.3	47.7
reddit2	5.8	0.53	2.0	68.7	0.55	71.1	45.6	5.5	54.2
arxiv-year	1.6	0.29	0.82	1.9	0.27	7.7	4.1	0.70	3.4
pokec	21.9	2.5‡	88.8	97.4	1.7	231.3	70.9	2.7	7.3
snap-patents	OOM	2.6	138.3	51.1	2.7	‡	‡	13.7	25.0
twitch-gamers	1.9	0.75	9.8	26.4	0.31	‡	56.1	0.38	1.7

addition to the range discussed in the main text, the unscaled objective ($\lambda_{\text{rel}} = \infty$) stays within 1.1 points of the default on every dataset. The number of pairing rounds changes accuracy by less than one point on every dataset and mainly sets the running time, which falls 2–3× from $J = 1$ to $J = 5$ (ogbn-products: 288 to 96 s).

E.5 Full results

Full results. Tables 9–12 report all five backbones of Appendix E.2 at reduction ratios 80, 90, 95 and 99%, in the format of Table 1, with one block of five backbone rows per dataset; Table 1 is the GCN row of each block in Table 10. The OOT, OOM and ‡ markers are given once per block, since the coarsened graph is shared by all backbones.

Table 9: Test accuracy (%) of every backbone on the coarsened graph at 80% reduction ratio, evaluated on the original graph. Best in **bold**, second best underlined, ranked among the methods with a result. ‡ marks a coarsening whose achieved ratio missed the $80 \pm 3\%$ tolerance, OOT and OOM one that exceeded the time limit or ran out of memory.

Backbone	Full graph	A-ConvMatch	SGBGC	UGC	Var-Edges	Var-Neigh	Ours
arxiv-year							
GCN	$45.97_{\pm.13}$	$39.42_{\pm1.44}$	$35.52_{\pm.25}$	$32.55_{\pm1.26}$	$39.56_{\pm.10}$	$39.86_{\pm.35}$	$\mathbf{42.12}_{\pm.07}$
GraphSAGE	$47.47_{\pm.45}$	$39.75_{\pm.63}$	$36.48_{\pm.99}$	$29.40_{\pm.56}$	$38.13_{\pm.30}$	$37.66_{\pm.82}$	$\mathbf{41.21}_{\pm.10}$
SGC	$39.47_{\pm.11}$	$37.14_{\pm.25}$	$36.97_{\pm.64}$	$36.09_{\pm.10}$	$38.35_{\pm.05}$	$37.99_{\pm.04}$	$\mathbf{38.91}_{\pm.10}$
SIGN	$44.64_{\pm.26}$	$\mathbf{40.07}_{\pm.23}$	$35.52_{\pm1.14}$	$37.18_{\pm.34}$	$38.65_{\pm.11}$	$38.30_{\pm.35}$	$39.83_{\pm.20}$
GPR-GNN	$41.95_{\pm.07}$	$39.35_{\pm.12}$	$37.46_{\pm.15}$	$35.63_{\pm.84}$	$40.22_{\pm.02}$	$39.78_{\pm.03}$	$\mathbf{40.68}_{\pm.19}$
igb-small							
GCN	$72.31_{\pm.01}$	$70.43_{\pm.07}$	$71.22_{\pm.05}$	$68.03_{\pm.42}$	$70.82_{\pm.08}$	$70.30_{\pm.13}$	$\mathbf{71.99}_{\pm.04}$
GraphSAGE	$75.49_{\pm.07}$	$72.57_{\pm.20}$	$73.97_{\pm.04}$	$69.31_{\pm.35}$	$73.97_{\pm.08}$	$73.19_{\pm.12}$	$\mathbf{74.23}_{\pm.08}$
SGC	$71.69_{\pm.00}$	$69.99_{\pm.00}$	$70.68_{\pm.00}$	$61.35_{\pm.01}$	$69.93_{\pm.01}$	$68.92_{\pm.01}$	$\mathbf{71.45}_{\pm.00}$
SIGN	$76.01_{\pm.03}$	$73.86_{\pm.06}$	$\mathbf{74.98}_{\pm.04}$	$72.35_{\pm.08}$	$74.55_{\pm.03}$	$74.12_{\pm.04}$	$74.40_{\pm.03}$
GPR-GNN	$75.50_{\pm.03}$	$73.75_{\pm.04}$	$\mathbf{74.96}_{\pm.03}$	$73.16_{\pm.02}$	$74.53_{\pm.03}$	$73.90_{\pm.02}$	$74.72_{\pm.05}$
ogbn-arxiv							
GCN	$71.90_{\pm.17}$	$66.60_{\pm.51}$	$66.73_{\pm.55}$	$53.01_{\pm1.82}$	$69.13_{\pm.43}$	$67.74_{\pm.53}$	$\mathbf{69.87}_{\pm.14}$
GraphSAGE	$71.73_{\pm.28}$	$\mathbf{66.73}_{\pm.60}$	$62.42_{\pm.48}$	$55.75_{\pm1.44}$	$62.74_{\pm.61}$	$63.12_{\pm.63}$	$66.48_{\pm.49}$
SGC	$68.60_{\pm.04}$	$66.58_{\pm.04}$	$64.09_{\pm.23}$	$46.85_{\pm2.10}$	$65.94_{\pm.10}$	$65.52_{\pm.16}$	$\mathbf{67.71}_{\pm.07}$
SIGN	$71.44_{\pm.13}$	$66.31_{\pm.22}$	$64.05_{\pm.37}$	$60.74_{\pm.23}$	$65.65_{\pm.47}$	$65.37_{\pm.42}$	$\mathbf{67.49}_{\pm.32}$
GPR-GNN	$69.05_{\pm.30}$	$\mathbf{68.91}_{\pm.20}$	$63.72_{\pm.73}$	$57.77_{\pm.45}$	$68.77_{\pm.20}$	$67.37_{\pm.15}$	$68.76_{\pm.37}$
ogbn-mag							
GCN	$35.64_{\pm.12}$	$31.89_{\pm.51}$	$29.17_{\pm.69}$	$20.24_{\pm.32}$	$33.10_{\pm.48}$	$28.99_{\pm.47}$	$\mathbf{35.85}_{\pm.33}$
GraphSAGE	$35.81_{\pm.15}$	$31.16_{\pm.61}$	$30.26_{\pm.45}$	$19.70_{\pm1.04}$	$30.66_{\pm.45}$	$28.32_{\pm.61}$	$\mathbf{33.73}_{\pm.24}$
SGC	$33.85_{\pm.45}$	$30.12_{\pm.08}$	$29.35_{\pm.10}$	$15.46_{\pm.89}$	$31.07_{\pm.12}$	$29.83_{\pm.23}$	$\mathbf{32.51}_{\pm.08}$
SIGN	$35.97_{\pm.10}$	$30.70_{\pm.23}$	$31.41_{\pm.33}$	$27.62_{\pm.37}$	$32.90_{\pm.29}$	$31.05_{\pm.23}$	$\mathbf{33.60}_{\pm.22}$
GPR-GNN	$31.40_{\pm.58}$	$30.73_{\pm.50}$	$28.38_{\pm.20}$	$22.91_{\pm.03}$	$\mathbf{32.27}_{\pm.29}$	$29.49_{\pm.15}$	$32.15_{\pm.43}$
ogbn-products							
GCN	$76.67_{\pm.68}$	$74.30_{\pm.23}$		$70.67_{\pm.53}$			$\mathbf{76.22}_{\pm.39}$
GraphSAGE	$79.61_{\pm.20}$	$74.13_{\pm.17}$		$67.77_{\pm.24}$			$\mathbf{75.14}_{\pm.28}$
SGC	$65.15_{\pm.00}$	$64.98_{\pm.05}$	OOT	$62.29_{\pm.15}$	OOT	OOT	$\mathbf{66.42}_{\pm.00}$
SIGN	$75.16_{\pm.06}$	$72.05_{\pm.05}$		$67.74_{\pm.33}$			$\mathbf{73.97}_{\pm.12}$
GPR-GNN	$73.65_{\pm.68}$	$73.97_{\pm.20}$		$67.08_{\pm.29}$			$\mathbf{75.21}_{\pm.31}$
pokec							
GCN	$72.50_{\pm.44}$	$51.52_{\pm.29}$	$51.52_{\pm.29}$	$54.19_{\pm1.58}$	$51.90_{\pm.08}$	$52.02_{\pm.20}$	$\mathbf{61.23}_{\pm2.34}$
GraphSAGE	$80.40_{\pm.24}$	$\mathbf{75.25}_{\pm.73}$	$73.03_{\pm2.01}$	$73.03_{\pm.37}$	$59.56_{\pm3.31}$	$66.67_{\pm3.21}$	$72.95_{\pm.41}$
SGC	$54.88_{\pm.20}$	$\mathbf{51.77}_{\pm.97}$	$50.78_{\pm.22}$	$51.35_{\pm.61}$	$51.01_{\pm.09}$	$51.01_{\pm.00}$	$50.66_{\pm.00}$
SIGN	$73.81_{\pm.53}$	$62.36_{\pm.23}$	$62.13_{\pm.32}$	$64.34_{\pm.21}$	$60.02_{\pm.11}$	$62.90_{\pm.61}$	$\mathbf{66.61}_{\pm.37}$
GPR-GNN	$75.98_{\pm.53}$	$55.98_{\pm5.14}$	—	$51.93_{\pm1.18}$	$57.51_{\pm1.61}$	$\mathbf{66.06}_{\pm3.40}$	$58.99_{\pm1.51}$
reddit2							
GCN	$95.34_{\pm.01}$	$94.23_{\pm.07}$	$93.26_{\pm.11}$	$45.32_{\pm6.43}$	$\mathbf{94.63}_{\pm.04}$	$93.74_{\pm.11}$	$93.73_{\pm.07}$
GraphSAGE	$96.64_{\pm.07}$	$94.69_{\pm.15}$	$94.53_{\pm.19}$	$53.14_{\pm1.66}$	$94.45_{\pm.23}$	$94.77_{\pm.08}$	$\mathbf{94.87}_{\pm.03}$
SGC	$94.62_{\pm.03}$	$93.71_{\pm.08}$	$93.86_{\pm.04}$	$73.87_{\pm1.38}$	$\mathbf{94.20}_{\pm.01}$	$93.73_{\pm.04}$	$92.90_{\pm.19}$
SIGN	$96.30_{\pm.05}$	$94.27_{\pm.09}$	$93.64_{\pm.09}$	$80.90_{\pm.24}$	$94.63_{\pm.08}$	$95.06_{\pm.05}$	$\mathbf{95.15}_{\pm.06}$
GPR-GNN	$96.15_{\pm.02}$	$93.83_{\pm.07}$	$92.91_{\pm.04}$	$83.38_{\pm.17}$	$\mathbf{95.33}_{\pm.07}$	$94.43_{\pm.06}$	$93.87_{\pm.07}$
snap-patents							
GCN	$41.63_{\pm.04}$			$30.35_{\pm.32}$			$\mathbf{34.87}_{\pm.17}$
GraphSAGE	$44.11_{\pm.17}$			$28.44_{\pm1.00}$			$\mathbf{38.37}_{\pm.16}$
SGC	$35.23_{\pm.01}$	OOM	OOT	$26.13_{\pm.48}$	‡	‡	$\mathbf{31.41}_{\pm.19}$
SIGN	$42.22_{\pm.06}$			$32.71_{\pm.21}$			$\mathbf{38.33}_{\pm.06}$
GPR-GNN	$32.84_{\pm.18}$			$24.09_{\pm.74}$			$\mathbf{29.11}_{\pm.10}$
twitch-gamers							
GCN	$63.68_{\pm.10}$	$60.02_{\pm.90}$	$59.64_{\pm.82}$	$60.56_{\pm.30}$	$60.10_{\pm.37}$	$55.21_{\pm.89}$	$\mathbf{61.37}_{\pm.24}$
GraphSAGE	$64.53_{\pm.04}$	$56.64_{\pm.13}$	$55.63_{\pm1.25}$	$60.84_{\pm.14}$	$53.80_{\pm.90}$	$54.24_{\pm.33}$	$\mathbf{62.93}_{\pm.15}$
SGC	$60.60_{\pm.10}$	$\mathbf{59.42}_{\pm.37}$	$54.68_{\pm.77}$	$56.89_{\pm.81}$	$56.06_{\pm.51}$	$58.48_{\pm.40}$	$58.69_{\pm.47}$
SIGN	$63.43_{\pm.12}$	$\mathbf{62.15}_{\pm.10}$	$59.38_{\pm.84}$	$60.28_{\pm.28}$	$59.59_{\pm.69}$	$58.60_{\pm.78}$	$61.56_{\pm.19}$
GPR-GNN	$60.50_{\pm.56}$	$\mathbf{60.24}_{\pm.07}$	$59.64_{\pm.05}$	$58.44_{\pm1.73}$	$59.55_{\pm.09}$	$59.82_{\pm.19}$	$59.11_{\pm.17}$

Table 10: Test accuracy (%) of every backbone on the coarsened graph at 90% reduction ratio, evaluated on the original graph; the GCN row of each block reproduces Table 1. Best in **bold**, second best underlined, ranked among the methods with a result. ‡ marks a coarsening whose achieved ratio missed the $90 \pm 3\%$ tolerance, OOT and OOM one that exceeded the time limit or ran out of memory.

Backbone	Full graph	A-ConvMatch	SGBGC	UGC	Var-Edges	Var-Neigh	Ours
arxiv-year							
GCN	$45.97_{\pm.13}$	$40.59_{\pm.29}$	$34.69_{\pm.77}$	$32.54_{\pm1.41}$	$37.42_{\pm.68}$	$38.80_{\pm.33}$	$\mathbf{40.62}_{\pm.18}$
GraphSAGE	$47.47_{\pm.45}$	$38.13_{\pm.30}$	$37.10_{\pm.56}$	$28.82_{\pm.10}$	$37.68_{\pm.06}$	$36.92_{\pm.84}$	$\mathbf{38.30}_{\pm.32}$
SGC	$39.47_{\pm.11}$	$37.49_{\pm.03}$	$36.41_{\pm.19}$	$31.16_{\pm1.76}$	$38.15_{\pm.06}$	$37.71_{\pm.04}$	$\mathbf{38.70}_{\pm.05}$
SIGN	$44.64_{\pm.26}$	$\mathbf{39.00}_{\pm.11}$	$34.74_{\pm.85}$	$37.36_{\pm.22}$	$36.52_{\pm.15}$	$34.94_{\pm.94}$	$38.77_{\pm.26}$
GPR-GNN	$41.95_{\pm.07}$	$39.26_{\pm.07}$	$36.97_{\pm.13}$	$29.04_{\pm.08}$	$39.26_{\pm.02}$	$39.38_{\pm.12}$	$\mathbf{39.80}_{\pm.21}$
igb-small							
GCN	$72.31_{\pm.01}$	$70.10_{\pm.09}$	$70.78_{\pm.06}$	$64.53_{\pm.91}$	$70.23_{\pm.03}$	$69.53_{\pm.14}$	$\mathbf{71.82}_{\pm.03}$
GraphSAGE	$75.49_{\pm.07}$	$71.50_{\pm.12}$	$73.14_{\pm.11}$	$64.88_{\pm.69}$	$73.45_{\pm.07}$	$72.30_{\pm.25}$	$\mathbf{73.69}_{\pm.05}$
SGC	$71.69_{\pm.00}$	$69.76_{\pm.01}$	$70.33_{\pm.00}$	$51.50_{\pm.05}$	$69.97_{\pm.02}$	$68.85_{\pm.03}$	$\mathbf{71.34}_{\pm.00}$
SIGN	$76.01_{\pm.03}$	$73.53_{\pm.06}$	$74.04_{\pm.04}$	$71.22_{\pm.29}$	$73.99_{\pm.04}$	$73.85_{\pm.05}$	$\mathbf{74.04}_{\pm.04}$
GPR-GNN	$75.50_{\pm.03}$	$73.12_{\pm.03}$	$74.06_{\pm.02}$	$72.00_{\pm.11}$	$74.00_{\pm.01}$	$73.50_{\pm.08}$	$\mathbf{74.48}_{\pm.01}$
ogbn-arxiv							
GCN	$71.90_{\pm.17}$	$64.86_{\pm.57}$	$64.00_{\pm.41}$	$33.41_{\pm1.38}$	$66.01_{\pm.58}$	$66.04_{\pm.51}$	$\mathbf{68.93}_{\pm.11}$
GraphSAGE	$71.73_{\pm.28}$	$\mathbf{63.97}_{\pm.37}$	$61.67_{\pm.47}$	$43.56_{\pm2.76}$	$58.92_{\pm1.12}$	$57.41_{\pm.89}$	$63.22_{\pm.59}$
SGC	$68.60_{\pm.04}$	$65.41_{\pm.13}$	$59.46_{\pm.42}$	$36.09_{\pm.87}$	$63.66_{\pm.37}$	$63.45_{\pm.88}$	$\mathbf{67.27}_{\pm.05}$
SIGN	$71.44_{\pm.13}$	$63.38_{\pm.32}$	$60.17_{\pm.50}$	$55.64_{\pm.81}$	$62.38_{\pm.38}$	$60.09_{\pm.48}$	$\mathbf{64.02}_{\pm.29}$
GPR-GNN	$69.05_{\pm.30}$	$67.94_{\pm.24}$	$60.11_{\pm.51}$	$54.44_{\pm.39}$	$67.00_{\pm.23}$	$65.86_{\pm.10}$	$\mathbf{68.29}_{\pm.20}$
ogbn-mag							
GCN	$35.64_{\pm.12}$	$31.97_{\pm.39}$	$25.92_{\pm.54}$	$13.44_{\pm.96}$	$28.46_{\pm.32}$	$27.35_{\pm.31}$	$\mathbf{35.11}_{\pm.19}$
GraphSAGE	$35.81_{\pm.15}$	$29.82_{\pm.59}$	$27.30_{\pm.38}$	$14.36_{\pm1.23}$	$28.16_{\pm.38}$	$25.67_{\pm.42}$	$\mathbf{31.70}_{\pm.22}$
SGC	$33.85_{\pm.45}$	$30.24_{\pm.09}$	$27.77_{\pm.19}$	$12.50_{\pm.80}$	$30.15_{\pm.04}$	$29.35_{\pm.11}$	$\mathbf{32.41}_{\pm.27}$
SIGN	$35.97_{\pm.10}$	$29.29_{\pm.23}$	$28.73_{\pm.47}$	$24.57_{\pm.24}$	$31.46_{\pm.48}$	$29.17_{\pm.19}$	$\mathbf{32.21}_{\pm.45}$
GPR-GNN	$31.40_{\pm.58}$	$30.52_{\pm.23}$	$27.10_{\pm.13}$	$19.68_{\pm.46}$	$31.16_{\pm.23}$	$29.78_{\pm.11}$	$\mathbf{33.55}_{\pm.25}$
ogbn-products							
GCN	$76.67_{\pm.68}$	$72.99_{\pm.30}$	$66.53_{\pm.34}$	$65.24_{\pm.30}$			$\mathbf{76.43}_{\pm.20}$
GraphSAGE	$79.61_{\pm.20}$	$70.65_{\pm.17}$	$67.30_{\pm.24}$	$64.71_{\pm.41}$			$\mathbf{74.11}_{\pm.33}$
SGC	$65.15_{\pm.00}$	$64.07_{\pm.01}$	$48.79_{\pm.06}$	$49.34_{\pm.35}$	OOT	OOT	$\mathbf{66.20}_{\pm.01}$
SIGN	$75.16_{\pm.06}$	$70.58_{\pm.07}$	$63.26_{\pm.36}$	$62.64_{\pm.21}$			$\mathbf{73.06}_{\pm.12}$
GPR-GNN	$73.65_{\pm.68}$	$73.77_{\pm.35}$	$64.57_{\pm.57}$	$62.09_{\pm.66}$			$\mathbf{76.18}_{\pm.36}$
pokec							
GCN	$72.50_{\pm.44}$	$51.06_{\pm.26}$	$52.06_{\pm.34}$	$51.79_{\pm.15}$	$51.90_{\pm.21}$	$50.86_{\pm.12}$	$\mathbf{58.47}_{\pm1.03}$
GraphSAGE	$80.40_{\pm.24}$	$\mathbf{71.50}_{\pm1.81}$	$70.46_{\pm.78}$	$65.82_{\pm1.30}$	$57.87_{\pm2.14}$	$63.59_{\pm1.82}$	$70.22_{\pm.49}$
SGC	$54.88_{\pm.20}$	$51.44_{\pm1.11}$	$50.92_{\pm.39}$	$\mathbf{51.87}_{\pm.37}$	$51.07_{\pm.09}$	$50.87_{\pm.14}$	$51.22_{\pm.63}$
SIGN	$73.81_{\pm.53}$	$61.72_{\pm.15}$	$61.86_{\pm.22}$	$61.99_{\pm.16}$	$59.53_{\pm.29}$	$62.06_{\pm.30}$	$\mathbf{65.17}_{\pm.23}$
GPR-GNN	$75.98_{\pm.53}$	$52.11_{\pm1.46}$	$\mathbf{58.84}_{\pm.70}$	$51.88_{\pm1.14}$	$51.46_{\pm.57}$	$55.02_{\pm1.07}$	$56.96_{\pm.66}$
reddit2							
GCN	$95.34_{\pm.01}$	$93.80_{\pm.06}$	$91.54_{\pm.17}$	$25.39_{\pm2.45}$	$\mathbf{93.88}_{\pm.12}$	$91.52_{\pm.56}$	$93.55_{\pm.05}$
GraphSAGE	$96.64_{\pm.07}$	$92.20_{\pm.17}$	$88.81_{\pm.68}$	$33.98_{\pm1.26}$	$91.97_{\pm.14}$	$90.96_{\pm.82}$	$\mathbf{93.17}_{\pm.09}$
SGC	$94.62_{\pm.03}$	$93.56_{\pm.02}$	$92.25_{\pm.49}$	$38.34_{\pm1.17}$	$\mathbf{94.14}_{\pm.01}$	$93.11_{\pm.08}$	$92.74_{\pm.35}$
SIGN	$96.30_{\pm.05}$	$92.15_{\pm.30}$	$90.03_{\pm.13}$	$58.46_{\pm1.85}$	$93.46_{\pm.07}$	$92.37_{\pm.08}$	$\mathbf{94.21}_{\pm.07}$
GPR-GNN	$96.15_{\pm.02}$	—	$90.34_{\pm.13}$	$65.76_{\pm.34}$	$\mathbf{94.79}_{\pm.04}$	$92.33_{\pm.05}$	$91.45_{\pm.22}$
snap-patents							
GCN	$41.63_{\pm.04}$		$29.82_{\pm.12}$	$26.21_{\pm.58}$			$\mathbf{34.71}_{\pm.22}$
GraphSAGE	$44.11_{\pm.17}$		$31.29_{\pm.34}$	$26.49_{\pm.64}$			$\mathbf{36.65}_{\pm.32}$
SGC	$35.23_{\pm.01}$	OOM	$27.92_{\pm.92}$	$24.70_{\pm.28}$	‡	‡	$\mathbf{32.08}_{\pm.08}$
SIGN	$42.22_{\pm.06}$		$30.93_{\pm.16}$	$29.75_{\pm.21}$			$\mathbf{37.35}_{\pm.06}$
GPR-GNN	$32.84_{\pm.18}$		$27.45_{\pm.07}$	$22.34_{\pm.35}$			$\mathbf{28.74}_{\pm.13}$
twitch-gamers							
GCN	$63.68_{\pm.10}$	$60.73_{\pm.49}$	$59.66_{\pm.70}$	$58.65_{\pm.74}$		$59.08_{\pm.46}$	$\mathbf{61.54}_{\pm.13}$
GraphSAGE	$64.53_{\pm.04}$	$54.66_{\pm.29}$	$55.76_{\pm1.62}$	$59.68_{\pm.60}$		$53.81_{\pm.71}$	$\mathbf{61.82}_{\pm.27}$
SGC	$60.60_{\pm.10}$	$\mathbf{59.94}_{\pm.05}$	$55.17_{\pm.66}$	$58.16_{\pm.51}$	‡	$55.92_{\pm1.30}$	$58.63_{\pm.02}$
SIGN	$63.43_{\pm.12}$	$\mathbf{62.04}_{\pm.08}$	$58.91_{\pm.31}$	$60.14_{\pm.26}$		$55.67_{\pm2.69}$	$61.43_{\pm.09}$
GPR-GNN	$60.50_{\pm.56}$	$60.31_{\pm.13}$	$\mathbf{60.90}_{\pm.01}$	$59.54_{\pm.21}$		$56.16_{\pm.57}$	$60.11_{\pm.02}$

Table 11: Test accuracy (%) of every backbone on the coarsened graph at 95% reduction ratio, evaluated on the original graph. Best in **bold**, second best underlined, ranked among the methods with a result. ‡ marks a coarsening whose achieved ratio missed the $95 \pm 3\%$ tolerance, OOT and OOM one that exceeded the time limit or ran out of memory.

Backbone	Full graph	A-ConvMatch	SGBGC	UGC	Var-Edges	Var-Neigh	Ours
arxiv-year							
GCN	$45.97_{\pm.13}$	$39.63_{\pm.17}$	$34.81_{\pm.50}$	$29.16_{\pm.73}$	$35.32_{\pm1.32}$	$33.17_{\pm1.03}$	**40.19**$_{\pm.19}$
GraphSAGE	$47.47_{\pm.45}$	**38.15**$_{\pm.13}$	$36.55_{\pm.34}$	$34.91_{\pm.65}$	$35.97_{\pm.34}$	$34.08_{\pm1.81}$	$36.93_{\pm.43}$
SGC	$39.47_{\pm.11}$	$37.50_{\pm.12}$	$35.00_{\pm.70}$	$29.58_{\pm2.14}$	$35.86_{\pm.04}$	$32.84_{\pm.25}$	**38.17**$_{\pm.07}$
SIGN	$44.64_{\pm.26}$	**38.41**$_{\pm.39}$	$31.93_{\pm1.15}$	$35.43_{\pm.25}$	$35.77_{\pm.24}$	$28.97_{\pm.35}$	$37.32_{\pm1.02}$
GPR-GNN	$41.95_{\pm.07}$	$38.75_{\pm.11}$	$36.30_{\pm.08}$	$29.43_{\pm.80}$	$35.84_{\pm.16}$	$23.49_{\pm.13}$	**39.08**$_{\pm.16}$
igb-small							
GCN	$72.31_{\pm.01}$	$69.96_{\pm.03}$	$70.24_{\pm.06}$	$54.38_{\pm1.32}$		$68.60_{\pm.21}$	**71.48**$_{\pm.04}$
GraphSAGE	$75.49_{\pm.07}$	$70.95_{\pm.04}$	$72.21_{\pm.26}$	$51.58_{\pm3.24}$		$71.38_{\pm.31}$	**73.12**$_{\pm.06}$
SGC	$71.69_{\pm.00}$	$69.91_{\pm.01}$	$69.63_{\pm.02}$	$44.19_{\pm.22}$	‡	$68.98_{\pm.00}$	**71.10**$_{\pm.00}$
SIGN	$76.01_{\pm.03}$	$73.30_{\pm.05}$	$73.30_{\pm.02}$	$69.08_{\pm.32}$		$73.35_{\pm.07}$	**73.61**$_{\pm.05}$
GPR-GNN	$75.50_{\pm.03}$	—	$73.17_{\pm.04}$	$70.49_{\pm.06}$		$73.02_{\pm.01}$	**74.02**$_{\pm.03}$
ogbn-arxiv							
GCN	$71.90_{\pm.17}$	$66.40_{\pm.19}$	$60.78_{\pm.23}$	$27.65_{\pm.49}$	$60.11_{\pm1.66}$	$63.23_{\pm.42}$	**68.01**$_{\pm.20}$
GraphSAGE	$71.73_{\pm.28}$	**62.02**$_{\pm.64}$	$58.34_{\pm.92}$	$43.19_{\pm2.83}$	$54.61_{\pm.89}$	$53.86_{\pm.40}$	$58.11_{\pm.51}$
SGC	$68.60_{\pm.04}$	$64.26_{\pm.12}$	$51.26_{\pm2.60}$	$29.82_{\pm.98}$	$58.33_{\pm.15}$	$59.14_{\pm.76}$	**66.20**$_{\pm.13}$
SIGN	$71.44_{\pm.13}$	**61.01**$_{\pm.38}$	$56.23_{\pm.39}$	$47.83_{\pm.66}$	$53.93_{\pm.45}$	$51.08_{\pm.84}$	$58.03_{\pm.47}$
GPR-GNN	$69.05_{\pm.30}$	$66.66_{\pm.65}$	$57.25_{\pm.22}$	$49.95_{\pm.75}$	$53.19_{\pm.10}$	$60.08_{\pm.80}$	**66.93**$_{\pm.30}$
ogbn-mag							
GCN	$35.64_{\pm.12}$	$30.78_{\pm.16}$	$23.97_{\pm.63}$	$10.36_{\pm1.20}$	$18.65_{\pm.94}$	$24.31_{\pm.47}$	**33.87**$_{\pm.25}$
GraphSAGE	$35.81_{\pm.15}$	**29.40**$_{\pm.33}$	$26.84_{\pm.38}$	$10.07_{\pm1.99}$	$25.35_{\pm.66}$	$24.16_{\pm1.12}$	$28.52_{\pm.12}$
SGC	$33.85_{\pm.45}$	$29.60_{\pm.15}$	$27.10_{\pm.10}$	$10.52_{\pm.20}$	$27.27_{\pm.08}$	$27.85_{\pm.29}$	**31.53**$_{\pm.05}$
SIGN	$35.97_{\pm.10}$	$29.52_{\pm.21}$	$26.64_{\pm.41}$	$19.56_{\pm.57}$	$28.24_{\pm.22}$	$26.25_{\pm.19}$	**29.87**$_{\pm.31}$
GPR-GNN	$31.40_{\pm.58}$	$30.06_{\pm.32}$	$26.12_{\pm.27}$	$5.40_{\pm1.93}$	$25.64_{\pm.25}$	$26.57_{\pm.05}$	**32.55**$_{\pm.17}$
ogbn-products							
GCN	$76.67_{\pm.68}$	$71.77_{\pm.53}$	$62.68_{\pm.46}$	—			**75.85**$_{\pm.20}$
GraphSAGE	$79.61_{\pm.20}$	$67.82_{\pm.05}$	$62.34_{\pm.40}$	$58.82_{\pm1.08}$			**71.86**$_{\pm.38}$
SGC	$65.15_{\pm.00}$	$63.67_{\pm.07}$	$45.17_{\pm.26}$	$34.30_{\pm.40}$	OOT	OOT	**65.29**$_{\pm.01}$
SIGN	$75.16_{\pm.06}$	$68.42_{\pm.07}$	$55.28_{\pm.56}$	—			**71.66**$_{\pm.08}$
GPR-GNN	$73.65_{\pm.68}$	$72.95_{\pm.17}$	$61.34_{\pm.57}$	$53.59_{\pm.43}$			**76.22**$_{\pm.30}$
pokec							
GCN	$72.50_{\pm.44}$	$51.44_{\pm.55}$	$52.43_{\pm.47}$	$50.75_{\pm.14}$		$51.61_{\pm.32}$	**58.09**$_{\pm.85}$
GraphSAGE	$80.40_{\pm.24}$	**68.66**$_{\pm1.31}$	$67.71_{\pm2.03}$	$62.73_{\pm.78}$		$61.42_{\pm1.53}$	$68.34_{\pm.69}$
SGC	$54.88_{\pm.20}$	**51.86**$_{\pm.55}$	$51.15_{\pm.77}$	$50.98_{\pm.24}$	‡	$51.09_{\pm.01}$	$51.10_{\pm.53}$
SIGN	$73.81_{\pm.53}$	$61.27_{\pm.10}$	$61.84_{\pm.11}$	$61.24_{\pm.08}$		$59.40_{\pm.20}$	**64.01**$_{\pm.19}$
GPR-GNN	$75.98_{\pm.53}$	$52.02_{\pm.69}$	$55.26_{\pm4.05}$	$51.32_{\pm.25}$		—	**57.15**$_{\pm1.37}$
reddit2							
GCN	$95.34_{\pm.01}$	**93.71**$_{\pm.10}$	$87.65_{\pm.51}$	$25.62_{\pm1.14}$	$81.77_{\pm.71}$	$90.40_{\pm.34}$	$93.40_{\pm.08}$
GraphSAGE	$96.64_{\pm.07}$	**89.12**$_{\pm.34}$	$87.69_{\pm.72}$	$24.63_{\pm1.10}$	$54.97_{\pm1.28}$	$88.69_{\pm.70}$	$87.94_{\pm.52}$
SGC	$94.62_{\pm.03}$	**93.54**$_{\pm.09}$	$90.57_{\pm.13}$	$32.69_{\pm.38}$	$80.08_{\pm.93}$	$91.38_{\pm.31}$	$92.34_{\pm.17}$
SIGN	$96.30_{\pm.05}$	$90.54_{\pm.35}$	$84.75_{\pm.27}$	$41.36_{\pm1.08}$	$77.44_{\pm.38}$	$86.77_{\pm.29}$	**92.64**$_{\pm.12}$
GPR-GNN	$96.15_{\pm.02}$	—	$87.38_{\pm.20}$	$54.20_{\pm1.70}$	$72.97_{\pm.67}$	**91.85**$_{\pm.39}$	$87.35_{\pm.10}$
snap-patents							
GCN	$41.63_{\pm.04}$		$29.83_{\pm.32}$	$26.81_{\pm.18}$			**34.64**$_{\pm.09}$
GraphSAGE	$44.11_{\pm.17}$		$31.36_{\pm.25}$	$25.44_{\pm.06}$			**34.41**$_{\pm.31}$
SGC	$35.23_{\pm.01}$	OOM	$26.42_{\pm.00}$	$22.54_{\pm1.66}$	‡	‡	**32.04**$_{\pm.02}$
SIGN	$42.22_{\pm.06}$		—	—			**35.86**$_{\pm.04}$
GPR-GNN	$32.84_{\pm.18}$		$27.78_{\pm.04}$	$21.20_{\pm.04}$			**28.97**$_{\pm.11}$
twitch-gamers							
GCN	$63.68_{\pm.10}$	$60.65_{\pm.36}$	$58.69_{\pm.21}$	$56.77_{\pm.17}$		$58.12_{\pm.37}$	**60.93**$_{\pm.26}$
GraphSAGE	$64.53_{\pm.04}$	$53.69_{\pm.50}$	$57.95_{\pm1.34}$	$58.54_{\pm.26}$		$55.24_{\pm1.51}$	**60.53**$_{\pm.09}$
SGC	$60.60_{\pm.10}$	**59.71**$_{\pm.39}$	$56.64_{\pm.32}$	$56.46_{\pm.48}$	‡	$55.92_{\pm1.30}$	$59.29_{\pm.21}$
SIGN	$63.43_{\pm.12}$	**61.41**$_{\pm.13}$	$57.38_{\pm1.24}$	$59.59_{\pm.37}$		$57.94_{\pm.10}$	$60.50_{\pm.07}$
GPR-GNN	$60.50_{\pm.56}$	$60.29_{\pm.04}$	**60.75**$_{\pm.09}$	$58.14_{\pm.07}$		—	$60.29_{\pm.05}$

Table 12: Test accuracy (%) of every backbone on the coarsened graph at 99% reduction ratio, evaluated on the original graph. Best in **bold**, second best underlined, ranked among the methods with a result. ‡ marks a coarsening whose achieved ratio missed the $99 \pm 3\%$ tolerance, OOT and OOM one that exceeded the time limit or ran out of memory.

Backbone	Full graph	A-ConvMatch	SGBGC	UGC	Var-Edges	Var-Neigh	Ours
arxiv-year							
GCN	$45.97_{\pm.13}$	$38.33_{\pm.23}$	$34.77_{\pm.38}$	$26.39_{\pm3.33}$		$33.93_{\pm.41}$	**38.95**$_{\pm.40}$
GraphSAGE	$47.47_{\pm.45}$	**37.00**$_{\pm.17}$	$36.22_{\pm.24}$	$26.45_{\pm3.13}$		$30.24_{\pm1.76}$	$35.93_{\pm.38}$
SGC	$39.47_{\pm.11}$	**37.21**$_{\pm.13}$	$35.63_{\pm.09}$	$26.04_{\pm2.34}$	‡	$27.48_{\pm.70}$	$37.10_{\pm.03}$
SIGN	$44.64_{\pm.26}$	$36.69_{\pm.28}$	$32.60_{\pm1.77}$	$29.33_{\pm.66}$		$28.83_{\pm.36}$	**37.52**$_{\pm.95}$
GPR-GNN	$41.95_{\pm.07}$	**37.76**$_{\pm.06}$	$36.14_{\pm.15}$	$30.75_{\pm.32}$		$22.73_{\pm.10}$	$37.27_{\pm.16}$
igb-small							
GCN	$72.31_{\pm.01}$	$69.32_{\pm.07}$	$69.77_{\pm.13}$	$25.98_{\pm.00}$		$69.50_{\pm.10}$	**70.39**$_{\pm.07}$
GraphSAGE	$75.49_{\pm.07}$	$69.95_{\pm.23}$	$71.01_{\pm.16}$	$35.98_{\pm4.63}$		$71.40_{\pm.17}$	**72.07**$_{\pm.04}$
SGC	$71.69_{\pm.00}$	$69.59_{\pm.00}$	$69.46_{\pm.01}$	$34.07_{\pm.62}$	‡	$69.18_{\pm.01}$	**70.14**$_{\pm.00}$
SIGN	$76.01_{\pm.03}$	**72.55**$_{\pm.02}$	$72.12_{\pm.11}$	$62.79_{\pm.31}$		$71.86_{\pm.08}$	$72.35_{\pm.03}$
GPR-GNN	$75.50_{\pm.03}$	—	$72.66_{\pm.00}$	$65.30_{\pm.21}$		$72.30_{\pm.01}$	**72.67**$_{\pm.05}$
ogbn-arxiv							
GCN	$71.90_{\pm.17}$	$61.84_{\pm1.67}$	$59.04_{\pm1.36}$	$23.45_{\pm1.54}$		$54.82_{\pm1.71}$	**64.30**$_{\pm.54}$
GraphSAGE	$71.73_{\pm.28}$	**58.90**$_{\pm.50}$	$49.76_{\pm3.95}$	$29.51_{\pm.44}$		$41.44_{\pm2.30}$	$52.33_{\pm.46}$
SGC	$68.60_{\pm.04}$	$62.29_{\pm.16}$	$51.31_{\pm.60}$	$26.56_{\pm2.05}$	‡	$57.65_{\pm.44}$	**65.31**$_{\pm.29}$
SIGN	$71.44_{\pm.13}$	**54.97**$_{\pm.48}$	$47.28_{\pm1.14}$	$39.99_{\pm1.11}$		$37.83_{\pm.62}$	$53.04_{\pm.49}$
GPR-GNN	$69.05_{\pm.30}$	**63.09**$_{\pm.46}$	$58.41_{\pm.67}$	$43.56_{\pm.09}$		$45.31_{\pm.37}$	$61.66_{\pm.22}$
ogbn-mag							
GCN	$35.64_{\pm.12}$	$29.96_{\pm.25}$	$21.61_{\pm1.11}$	$5.54_{\pm.82}$	$16.28_{\pm.89}$	$17.33_{\pm.89}$	**31.59**$_{\pm.28}$
GraphSAGE	$35.81_{\pm.15}$	**27.53**$_{\pm.54}$	$24.10_{\pm.78}$	$4.31_{\pm.60}$	$22.67_{\pm.26}$	$18.04_{\pm.85}$	$24.84_{\pm.87}$
SGC	$33.85_{\pm.45}$	$29.22_{\pm.04}$	$26.92_{\pm.30}$	$8.41_{\pm1.67}$	$21.77_{\pm.29}$	$20.45_{\pm.91}$	**29.60**$_{\pm.06}$
SIGN	$35.97_{\pm.10}$	**27.63**$_{\pm.31}$	$23.46_{\pm.44}$	$11.47_{\pm.41}$	$19.57_{\pm.51}$	$16.31_{\pm1.11}$	$25.11_{\pm.37}$
GPR-GNN	$31.40_{\pm.58}$	$29.22_{\pm.19}$	$24.54_{\pm.34}$	$3.77_{\pm.94}$	$22.10_{\pm.34}$	$15.06_{\pm.24}$	**29.22**$_{\pm.31}$
ogbn-products							
GCN	$76.67_{\pm.68}$	$69.56_{\pm.25}$	$51.82_{\pm.11}$	$26.02_{\pm1.88}$			**73.45**$_{\pm.25}$
GraphSAGE	$79.61_{\pm.20}$	$61.42_{\pm.18}$	$47.71_{\pm1.36}$	$26.73_{\pm1.40}$			**67.94**$_{\pm.20}$
SGC	$65.15_{\pm.00}$	**63.36**$_{\pm.06}$	$41.30_{\pm.07}$	$13.47_{\pm.11}$	OOT	OOT	$63.34_{\pm.09}$
SIGN	$75.16_{\pm.06}$	$65.03_{\pm.11}$	$50.29_{\pm.44}$	$38.53_{\pm.39}$			**68.70**$_{\pm.20}$
GPR-GNN	$73.65_{\pm.68}$	—	$60.29_{\pm.46}$	$33.20_{\pm1.01}$			**74.16**$_{\pm.74}$
pokec							
GCN	$72.50_{\pm.44}$	$50.82_{\pm.16}$	$52.14_{\pm.15}$	$50.95_{\pm.21}$		$51.79_{\pm.19}$	**55.01**$_{\pm.37}$
GraphSAGE	$80.40_{\pm.24}$	$59.05_{\pm1.27}$	**65.03**$_{\pm.46}$	$55.77_{\pm3.61}$		$54.06_{\pm.58}$	$63.12_{\pm1.36}$
SGC	$54.88_{\pm.20}$	$51.87_{\pm.88}$	**52.56**$_{\pm.06}$	$52.30_{\pm1.11}$	‡	$51.19_{\pm.01}$	$51.04_{\pm.06}$
SIGN	$73.81_{\pm.53}$	$60.79_{\pm.13}$	$60.25_{\pm.24}$	$59.30_{\pm.21}$		$58.93_{\pm.48}$	**61.33**$_{\pm.12}$
GPR-GNN	$75.98_{\pm.53}$	—	$53.92_{\pm1.24}$	$50.95_{\pm.26}$		—	**57.21**$_{\pm.14}$
reddit2							
GCN	$95.34_{\pm.01}$		$56.84_{\pm1.54}$	$23.72_{\pm1.45}$		$65.41_{\pm.40}$	**92.34**$_{\pm.10}$
GraphSAGE	$96.64_{\pm.07}$		$69.00_{\pm1.52}$	$18.87_{\pm1.40}$		**78.67**$_{\pm.94}$	$69.72_{\pm.82}$
SGC	$94.62_{\pm.03}$	OOM	$71.60_{\pm.64}$	$28.12_{\pm.42}$	‡	$73.31_{\pm.95}$	**91.48**$_{\pm.16}$
SIGN	$96.30_{\pm.05}$		$60.47_{\pm.30}$	$19.09_{\pm.89}$		$62.06_{\pm.56}$	**86.58**$_{\pm.23}$
GPR-GNN	$96.15_{\pm.02}$		$59.07_{\pm2.52}$	$26.62_{\pm.33}$		$60.28_{\pm4.84}$	**78.72**$_{\pm.51}$
snap-patents							
GCN	$41.63_{\pm.04}$			$22.06_{\pm.35}$			**33.55**$_{\pm.04}$
GraphSAGE	$44.11_{\pm.17}$			$26.09_{\pm.68}$			**32.21**$_{\pm.13}$
SGC	$35.23_{\pm.01}$	OOM	‡	$22.13_{\pm.70}$	‡	‡	**31.44**$_{\pm.01}$
SIGN	$42.22_{\pm.06}$			$23.36_{\pm.39}$			**32.71**$_{\pm.05}$
GPR-GNN	$32.84_{\pm.18}$			$23.63_{\pm.97}$			**29.23**$_{\pm.01}$
twitch-gamers							
GCN	$63.68_{\pm.10}$	**60.64**$_{\pm.22}$	$53.75_{\pm.35}$	$53.10_{\pm.00}$		$55.43_{\pm.54}$	$59.85_{\pm.17}$
GraphSAGE	$64.53_{\pm.04}$	$56.24_{\pm.72}$	$55.02_{\pm1.10}$	$57.71_{\pm.53}$		$53.99_{\pm.49}$	**60.64**$_{\pm.13}$
SGC	$60.60_{\pm.10}$	**60.53**$_{\pm.02}$	$58.46_{\pm.06}$	$54.65_{\pm1.60}$	‡	$57.89_{\pm.17}$	$59.55_{\pm.39}$
SIGN	$63.43_{\pm.12}$	**61.01**$_{\pm.10}$	$56.14_{\pm1.34}$	$58.28_{\pm.43}$		$58.98_{\pm.47}$	$60.53_{\pm.07}$
GPR-GNN	$60.50_{\pm.56}$	**60.18**$_{\pm.04}$	$59.96_{\pm.22}$	$54.49_{\pm.56}$		$57.41_{\pm3.73}$	$60.18_{\pm.03}$

F SIGNIFICANCE TESTING

Table 13 reports the accuracy difference between our coarsening and each baseline of Table 1, dataset by dataset, and tests it at three levels (Demšar, 2006). Within a cell, the training seeds of the two methods are compared by Welch's two-sided t-test, Holm-corrected over the 57 cells in which both produced a graph; both coarsenings are fixed, so this test covers training randomness only. Across a column, the per-dataset differences are compared by the exact two-sided Wilcoxon signed-rank test, with the dataset as the unit of replication and Holm correction over the seven baselines. Across all eight coarseners, the methods are ranked within each dataset and compared by the Friedman test with the Iman–Davenport correction, followed by a Holm-corrected post-hoc comparison of each baseline's mean rank against ours.

Table 13: Accuracy difference (percentage points) between our coarsening and each baseline at ratio 90, from the training seeds behind Table 1; positive favors ours. **Bold** marks a difference significant in our favour (Welch's t-test, Holm-corrected over all tested cells), underline one significant in the baseline's favour, and ‡ marks a coarsening whose achieved ratio missed the $90 \pm 3\%$ tolerance, counted as a win but not tested. *Best* is ours minus the strongest baseline on each dataset. The middle block gives the Wilcoxon signed-rank test of each column, the lower block the mean rank of each coarsener over all nine datasets, a missing result sharing the last places (unlike the mean rank of Table 1, which skips such cells), and its Friedman post-hoc comparison with our mean rank of 1.22.

Dataset	A-ConvMatch	Random	SGBGC	STPGC	UGC	Var-Edges	Var-Neigh	Best
arxiv-year	+0.03	**+2.17**	**+5.93**	**+3.77**	**+8.09**	**+3.21**	**+1.83**	+0.03
igb-small	**+1.73**	**+1.33**	**+1.04**	**+3.21**	**+7.29**	**+1.59**	**+2.29**	+1.04
ogbn-arxiv	**+4.06**	**+9.23**	**+4.93**	**+4.52**	**+35.52**	**+2.92**	**+2.89**	+2.89
ogbn-mag	**+3.14**	**+20.41**	**+9.19**	**+5.35**	**+21.67**	**+6.64**	**+7.76**	+3.14
ogbn-products	**+3.44**	**+41.12**	**+9.90**	**+16.17**	**+11.20**	‡	**+7.16**	+3.44
pokec	**+7.41**	‡	**+6.41**	**+6.70**	**+6.68**	**+6.57**	**+7.61**	+6.41
reddit2	−0.24	**+1.65**	**+2.01**	**+18.83**	**+68.16**	−0.33	**+2.03**	−0.33
snap-patents	‡	**+10.16**	**+4.88**	**+2.39**	**+8.50**	‡	‡	+2.39
twitch-gamers	**+0.81**	**+3.28**	**+1.88**	**+4.16**	**+2.89**	‡	**+2.46**	+0.81
Datasets n	8	8	9	9	9	6	8	9
Ours ahead	7	8	9	9	9	5	8	8
Wilcoxon p	0.023	0.008	0.004	0.004	0.004	0.062	0.008	0.012
Holm, $\alpha = 0.05$	✓	✓	✓	✓	✓	✗	✓	✓
Mean rank	3.56	5.67	4.22	5.33	6.44	4.78	4.78	
Post-hoc p	0.043	<0.001	0.009	<0.001	<0.001	0.002	0.002	
Holm, $\alpha = 0.05$	✓	✓	✓	✓	✓	✓	✓	

Our coarsening is ahead in 55 of the 57 tested cells, significantly so in 54; the exceptions are a tie with A-ConvMatch on arxiv-year and two losses on Reddit2, to Variation-Edges and A-ConvMatch, both below 0.35 pp. Training variance is small, with a median per-cell standard deviation of 0.20 pp, so differences of a quarter of a point are significant in either direction. Across datasets, six of the seven columns survive the Holm correction. Variation-Edges does not, because it completed only six datasets, and with $n = 6$ the exact signed-rank test cannot return a two-sided p below 0.031 once a single loss is present. Against the strongest baseline on each dataset our coarsening method leads on eight of nine with a median margin of 2.39 pp ($p = 0.012$).

Ranking all eight coarseners within each dataset, our mean rank is 1.22, first on eight datasets and third on Reddit2; the Friedman test rejects equal performance ($p = 4 \times 10^{-4}$), and the post-hoc comparison separates ours from every baseline, the closest being A-ConvMatch at $p = 0.043$. This level also separates Variation-Edges, since its three failures at this budget count as last places, whereas the paired test sees only the datasets it completed. Restricted to the five datasets on which all eight coarseners completed, the omnibus test still rejects ($p = 0.002$), but with $N = 5$ the post-hoc comparison separates only STPGC and UGC, and the cell-level and paired tests carry the comparison with A-ConvMatch and SGBGC.

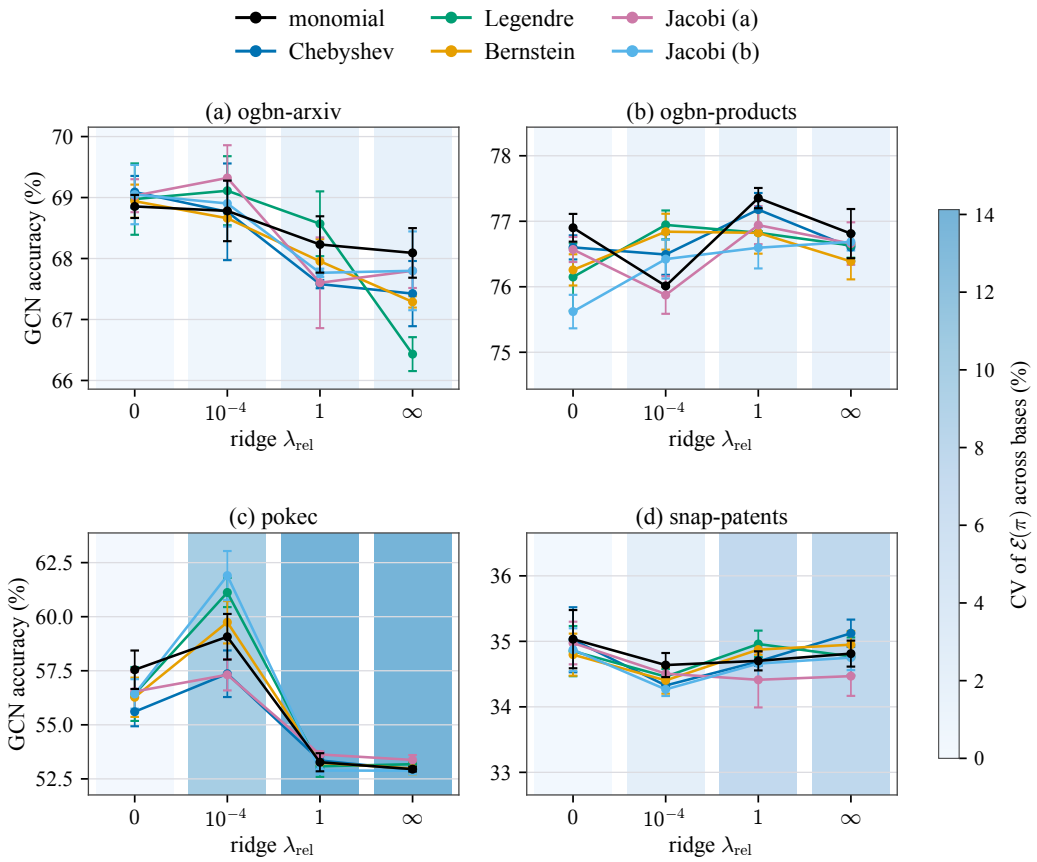


Figure 6: Basis sweep: GCN accuracy at ratio 90 against the ridge λ_{rel}, one line per polynomial basis, one panel per dataset. The band behind each ridge is tinted by the coefficient of variation, across the six bases at that ridge, of the objective $\mathcal{E}(\pi)$ (color bar).

G Basis and operator invariance

As summarized in Table 4, the merge criterion is invariant to the polynomial basis and to any affine reparametrization of the operator, and may change under an operator with other eigenvectors or a non-affine spectral reweighting. We evaluate this empirically on ogbn-arxiv, ogbn-products, pokec and snap-patents at ratio 90: Appendix G.1 checks the robustness of the coarsening across bases, and Appendix G.2 measures how downstream accuracy changes under different operators.

G.1 Basis

We build the trajectory in six bases of the degree-3 polynomials, monomial, Chebyshev, Legendre, Bernstein and two Jacobi families, at each of $\lambda_{\text{rel}} \in \{0, 10^{-4}, 1, \infty\}$, where ∞ denotes the objective $\mathcal{E}$. Figure 6 plots the resulting accuracy against the ridge, one line per basis, and tints the band behind each ridge by the coefficient of variation, across the six partitions, of the objective $\mathcal{E}(\pi)$.

At $\lambda_{\text{rel}} = 0$ the merge operates on an orthonormal basis Q of the trajectory's column space, which every polynomial basis shares, so the criterion is invariant. For $\lambda_{\text{rel}} > 0$ the weights $\phi_i^2 = \rho_i^2/(\rho_i^2 + \lambda_{\text{rel}})$ of Equation (16) are functions of the singular values of the trajectory. A change of basis multiplies the trajectory by the invertible matrix $\Psi^\top$ of Equation (12); this preserves its column space but, unless Ψ is orthogonal, alters its singular values and with them the weights. Every basis therefore presents the merge with the same subspace under a different weighting of its directions, and the merge selects a different partition. The same mechanism explains why polynomial bases of

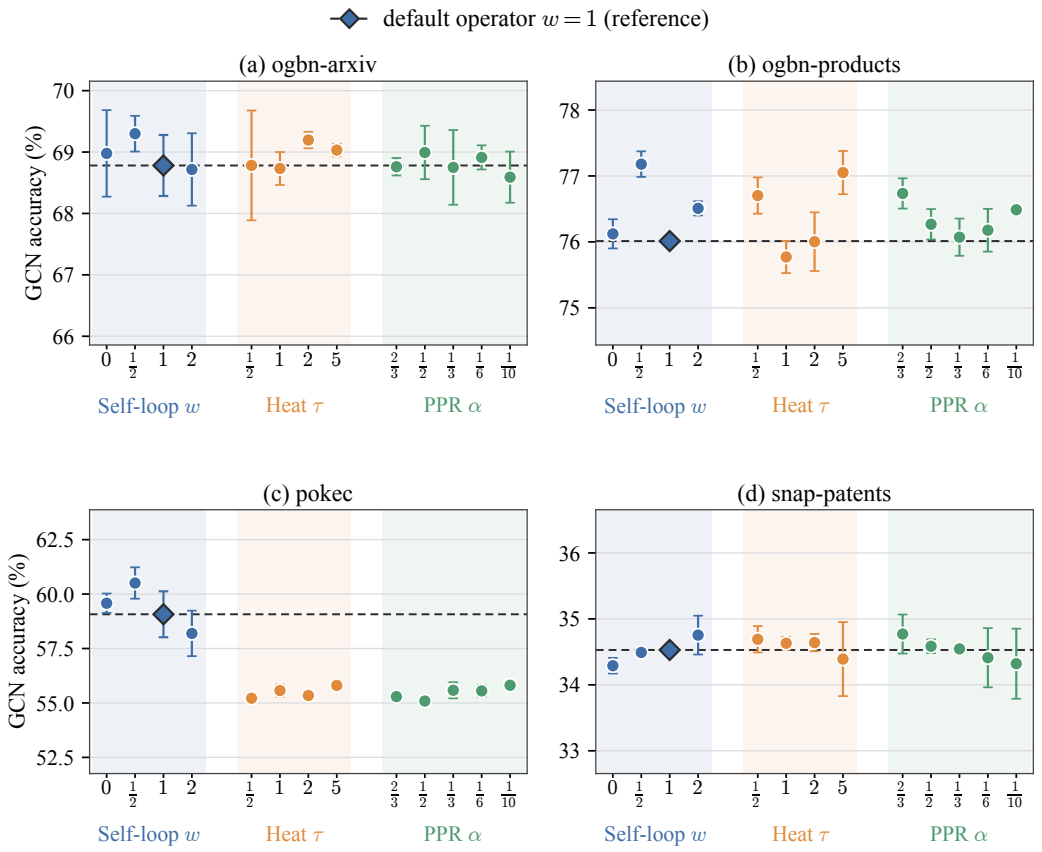


Figure 7: Operator sweep: GCN accuracy at ratio 90 and $\lambda_{\text{rel}} = 10^{-4}$ per operator variant of Table 14, grouped by family with the parameter value under each point, one panel per dataset. The dashed line and the diamond mark the default operator $w = 1$, which is the reference of every comparison.

equal expressive power condition spectral GNN training differently (Wang & Zhang, 2022; Guo & Wei, 2023).

Empirically, we can see this in Figure 6. At $\lambda_{\text{rel}} = 0$ the six bases reach the same objective, with a coefficient of variation below 0.4% on every graph, and their accuracies agree to within two points, comparable to the seed noise. For $\lambda_{\text{rel}} > 0$ the objective spreads to 0.7–14%, but accuracy moves with it mainly on pokec: at the default 10^{-4} the bases span 4.6 points there (57.3 to 61.9, monomial 59.1 ± 1.1), against at most 1.1 points on the other graphs. The spread on pokec closes again at $\lambda_{\text{rel}} \geq 1$, where the ridge stops selecting directions. The basis therefore matters only where the ridge is active, and beyond seed noise on one of the four graphs.

The ridge itself moves accuracy more than the basis does, and its best value depends on the graph: ogbn-arxiv and snap-patents are best at $\lambda_{\text{rel}} \leq 10^{-4}$, ogbn-products at 1, and pokec at the default 10^{-4}, where it gains 1.5 points over 0 and 5.8 over 1. This is consistent with Table 4: the basis changes the coordinates of the objective, whereas the ridge changes its geometry.

G.2 Operator

We coarsen with the thirteen propagation operators of Table 14: the renormalized adjacency with self-loop weight w, and the two graph diffusions of Klicpera et al. (2019b) built on it, the heat kernel and personalized PageRank (PPR). The default operator $\hat{A}$ is the variant $w = 1$ and the reference of every comparison. Figure 7 plots accuracy per operator, grouped by family.

Table 14: Propagation operators of the sweep. $\hat{A} = \hat{A}_1$ is the default operator; the last column states whether the eigenvectors of $\hat{A}$ are retained.

Family	Operator	Parameter	Eigenvectors
Self-loop weight	$\hat{A}_w = (D + wI)^{-1/2}(A + wI)(D + wI)^{-1/2}$	$w \in \{0, \frac{1}{2}, 1, 2\}$	change with w
Heat kernel	$e^{-\tau(I-\hat{A})}$	$\tau \in \{\frac{1}{2}, 1, 2, 5\}$	retained
PPR	$\alpha\,(I - (1-\alpha)\hat{A})^{-1}$	$\alpha \in \{\frac{2}{3}, \frac{1}{2}, \frac{1}{3}, \frac{1}{6}, \frac{1}{10}\}$	retained

On ogbn-arxiv, ogbn-products and snap-patents the thirteen operators agree to within 1.4 points, comparable to the seed noise. Pokec splits by family: the self-loop weights stay within 1.4 points of the reference (59.1 ± 1.1), while every heat and PPR variant falls to 55.1–55.8, a loss of about four points that no choice of τ or α recovers. The same partitions, scored on the reference trajectory, have objectives 3–9% above the reference partition, against 0–4% for the self-loop variants. On pokec the coarsening operator therefore matters, and the two diffusion families cost accuracy where the self-loop weight does not.

H KRYLOV APPROXIMATION CONSTANT AND OBJECTIVE OF EVERY COARSENING

Table 15 evaluates the Krylov approximation constant $\mathcal{E}^\star(\pi)$ of Definition 2, reported squared, and the objective $\mathcal{E}(\pi)$ of Equation (5) on the partition of every method at ratio 90, with $K = 3$ and the coarse operator $S_c = RSR^+$ of Theorem 1. Features are scaled to $\|X\|_F = 1$, as in Definition 2. Because S_c is determined by the partition, the table compares partitions only; the edges a method keeps do not enter it. Our two columns differ for this reason alone: edge selection (Section 3.3) runs between merge levels and so changes the partition that later levels produce.

Table 15: Squared Krylov approximation constant and objective of each coarsening, reported as $\mathcal{E}^\star(\pi)^2 \,/\, \mathcal{E}(\pi)$, at ratio 90 with $K = 3$, $S_c = RSR^+$ and $\|X\|_F = 1$; lower is better, and the leading zero is omitted. Within each quantity, best in **bold**, second best underlined. ‡, OOT and OOM mark the cells that Table 1 does not report, for the reasons given there; an off-budget partition keeps more supernodes than the budget allows, which would lower both quantities, so its values are withheld. "—" marks a partition not available for this evaluation.

	Baselines							Ours	
Dataset	A-ConvMatch	Random	SGBGC	STPGC	UGC	Var-Edges	Var-Neigh	pruned	unpruned
arxiv-year	.158 / **.301**	.210 / .647	.184 / .567	.181 / .564	.213 / .458	.175 / .525	.175 / .518	.148 / .352	**.145** / .347
igb-small	.540 / .797	.682 / 1.142	.677 / 1.124	.669 / 1.122	.858 / 1.636	.603 / .941	.585 / .896	.470 / .680	**.427** / **.601**
ogbn-arxiv	.158 / **.301**	.210 / .647	.182 / .558	.180 / .563	.211 / .455	.175 / .525	.175 / .518	.148 / .349	**.144** / .347
ogbn-mag	.126 / **.210**	.150 / .392	.150 / .403	.150 / .404	.205 / .498	.131 / .333	.135 / .346	.113 / .265	**.108** / .251
ogbn-products	.475 / .539	.898 / 1.229	.831 / 1.048	—	.621 / .892	OOT	OOT	.214 / .256	**.206** / **.247**
pokec	.130 / .174	‡	.195 / .588	.192 / .577	.172 / .299	.195 / .541	.182 / .487	.045 / .156	**.042** / **.147**
reddit2	—	.776 / .856	.839 / .956	.818 / .902	.841 / .991	.743 / .775	.778 / .848	.551 / .596	**.535** / **.578**
snap-patents	OOM	.664 / 1.379	.528 / .923	.563 / 1.042	.646 / 1.316	‡	‡	.336 / .606	**.310** / **.565**
twitch-gamers	.353 / .379	.822 / 1.047	.853 / 1.034	.843 / 1.017	.780 / .948	‡	.829 / 1.025	.134 / .166	**.127** / **.159**

Our coarsening attains the smallest $\mathcal{E}^\star(\pi)$ on all nine datasets, unpruned first and pruned second. On $\mathcal{E}(\pi)$ it is first on six of the nine; on ogbn-arxiv, arxiv-year and ogbn-mag, A-ConvMatch is lower (0.301 against 0.347 on ogbn-arxiv, 0.210 against 0.251 on ogbn-mag). The measured ratio $\mathcal{E}(\pi)/\mathcal{E}^\star(\pi)^2$ lies in $[1.04,\ 3.49]$ against the admissible $[1, 4]$ of Theorem 1(i), so neither side of the bound is tight uniformly across methods and datasets.

Across the degree. Figure 8 repeats the evaluation of Table 15 for $K = 1, \ldots, 5$ on the unpruned partition of every dataset, with the same S_c and scaling, to show where $\mathcal{E}^\star(\pi)^2$ falls within the admissible interval as K grows. On eight datasets the transfer error is largest at hop 0, so $\mathcal{E}^\star(\pi)$ is independent of K while $\mathcal{E}(\pi)$ gains one residual term per hop; the ratio $\mathcal{E}(\pi)/\mathcal{E}^\star(\pi)^2$ therefore

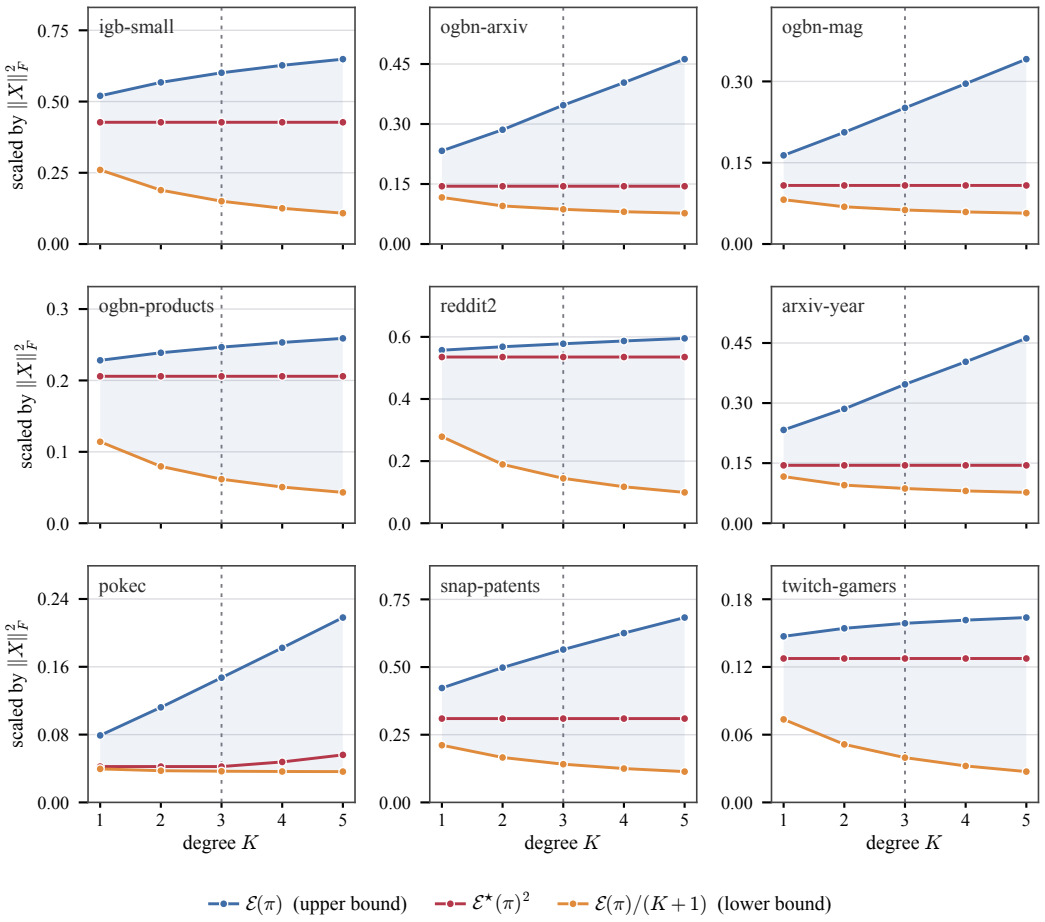


Figure 8: The chain $\mathcal{E}(\pi)/(K+1) \le \mathcal{E}^\star(\pi)^2 \le \mathcal{E}(\pi)$ of Theorem 1(i) for $K = 1, \dots, 5$, on the unpruned partition of our coarsening at ratio 90, with linear propagation, $S_c = RSR^+$ and $\|X\|_F = 1$. The shaded band is the admissible interval; the dotted rule marks the default $K = 3$, at which the values are those of Table 15 (unpruned).

increases with K but remains below $K + 1$. On pokec a later hop attains the maximum, so both bounds move with K and $\mathcal{E}^\star(\pi)^2$ stays close to the lower one.