FREQUENCY ORDERS AND RESPONSE FLOORS OF FIXED FILTERS IN DECOUPLED GRAPH MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Decoupled graph models classify nodes by filtering node features once with a fixed filter and training only a classifier. The filter comes from a propagation scheme, adjacency powers (SGC) or personalized PageRank (PPR), and only its hyperparameter is tuned. Dataset-level metrics and block-model analyses of specific operators leave open which scheme to use and how much it can lose to no filter. Since a filter rescales each graph frequency separately, we characterize a scheme by its frequency order. PPR orders frequencies by eigenvalue and SGC by magnitude. In a linear feature model, a filter raises the signal-to-noise ratio (SNR) exactly when its squared response covaries positively with the excess of signal over noise across frequencies. Stochastic dominance between signal and noise along a scheme's order decides the whole scheme. Every PPR filter has a response floor bounding its SNR loss, whereas adjacency powers have none. On contextual stochastic block models, the criterion predicts the sign of every significant gain. On 30 real graphs with equal tuning budgets, neither scheme dominates, and PPR significantly beats SGC wherever SGC is significantly below an MLP. The results support searching both frequency orders and keeping the identity in schemes without a floor.

1 INTRODUCTION

Graph neural networks (GNNs) are the standard model for node classification (Kipf & Welling, 2017). Message passing interleaves propagation over the graph with learned transformations, so the cost of training grows with the graph. Decoupled models separate the two steps (Wu et al., 2019; Gasteiger et al., 2019a; Frasca et al., 2020). Precomputed variants such as SGC and SIGN compute $g(\mathbf{W})\mathbf{X}$ once, where $\mathbf{W}$ is a normalized adjacency matrix, $\mathbf{X}$ holds the node features and g is a fixed filter. A classifier is then trained on its rows at the cost of a multilayer perceptron (MLP).

The filter is drawn from a propagation scheme, and only its hyperparameter is tuned. SGC uses the powers $\mathbf{W}^K$ (Wu et al., 2019). Personalized PageRank (PPR), $g_\alpha(\mathbf{W}) = \alpha(\mathbf{I} - (1-\alpha)\mathbf{W})^{-1}$, propagates the predictions of APPNP and is one of the diffusions of GDC (Gasteiger et al., 2019a;b). We apply it to the features. Both schemes are accurate on homophilous graphs. On heterophilous graphs, an MLP on the raw features is often competitive (Zhu et al., 2020; Lim et al., 2021). Whether to filter at all, and with which scheme, is a genuine modelling decision.

Existing guidance for this decision is of two kinds. Dataset-level metrics indicate whether graph-aware models help. Examples are edge and adjusted homophily, label informativeness and classifier-based distinguishability (Platonov et al., 2023a; Luan et al., 2023). Feature-level and node-level analyses refine them (Zheng et al., 2025; Mao et al., 2023). Block-model analyses characterize one or two operators at a time: one or several graph convolutions (Baranwal et al., 2021; 2023), PPR against graph convolution (Wu et al., 2023), or SGC under heterophily (Chanpuriya & Musco, 2022). To our knowledge, none of these analyses gives one verdict for a whole propagation scheme or bounds how much a scheme can lose to no filter.

A fixed filter acts on every graph frequency separately. It multiplies the component of the features along each eigenvector of $\mathbf{W}$ by $g(\mu)$, where μ is the eigenvalue. The benefit of a filter should therefore depend on where the class signal and the noise lie in the spectrum. A scheme should be judged by the order in which its filters weight the spectrum (Figure 1).

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.

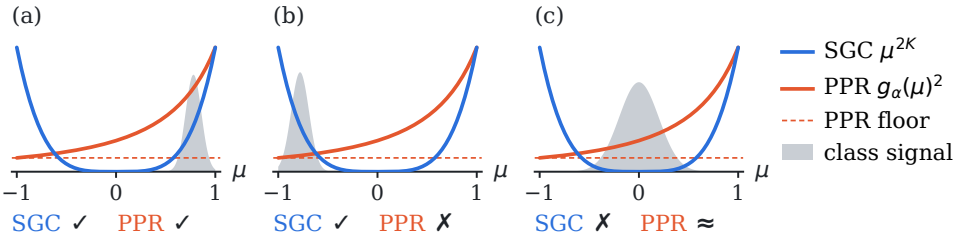

Figure 1: Squared responses of SGC (μ^{2K}, ordered by $|\mu|$) and PPR ($g_\alpha(\mu)^2$, ordered by μ; dashed: its floor) for three placements of the class signal (grey) under white noise. (a) Signal near $\mu = 1$ favours both schemes for moderate K and α. (b) Signal near $\mu = -1$ is weighted by every power as if near $\mu = 1$ (up to sign for odd K) and damped by every PPR filter. (c) Signal in the middle of the spectrum is removed by SGC, whereas the floor bounds the loss of PPR.

We make this precise for a linear feature model with class means and an arbitrary noise covariance. A filter rescales the class signal and the noise of each frequency by the same factor $g(\mu)^2$. The *signal excess* measures how far the class signal exceeds the noise at each frequency. The signal-to-noise ratio (SNR) of the filtered features is at least that of the raw features if and only if $g(\mu)^2$ has non-negative covariance with the signal excess (Theorem 1). Every PPR filter weights frequencies in the order of μ, and every SGC filter in the order of $|\mu|$. If the class signal first-order stochastically dominates the noise along a scheme's order, no member of the scheme loses to no filter (Theorem 2). This requires the top eigenspace of the order to carry at least as large a share of the class signal as of the noise. Because the orders differ, the two schemes can reach opposite verdicts on the same graph (Corollary 1). A PPR filter also passes every frequency with gain at least $\alpha/(2-\alpha)$, which bounds its SNR ratio to the identity from both sides. Powers of $\mathbf{W}$ can lose the whole SNR (Proposition 2). SGC and PPR are chosen deliberately: they realize both orders and both floor cases, and heat kernels, lazy and high-pass powers and DGC filters share one of these orders or its reverse (Appendix E). A label-aligned refinement accounts for filters that mix classes, with a one-step homophily threshold of order $1/\sqrt{D}$ on D-regular graphs and a floor for every α (Propositions 3 and 4).

On contextual stochastic block models, where the feature model holds, the sign of the criterion agrees with the realized gain in all 40 decisive comparisons of 44, those whose 95% interval excludes zero (Section 4.1). On 30 real graphs (16 standard benchmarks and 14 additional graphs), we tune SGC, PPR and an MLP in three rounds of splits, with equal budgets of 60 configurations and per-split validation selection (Section 4.2). Neither scheme dominates. PPR is significantly better on 15 graphs and SGC on 11, and over all rounds the largest margins are 12.93 accuracy points for PPR and 14.76 ROC-AUC points for SGC. In all seven graph and round pairs (four graphs) where SGC falls significantly below the MLP, PPR is significantly better than SGC. The worst loss of the selected PPR filter to the MLP, 1.74 points, is less than a quarter of that of SGC, 8.25 points.

Central finding. Choosing a propagation scheme amounts to choosing an order on the graph spectrum. In the feature model, one covariance decides whether a filter beats no filter. When the class signal dominates the noise along the order, the verdict holds for the whole scheme, and a response floor bounds how much the scheme can lose. This account adds the frequency orders and the floor asymmetry to two classical facts: the power spectral density of stationary graph signals (Perraudin & Vandergheynst, 2017) and first-order stochastic dominance. The theory is exact in the model, and block models confirm it. On real graphs its design consequences hold: no fixed scheme is safe everywhere, PPR stays close to no filter, and the identity removes every significant loss of SGC.

Contributions. (1) *Spectral criterion and frequency orders.* An exact spectral criterion for when a fixed filter lowers the SNR in a linear feature model. A certificate decides a whole scheme when the class signal dominates the noise along its frequency order. Both apply to any set of filters ordered by a score. (2) *Response floors and design rules.* Every PPR filter has a response floor that bounds its loss in SNR and in a label-aligned ratio, whereas adjacency powers have none. Two rules follow for decoupled models: search both frequency orders, and keep the identity in any scheme without a floor. (3) *Equal-budget evidence.* On block models, the criterion predicts the sign of every significant gain. On real graphs from several domains, with equal budgets, per-split validation and replication

on fresh splits and additional graphs, neither scheme dominates and PPR stays close to no filter. Validation over the MLP, SGC and PPR grids together is never significantly below any of them.

2 Related work

Decoupled and fixed-filter graph models. SGC, APPNP and SIGN separate a fixed propagation from the trained transformation (Wu et al., 2019; Gasteiger et al., 2019a; Frasca et al., 2020), and GDC replaces the input graph of a GNN by a precomputed diffusion (Gasteiger et al., 2019b). SGC and SIGN precompute the propagation, and a fixed low-pass filter followed by a linear classifier or a small MLP is competitive on homophilous benchmarks (Wu et al., 2019; NT et al., 2021). GraphAny combines several fixed linear filters with a learned attention (Zhao et al., 2025). In APPNP, the teleport term of PPR keeps a share of each node's own prediction (Gasteiger et al., 2019a). Initial residual connections play the same role in deep GNNs (Chen et al., 2020; Scholkemper et al., 2025), and SIGN and GPR-GNN include an unpropagated channel (Frasca et al., 2020; Chien et al., 2021). Our floor results quantify which schemes need the identity and why.

Learned filters and heterophily. Heterophily exposes the low-pass bias of message passing (Pei et al., 2020; Zhu et al., 2020). S^2GC and DGC analyse fixed SGC-type diffusions spectrally and retain part of each node's own features (Zhu & Koniusz, 2021; Wang et al., 2021). Adaptive models learn the frequency response through identity, low-pass and high-pass channels (Bo et al., 2021; Luan et al., 2022) or through polynomial and spectral parameterizations (Chien et al., 2021; He et al., 2021; 2022; Wang & Zhang, 2022; Bo et al., 2023), and GPR-GNN contains both SGC-type monomials and PPR. We explain why the fixed corners of this space behave differently.

When does propagation help? Dataset-level metrics indicate when graph filtering helps: feature and label smoothness (Hou et al., 2020), adjusted homophily and label informativeness (Platonov et al., 2023a), and classifier-based distinguishability (Luan et al., 2023). Feature-level and node-level analyses refine this question (Zheng et al., 2025; Mao et al., 2023). Homophily is not necessary for graph convolution to help (Ma et al., 2022), and on new heterophilous benchmarks standard GNNs almost always outperform specialized models (Platonov et al., 2023b). On the contextual stochastic block model (Deshpande et al., 2018) and its variants, several works characterize the separability that one or more graph convolutions yield (Baranwal et al., 2021; 2023; Wang et al., 2024). Wu et al. (2023) show that PPR propagation delays oversmoothing and can nevertheless lose to graph convolution. Chanpuriya & Musco (2022) show that even powers of the expected normalized adjacency of a two-block model preserve the class means in the limit of complete heterophily.

Spectral accounts. NT et al. (2021) argue that a low-pass filter denoises features when the class signal is low-frequency and the noise is white. Chen et al. (2022) lower-bound the error of spectral GNNs through the response efficiency of the filter on the label difference. In graph signal processing, stationary processes have a power spectral density in the eigenbasis of the operator, and Wiener filters weight each frequency by an increasing function of its signal-to-noise ratio (Perraudin & Vandergheynst, 2017; Marques et al., 2017). Our criterion uses these densities to compare a *given* fixed filter with the identity, and under its dominance condition it decides whole schemes.

3 When a fixed filter beats no filter

3.1 A spectral criterion

Operators and filters. Let $\mathbf{A}$ be the adjacency matrix of an undirected graph with n nodes and degree matrix $\mathbf{D}$. We consider the symmetric normalization $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ and the renormalized operator $\hat{\mathbf{W}} = \tilde{\mathbf{D}}^{-1/2}(\mathbf{A} + \mathbf{I})\tilde{\mathbf{D}}^{-1/2}$ with $\tilde{\mathbf{D}} = \mathbf{D} + \mathbf{I}$ (Kipf & Welling, 2017). Both are symmetric with spectrum in $[-1, 1]$ (Lemma 1 in Appendix A). We write $\mathbf{W}$ for either operator and $\mathbf{W} = \mathbf{U}\,\mathrm{diag}(\mu)\mathbf{U}^\top$ for an eigendecomposition with orthonormal eigenvectors $\mathbf{u}_k$. A filter is a function g on $[-1, 1]$, applied as $g(\mathbf{W}) = \mathbf{U}\,\mathrm{diag}(g(\mu))\mathbf{U}^\top$. A decoupled model computes $\mathbf{Z} = g(\mathbf{W})\mathbf{X}$ once and trains a head on the rows of $\mathbf{Z}$. We study three propagation schemes, each a one-parameter set of filters. SGC uses powers $g(\mu) = \mu^K$ (Wu et al., 2019). Personalized PageRank (PPR) uses $g_\alpha(\mu) = \alpha/(1 - (1 - \alpha)\mu)$, so that $g_\alpha(\mathbf{W}) = \alpha(\mathbf{I} - (1 - \alpha)\mathbf{W})^{-1}$; APPNP approximately applies it to a network's predictions (Gasteiger et al., 2019a). High-pass powers use $g(\mu) = ((1 - \mu)/2)^K$. The identity is $K = 0$ in the first and third scheme and $\alpha = 1$ in the second.

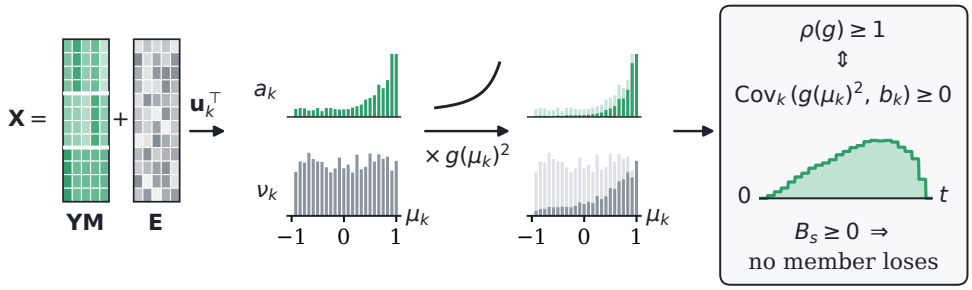

Figure 2: The criterion (illustrative spectrum). Projecting $\mathbf{X} = \mathbf{YM}+\mathbf{E}$ onto the eigenvectors $\mathbf{u}_k$ of $\mathbf{W}$ gives a density a_k of the centred class signal $\mathbf{S} = \mathbf{Y}_c\mathbf{M}$ and a noise density ν_k at each eigenvalue μ_k. A fixed filter reweights both by $g(\mu_k)^2$ (light bars: before filtering) and beats no filter exactly when $g(\mu_k)^2$ covaries positively with the signal excess b_k (Theorem 1). A non-negative cumulative excess B_s guarantees that no filter of an s-ordered set loses to no filter (Theorem 2).

Signal and noise. Let $\mathbf{Y} \in \{0,1\}^{n\times C}$ hold one-hot labels over C classes with class frequencies $\boldsymbol{\pi}$, and let $\mathbf{M} \in \mathbb{R}^{C\times d}$ hold the class means of the d features. We model the features as

$$\mathbf{X} = \mathbf{YM} + \mathbf{E}, \tag{1}$$

where the d columns of $\mathbf{E}$ are independent, have mean zero and share an arbitrary positive semidefinite node covariance $\sigma^2\mathbf{N}$ with $\sigma^2 > 0$ and $\operatorname{tr}\mathbf{N} = n$. The centred class signal is $\mathbf{S} = \mathbf{Y}_c\mathbf{M}$ with $\mathbf{Y}_c = \mathbf{Y} - \mathbf{1}\boldsymbol{\pi}^\top$. The remaining term $g(\mathbf{W})\mathbf{1}\boldsymbol{\pi}^\top\mathbf{M}$ moves every row along the single direction $\mathbf{M}^\top\boldsymbol{\pi}$ and involves the labels only through $\boldsymbol{\pi}$; we exclude it from the signal. A head could exploit it where degree correlates with the class; the SNR ignores this. For each eigen-index k we define the *signal density* $a_k = \|\mathbf{u}_k^\top\mathbf{S}\|_2^2$ and the *noise density* $\nu_k = (\mathbf{U}^\top\mathbf{N}\mathbf{U})_{kk}$, with mean $\bar\nu = 1$; $\bar a$ is the mean of the a_k. The *signal excess* is

$$b_k = a_k - (\bar a/\bar\nu)\,\nu_k, \qquad \textstyle\sum_k b_k = 0.$$

Individual a_k, ν_k and b_k depend on the choice of basis inside an eigenspace, but their sums over an eigenspace do not. Every result below depends on them only through such sums. We assume $\mathbf{S} \neq 0$.

Proposition 1 (SNR of a filtered representation)**.** *Let g satisfy $\sum_k g(\mu_k)^2\nu_k > 0$ and define* $\mathrm{SNR}(g) = \|g(\mathbf{W})\mathbf{S}\|_F^2/\mathbb{E}\|g(\mathbf{W})\mathbf{E}\|_F^2$. *Then*

$$\mathrm{SNR}(g) = \frac{1}{\sigma^2 d}\,\frac{\sum_k g(\mu_k)^2 a_k}{\sum_k g(\mu_k)^2\nu_k}.$$

A fixed filter reweights the signal and the noise at every frequency by the same factor $g(\mu_k)^2$ (Figure 2). Under white noise ($\mathbf{N} = \mathbf{I}$), $\mathrm{SNR}(g)$ is $\|\mathbf{S}\|_F^2/(\sigma^2 d)$ times the response efficiency (Chen et al., 2022) of g^2 on the class signal. A low-pass filter helps when the signal is low-frequency (NT et al., 2021). We compare a filter with no filter through the ratio $\rho(g) = \mathrm{SNR}(g)/\mathrm{SNR}(\mathbf{I})$.

Theorem 1 (Fixed filter versus no filter)**.** *Let g satisfy $\sum_k g(\mu_k)^2\nu_k > 0$. Then $\rho(g) \geq 1$ if and only if $\mathrm{Cov}_k\left(g(\mu_k)^2, b_k\right) \geq 0$, where the covariance is over an index k drawn uniformly from $\{1,\dots,n\}$. Equality holds on one side if and only if it holds on the other.*

3.2 One verdict for a whole scheme

Theorem 1 decides one filter at a time. Each standard scheme has one frequency order, and when the class signal dominates the noise along it, one verdict holds for all its hyperparameters.

Definition 1 (s-ordered filter set)**.** *A* score *is a function $s : [-1,1] \to \mathbb{R}$. A set $\mathcal{F}$ of filters is* s-ordered *if for every $g \in \mathcal{F}$ there is a non-decreasing function φ_g with $g(\mu_k)^2 = \varphi_g(s(\mu_k))$ for all k. For a value t of the score on the spectrum, $\bar b_s(t)$ is the average of b_k over $\{k : s(\mu_k) = t\}$. The* cumulative excess *is $B_s(t) = \sum_{k:\, s(\mu_k)\geq t} b_k$ for $t \in \mathbb{R}$.*

Each level set of s is a union of eigenspaces, so $\bar b_s$ and B_s do not depend on the basis. The cumulative excess compares two distributions over the score: $B_s(t) = \|\mathbf{S}\|_F^2\,\{P_a(s \geq t) - P_\nu(s \geq t)\}$, where P_a and P_ν put mass proportional to a_k and ν_k on eigen-index k.

Theorem 2 (Uniform verdict over an ordered filter set)**.** *Let $\mathcal{F}$ be s-ordered (Definition 1) and let $g \in \mathcal{F}$ satisfy $\sum_k g(\mu_k)^2 \nu_k > 0$.* ***(a)*** *If $B_s(t) \geq 0$ for every t, then $\mathrm{SNR}(g) \geq \mathrm{SNR}(\mathbf{I})$. This holds in particular if $\bar{b}_s$ is non-decreasing.* ***(b)*** *If $B_s(t) \leq 0$ for every t, then $\mathrm{SNR}(g) \leq \mathrm{SNR}(\mathbf{I})$. This holds in particular if $\bar{b}_s$ is non-increasing. In (a) and (b) the inequality is strict if φ_g is strictly increasing on the values of s on the spectrum and $\bar{b}_s \not\equiv 0$.* ***(c)*** *Conversely, if $B_s(t) < 0$ for some t, then some filter $\tilde{g}$ with $\tilde{g}^2 = \varphi \circ s$ on the spectrum, for a non-decreasing φ, has $\mathrm{SNR}(\tilde{g}) < \mathrm{SNR}(\mathbf{I})$.*

The condition in (a) says that the class signal is stochastically larger than the noise along the score. Part (c) shows that no weaker condition guarantees a win for every filter of this form. Since $\mathrm{SNR}(g)/\mathrm{SNR}(\mathbf{I}) = \mathbb{E}_{P_a}[g^2]/\mathbb{E}_{P_\nu}[g^2]$, parts (a) and (c) restate first-order stochastic dominance of P_a over P_ν along s. Our contribution is to identify the order of each scheme, which turns this classical fact into one verdict for all hyperparameters of SGC or PPR.

Corollary 1 (SGC, PPR and high-pass schemes)**.** *Consider filters with $\sum_k g(\mu_k)^2 \nu_k > 0$.* ***(i)*** *PPR is μ-ordered, SGC is $|\mu|$-ordered, and high-pass powers are $(-\mu)$-ordered. The maps φ_g are strictly increasing for $\alpha \in (0, 1)$ and $K \geq 1$.* ***(ii)*** *If $B_\mu(t) \leq 0$ for every t, no PPR filter beats no filter and no high-pass filter loses to it. The inequalities are strict for $\alpha < 1$ and $K \geq 1$ unless $\bar{b}_\mu \equiv 0$. The reverse verdicts hold if $B_\mu(t) \geq 0$ for every t.* ***(iii)*** *If $B_\mu(t) \leq 0$ and $B_{|\mu|}(t) \geq 0$ for every t, no SGC filter loses to no filter and no PPR filter beats it. For $K \geq 1$ and $\alpha < 1$ the inequalities are strict unless $\bar{b}_{|\mu|} \equiv 0$ or $\bar{b}_\mu \equiv 0$, respectively. Both hypotheses hold, for example, for $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ on a connected regular bipartite graph whose two sides are the two classes, under white noise.*

Part (iii) is consistent with the observation that graph convolution can help under strong heterophily (Ma et al., 2022; Chanpuriya & Musco, 2022): signal at $\mu = -1$ survives in μ^{2K} and is damped by every PPR filter. A scheme-wide win is harder to certify than a scheme-wide loss. On a connected graph that is not bipartite, the top level set of both scores is the single direction at $\mu = 1$, namely $\mathbf{D}^{1/2}\mathbf{1}$ for $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ and $\tilde{\mathbf{D}}^{1/2}\mathbf{1}$ for $\hat{\mathbf{W}}$. If this direction carries a smaller share of the class signal than of the noise, then $B_\mu(1) = B_{|\mu|}(1) < 0$. PPR as $\alpha \to 0$ and SGC as $K \to \infty$ converge to this direction and lose (Proposition 2(b,c) and Appendix A). On such graphs, Theorem 2 can certify a scheme-wide loss, as for PPR in Corollary 1(ii), but no scheme-wide win. The condition can be checked exactly from the spectrum. On Cora with $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ under white noise, $B_\mu(t) \geq 0$ and $B_{|\mu|}(t) \geq 0$ for every t, so no SGC or PPR filter loses to no filter in SNR. The 78 connected components of Cora contribute to this: the eigenspace at $\mu = 1$ has dimension 78 and carries 7.3% of the class signal and 2.9% of the white noise, so $B_\mu(1) > 0$.

3.3 Floors: how much a scheme can lose

Proposition 2 (Floors)**.** *Let $m_g = \min_k g(\mu_k)^2$ and $M_g = \max_k g(\mu_k)^2$.* ***(a)*** *If $m_g > 0$, then $m_g/M_g \leq \mathrm{SNR}(g)/\mathrm{SNR}(\mathbf{I}) \leq M_g/m_g$ for every signal and noise density.* ***(b)*** *For PPR with $\alpha \in (0, 1]$, $g_\alpha(\mu) \in [\alpha/(2-\alpha), 1]$ on $[-1, 1]$, hence*

$$\Big(\frac{\alpha}{2-\alpha}\Big)^2 \leq \frac{\mathrm{SNR}(g_\alpha)}{\mathrm{SNR}(\mathbf{I})} \leq \Big(\frac{2-\alpha}{\alpha}\Big)^2, \qquad \sup_{\mu \in [-1,1]} |g_\alpha(\mu) - 1| = \frac{2(1-\alpha)}{2-\alpha}.$$

If $\mathcal{T}_1$ is the set of indices with $\mu_k = 1$ and $\sum_{k \in \mathcal{T}_1} \nu_k > 0$, then $\mathrm{SNR}(g_\alpha) \to \sum_{k \in \mathcal{T}_1} a_k/(\sigma^2 d \sum_{k \in \mathcal{T}_1} \nu_k)$ as $\alpha \to 0$. ***(c)*** *For SGC with $K \geq 1$ under white noise, the infimum of $\rho(\mathbf{W}^K)$ over all non-negative densities $a \neq 0$, realizable or not, equals $n \min_k \mu_k^{2K} / \sum_k \mu_k^{2K}$. It is zero if 0 is an eigenvalue and tends to zero as $K \to \infty$ unless all $|\mu_k|$ are equal. Moreover, for any noise, if $\mathcal{T}$ is the set of indices with $|\mu_k| = \max_j |\mu_j|$ and $\sum_{k \in \mathcal{T}} \nu_k > 0$, then $\mathrm{SNR}(\mathbf{W}^K) \to \sum_{k \in \mathcal{T}} a_k/(\sigma^2 d \sum_{k \in \mathcal{T}} \nu_k)$ as $K \to \infty$.*

Part (c) makes the lack of a floor concrete. Actual class signals attain the value zero: on a complete bipartite graph whose two classes split both sides in half, the label vector $\mathbf{y} \in \{\pm 1\}^n$ satisfies $\mathbf{W}\mathbf{y} = 0$ for $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, so $\mathrm{SNR}(\mathbf{W}^K) = 0$ for every $K \geq 1$ (Appendix A). On a connected, regular, non-bipartite graph, $\mathcal{T} = \{k : \mathbf{u}_k \propto \mathbf{1}\}$ and $\mathbf{1}^\top \mathbf{S} = 0$. Hence $\mathrm{SNR}(\mathbf{W}^K) \to 0$ under white noise while $\mathrm{SNR}(\mathbf{I}) > 0$. This is oversmoothing seen through the SNR (Li et al., 2018; Oono & Suzuki, 2020; Keriven, 2022). High-pass powers also have no floor, since $g(1) = 0$.

Consequence for the search grid. In SNR, a grid that contains the identity never loses to no filter, and the best member of a PPR grid reaching $\alpha_{\max} = 0.9$ keeps a log ratio of at least $2\ln(9/11) \approx -0.401$. An SGC grid with $K \geq 1$ has no such guarantee.

3.4 Class mixing: a label-aligned refinement

The SNR counts all filtered class energy as signal. A filter can move the class signal of one node onto nodes of another class. The label-aligned ratio below counts as signal only the part of the filtered class signal that each node's own label explains. We therefore split the filtered signal into the part explained by each node's own label and a residual.

Definition 2 (Label-aligned ratio). *Let $\mathbf{P} = \mathbf{Y}_c(\mathbf{Y}_c^\top \mathbf{Y}_c)^+ \mathbf{Y}_c^\top$ be the orthogonal projection onto the column space of $\mathbf{Y}_c$, where $^+$ is the pseudo-inverse. With $\mathcal{A}(g) = \|\mathbf{P}g(\mathbf{W})\mathbf{S}\|_F^2$ and $\mathbf{R}(g) = (\mathbf{I} - \mathbf{P})g(\mathbf{W})\mathbf{S}$,*

$$F(g) = \frac{\mathcal{A}(g)}{\|\mathbf{R}(g)\|_F^2 + \mathbb{E}\|g(\mathbf{W})\mathbf{E}\|_F^2}.$$

The ratio F is a trace-ratio (Fisher-type) between-to-within criterion for the filtered rows. It satisfies $F(\mathbf{I}) = \mathrm{SNR}(\mathbf{I})$ because $\mathbf{PS} = \mathbf{S}$.

Proposition 3 (One step, two classes). *Let $C = 2$ with balanced classes, class means $\pm\gamma\mathbf{m}$ with $\|\mathbf{m}\| = 1$, and $\mathbf{N} = \mathbf{I}$. Let $\mathbf{y} \in \{\pm 1\}^n$ encode the labels and set $\beta = \mathbf{y}^\top \mathbf{W}\mathbf{y}/n$, $\xi^2 = \|\mathbf{W}\mathbf{y}\|^2/n - \beta^2$, $\tau = \mathrm{tr}(\mathbf{W}^2)/n$ and $\varepsilon = \gamma^2/(\sigma^2 d)$. Then $F(\mathbf{W}) \geq F(\mathbf{I})$ if and only if $\beta^2 \geq \tau + \varepsilon\xi^2$. Consequently $\beta^2 < \tau$ implies $F(\mathbf{W}) < F(\mathbf{I})$, and $\beta^2 \geq (\tau + \varepsilon)/(1 + \varepsilon)$ implies $F(\mathbf{W}) \geq F(\mathbf{I})$. On a D-regular graph with $\mathbf{W} = \mathbf{A}/D$ and edge homophily h, $\beta = 2h - 1$ and $\tau = 1/D$ exactly.*

For weak signals ($\varepsilon \ll 1$), one propagation step therefore helps essentially when $|2h - 1| \geq 1/\sqrt{D}$. On a two-class block model with within-class and between-class edge probabilities p_{in} and p_{out}, $\beta \approx (p_{\mathrm{in}} - p_{\mathrm{out}})/(p_{\mathrm{in}} + p_{\mathrm{out}}) = 2h - 1$, so this threshold has the $1/\sqrt{D}$ scaling of the separability gain of graph convolution (Baranwal et al., 2021), and the same parameter governs SGC under heterophily (Chanpuriya & Musco, 2022). In contrast, $\mathrm{SNR}(\mathbf{W})/\mathrm{SNR}(\mathbf{I}) = (\beta^2 + \xi^2)/\tau$, which is close to one when $\beta \approx 0$ on a random graph (Appendix B). In our block-model experiment at $h = 0.5$ with $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, SGC with $K = 2$ selected on validation loses 6.2 points $[-9.8, -3.8]$, although its log SNR ratio is only -0.05. Its log ratio $\log F(\mathbf{W}^2)/F(\mathbf{I})$ is -0.71 (Appendix B).

The floor of PPR also holds for the label-aligned ratio, because every PPR filter satisfies $g_\alpha(\mathbf{W}) \succeq \ell_\alpha \mathbf{I}$ with $\ell_\alpha = \alpha/(2 - \alpha)$.

Proposition 4 (Label-aligned floor). *Let $\sigma^2 > 0$ and $\ell_\alpha = \alpha/(2 - \alpha)$. For every noise covariance, every number of classes and $\alpha \in (0, 1]$,*

$$\frac{F(g_\alpha)}{F(\mathbf{I})} \geq \frac{\ell_\alpha^2}{1 + \frac{1}{4}(1 - \ell_\alpha)^2 F(\mathbf{I})}.$$

No such bound holds for SGC: $\mathcal{A}(\mathbf{W}^K) = 0$ is possible for $K \geq 1$.

4 Experiments

We examine the theory in two ways. **(C1)** On block models, where the feature model holds, the criterion predicts the sign of the gain over no filter, and the two schemes can reach opposite verdicts. **(C2)** On real graphs we test three design consequences: no fixed scheme is safe on every graph, PPR stays close to no filter, and the identity protects a scheme without a floor.

4.1 Controlled test: the two schemes on block models

We sample two-class contextual stochastic block models (Deshpande et al., 2018) with $n = 2000$ nodes, expected degree 10, edge homophily $h \in \{0.02, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9, 0.95, 0.98\}$ and Gaussian features $\mathbf{X} = \mathbf{Y}\mathbf{M} + \mathbf{E}$ in $d = 64$ dimensions. The class separation gives a linear head on raw features about 70% accuracy (Appendix C). For each scheme we select the hyperparameter ($K \in \{1, 2, 3, 4\}$ for SGC,

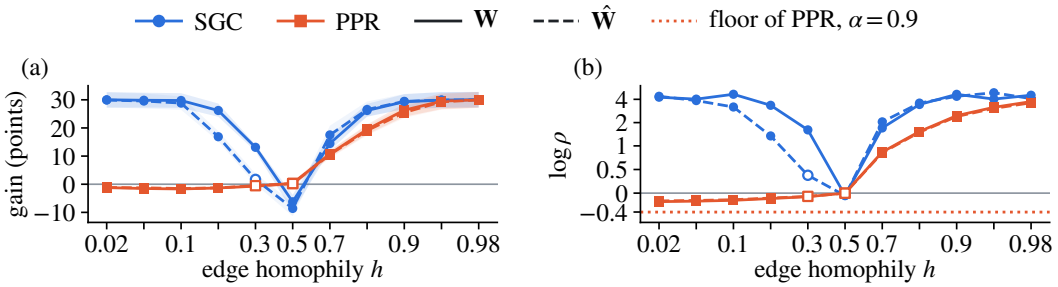

Figure 3: The two schemes on contextual stochastic block models (five instances per cell; solid: $\mathbf{W}$, dashed: $\hat{\mathbf{W}}$). (a) Gain over raw features of the validation-selected SGC and PPR filters, with 95% bootstrap bands. (b) Log SNR ratio $\log \rho$ of the selected filter under white noise (symmetric log scale), averaged over the instances; dotted: the floor $2\ln(9/11)$ of PPR with $\alpha = 0.9$. Hollow markers: cells whose interval contains zero. In every other cell, the signs in (a) and (b) agree.

$\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ for PPR) and the regularization of a logistic head on validation, with five instances per cell.

Figure 3 contrasts the two schemes. We call a cell *decisive* when the 95% bootstrap interval of its gain over raw features across the five instances excludes zero. On near-bipartite graphs ($h \leq 0.2$) the class signal sits at the negative end of the spectrum, which every PPR filter damps (Corollary 1(iii)). There SGC with $K \leq 4$ gains 16.9 to 30.0 points over raw features, on all four cells and for both normalizations. In every such instance, every PPR filter of the grid has a lower SNR than the raw features. The validation-selected PPR filter is always $\alpha = 0.9$, the one closest to the identity, and it loses 1.2 to 1.7 points. Its log SNR ratio, between -0.18 and -0.10, stays above the floor $2\ln(9/11) \approx -0.401$ of Proposition 2(b). On homophilous graphs ($h \geq 0.7$) the class signal lies in the positive half of the spectrum, where the two orders agree. Both schemes then gain significantly in every cell, SGC by 14.5 to 30.0 points and PPR by 10.5 to 30.0 points, and PPR selects $\alpha = 0.1$ in every instance. At $h = 0.7$, PPR with $\mathbf{W}$ gains 10.8 points, whereas at the mirror value $h = 0.3$ it selects $\alpha = 0.9$ and loses 0.7 points. Across the 40 decisive cells (of 44) of both schemes and both operators, the sign of the criterion at the selected filter agrees with the sign of the gap. At the level of single instances it agrees in 207 of 220, and every disagreement occurs at $h \in \{0.3, 0.5\}$. With $\hat{\mathbf{W}}$ the contrast is weaker, as Remark 1 (Appendix A) predicts: at $h = 0.2$ SGC gains 16.9 points against 26.2 with $\mathbf{W}$. At $h = 0.5$ the graph carries no class structure, and the loss of SGC is the class-mixing case of Section 3.4, which the label-aligned ratio predicts.

4.2 Real graphs

Datasets. We use 30 node-classification graphs. The 16 standard benchmarks are Cora, CiteSeer and PubMed (Yang et al., 2016); Amazon-Computers, Amazon-Photo, Coauthor-CS and Coauthor-Physics (Shchur et al., 2018); WikiCS (Mernyei & Cangea, 2020); roman-empire, amazon-ratings, minesweeper, tolokers and questions (Platonov et al., 2023b); the filtered chameleon and squirrel graphs of Platonov et al. (2023b), originally from Rozemberczki et al. (2021); and Actor (Pei et al., 2020). The 14 additional graphs are five Facebook100 social networks with gender labels (Traud et al., 2012; Lim et al., 2021), six Twitch networks and the GitHub network (Rozemberczki et al., 2021), the DBLP citation graph (Bojchevski & Günnemann, 2018) and the Wiki hyperlink graph (Yang et al., 2015). Every graph is made undirected without self-loops. Graphs with binary labels are scored by ROC-AUC and the others by accuracy (Table 2, Appendix C).

Protocol. Let Λ be the set of labelled nodes of a graph and let $r = 1, \ldots, R$ with $R = 10$ index splits of Λ into training, validation and test sets (T_r, V_r, Q_r). An arm $\omega \in \{\mathrm{MLP}, \mathrm{SGC}, \mathrm{PPR}\}$ has a candidate set Θ_ω of 60 configurations $\theta = (\eta, g)$ of a head η and a filter g. The MLP uses $g \equiv 1$, that is, $g(\mathbf{W}) = \mathbf{I}$. SGC uses $g(\mu) = \mu^K$ with $K \in \{1, 2, 3, 4, 6, 8\}$, and PPR uses g_α with $\alpha \in \{0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9\}$, each with the symmetric normalization $\mathbf{W}$ or with $\hat{\mathbf{W}}$. The head η is an MLP of depth 2 or 3. The 60 heads are the same for the three arms. The filter of each SGC or PPR configuration is drawn at random from its grid, so every arm trains the same number

Table 1: Gaps $\Delta(\omega, \omega')$ in points (accuracy; ROC-AUC$\times$100 for †); grey entries are not significant. Red rows: SGC significantly below the MLP. Blue rows: SGC significantly above PPR by at least 5 points. Grey column: both design rules together, by validation over the MLP, SGC and PPR grids (180 candidates). Intervals and absolute scores: Appendix D.

(a) Round II	Gap to the MLP				(b) Round III	Gap to the MLP			
Graph	SGC	PPR	Rules	PPR−SGC	Graph	SGC	PPR	Rules	PPR−SGC
Cora	+10.9	+12.1	+11.8	+1.1	Penn94†	+6.4	+0.4	+6.4	−6.1
CiteSeer	+3.0	+3.4	+3.4	+0.4	Cornell5†	+8.6	+1.8	+8.6	−6.8
PubMed	+0.4	+1.9	+1.9	+1.5	JohnsHopkins55†	+14.6	+5.4	+14.6	−9.1
Computers	+6.1	+5.8	+6.1	−0.3	Amherst41†	+21.2	+6.4	+21.2	−14.8
Photo	+2.8	+3.7	+3.7	+0.9	Reed98†	+8.5	−1.2	+8.5	−9.7
Coauthor-CS	−0.8	+1.1	+1.1	+1.9	Twitch-DE†	+4.5	+3.1	+4.5	−1.4
Physics	+0.4	+0.8	+0.8	+0.5	Twitch-EN†	+2.4	+2.5	+2.5	+0.1
WikiCS	+3.5	+3.9	+3.9	+0.5	Twitch-ES†	+4.2	+2.4	+4.2	−1.8
roman-empire	−5.5	+7.3	+7.3	+12.8	Twitch-FR†	+4.6	+2.1	+4.6	−2.5
amazon-ratings	+1.3	+2.8	+2.8	+1.6	Twitch-PT†	+5.8	+3.7	+5.5	−2.1
minesweeper†	+35.6	+37.6	+37.6	+2.0	Twitch-RU†	+2.0	+1.4	+1.9	−0.6
tolokers†	+5.3	+6.1	+6.0	+0.8	GitHub†	+1.6	+1.7	+1.7	+0.1
questions†	+5.8	+6.6	+6.5	+0.8	DBLP	+8.3	+8.8	+8.8	+0.5
chameleon-f.	−1.3	−0.3	−0.6	+1.1	Wiki	−4.7	+2.7	+2.7	+7.4
squirrel-f.	+4.7	+1.7	+4.7	−3.0					
Actor	−8.0	−1.6	+0.0	+6.4					
Minimum	−8.0	−1.6	−0.6	−3.0	Minimum	−4.7	−1.2	**+1.7**	−14.8
# sig. < 0	3	1	**0**	2	# sig. < 0	1	0	**0**	9

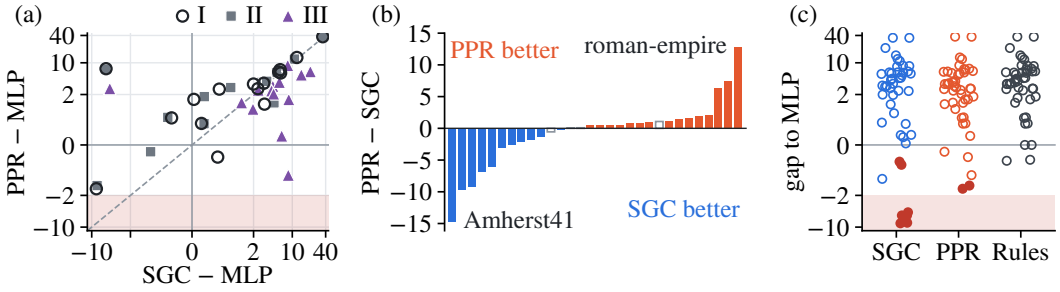

Figure 4: (a) Gain of tuned SGC and PPR over the tuned MLP per graph and round, in points (46 pairs; symmetric log scale). No PPR point lies below -2 points (shaded), whereas SGC reaches -8.25. (b) PPR$-$SGC on the 30 graphs of Rounds II and III, sorted; filled bars are significant. (c) Gap to the MLP of SGC, PPR and both design rules together (Rules, as in Table 1) over the 46 pairs, on the scale of (a); red: significantly negative. Rules has none.

of configurations. Let $\mathrm{val}_r(\theta)$ and $\mathrm{test}_r(\theta)$ be the validation and test metrics of θ trained on T_r. On each split, arm ω selects a configuration from that split's validation metric alone. The gap between arms ω and ω' averages the test differences:

$$\hat{\theta}_{\omega,r} \in \arg\max_{\theta \in \Theta_\omega} \mathrm{val}_r(\theta), \qquad \Delta(\omega, \omega') = \frac{1}{R} \sum_{r=1}^{R} \big(\mathrm{test}_r(\hat{\theta}_{\omega,r}) - \mathrm{test}_r(\hat{\theta}_{\omega',r})\big).$$

Ties are broken by a fixed order, and gaps are in points. Appendix C documents the implementation details, including the split, initialization and bootstrap seeds. The interval of $\Delta(\omega, \omega')$ is the 95% percentile interval of 10^4 paired bootstrap resamples of the R splits, and $\Delta(\omega, \omega')$ is *significant* when this interval excludes zero. The labels of Q_r never enter the selection on split r, and selecting one configuration per arm by its mean validation metric over the splits confirms the scheme comparisons (Appendix D). **Round I** evaluates the 16 benchmarks on the ten official splits of the seven graphs of Platonov et al. (2023b) and on random 50/25/25 splits of the other nine. **Round II** repeats Round I on fresh random 50/25/25 splits of the same graphs. **Round III** evaluates the 14 additional graphs on random 50/25/25 splits of their labelled nodes. Table 1 reports Rounds II and III (Round I: Appendix D).

The criterion on real graphs. The spectral criterion is an SNR diagnostic within the feature model, computed with the class means of all labelled nodes (Appendix C). We average $\log \rho(g)$ under white noise over the filters selected on the ten splits. Its sign agrees with the sign of the significant gap to the MLP in 81 of 86 cells, against 77 for the constant prediction that filtering helps. It misses five cells: SGC on Coauthor-CS in Rounds I and II and on Wiki in Round III, and PPR on roman-empire in Rounds I and II. The difference of the two schemes' $\log \rho$ orders them correctly in 15 of the 38 cells where they differ significantly. Our real-graph evaluation therefore targets the design consequences of the theory, and the criterion itself serves as a diagnostic.

No fixed scheme is safe on every graph. Table 1 and Figure 4 report the gaps. Over the 30 graphs of Rounds II and III, PPR is significantly better than SGC on 15 and SGC than PPR on 11, by at least one point on 8 and 10 of them. The largest margins point in opposite directions: PPR leads by 12.93 points on roman-empire (Round I), and SGC leads by 14.76 points on Amherst41. On all five Facebook100 graphs, SGC exceeds PPR by 6.1 to 14.8 points and the MLP by 6.4 to 21.2 points. In Rounds II and III, PPR leads by 6.4 to 12.8 points on roman-empire, Actor and Wiki.

The two orders on Facebook100. In these networks, gender shows weak homophily but strong monophily, which makes friends of friends similar in gender (Altenburger & Ugander, 2018). Let $\mathbf{\Pi}_-$ be the orthogonal projector onto the eigenvectors of $\mathbf{W}$ with $\mu_k < 0$ and let $\mathbf{X}^- = \mathbf{\Pi}_-\mathbf{X}$. Every filter commutes with $\mathbf{\Pi}_-$, so Theorem 1 decides each filter on $\mathbf{X}^-$ with a_k and ν_k restricted to $\mu_k < 0$ (Appendix D). On this band every PPR response lies in $[\alpha/(2-\alpha), \alpha)$, whereas μ^{2K} grows towards $\mu = -1$. On the five Facebook100 graphs under white noise, the exact spectra give $\rho(g_\alpha) < 1$ on $\mathbf{X}^-$ for all seven α, whereas the SGC grid contains powers with $\rho(\mathbf{W}^K) > 1$ ($K = 1, 2$) on every graph, so only the SGC grid can raise the SNR. Retrained on $\mathbf{X}^-$ with the operator $\mathbf{W}$ (Table 8), SGC is significantly above the MLP on all five graphs, by 4.34 to 15.48 points, and significantly above PPR. PPR is significantly above the MLP on the original features of four of these graphs, but on $\mathbf{X}^-$ of none. There, validation selects $\alpha = 0.9$ on 29 of the 50 splits, and this filter keeps the SNR within 2% of the identity. On $\mathbf{X}^+ = \mathbf{X} - \mathbf{X}^-$, where the two orders agree, SGC gains significantly on all five graphs and PPR on four.

PPR stays close to no filter. Over all three rounds (46 graph and round pairs on 30 graphs), the validation-selected PPR filter falls at most 1.74 points below the MLP (Actor, Round I). SGC falls by up to 8.25 points on the same graph. SGC is significantly below the MLP in seven pairs on four graphs: Coauthor-CS, roman-empire and Actor in Rounds I and II, and Wiki in Round III. In each of them PPR is significantly better than SGC, by 1.7 to 12.9 points. These losses persist when each scheme's most frequently selected filter gets all 60 heads, as the MLP does: SGC stays significantly below the MLP in all seven pairs, and no significant scheme comparison changes sign (Table 7).

The identity protects a scheme without a floor. By Proposition 2(b), a PPR grid that reaches $\alpha = 0.9$ contains a filter with response at least $9/11 \approx 0.818$ at every frequency, whereas every power of $\mathbf{W}$ vanishes at $\mu = 0$. For SGC, the remedy at the level of the grid is to keep the identity among the candidates. For a scheme f we select on validation over $\Theta_f \cup \Theta_{\mathrm{MLP}}$, that is, 120 candidates. Over the 92 scheme, graph and round cells, this search removes all 9 significant losses to the MLP, and its only significant cost against the scheme alone is 0.29 points (Appendix D). Validation over $\Theta_{\mathrm{MLP}} \cup \Theta_{\mathrm{SGC}} \cup \Theta_{\mathrm{PPR}}$, at three times the budget of one arm, is significantly above the MLP in 42 of the 46 pairs and never significantly below the MLP, SGC or PPR (Table 1; Appendix D).

5 Conclusion

Choosing a propagation scheme amounts to choosing an order on the graph spectrum. Within a linear feature model, whether a fixed filter beats no filter in signal-to-noise ratio is decided by one covariance between its squared response and the excess of class signal over noise. Whenever the class signal dominates the noise along a scheme's order, this covariance gives one verdict for all its hyperparameters. Adjacency powers and personalized PageRank order the spectrum differently, so in the model neither can dominate the other. Of the two, only personalized PageRank has a response floor that bounds its signal-to-noise loss. Block models and a spectral intervention on social networks show the two orders at work. Real graphs from several domains agree with the resulting design rules: search both frequency orders, and keep the identity in any scheme without a floor. For representation learning on graphs, these results provide a spectral lens for whole propagation schemes and a reproducible equal-budget evaluation of fixed filters.

REPRODUCIBILITY STATEMENT

All proofs are in Appendices A and B, and all protocols, grids, seeds and selection rules are in Appendices C and D. The supplementary material contains the code for data loading, feature precomputation, training, selection and analysis. Its scripts run every experiment of Section 4 and Appendix D, including the matched head budget and the band intervention. They also compute the spectral quantities of Sections 3.2 and 4.2 and produce the tables and figures that report the experiments.

AI USE STATEMENT

As guided by the conference AI disclosure policy, in this work, we used generative AI tools to assist in the writing of proofs, design or provide feedback on research methodology or experiments, implement methods, assist with translation, clean and reformat datasets, formulate mathematical claims, interpret results, and propose or refine hypotheses. We have not used generative AI tools to generate synthetic datasets, help develop theoretical models or conceptual frameworks, or provide critical ingredients for proving mathematical claims. Additionally, we also used generative AI tools to search related works, help authors create or modify scientific figures or images, help authors build experiment pipelines, help improve the writing quality and clarity of the manuscript, edit a research paper to improve readability, sourcing/searching for information and check for potential plagiarism, brainstorming, identify relevant literature, format references. Parts of the manuscript were first drafted with generative AI tools; all text was reviewed and edited by human authors. And all AI-assisted work has been reviewed and verified by human authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Kristen M. Altenburger and Johan Ugander. Monophily in social networks introduces similarity among friends-of-friends. *Nature Human Behaviour*, 2(4):284–290, 2018. doi: 10.1038/s41562-018-0321-8.

Aseem Baranwal, Kimon Fountoulakis, and Aukosh Jagannath. Graph convolution for semi-supervised classification: Improved linear separability and out-of-distribution generalization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 684–693, 2021. URL https://proceedings.mlr.press/v139/baranwal21a.html.

Aseem Baranwal, Kimon Fountoulakis, and Aukosh Jagannath. Effects of graph convolutions in multi-layer networks. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=P-73JPgRs0R.

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 3950–3957, 2021.

Deyu Bo, Chuan Shi, Lele Wang, and Renjie Liao. Specformer: Spectral graph neural networks meet transformers. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=0pdSt3oyJa1.

Aleksandar Bojchevski and Stephan Günnemann. Deep Gaussian embedding of graphs: Unsupervised inductive learning via ranking. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=r1ZdKJ-0W.

Sudhanshu Chanpuriya and Cameron Musco. Simplified graph convolution with heterophily. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27184–27197, 2022.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1725–1735, 2020.

Zhixian Chen, Tengfei Ma, and Yang Wang. When does a spectral graph neural network fail in node classification? arXiv preprint arXiv:2202.07902, 2022. URL https://arxiv.org/abs/2202.07902.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized PageRank graph neural network. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=n6jl7fLxrP.

Yash Deshpande, Subhabrata Sen, Andrea Montanari, and Elchanan Mossel. Contextual stochastic block models. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Fabrizio Frasca, Emanuele Rossi, Davide Eynard, Ben Chamberlain, Michael Bronstein, and Federico Monti. SIGN: Scalable inception graph neural networks. arXiv preprint arXiv:2004.11198, 2020. URL https://arxiv.org/abs/2004.11198.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized PageRank. In *International Conference on Learning Representations*, 2019a.

Johannes Gasteiger, Stefan Weißenberger, and Stephan Günnemann. Diffusion improves graph learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019b.

Mingguo He, Zhewei Wei, Zengfeng Huang, and Hongteng Xu. BernNet: Learning arbitrary graph spectral filters via Bernstein approximation. In *Advances in Neural Information Processing Systems*, volume 34, 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 *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/2f9b3ee2bcea04b327c09d7e3145bd1e-Abstract-Conference.html.

Yifan Hou, Jian Zhang, James Cheng, Kaili Ma, Richard T. B. Ma, Hongzhi Chen, and Ming-Chang Yang. Measuring and improving the use of graph information in graph neural networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rkeIIkHKvS.

Nicolas Keriven. Not too little, not too much: a theoretical analysis of graph (over)smoothing. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2268–2281, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/0f956ca6f667c62e0f71511773c86a59-Abstract-Conference.html.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11604.

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 *Advances in Neural Information Processing Systems*, volume 34, pp. 20887–20902, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/ae816a80e4c1c56caa2eb4e1819cbb2f-Abstract.html.

Sitao Luan, Chenqing Hua, Qincheng Lu, Jiaqi Zhu, Mingde Zhao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. Revisiting heterophily for graph neural networks. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/092359ce5cf60a80e882378944bf1be4-Abstract-Conference.html.

Sitao Luan, Chenqing Hua, Minkai Xu, Qincheng Lu, Jiaqi Zhu, Xiao-Wen Chang, Jie Fu, Jure Leskovec, and Doina Precup. When do graph neural networks help with node classification? Investigating the homophily principle on node distinguishability. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/5ba11de4c74548071899cf41dec078bf-Abstract-Conference.html.

Yao Ma, Xiaorui Liu, Neil Shah, and Jiliang Tang. Is homophily a necessity for graph neural networks? In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=ucASPPD9GKN.

Haitao Mao, Zhikai Chen, Wei Jin, Haoyu Han, Yao Ma, Tong Zhao, Neil Shah, and Jiliang Tang. Demystifying structural disparity in graph neural networks: Can one size fit all? In *Advances in Neural Information Processing Systems*, volume 36, pp. 37013–37067, 2023.

Antonio G. Marques, Santiago Segarra, Geert Leus, and Alejandro Ribeiro. Stationary graph processes and spectral estimation. *IEEE Transactions on Signal Processing*, 65(22):5911–5926, 2017.

Péter Mernyei and Cătălina Cangea. Wiki-CS: A Wikipedia-based benchmark for graph neural networks. arXiv preprint arXiv:2007.02901, 2020. URL https://arxiv.org/abs/2007.02901.

Hoang NT, Takanori Maehara, and Tsuyoshi Murata. Revisiting graph neural networks: Graph filtering perspective. In *25th International Conference on Pattern Recognition (ICPR)*, pp. 8376–8383, 2021. doi: 10.1109/ICPR48806.2021.9412278.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1ldO2EFPr.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-GCN: Geometric graph convolutional networks. In *International Conference on Learning Representations*, 2020.

Nathanaël Perraudin and Pierre Vandergheynst. Stationary signal processing on graphs. *IEEE Transactions on Signal Processing*, 65(13):3462–3477, 2017.

Oleg Platonov, Denis Kuznedelev, Artem Babenko, and Liudmila Prokhorenkova. Characterizing graph datasets for node classification: Homophily-heterophily dichotomy and beyond. In *Advances in Neural Information Processing Systems*, volume 36, 2023a. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/01b681025fdbda8e935a66cc5bb6e9de-Abstract-Conference.html.

Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of GNNs under heterophily: Are we really making progress? In *International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=tJbbQfw-5wv.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2):cnab014, 2021.

Michael Scholkemper, Xinyi Wu, Ali Jadbabaie, and Michael Schaub. Residual connections and normalization can provably prevent oversmoothing in GNNs. In *International Conference on Learning Representations*, 2025.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. arXiv preprint arXiv:1811.05868, 2018. URL https://arxiv.org/abs/1811.05868.

Amanda L. Traud, Peter J. Mucha, and Mason A. Porter. Social structure of Facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180, 2012. doi: 10.1016/j.physa.2011.12.021.

Junfu Wang, Yuanfang Guo, Liang Yang, and Yunhong Wang. Understanding heterophily for graph neural networks. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 50489–50529, 2024. URL https://proceedings.mlr.press/v235/wang24u.html.

Xiyuan Wang and Muhan Zhang. How powerful are spectral graph neural networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 2022. URL https://proceedings.mlr.press/v162/wang22am.html.

Yifei Wang, Yisen Wang, Jiansheng Yang, and Zhouchen Lin. Dissecting the diffusion process in linear graph convolutional networks. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International Conference on Machine Learning*, pp. 6861–6871, 2019.

Xinyi Wu, Zhengdao Chen, William Wei Wang, and Ali Jadbabaie. A non-asymptotic analysis of oversmoothing in graph neural networks. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=CJd-BtnwtXq.

Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y. Chang. Network representation learning with rich text information. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2111–2117, 2015. URL https://www.ijcai.org/Abstract/15/299.

Renchi Yang, Jieming Shi, Xiaokui Xiao, Yin Yang, Juncheng Liu, and Sourav S. Bhowmick. Scaling attributed network embedding to massive graphs. *Proceedings of the VLDB Endowment*, 14 (1):37–49, 2021. doi: 10.14778/3421424.3421430.

Zhilin Yang, William Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In *International Conference on Machine Learning*, volume 48, pp. 40–48, 2016.

Jianan Zhao, Zhaocheng Zhu, Mikhail Galkin, Hesham Mostafa, Michael Bronstein, and Jian Tang. Fully-inductive node classification on arbitrary graphs. In *International Conference on Learning Representations*, pp. 23968–23986, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/3bdeb28a531f7af94b56bcdf8ee88f17-Abstract-Conference.html.

Yilun Zheng, Xiang Li, Sitao Luan, Xiaojiang Peng, and Lihui Chen. Let your features tell the differences: Understanding graph convolution by feature splitting. In *International Conference on Learning Representations*, 2025.

Hao Zhu and Piotr Koniusz. Simple spectral graph convolution. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=CYO5T-YjWZV.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

A PROOFS

Throughout, $x_k = g(\mu_k)^2 \geq 0$ and $\kappa = \bar{a}/\bar{\nu}$. We use $\sum_k \nu_k = \operatorname{tr} \mathbf{N} = n > 0$ and $\sum_k a_k = \|\mathbf{S}\|_F^2 > 0$. For a set $\mathcal{L}$ of eigen-indices that is a union of eigenspaces, let $\mathbf{\Pi}_{\mathcal{L}} = \sum_{k \in \mathcal{L}} \mathbf{u}_k \mathbf{u}_k^\top$ be the orthogonal projection onto their span. This projection does not depend on the basis chosen inside the eigenspaces, and

$$\sum_{k \in \mathcal{L}} a_k = \|\mathbf{\Pi}_{\mathcal{L}} \mathbf{S}\|_F^2, \qquad \sum_{k \in \mathcal{L}} \nu_k = \operatorname{tr}(\mathbf{\Pi}_{\mathcal{L}} \mathbf{N}), \qquad \sum_{k \in \mathcal{L}} b_k = \|\mathbf{\Pi}_{\mathcal{L}} \mathbf{S}\|_F^2 - \kappa \operatorname{tr}(\mathbf{\Pi}_{\mathcal{L}} \mathbf{N}). \tag{2}$$

Hence every sum of a_k, ν_k or b_k over a union of eigenspaces is basis-free.

A.1 SPECTRUM OF THE OPERATORS

Lemma 1 (Spectrum). ***(a)*** *Let $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, with zero rows and columns for isolated nodes. Then $\mathbf{W}$ is symmetric with spectrum in $[-1, 1]$. If the graph is connected with at least one edge, the eigenvalue 1 is simple with eigenvector $\mathbf{D}^{1/2}\mathbf{1}$, and -1 is an eigenvalue if and only if the graph is bipartite; it is then simple.* ***(b)*** *Let $\hat{\mathbf{W}} = \tilde{\mathbf{D}}^{-1/2}(\mathbf{A} + \mathbf{I})\tilde{\mathbf{D}}^{-1/2}$ with $\tilde{\mathbf{D}} = \mathbf{D} + \mathbf{I}$. Its spectrum lies in $[-1 + 2/\tilde{\Delta}, 1]$, where $\tilde{\Delta} = \max_i \tilde{\mathbf{D}}_{ii}$, and $\tilde{\mathbf{D}}^{1/2}\mathbf{1}$ is an eigenvector with eigenvalue 1. If the graph is connected, this eigenvalue is simple.*

Proof. (a) Symmetry is clear. On the non-isolated nodes, $\mathbf{W}$ is similar to the row-stochastic matrix $\mathbf{D}^{-1}\mathbf{A}$, whose spectral radius is 1. The isolated nodes contribute the eigenvalue 0. Hence the spectrum lies in $[-1, 1]$. The identity $\mathbf{W}\mathbf{D}^{1/2}\mathbf{1} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{1} = \mathbf{D}^{1/2}\mathbf{1}$ gives the eigenvalue 1. For a connected graph, $\mathbf{W}$ is nonnegative and irreducible. By the Perron-Frobenius theorem, the eigenvalue 1 is simple, and -1 is an eigenvalue if and only if $\mathbf{W}$ has period 2, that is, if and only if the graph is bipartite. This eigenvalue is then simple. (b) The same similarity argument, applied to the row-stochastic matrix $\tilde{\mathbf{D}}^{-1}(\mathbf{A} + \mathbf{I})$, bounds the spectrum by 1 in absolute value, and $\hat{\mathbf{W}}\tilde{\mathbf{D}}^{1/2}\mathbf{1} = \tilde{\mathbf{D}}^{1/2}\mathbf{1}$. Simplicity for a connected graph follows from Perron-Frobenius as in (a). For the lower bound, let $\mathbf{x} \neq 0$ and $\mathbf{z} = \tilde{\mathbf{D}}^{-1/2}\mathbf{x}$. Then

$$\mathbf{x}^\top(\mathbf{I} + \hat{\mathbf{W}})\mathbf{x} = \mathbf{z}^\top(\tilde{\mathbf{D}} + \mathbf{A} + \mathbf{I})\mathbf{z} = \mathbf{z}^\top(\mathbf{D} + \mathbf{A})\mathbf{z} + 2\|\mathbf{z}\|^2 = \sum_{\{i,j\}\text{ edge}} (z_i + z_j)^2 + 2\|\mathbf{z}\|^2 \geq 2\|\mathbf{z}\|^2.$$

Since $\|\mathbf{z}\|^2 = \mathbf{x}^\top\tilde{\mathbf{D}}^{-1}\mathbf{x} \geq \|\mathbf{x}\|^2/\tilde{\Delta}$, every eigenvalue satisfies $1 + \mu \geq 2/\tilde{\Delta}$. □

Remark 1 (Renormalization). *The self-loops of $\hat{\mathbf{W}}$ lift its smallest eigenvalue: $\mu_{\min}(\hat{\mathbf{W}}) \geq -1 + 2/\tilde{\Delta}$, where $\tilde{\Delta}$ is the largest diagonal entry of $\tilde{\mathbf{D}}$ (Lemma 1). This is the spectral shrinking noted by Wu et al. (2019). It compresses the bipartite end of the spectrum toward zero. A class signal that $\mathbf{W}$ places near $\mu = -1$ therefore receives less weight from $\hat{\mathbf{W}}^K$ relative to the top eigenvector, and we expect the contrast of Corollary 1(iii) to weaken. It also raises the PPR response floor of Proposition 2(b) from $\alpha/(2-\alpha)$ to $\alpha/(1-(1-\alpha)\mu_{\min})$, and hence the SNR bound to $m_g/M_g = (\alpha/(1-(1-\alpha)\mu_{\min}))^2$.*

The bound on $\mu_{\min}(\hat{\mathbf{W}})$ in Remark 1 is Lemma 1(b). Since $\mu_{\max}(\hat{\mathbf{W}}) = 1$, the PPR response on the spectrum of $\hat{\mathbf{W}}$ lies between $g_\alpha(\mu_{\min})$ and 1, and Proposition 2(a) gives the SNR bound stated in the remark.

A.2 PROOF OF PROPOSITION 1

Since $\mathbf{U}$ is orthogonal, $\|g(\mathbf{W})\mathbf{S}\|_F^2 = \|\operatorname{diag}(g(\mu))\mathbf{U}^\top\mathbf{S}\|_F^2 = \sum_k g(\mu_k)^2\|\mathbf{u}_k^\top\mathbf{S}\|_2^2 = \sum_k x_k a_k$. The d columns of $\mathbf{E}$ have mean zero and covariance $\sigma^2\mathbf{N}$, so $\mathbb{E}[\mathbf{E}\mathbf{E}^\top] = \sigma^2 d\,\mathbf{N}$. By linearity and cyclicity of the trace,

$$\mathbb{E}\|g(\mathbf{W})\mathbf{E}\|_F^2 = \operatorname{tr}\left(g(\mathbf{W})^2\,\mathbb{E}[\mathbf{E}\mathbf{E}^\top]\right) = \sigma^2 d\,\operatorname{tr}\left(\operatorname{diag}(x)\,\mathbf{U}^\top\mathbf{N}\mathbf{U}\right) = \sigma^2 d\sum_k x_k\nu_k.$$

Only the diagonal of $\mathbf{U}^\top\mathbf{N}\mathbf{U}$ enters. The denominator is positive by assumption, and dividing gives the formula. Independence of the columns is not needed for this identity; it matters only for the interpretation of $\mathbf{N}$ as a node covariance shared by all features. □

A.3 PROOF OF THEOREM 1

By Proposition 1, $\mathrm{SNR}(\mathbf{I}) = \sum_k a_k/(\sigma^2 d\sum_k \nu_k) = \kappa/(\sigma^2 d)$. Since $\sum_k x_k\nu_k > 0$,

$$\mathrm{SNR}(g) \geq \mathrm{SNR}(\mathbf{I}) \iff \sum_k x_k a_k \geq \kappa\sum_k x_k\nu_k \iff \sum_k x_k b_k \geq 0,$$

and equality holds in each statement exactly when it holds in the others. Because $\sum_k b_k = 0$, we have $\sum_k x_k b_k = n\ \mathrm{Cov}_k(x_k, b_k)$. □

A.4 Proof of Theorem 2

Let $t_1 < \cdots < t_L$ be the distinct values of $s(\mu_k)$, let $\mathcal{L}_j = \{k : s(\mu_k) = t_j\}$ with $n_j = |\mathcal{L}_j|$, and let $c_j = \sum_{k\in\mathcal{L}_j} b_k = n_j\,\bar{b}_s(t_j)$. Each $\mathcal{L}_j$ is a union of eigenspaces because $s(\mu_k)$ depends on k only through μ_k. By Eq. (2), c_j, $\bar{b}_s$ and B_s are basis-free. Put $B_j = \sum_{i\geq j} c_i$. Then $B_1 = \sum_k b_k = 0$, $B_s(t) = B_j$ for $t \in (t_{j-1}, t_j]$ with $j \geq 2$, and $B_s(t) = 0$ for $t \leq t_1$ and for $t > t_L$. The hypothesis of (a) is therefore equivalent to $B_j \geq 0$ for $j = 2, \ldots, L$, and similarly for (b).

Summation by parts. Let $g \in \mathcal{F}$ and $\varphi_j = \varphi_g(t_j)$, so that $x_k = \varphi_j$ for $k \in \mathcal{L}_j$. With $c_j = B_j - B_{j+1}$ and $B_{L+1} = 0$,

$$\sum_k x_k b_k = \sum_{j=1}^{L} \varphi_j c_j = \varphi_1 B_1 + \sum_{j=2}^{L} (\varphi_j - \varphi_{j-1}) B_j = \sum_{j=2}^{L} (\varphi_j - \varphi_{j-1}) B_j. \tag{3}$$

Since φ_g is non-decreasing, every factor $\varphi_j - \varphi_{j-1}$ is non-negative.

(a) and (b). If $B_j \geq 0$ for all $j \geq 2$, then Eq. (3) is non-negative, and Theorem 1 gives $\mathrm{SNR}(g) \geq \mathrm{SNR}(\mathbf{I})$. Part (b) is identical with the signs reversed.

Monotone averages. Suppose $\bar{b}_s$ is non-decreasing and $B_j < 0$ for some $j \geq 2$. The weighted average of $\bar{b}_s(t_i)$ over $i \geq j$ is then negative, so its smallest term $\bar{b}_s(t_j)$ is negative. Monotonicity gives $\bar{b}_s(t_i) \leq \bar{b}_s(t_j) < 0$ for all $i < j$. Hence $B_1 = \sum_{i<j} n_i \bar{b}_s(t_i) + B_j < 0$, which contradicts $B_1 = 0$. So a non-decreasing $\bar{b}_s$ satisfies the hypothesis of (a). The non-increasing case follows by applying this argument to $-b$.

Strictness. Let φ_g be strictly increasing on $\{t_1, \ldots, t_L\}$, so every factor $\varphi_j - \varphi_{j-1}$ in Eq. (3) is positive. Let $\bar{b}_s \not\equiv 0$, so some $c_j \neq 0$. If $B_j = 0$ for all $j \geq 2$, then $c_j = B_j - B_{j+1} = 0$ for $j \geq 2$ and $c_1 = B_1 - B_2 = 0$, a contradiction. Hence some B_j with $j \geq 2$ is nonzero, and under the sign hypothesis of (a) or (b) the sum in Eq. (3) is nonzero. Theorem 1 then gives a strict inequality.

(c). Let $B_s(t) < 0$ and fix $\lambda > 0$. Let $\chi_t(u) = 1$ for $u \geq t$ and $\chi_t(u) = 0$ otherwise. The filter $\tilde{g}(\mu) = (\lambda + \chi_t(s(\mu)))^{1/2}$ satisfies $\tilde{g}^2 = \varphi \circ s$ with the non-decreasing map $\varphi = \lambda + \chi_t$. With $x_k = \tilde{g}(\mu_k)^2$, it satisfies $\sum_k x_k \nu_k \geq \lambda n > 0$ and $\sum_k x_k b_k = \lambda \sum_k b_k + B_s(t) = B_s(t) < 0$. Theorem 1 gives $\mathrm{SNR}(\tilde{g}) < \mathrm{SNR}(\mathbf{I})$. □

Distributional form. Summing b_k over $\{k : s(\mu_k) \geq t\}$ and using $\kappa = \sum_k a_k / \sum_k \nu_k$ gives $B_s(t) = \|\mathbf{S}\|_F^2 \{P_a(s \geq t) - P_\nu(s \geq t)\}$ with $P_a(k) = a_k / \sum_i a_i$ and $P_\nu(k) = \nu_k / \sum_i \nu_i$.

A.5 Proof of Corollary 1

(i) For PPR with $\alpha \in (0, 1]$ and $\mu \in [-1, 1]$, the denominator $1 - (1-\alpha)\mu$ lies in $[\alpha, 2-\alpha]$ and is positive. Hence $g_\alpha > 0$, and $g'_\alpha(\mu) = \alpha(1-\alpha)/(1-(1-\alpha)\mu)^2$ is positive for $\alpha < 1$. So $g_\alpha^2 = \varphi(\mu)$ with $\varphi = g_\alpha^2$ strictly increasing on $[-1, 1]$; for $\alpha = 1$, $\varphi \equiv 1$. For SGC, $\mu^{2K} = \varphi(|\mu|)$ with $\varphi(u) = u^{2K}$, which is strictly increasing on $[0, 1]$ for $K \geq 1$. For high-pass powers, $((1-\mu)/2)^{2K} = \varphi(-\mu)$ with $\varphi(u) = ((1+u)/2)^{2K}$, which is strictly increasing on $[-1, 1]$ for $K \geq 1$ because $1 + u \geq 0$ there.

(ii) The PPR statement is Theorem 2(b) with $s = \mu$. For the high-pass statement, note that for every t,

$$B_{-\mu}(t) = \sum_{k:\,\mu_k \leq -t} b_k = -\sum_{k:\,\mu_k > -t} b_k,$$

because $\sum_k b_k = 0$. The last sum equals $B_\mu(t')$ for the smallest eigenvalue $t' > -t$, or 0 if there is none. Hence $B_\mu \leq 0$ everywhere implies $B_{-\mu} \geq 0$ everywhere, and Theorem 2(a) applies with $s = -\mu$. The level sets of μ and $-\mu$ coincide, so $\bar{b}_{-\mu} \not\equiv 0$ if and only if $\bar{b}_\mu \not\equiv 0$, and strictness follows from (i). The reverse case is symmetric.

(iii) The first statement combines (ii) with Theorem 2(a) for $s = |\mu|$. Strictness follows from (i) and the strictness clause of Theorem 2, as in (ii). For the example, let the graph be connected, D-regular and bipartite, with sides $\mathcal{V}_1$ and $\mathcal{V}_2$ that are the two classes. Regularity forces $|\mathcal{V}_1| = |\mathcal{V}_2|$, so $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})$. Let $\mathbf{y} = \mathbf{1}_{\mathcal{V}_1} - \mathbf{1}_{\mathcal{V}_2}$. Then $\mathbf{Y}_c = \frac{1}{2}[\mathbf{y}, -\mathbf{y}]$ and $\mathbf{S} = \frac{1}{2}\mathbf{y}(\mathbf{M}_{1\cdot} - \mathbf{M}_{2\cdot})^\top$, which is

nonzero by assumption. The operator is $\mathbf{W} = \mathbf{A}/D$, and $\mathbf{W}\mathbf{y} = -\mathbf{y}$. By Lemma 1, the eigenvalue -1 is simple, so $\mathbf{u}_{k_0} = \pm\mathbf{y}/\sqrt{n}$ for the index k_0 with $\mu_{k_0} = -1$. Hence a is concentrated on k_0, with $a_{k_0} = \|\mathbf{S}\|_F^2$. Under white noise, $\nu \equiv 1$, so $b_{k_0} = \|\mathbf{S}\|_F^2(1 - 1/n)$ and $b_k = -\|\mathbf{S}\|_F^2/n$ for $k \neq k_0$. For $\mathbf{W}$, $\mu_{k_0} = -1$ is the smallest eigenvalue, so $B_\mu(t) = -\|\mathbf{S}\|_F^2\,|\{k : \mu_k \geq t\}|/n \leq 0$ for $t > -1$ and $B_\mu(t) = 0$ otherwise. Also $|\mu_{k_0}| = 1$ is the largest value of $|\mu|$, so every nonzero $B_{|\mu|}(t)$ has the form $\|\mathbf{S}\|_F^2(1 - |\{k : |\mu_k| \geq t\}|/n) \geq 0$. □

The example needs $\mathbf{W}$. For $\hat{\mathbf{W}}$ the eigenvalue 1 is the only one of modulus 1, its eigenvector $\mathbf{1}/\sqrt{n}$ carries no signal, and $B_{|\mu|}(1) = -\|\mathbf{S}\|_F^2/n < 0$.

A.6 Proof of Proposition 2

(a) Since $m_g \leq x_k \leq M_g$ and $a, \nu \geq 0$, we have $\sum_k x_k a_k \geq m_g \sum_k a_k$ and $\sum_k x_k \nu_k \leq M_g \sum_k \nu_k$. Dividing and using Proposition 1 gives $\mathrm{SNR}(g) \geq (m_g/M_g)\mathrm{SNR}(\mathbf{I})$. The upper bound follows from $\sum_k x_k a_k \leq M_g \sum_k a_k$ and $\sum_k x_k \nu_k \geq m_g \sum_k \nu_k$.

(b) By the proof of Corollary 1(i), g_α is non-decreasing on $[-1, 1]$ with $g_\alpha(-1) = \alpha/(2-\alpha)$ and $g_\alpha(1) = 1$. On the spectrum, $m_g \geq (\alpha/(2-\alpha))^2$ and $M_g \leq 1$, and (a) gives the two-sided bound. Next,

$$g_\alpha(\mu) - 1 = \frac{\alpha - 1 + (1-\alpha)\mu}{1 - (1-\alpha)\mu} = -\frac{(1-\alpha)(1-\mu)}{1-(1-\alpha)\mu},$$

so $1 - g_\alpha(\mu) = (1-\alpha)(1-\mu)/(1-(1-\alpha)\mu)$ is non-negative and decreasing in μ, because the derivative of $(1-\mu)/(1-(1-\alpha)\mu)$ is $-\alpha/(1-(1-\alpha)\mu)^2$. The supremum is therefore attained at $\mu = -1$ and equals $2(1-\alpha)/(2-\alpha)$. Finally, as $\alpha \to 0$, $g_\alpha(\mu_k) \to 0$ for every $\mu_k < 1$ because $1 - \mu_k > 0$, while $g_\alpha(1) = 1$. Hence $\sum_k x_k a_k \to \sum_{k \in \mathcal{T}_1} a_k$ and $\sum_k x_k \nu_k \to \sum_{k \in \mathcal{T}_1} \nu_k > 0$, which gives the limit.

(c) Let $\mu_* = \max_k |\mu_k| > 0$ and $x_k = \mu_k^{2K}$. Under white noise, $\sum_k \nu_k = n$ and

$$\frac{\mathrm{SNR}(\mathbf{W}^K)}{\mathrm{SNR}(\mathbf{I})} = \frac{n \sum_k x_k a_k}{\sum_k x_k \sum_k a_k} \geq \frac{n \min_k x_k}{\sum_k x_k},$$

with equality when a is concentrated on an index attaining $\min_k |\mu_k|$. The bound is 0 if some $\mu_k = 0$. Otherwise it is at most $n(\min_k |\mu_k|/\mu_*)^{2K}$, which tends to 0 as $K \to \infty$ whenever $\min_k |\mu_k| < \mu_*$. For the limit with general noise, divide numerator and denominator of Proposition 1 by μ_*^{2K}. Since $(\mu_k/\mu_*)^{2K}$ tends to 1 for $k \in \mathcal{T}$ and to 0 otherwise, and $\sum_{k \in \mathcal{T}} \nu_k > 0$, the ratio converges to $\sum_{k \in \mathcal{T}} a_k/(\sigma^2 d \sum_{k \in \mathcal{T}} \nu_k)$.

A realizable zero. Let the graph be the complete bipartite graph with sides of even size m, so $\mathbf{W} = \mathbf{A}/m$. Let each class contain half of each side, and let $\mathbf{y} = \pm 1$ encode the classes. For a node on one side, $(\mathbf{W}\mathbf{y})_i = \frac{1}{m}\sum_j y_j$ over the other side, which is 0. Hence $\mathbf{y}$ lies in the eigenspace of $\mu = 0$. With two balanced classes $\mathbf{S} = \frac{1}{2}\mathbf{y}(\mathbf{M}_{1\cdot} - \mathbf{M}_{2\cdot})^\top$, so $\mathbf{W}^K\mathbf{S} = 0$ and $\mathrm{SNR}(\mathbf{W}^K) = 0$ for every $K \geq 1$, while $\mathrm{SNR}(\mathbf{I}) > 0$. Under white noise the denominator is $\sigma^2 d \sum_k \mu_k^{2K} = 2\sigma^2 d > 0$, so the ratio is defined.

Regular graphs. Let the graph be connected, regular and not bipartite. By Lemma 1, for either operator $\mathcal{T}$ consists of the index of the simple eigenvalue 1, with eigenvector $\mathbf{1}/\sqrt{n}$. Since $\mathbf{1}^\top\mathbf{Y}_c = n\boldsymbol{\pi}^\top - n\boldsymbol{\pi}^\top = 0$, the top eigenspace carries no signal, and $\mathrm{SNR}(\mathbf{W}^K) \to 0$ under white noise. □

Limits at the top of the score. Suppose the eigenspace at the top of the score carries a smaller share of the signal than of the noise, so its excess is negative: $\sum_{k \in \mathcal{T}_1} b_k < 0$ for (b) and $\sum_{k \in \mathcal{T}} b_k < 0$ for (c). The limits in (b) and (c) are then smaller than $\mathrm{SNR}(\mathbf{I}) = \kappa/(\sigma^2 d)$, because $\sum_k a_k < \kappa \sum_k \nu_k$ over the corresponding set. Hence PPR loses for all small enough α and SGC loses for all large enough K, as used after Corollary 1.

B Label-aligned refinement: details

The aligned signal. The projection $\mathbf{P}$ of Definition 2 is the least-squares fit of the filtered signal on the centred labels: $\mathbf{P}g(\mathbf{W})\mathbf{S} = \mathbf{Y}_c\hat{\mathbf{B}}(g)$ with $\hat{\mathbf{B}}(g) = (\mathbf{Y}_c^\top\mathbf{Y}_c)^+\mathbf{Y}_c^\top g(\mathbf{W})\mathbf{S}$, and $\mathbf{Y}_c^\top\mathbf{R}(g) = 0$. The pseudo-inverse is needed because $\mathbf{Y}_c\mathbf{1}_C = 0$, so $\mathbf{Y}_c^\top\mathbf{Y}_c$ is singular.

The aligned energy depends on g through the label-space operator $\mathbf{Y}_c^\top g(\mathbf{W})\mathbf{Y}_c$, since $\mathcal{A}(g) = \operatorname{tr}\left(\mathbf{M}^\top \mathbf{Y}_c^\top g(\mathbf{W})\mathbf{Y}_c(\mathbf{Y}_c^\top \mathbf{Y}_c)^+ \mathbf{Y}_c^\top g(\mathbf{W})\mathbf{Y}_c\mathbf{M}\right)$. For $g(\mu) = \mu$ this operator is a degree-normalized class-compatibility matrix. The residual $\mathbf{R}(g)$ is filtered class signal that is not explained by the node's own class mean, and F treats it as additional within-class variation.

$F(\mathbf{I}) = \mathrm{SNR}(\mathbf{I})$**.** The columns of $\mathbf{S} = \mathbf{Y}_c\mathbf{M}$ lie in the column space of $\mathbf{Y}_c$, so $\mathbf{PS} = \mathbf{S}$, $\mathbf{R}(\mathbf{I}) = 0$ and $\mathcal{A}(\mathbf{I}) = \|\mathbf{S}\|_F^2$. The denominator is $\mathbb{E}\|\mathbf{E}\|_F^2$, and the two ratios coincide.

B.1 Proof of Proposition 4

We assume $\sigma^2 > 0$ and write $\mathbf{G} = g_\alpha(\mathbf{W})$ and $\ell = \ell_\alpha$. The function g_α is increasing on $[-1, 1]$ with $g_\alpha(-1) = \ell$ and $g_\alpha(1) = 1$, so the symmetric matrix $\mathbf{G}$ satisfies $\ell\mathbf{I} \preceq \mathbf{G} \preceq \mathbf{I}$. The columns of $\mathbf{S} = \mathbf{Y}_c\mathbf{M}$ lie in the column space of $\mathbf{Y}_c$, so $\mathbf{PS} = \mathbf{S}$ and

$$\langle \mathbf{PGS}, \mathbf{S}\rangle_F = \langle \mathbf{GS}, \mathbf{PS}\rangle_F = \langle \mathbf{GS}, \mathbf{S}\rangle_F \;\geq\; \ell\|\mathbf{S}\|_F^2.$$

The Cauchy-Schwarz inequality gives $\|\mathbf{PGS}\|_F \geq \ell\|\mathbf{S}\|_F$, that is, $\mathcal{A}(g_\alpha) \geq \ell^2\|\mathbf{S}\|_F^2$. For the residual, $(\mathbf{I} - \mathbf{P})\mathbf{S} = 0$, so with $c = (1 + \ell)/2$ we have $\mathbf{R}(g_\alpha) = (\mathbf{I} - \mathbf{P})\mathbf{GS} = (\mathbf{I} - \mathbf{P})(\mathbf{G} - c\mathbf{I})\mathbf{S}$. The projection $\mathbf{I} - \mathbf{P}$ has spectral norm at most one and the eigenvalues of $\mathbf{G} - c\mathbf{I}$ lie in $[-(1 - \ell)/2, (1 - \ell)/2]$, hence $\|\mathbf{R}(g_\alpha)\|_F \leq \frac{1}{2}(1 - \ell)\|\mathbf{S}\|_F$. By the proof of Proposition 1, $\mathbb{E}\|\mathbf{GE}\|_F^2 = \sigma^2 d\sum_k g_\alpha(\mu_k)^2\nu_k \leq \sigma^2 d\sum_k \nu_k = \sigma^2 dn$, where $\nu_k \geq 0$ because $\mathbf{N} \succeq 0$. Since $F(\mathbf{I}) = \mathrm{SNR}(\mathbf{I}) = \|\mathbf{S}\|_F^2/(\sigma^2 dn)$,

$$F(g_\alpha) \;\geq\; \frac{\ell^2\|\mathbf{S}\|_F^2}{\frac{1}{4}(1 - \ell)^2\|\mathbf{S}\|_F^2 + \sigma^2 dn} \;=\; \frac{\ell^2\, F(\mathbf{I})}{1 + \frac{1}{4}(1 - \ell)^2 F(\mathbf{I})}.$$

The bound equals one at $\alpha = 1$ and is positive for every $\alpha > 0$. For SGC, the complete bipartite example in the proof of Proposition 2 has $\mathbf{Wy} = 0$, hence $\mathbf{W}^K\mathbf{S} = 0$ and $\mathcal{A}(\mathbf{W}^K) = 0$ for every $K \geq 1$. For the renormalized operator, the complete graph with two balanced classes has $\hat{\mathbf{W}} = \mathbf{1}\mathbf{1}^\top/n$, so $\hat{\mathbf{W}}\mathbf{y} = 0$ and again $\mathcal{A}(\hat{\mathbf{W}}^K) = 0$.

B.2 Proof of Proposition 3

Write the classes as $\mathbf{y} = +1$ and $\mathbf{y} = -1$. Balance gives $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})$ and $\mathbf{Y}_c = \frac{1}{2}[\mathbf{y}, -\mathbf{y}]$, so the column space of $\mathbf{Y}_c$ is spanned by $\mathbf{y}$ and $\mathbf{P} = \mathbf{y}\mathbf{y}^\top/n$. With class means $\pm\gamma\mathbf{m}$, $\mathbf{S} = \gamma\,\mathbf{y}\mathbf{m}^\top$. (For arbitrary means $\mathbf{M}_{1\cdot}, \mathbf{M}_{2\cdot}$ one takes $\gamma\mathbf{m} = (\mathbf{M}_{1\cdot} - \mathbf{M}_{2\cdot})/2$; the common part cancels in $\mathbf{S}$.) Then

$$\mathbf{PWS} = \gamma\beta\,\mathbf{y}\mathbf{m}^\top, \qquad \mathbf{R}(\mathbf{W}) = \gamma(\mathbf{Wy} - \beta\mathbf{y})\mathbf{m}^\top,$$

so $\mathcal{A}(\mathbf{W}) = n\gamma^2\beta^2$ and $\|\mathbf{R}(\mathbf{W})\|_F^2 = \gamma^2(\|\mathbf{Wy}\|^2 - 2\beta\,\mathbf{y}^\top\mathbf{Wy} + n\beta^2) = n\gamma^2\xi^2$. With $\mathbf{N} = \mathbf{I}$, $\mathbb{E}\|\mathbf{WE}\|_F^2 = \sigma^2 d\operatorname{tr}(\mathbf{W}^2) = n\sigma^2 d\tau$. Hence

$$F(\mathbf{W}) = \frac{\beta^2\gamma^2}{\gamma^2\xi^2 + \sigma^2 d\tau}, \qquad F(\mathbf{I}) = \frac{\gamma^2}{\sigma^2 d}.$$

We have $\tau > 0$ because $\mathbf{W} \neq 0$. Cross-multiplying positive denominators, $F(\mathbf{W}) \geq F(\mathbf{I})$ if and only if $\beta^2\sigma^2 d \geq \gamma^2\xi^2 + \sigma^2 d\tau$, that is, $\beta^2 \geq \tau + \varepsilon\xi^2$. For the two consequences, $\xi^2 = \|(\mathbf{I} - \mathbf{P})\mathbf{Wy}\|^2/n \geq 0$, so $\beta^2 < \tau$ gives $\beta^2 < \tau + \varepsilon\xi^2$. Also $\|\mathbf{Wy}\|^2 \leq \|\mathbf{y}\|^2 = n$ because the spectrum lies in $[-1, 1]$, so $\xi^2 \leq 1 - \beta^2$. If $\beta^2 \geq (\tau + \varepsilon)/(1 + \varepsilon)$, then $\beta^2 \geq \tau + \varepsilon(1 - \beta^2) \geq \tau + \varepsilon\xi^2$. For a D-regular graph, $\mathbf{W} = \mathbf{A}/D$. Let $e_=$ and $e_\neq$ count edges within and across classes, with $e_= + e_\neq = nD/2$ and $h = e_=/(e_= + e_\neq)$. Then $\mathbf{y}^\top\mathbf{Ay} = 2(e_= - e_\neq) = nD(2h - 1)$, so $\beta = 2h - 1$. Also $\operatorname{tr}(\mathbf{A}^2) = \sum_i D = nD$, so $\tau = 1/D$. □

The SNR in the same setting. By Proposition 1, $\mathrm{SNR}(\mathbf{W}) = \gamma^2\|\mathbf{Wy}\|^2/(n\sigma^2 d\tau)$, so $\mathrm{SNR}(\mathbf{W})/\mathrm{SNR}(\mathbf{I}) = (\beta^2 + \xi^2)/\tau$. The SNR thus counts the misaligned energy ξ^2 as signal.

Block-model approximation. Consider a balanced two-class stochastic block model without self-loops, with edge homophily h and expected degree $\bar{D}$, the setting of Baranwal et al. (2021), and the operator $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. For a nearly regular graph, $\beta \approx 2h - 1$ and $\tau \approx 1/\bar{D}$. Each entry of $\mathbf{Wy}$ averages about $\bar{D}$ neighbour labels, which suggests $\xi^2 \approx (1 - \beta^2)/\bar{D}$. To leading order in γ^2/d, the condition $\beta^2 \geq \tau + \varepsilon\xi^2$ then becomes

$$\beta^2 \geq \frac{1 + \varepsilon}{\bar{D} + \varepsilon} \approx \frac{1}{\bar{D}} \qquad (\varepsilon \ll 1),$$

and the SNR ratio becomes $\bar{D}\beta^2 + 1 - \beta^2$, which equals one at $\beta = 0$. On the 30 block-model instances of Section 4.1 with $h \leq 0.5$ and the operator $\mathbf{W}$, the computed $\log\{F(\mathbf{W})/F(\mathbf{I})\}$ and this closed form differ by at most 0.003 and agree in sign on all 30. The exact identity $\mathcal{A}(\mathbf{W}) = n\gamma^2\beta^2$ holds on every instance, and $\mathrm{tr}(\mathbf{W}^2)/n$ differs from $1/\bar{D}$ by at most 0.0004. At $h = 0.02$ the exact ξ^2 is 2.6 to 2.8 times $(1 - \beta^2)/\bar{D}$, which we attribute to degree heterogeneity; the comparison is unaffected, because $\varepsilon\xi^2$ is about $0.006\,\xi^2$ there and hence negligible next to τ.

The counterexample in detail. The block models have $n = 2000$, $\bar{D} \approx 10$, $d = 64$ and $\gamma = 0.6$ (Appendix C). At $h = 0.5$ the graph carries no class structure. With the operator $\mathbf{W}$, SGC with $K \in \{1, 2, 3, 4\}$ selected on validation picks $K = 2$ in all five instances and loses 6.2 points against raw features, with 95% interval $[-9.8, -3.8]$. The log SNR ratio of the selected filter is -0.05, which predicts almost no effect. The label-aligned ratio gives $\log\{F(\mathbf{W}^2)/F(\mathbf{I})\} = -0.71$, negative in all five instances, and -6.8 for one step, where $\beta \approx 0$. With $\hat{\mathbf{W}}$, the loss is 8.6 points $[-11.4, -6.6]$ and the label-aligned log ratio is -0.80. Across the 20 decisive cells of the block-model experiment with $h \leq 0.5$, both the SNR and F predict the sign of the gap correctly.

C Protocols

Graphs and operators. Every graph is made undirected; self-loops and duplicate edges are removed and all edge weights are one. We use the symmetric normalization $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, with zero rows for isolated nodes, and the renormalized operator $\hat{\mathbf{W}} = \tilde{\mathbf{D}}^{-1/2}(\mathbf{A} + \mathbf{I})\tilde{\mathbf{D}}^{-1/2}$ of Kipf & Welling (2017). For either operator, written $\mathbf{W}$ in the rest of this paragraph, SGC features are $\mathbf{W}^K\mathbf{X}$. PPR features are the closed form $\alpha(\mathbf{I} - (1-\alpha)\mathbf{W})^{-1}\mathbf{X}$, computed by the fixed-point iteration $\mathbf{Z}_{i+1} = \alpha\mathbf{X} + (1-\alpha)\mathbf{W}\mathbf{Z}_i$, $\mathbf{Z}_0 = \alpha\mathbf{X}$, until $\frac{1-\alpha}{\alpha}\|\mathbf{Z}_{i+1} - \mathbf{Z}_i\|_F \leq 10^{-6}\|\mathbf{Z}_{i+1}\|_F$. Since $\|\mathbf{W}\|_2 \leq 1$, this is an a-posteriori bound on the relative error. Filtered features are computed once in double precision and stored in single precision. Table 2 lists the graphs.

Additional graphs. The Facebook100 graphs are loaded with the PyTorch Geometric `LINKXDataset`. Their label is the binary gender attribute of the public release, used only as a benchmark target as in Lim et al. (2021). Nodes of unknown gender stay in the graph and in every filter but belong to no split. The Twitch and GitHub graphs are read from the SNAP release of Rozemberczki et al. (2021), with binary multi-hot features over the listed feature identifiers. The Twitch label is the `mature` flag and the GitHub label is `ml_target`. Two users are listed twice in the Twitch-FR label file with the same label, and we keep one row each. DBLP and Wiki are loaded with the PyTorch Geometric `CitationFull` and `AttributedGraphDataset` classes. PyTorch Geometric attributes the latter collection to Yang et al. (2021).

Splits and rounds. Let Λ be the set of labelled nodes, listed in increasing order. For a given seed, `numpy.random.default_rng(seed).permutation(`$|\Lambda|$`)` orders the nodes of Λ; the first $\lfloor |\Lambda|/2 \rfloor$ nodes in this order form the training set, the next $\lfloor |\Lambda|/4 \rfloor$ the validation set and the rest the test set. Round I uses the ten official splits of roman-empire, amazon-ratings, minesweeper, tolokers, questions and the filtered chameleon and squirrel graphs (Platonov et al., 2023b), and the random splits with seeds 0 to 9 for the other nine benchmarks. Round II uses seeds 10 to 19 for all sixteen benchmarks, and Round III uses seeds 20 to 29 for the fourteen additional graphs. The configurations, initialization seeds and training procedure are identical in the three rounds.

Metrics. For a split with test set Q, accuracy is $|\{i \in Q : \hat{y}_i = y_i\}|/|Q|$. For binary labels, the score of node i is the softmax probability p_i of class 1. The ROC-AUC is $\big(\sum_{i \in Q_+} \mathrm{rank}(p_i) - |Q_+|(|Q_+| + 1)/2\big)/(|Q_+||Q_-|)$, where $Q_\pm$ are the positive and negative test nodes and ranks are taken over Q with ties assigned their average rank. Both metrics are multiplied by 100. The same metric is used for validation.

Heads, grids and training. A head has hidden width in $\{128, 256, 512\}$, depth (number of linear layers) in $\{2, 3\}$, learning rate in $\{10^{-3}, 10^{-2}\}$, weight decay in $\{0, 5 \cdot 10^{-4}\}$, dropout in $\{0.3, 0.5, 0.7\}$ and LayerNorm or BatchNorm after each hidden linear layer. The 60 heads are drawn without replacement from these 144 combinations with `numpy.random.default_rng(20260924)` and are shared by the three arms. The filter of the i-th SGC or PPR configuration is drawn uniformly and independently from its grid of 12 or 14 filters (hyperparameter and operator), with seeds 1 and 2. Training minimizes the cross-entropy on T_r with Adam for at most 500 epochs. For each split separately, training stops after 50 epochs with-

Table 2: Graphs after symmetrization and removal of self-loops and duplicate edges. Edges are undirected, and the mean degree is $2|\mathcal{E}|/n$ for the edge set $\mathcal{E}$. Edge homophily is the fraction of edges whose two labelled endpoints share a class. *Nodes include those of unknown gender, which belong to no split: 2,739, 1,838, 418, 203 and 97 nodes, respectively.

Graph	Nodes	Features	Classes	Edges	Mean degree	Edge homophily	Metric	Rounds
Cora	2,708	1,433	7	5,278	3.9	0.81	accuracy	I, II
CiteSeer	3,327	3,703	6	4,552	2.7	0.74	accuracy	I, II
PubMed	19,717	500	3	44,324	4.5	0.80	accuracy	I, II
Amazon-Computers	13,752	767	10	245,861	35.8	0.78	accuracy	I, II
Amazon-Photo	7,650	745	8	119,081	31.1	0.83	accuracy	I, II
Coauthor-CS	18,333	6,805	15	81,894	8.9	0.81	accuracy	I, II
Coauthor-Physics	34,493	8,415	5	247,962	14.4	0.93	accuracy	I, II
WikiCS	11,701	300	10	215,603	36.9	0.65	accuracy	I, II
roman-empire	22,662	300	18	32,927	2.9	0.05	accuracy	I, II
amazon-ratings	24,492	300	5	93,050	7.6	0.38	accuracy	I, II
minesweeper	10,000	7	2	39,402	7.9	0.68	ROC-AUC	I, II
tolokers	11,758	10	2	519,000	88.3	0.59	ROC-AUC	I, II
questions	48,921	301	2	153,540	6.3	0.84	ROC-AUC	I, II
chameleon (filtered)	890	2,325	5	8,854	19.9	0.24	accuracy	I, II
squirrel (filtered)	2,223	2,089	5	46,998	42.3	0.21	accuracy	I, II
Actor	7,600	932	5	26,659	7.0	0.22	accuracy	I, II
Penn94	41,554*	4,814	2	1,362,229	65.6	0.51	ROC-AUC	III
Cornell5	18,660*	4,735	2	790,777	84.8	0.55	ROC-AUC	III
JohnsHopkins55	5,180*	2,406	2	186,586	72.0	0.55	ROC-AUC	III
Amherst41	2,235*	1,193	2	90,954	81.4	0.53	ROC-AUC	III
Reed98	962*	745	2	18,812	39.1	0.52	ROC-AUC	III
Twitch-DE	9,498	3,170	2	153,138	32.2	0.63	ROC-AUC	III
Twitch-EN	7,126	3,170	2	35,324	9.9	0.56	ROC-AUC	III
Twitch-ES	4,648	3,170	2	59,382	25.6	0.58	ROC-AUC	III
Twitch-FR	6,549	3,170	2	112,666	34.4	0.56	ROC-AUC	III
Twitch-PT	1,912	3,169	2	31,299	32.7	0.57	ROC-AUC	III
Twitch-RU	4,385	3,170	2	37,304	17.0	0.62	ROC-AUC	III
GitHub	37,700	4,005	2	289,003	15.3	0.85	ROC-AUC	III
DBLP	17,716	1,639	4	52,867	6.0	0.83	accuracy	III
Wiki	2,405	4,973	17	11,596	9.6	0.61	accuracy	III

out improvement of the validation metric. The reported test metric is taken at the first epoch with the best validation metric. Configuration i uses the initialization seed $1000 + i$ in every arm.

Selection and intervals. Within an arm, configurations are ordered lexicographically by four keys. They are the validation metric val_r of the split (higher first), the hidden width (smaller first), the filter's distance from the identity (smaller K, larger α), and the configuration index (smaller first). The identity counts as $K = 0$ and $\alpha = 1$, and $\hat{\theta}_{\omega,r}$ is the first configuration in this order on split r. For two arms ω, ω' let $\delta_r = \mathrm{test}_r(\hat{\theta}_{\omega,r}) - \mathrm{test}_r(\hat{\theta}_{\omega',r})$. For $j = 1, \ldots, n_{\mathrm{boot}}$ with $n_{\mathrm{boot}} = 10^4$, draw $r_1^{(j)}, \ldots, r_R^{(j)}$ independently and uniformly from $\{1, \ldots, R\}$ with `numpy.random.default_rng(0)`, restarted for every contrast, and set $\Delta^{(j)} = R^{-1} \sum_l \delta_{r_l^{(j)}}$. The interval is formed by the empirical 2.5% and 97.5% quantiles of $\{\Delta^{(j)}\}$. It excludes zero when its lower end exceeds 10^{-9} or its upper end is below -10^{-9}.

Identity added to a scheme, and the union. For a scheme $f \in \{\mathrm{SGC}, \mathrm{PPR}\}$ the extended candidate set is $\Theta_f \cup \Theta_{\mathrm{MLP}}$, with the MLP configurations treated as the identity filter in the ordering above. In the union $\Theta_{\mathrm{MLP}} \cup \Theta_{\mathrm{SGC}} \cup \Theta_{\mathrm{PPR}}$ of Section 4.2, the third key is 0 for the MLP, K for SGC and $2 - \alpha$ for PPR. Selection and intervals are as for the arms.

Spectral quantities. The spectral quantities of real graphs (the Cora example of Section 3.2 and the diagnostic of Section 4.2) take $\mathbf{M}$ as the class means over all labelled nodes, test nodes included, and $\boldsymbol{\pi}$ as their class frequencies. The class row of an unlabelled node is set to $\boldsymbol{\pi}$, so its centred row in $\mathbf{S} = \mathbf{Y}_c\mathbf{M}$ is zero. These quantities never enter the selection. Under white noise, and with $\log$ the natural logarithm throughout, Proposition 1 gives $\log \rho(g) = \log\left(\|g(\mathbf{W})\mathbf{S}\|_F^2 / \operatorname{tr} g(\mathbf{W})^2\right) - \log\left(\|\mathbf{S}\|_F^2/n\right)$. We compute $\|g(\mathbf{W})\mathbf{S}\|_F^2$ from the filtered matrices and $\operatorname{tr} g(\mathbf{W})^2 = \sum_k g(\mu_k)^2$ from the exact eigenvalues of a dense eigendecomposition, for all 30 graphs. The diagnostic of Section 4.2 averages $\log \rho$ over the filters selected on the ten splits and compares its sign with the significant gaps; the cumulative excess B_s of Cora uses its eigenvectors.

Block models. Each instance draws its randomness from `numpy.random.default_rng` seeded with the root 7002 and its instance index, and its graph additionally with the position of h in the grid. Labels are balanced over two classes. Edges are sampled independently with probability $0.01h$ within a class and $0.01(1-h)$ across classes, which gives an expected degree of about 10. Graphs with an isolated node are resampled. Features are $\mathbf{X} = \mathbf{Y}\mathbf{M} + \mathbf{E}$ with class means $\pm\gamma\mathbf{m}$, a random unit vector $\mathbf{m}$ per instance, $\gamma = 0.6$, and standard Gaussian noise in $d = 64$ dimensions. Labels, means, noise and the $50/25/25$ split depend only on the instance, so the raw-feature arm is identical across h. The head is a logistic regression with inverse regularization strength (the parameter of scikit-learn) in $\{0.01, 0.1, 1, 10\}$ on standardized inputs. With $n = 2000$ nodes, the raw-feature head reaches 69.96% test accuracy on average. The interval of a cell is the 95% percentile interval of 10^4 bootstrap resamples of its five instances, drawn from one `numpy.random.default_rng(0)` stream over the cells in a fixed order. Filters are applied exactly through the eigendecomposition of the operator.

Compute. Rounds I and II each train $28{,}800$ (configuration, split) pairs and Round III trains $25{,}200$, each on eight RTX 5090 GPUs in about half an hour. The matched head budget of Appendix D trains $55{,}200$ pairs and the band intervention $27{,}000$ pairs on the same hardware, $165{,}000$ training runs in total. The block models take a few CPU minutes.

D Additional results

Table 3 lists every gap of the three rounds with its interval, Table 6 the test metric of each arm, and Table 5 the filters selected on validation. The signs of the SGC, PPR and PPR$-$SGC entries of the red rows of Table 1(a) are the same in Round I (Table 3). Table 4 lists the scheme, graph and round cells in which a scheme is significantly below the MLP, together with the extended search of Section 4.2. In every such cell the extended search selects an MLP configuration on every split, and the loss disappears. Over all 92 cells, the only significantly negative contrast of the extended search is -0.29 points $[-0.58, -0.03]$, against SGC alone on PubMed in Round II. **Pooled selection.** Selecting one configuration per arm by the mean validation metric over the ten splits gives the same conclusions for both schemes and for the identity extension. Over the 30 graphs of Rounds II and III, PPR is significantly better than SGC on 13 and SGC on 11. At worst, PPR falls 1.92 points below the MLP and SGC 8.76 points. SGC is significantly below the MLP in the same seven pairs, and PPR is significantly better than SGC in each. The identity extension removes all 9 losses, with a largest cost of 0.27 points. **Union search.** Validation over $\Theta_{\mathrm{MLP}} \cup \Theta_{\mathrm{SGC}} \cup \Theta_{\mathrm{PPR}}$ (Section 4.2) is significantly above the MLP in 42 of the 46 graph and round pairs, above SGC in 27 and above PPR in 15. It is never significantly below any of the three arms, and its lowest gaps are -0.63 points to the MLP, -1.42 to SGC and -0.37 to PPR. **Head budget per filter.** In the protocol of Section 4.2, the MLP trains all 60 heads, whereas a single SGC or PPR filter is paired with 2 to 8 of them. For every round, graph and scheme, we therefore also train the filter selected most often over the ten splits (Table 5) with all 60 heads. Each head uses the initialization seed of the corresponding MLP head, and we select per split as before (Table 7). SGC remains significantly below the MLP in all seven pairs, by 0.57 to 8.05 points. No significant comparison between the two schemes changes sign. Over the 30 graphs of Rounds II and III, PPR is significantly better than SGC on 13 and SGC on 11, against 15 and 11 in Section 4.2. PPR falls at most 1.53 points below the MLP. **Band intervention on Facebook100.** Let $\mathbf{W} = \mathbf{U}\,\mathrm{diag}(\mu)\mathbf{U}^\top$ and $\mathbf{\Pi}_- = \sum_{k:\,\mu_k<0} \mathbf{u}_k\mathbf{u}_k^\top$, a sum over whole eigenspaces and hence independent of the basis. For features $\mathbf{X}$ we set $\mathbf{X}^- = \mathbf{\Pi}_-\mathbf{X}$ and $\mathbf{X}^+ = \mathbf{X} - \mathbf{X}^-$. Since $g(\mathbf{W})\mathbf{\Pi}_- = \mathbf{\Pi}_- g(\mathbf{W}) = \sum_{k:\,\mu_k<0} g(\mu_k)\mathbf{u}_k\mathbf{u}_k^\top$, filtering $\mathbf{X}^-$ under the model (1) gives the centred class signal $g(\mathbf{W})\mathbf{\Pi}_-\mathbf{S}$ and noise columns with node covariance $\sigma^2 g(\mathbf{W})\mathbf{\Pi}_-\mathbf{N}\mathbf{\Pi}_- g(\mathbf{W})$. The term $g(\mathbf{W})\mathbf{\Pi}_-\mathbf{1}\boldsymbol{\pi}^\top\mathbf{M}$ moves every row along $\mathbf{M}^\top\boldsymbol{\pi}$ and is excluded as in Section 3.1. Proposition 1 and Theorem 1 therefore hold on $\mathbf{X}^-$ with the densities a_k and ν_k for $\mu_k < 0$ and 0 otherwise, and on $\mathbf{X}^+$ with the complementary band. Since g_α is increasing with $g_\alpha(-1) = \alpha/(2-\alpha)$ and $g_\alpha(0) = \alpha$, Proposition 2(a) gives $(2-\alpha)^{-2} \le \rho(g_\alpha) \le (2-\alpha)^2$ on $\mathbf{X}^-$. We compute $\mathbf{U}$ by a dense eigendecomposition in double precision (LAPACK `dsyevd`, and `dsyevr` for Penn94) and apply every filter exactly, as $g(\mathbf{W})\mathbf{X}^- = \sum_{k:\,\mu_k<0} g(\mu_k)\mathbf{u}_k\mathbf{u}_k^\top\mathbf{X}$ and analogously for $\mathbf{X}^+$. Every configuration keeps its head and its K or α and uses the operator $\mathbf{W}$, so that every filter is exactly band-limited. Splits, seeds, training and per-split selection are those of Round III, and the construction of $\mathbf{X}^\pm$ uses no labels. The spectral quantities use the class means over all labelled nodes, as the diagnostic of Section 4.2. Under white noise, $\log\rho(g_\alpha)$ on $\mathbf{X}^-$ lies in

$[-0.142, -0.007]$ for all seven α on the five graphs, and in $[-0.018, -0.007]$ for $\alpha = 0.9$. On $\mathbf{X}^-$, $\log \rho(\mathbf{W}^K)$ lies in $[0.359, 1.101]$ for $K = 1$ and in $[0.145, 1.241]$ for $K = 2$, and for $K \geq 4$ it is negative on every graph except Amherst41. Table 8 reports the gaps.

Table 3: All gaps $\Delta(\omega, \omega')$ with their paired 95% bootstrap intervals, in points (†: ROC-AUC×100). Grey entries are not significant. Red and blue rows as in Table 1.

Graph	SGC − MLP	PPR − MLP	PPR − SGC
Round I			
Cora	+12.10 [+11.37, +12.87]	+13.10 [+12.57, +13.68]	+1.00 [+0.65, +1.36]
CiteSeer	+2.39 [+1.24, +3.63]	+2.69 [+1.62, +3.81]	+0.30 [−0.37, +1.20]
PubMed	+0.06 [−0.16, +0.31]	+1.81 [+1.56, +2.02]	+1.75 [+1.47, +2.00]
Computers	+6.15 [+5.84, +6.44]	+5.95 [+5.65, +6.21]	−0.20 [−0.40, +0.03]
Photo	+2.00 [+1.60, +2.36]	+3.40 [+3.26, +3.56]	+1.41 [+1.16, +1.68]
Coauthor-CS	−0.66 [−0.84, −0.49]	+1.07 [+0.88, +1.27]	+1.74 [+1.61, +1.86]
Physics	+0.31 [+0.22, +0.40]	+0.85 [+0.76, +0.96]	+0.54 [+0.44, +0.64]
WikiCS	+3.06 [+2.49, +3.63]	+3.54 [+3.15, +3.94]	+0.49 [+0.08, +0.94]
roman-empire	−5.52 [−5.90, −5.12]	+7.41 [+7.17, +7.66]	+12.93 [+12.58, +13.30]
amazon-ratings	+0.89 [+0.45, +1.35]	+2.63 [+2.14, +3.11]	+1.74 [+1.30, +2.26]
minesweeper†	+35.47 [+34.67, +36.28]	+37.50 [+36.69, +38.41]	+2.03 [+1.67, +2.38]
tolokers†	+6.04 [+5.30, +6.86]	+7.06 [+6.55, +7.60]	+1.02 [+0.59, +1.49]
questions†	+5.68 [+5.19, +6.23]	+6.32 [+5.73, +6.92]	+0.65 [+0.28, +1.00]
chameleon-f.	+0.83 [−0.87, +2.45]	−0.48 [−1.39, +0.40]	−1.32 [−3.12, +0.61]
squirrel-f.	+3.15 [+1.95, +4.32]	+1.62 [+0.19, +2.81]	−1.53 [−3.35, +0.14]
Actor	−8.25 [−9.02, −7.62]	−1.74 [−2.42, −1.08]	+6.51 [+5.87, +7.17]
Round II			
Cora	+10.92 [+10.12, +11.62]	+12.05 [+11.27, +12.76]	+1.14 [+0.78, +1.46]
CiteSeer	+2.99 [+1.88, +4.03]	+3.42 [+2.32, +4.45]	+0.43 [+0.07, +0.72]
PubMed	+0.40 [+0.12, +0.69]	+1.92 [+1.69, +2.14]	+1.52 [+1.26, +1.79]
Computers	+6.14 [+5.65, +6.56]	+5.85 [+5.53, +6.12]	−0.29 [−0.51, −0.04]
Photo	+2.76 [+2.31, +3.23]	+3.69 [+3.37, +4.00]	+0.93 [+0.66, +1.21]
Coauthor-CS	−0.79 [−0.97, −0.60]	+1.09 [+0.87, +1.29]	+1.88 [+1.74, +2.03]
Physics	+0.39 [+0.27, +0.52]	+0.84 [+0.74, +0.95]	+0.45 [+0.34, +0.55]
WikiCS	+3.45 [+3.11, +3.83]	+3.91 [+3.68, +4.14]	+0.46 [+0.11, +0.83]
roman-empire	−5.53 [−5.97, −5.06]	+7.26 [+6.89, +7.63]	+12.79 [+12.41, +13.22]
amazon-ratings	+1.26 [+0.81, +1.74]	+2.85 [+2.29, +3.37]	+1.59 [+1.11, +2.09]
minesweeper†	+35.56 [+34.85, +36.31]	+37.60 [+37.10, +38.17]	+2.05 [+1.54, +2.50]
tolokers†	+5.29 [+4.54, +6.14]	+6.11 [+5.49, +6.73]	+0.82 [+0.37, +1.25]
questions†	+5.77 [+5.13, +6.48]	+6.58 [+6.05, +7.12]	+0.82 [+0.28, +1.34]
chameleon-f.	−1.35 [−3.18, +0.72]	−0.27 [−2.69, +1.88]	+1.08 [−1.48, +3.45]
squirrel-f.	+4.67 [+2.89, +6.34]	+1.67 [+0.22, +3.29]	−3.00 [−4.34, −1.67]
Actor	−8.01 [−8.59, −7.45]	−1.61 [−2.53, −0.91]	+6.39 [+5.73, +7.06]
Round III			
Penn94†	+6.44 [+6.26, +6.62]	+0.36 [+0.22, +0.50]	−6.08 [−6.22, −5.95]
Cornell5†	+8.64 [+8.19, +9.01]	+1.80 [+1.55, +2.03]	−6.83 [−7.05, −6.54]
JohnsHopkins55†	+14.57 [+13.54, +15.57]	+5.42 [+4.98, +5.84]	−9.14 [−9.93, −8.35]
Amherst41†	+21.17 [+19.79, +22.48]	+6.41 [+5.20, +7.58]	−14.76 [−15.74, −13.84]
Reed98†	+8.47 [+4.90, +12.11]	−1.20 [−3.83, +1.54]	−9.67 [−13.71, −5.35]
Twitch-DE†	+4.46 [+4.11, +4.81]	+3.06 [+2.71, +3.43]	−1.40 [−1.85, −0.91]
Twitch-EN†	+2.43 [+1.76, +3.07]	+2.51 [+1.75, +3.22]	+0.08 [−0.30, +0.53]
Twitch-ES†	+4.23 [+3.02, +5.39]	+2.43 [+1.47, +3.33]	−1.80 [−2.90, −0.78]
Twitch-FR†	+4.60 [+3.61, +5.60]	+2.07 [+1.36, +2.72]	−2.53 [−3.41, −1.87]
Twitch-PT†	+5.84 [+4.43, +7.07]	+3.73 [+2.88, +4.63]	−2.11 [−3.02, −1.05]
Twitch-RU†	+1.98 [+0.81, +3.44]	+1.42 [+0.10, +2.71]	−0.56 [−1.91, +0.67]
GitHub†	+1.61 [+1.47, +1.77]	+1.68 [+1.59, +1.76]	+0.07 [−0.03, +0.18]
DBLP	+8.28 [+7.67, +8.91]	+8.76 [+8.23, +9.33]	+0.49 [+0.33, +0.61]
Wiki	−4.72 [−5.45, −3.99]	+2.71 [+1.59, +3.79]	+7.43 [+6.23, +8.55]

Table 4: Cells in which a scheme is significantly below the MLP, and the search over the scheme grid extended by the identity (120 candidates; grey column). A difference of exactly zero means that both sides selected the same run. Grey entries are not significant.

Round	Graph	Scheme	Scheme − MLP	Extended − scheme	Extended − MLP
I	Actor	SGC	−8.25 [−9.02, −7.62]	+8.25 [+7.62, +9.02]	+0.00 [+0.00, +0.00]
I	Actor	PPR	−1.74 [−2.42, −1.08]	+1.74 [+1.08, +2.42]	+0.00 [+0.00, +0.00]
I	Coauthor-CS	SGC	−0.66 [−0.84, −0.49]	+0.66 [+0.49, +0.84]	+0.00 [+0.00, +0.00]
I	roman-empire	SGC	−5.52 [−5.90, −5.12]	+5.52 [+5.12, +5.90]	+0.00 [+0.00, +0.00]
II	Actor	SGC	−8.01 [−8.59, −7.45]	+8.01 [+7.45, +8.59]	+0.00 [+0.00, +0.00]
II	Actor	PPR	−1.61 [−2.53, −0.91]	+1.61 [+0.91, +2.53]	+0.00 [+0.00, +0.00]
II	Coauthor-CS	SGC	−0.79 [−0.97, −0.60]	+0.79 [+0.60, +0.97]	+0.00 [+0.00, +0.00]
II	roman-empire	SGC	−5.53 [−5.97, −5.06]	+5.53 [+5.06, +5.97]	+0.00 [+0.00, +0.00]
III	Wiki	SGC	−4.72 [−5.45, −3.99]	+4.72 [+3.99, +5.45]	+0.00 [+0.00, +0.00]

Table 5: Most frequently selected filter over the ten splits, with the number of splits in parentheses: K and operator for SGC, α and operator for PPR. Ties are broken towards the identity (smaller K, larger α), then towards $\hat{\mathbf{W}}$. Red and blue rows as in Table 1.

	Round I		Round II			Round III	
Graph	SGC	PPR	SGC	PPR	Graph	SGC	PPR
Cora	2, $\hat{\mathbf{W}}$ (9)	0.3, $\mathbf{W}$ (4)	2, $\hat{\mathbf{W}}$ (6)	0.3, $\mathbf{W}$ (3)	Penn94†	2, $\mathbf{W}$ (10)	0.2, $\mathbf{W}$ (5)
CiteSeer	2, $\hat{\mathbf{W}}$ (7)	0.5, $\hat{\mathbf{W}}$ (4)	2, $\hat{\mathbf{W}}$ (7)	0.5, $\hat{\mathbf{W}}$ (6)	Cornell5†	2, $\mathbf{W}$ (10)	0.2, $\hat{\mathbf{W}}$ (6)
PubMed	1, $\hat{\mathbf{W}}$ (6)	0.3, $\hat{\mathbf{W}}$ (4)	2, $\hat{\mathbf{W}}$ (7)	0.3, $\hat{\mathbf{W}}$ (7)	JohnsHopkins55†	2, $\mathbf{W}$ (10)	0.2, $\hat{\mathbf{W}}$ (3)
Computers	2, $\hat{\mathbf{W}}$ (10)	0.2, $\mathbf{W}$ (6)	2, $\hat{\mathbf{W}}$ (10)	0.2, $\mathbf{W}$ (6)	Amherst41†	2, $\mathbf{W}$ (8)	0.05, $\mathbf{W}$ (3)
Photo	2, $\hat{\mathbf{W}}$ (7)	0.3, $\hat{\mathbf{W}}$ (4)	2, $\hat{\mathbf{W}}$ (7)	0.3, $\hat{\mathbf{W}}$ (4)	Reed98†	2, $\mathbf{W}$ (6)	0.7, $\hat{\mathbf{W}}$ (2)
Coauthor-CS	1, $\hat{\mathbf{W}}$ (9)	0.5, $\hat{\mathbf{W}}$ (5)	1, $\hat{\mathbf{W}}$ (10)	0.5, $\hat{\mathbf{W}}$ (7)	Twitch-DE†	2, $\hat{\mathbf{W}}$ (6)	0.2, $\hat{\mathbf{W}}$ (5)
Physics	1, $\hat{\mathbf{W}}$ (5)	0.5, $\mathbf{W}$ (4)	1, $\hat{\mathbf{W}}$ (8)	0.5, $\mathbf{W}$ (5)	Twitch-EN†	2, $\hat{\mathbf{W}}$ (8)	0.2, $\mathbf{W}$ (3)
WikiCS	2, $\hat{\mathbf{W}}$ (10)	0.3, $\mathbf{W}$ (5)	2, $\hat{\mathbf{W}}$ (10)	0.3, $\mathbf{W}$ (5)	Twitch-ES†	2, $\hat{\mathbf{W}}$ (5)	0.2, $\mathbf{W}$ (4)
roman-empire	2, $\mathbf{W}$ (10)	0.3, $\hat{\mathbf{W}}$ (6)	2, $\mathbf{W}$ (10)	0.3, $\hat{\mathbf{W}}$ (4)	Twitch-FR†	2, $\hat{\mathbf{W}}$ (4)	0.1, $\hat{\mathbf{W}}$ (3)
amazon-ratings	2, $\mathbf{W}$ (6)	0.3, $\mathbf{W}$ (9)	2, $\mathbf{W}$ (6)	0.3, $\mathbf{W}$ (8)	Twitch-PT†	3, $\mathbf{W}$ (5)	0.3, $\hat{\mathbf{W}}$ (2)
minesweeper†	1, $\mathbf{W}$ (10)	0.7, $\mathbf{W}$ (9)	1, $\mathbf{W}$ (10)	0.7, $\mathbf{W}$ (8)	Twitch-RU†	2, $\hat{\mathbf{W}}$ (3)	0.3, $\hat{\mathbf{W}}$ (3)
tolokers†	2, $\mathbf{W}$ (7)	0.3, $\hat{\mathbf{W}}$ (5)	2, $\mathbf{W}$ (9)	0.2, $\hat{\mathbf{W}}$ (3)	GitHub†	2, $\hat{\mathbf{W}}$ (9)	0.1, $\hat{\mathbf{W}}$ (8)
questions†	2, $\hat{\mathbf{W}}$ (7)	0.05, $\mathbf{W}$ (4)	4, $\hat{\mathbf{W}}$ (7)	0.05, $\hat{\mathbf{W}}$ (8)	DBLP	6, $\hat{\mathbf{W}}$ (6)	0.05, $\hat{\mathbf{W}}$ (8)
chameleon-f.	2, $\mathbf{W}$ (3)	0.7, $\mathbf{W}$ (2)	2, $\hat{\mathbf{W}}$ (3)	0.9, $\hat{\mathbf{W}}$ (2)	Wiki	1, $\hat{\mathbf{W}}$ (5)	0.9, $\hat{\mathbf{W}}$ (3)
squirrel-f.	2, $\mathbf{W}$ (4)	0.05, $\mathbf{W}$ (3)	4, $\hat{\mathbf{W}}$ (2)	0.1, $\mathbf{W}$ (4)			
Actor	2, $\hat{\mathbf{W}}$ (5)	0.9, $\mathbf{W}$ (7)	2, $\hat{\mathbf{W}}$ (7)	0.9, $\mathbf{W}$ (3)			

Table 6: Test metric of each arm with per-split selection: mean and standard deviation over the ten splits, in points (accuracy; ROC-AUC$\times$100 for †). The best arm of each row is in bold; red and blue rows as in Table 1.

Graph	MLP	SGC	PPR
Round I			
Cora	75.91 ± 1.30	88.01 ± 0.85	**89.01** ± 0.79
CiteSeer	73.47 ± 1.35	75.86 ± 1.50	**76.16** ± 1.32
PubMed	88.04 ± 0.38	88.10 ± 0.39	**89.85** ± 0.37
Computers	85.79 ± 0.48	**91.94** ± 0.55	91.75 ± 0.36
Photo	92.43 ± 0.51	94.42 ± 0.37	**95.83** ± 0.42
Coauthor-CS	94.66 ± 0.34	94.00 ± 0.38	**95.73** ± 0.32
Physics	96.04 ± 0.09	96.35 ± 0.13	**96.89** ± 0.15
WikiCS	81.22 ± 1.01	84.28 ± 0.71	**84.76** ± 0.66
roman-empire	66.33 ± 0.45	60.82 ± 0.50	**73.74** ± 0.41
amazon-ratings	49.45 ± 0.50	50.34 ± 0.62	**52.08** ± 0.58
minesweeper†	50.65 ± 1.27	86.12 ± 0.79	**88.15** ± 0.87
tolokers†	74.53 ± 0.83	80.57 ± 1.14	**81.59** ± 1.08
questions†	70.65 ± 0.98	76.33 ± 0.95	**76.98** ± 0.97
chameleon-f.	43.57 ± 1.93	**44.40** ± 2.98	43.09 ± 2.57
squirrel-f.	39.28 ± 1.93	**42.43** ± 2.02	40.90 ± 2.08
Actor	**38.19** ± 0.95	29.95 ± 0.76	36.46 ± 1.03
Round II			
Cora	76.51 ± 1.55	87.43 ± 1.31	**88.57** ± 1.11
CiteSeer	73.70 ± 1.62	76.69 ± 1.09	**77.12** ± 1.27
PubMed	87.57 ± 0.48	87.97 ± 0.49	**89.48** ± 0.47
Computers	85.70 ± 0.70	**91.84** ± 0.56	91.55 ± 0.43
Photo	91.72 ± 0.76	94.48 ± 0.70	**95.41** ± 0.51
Coauthor-CS	94.89 ± 0.33	94.10 ± 0.22	**95.98** ± 0.17
Physics	96.09 ± 0.21	96.48 ± 0.18	**96.93** ± 0.14
WikiCS	80.99 ± 0.77	84.45 ± 0.32	**84.90** ± 0.68
roman-empire	66.61 ± 0.80	61.08 ± 0.87	**73.87** ± 0.60
amazon-ratings	49.35 ± 0.86	50.61 ± 0.68	**52.20** ± 0.98
minesweeper†	50.09 ± 1.28	85.65 ± 0.53	**87.69** ± 1.09
tolokers†	75.43 ± 0.89	80.72 ± 0.90	**81.54** ± 0.72
questions†	70.29 ± 1.67	76.06 ± 1.58	**76.87** ± 1.52
chameleon-f.	**43.81** ± 3.99	42.47 ± 2.72	43.54 ± 3.12
squirrel-f.	38.08 ± 2.18	**42.75** ± 1.31	39.75 ± 1.91
Actor	**37.78** ± 1.09	29.77 ± 1.24	36.17 ± 1.18
Round III			
Penn94†	84.14 ± 0.27	**90.58** ± 0.31	84.50 ± 0.26
Cornell5†	80.66 ± 0.55	**89.29** ± 0.59	82.46 ± 0.44
JohnsHopkins55†	73.04 ± 1.37	**87.60** ± 0.73	78.46 ± 1.18
Amherst41†	68.13 ± 2.33	**89.30** ± 1.18	74.55 ± 1.39
Reed98†	60.56 ± 4.07	**69.04** ± 4.05	59.37 ± 4.18
Twitch-DE†	68.61 ± 0.89	**73.06** ± 0.47	71.66 ± 1.09
Twitch-EN†	62.11 ± 0.99	64.54 ± 0.93	**64.63** ± 1.30
Twitch-ES†	65.05 ± 1.73	**69.28** ± 1.97	67.48 ± 1.99
Twitch-FR†	63.98 ± 1.40	**68.58** ± 1.14	66.04 ± 1.26
Twitch-PT†	63.01 ± 1.67	**68.86** ± 2.25	66.75 ± 1.85
Twitch-RU†	53.36 ± 1.78	**55.34** ± 1.95	54.78 ± 2.31
GitHub†	89.66 ± 0.36	91.27 ± 0.36	**91.34** ± 0.31
DBLP	77.43 ± 0.99	85.71 ± 0.37	**86.20** ± 0.41
Wiki	81.41 ± 1.23	76.69 ± 1.68	**84.12** ± 1.47

Table 7: Gaps $\Delta(\omega, \omega')$ when each scheme's most frequently selected filter (Table 5; K or α and operator) is trained with all 60 heads of the MLP, in points (†: ROC-AUC$\times$100). Grey entries are not significant. Red and blue rows as in Table 1, from these gaps.

Graph	SGC filter	SGC − MLP	PPR filter	PPR − MLP	PPR − SGC
Round I					
Cora	2, $\hat{\mathbf{W}}$	+12.29	0.3, $\mathbf{W}$	+12.75	+0.46
CiteSeer	2, $\hat{\mathbf{W}}$	+2.95	0.5, $\hat{\mathbf{W}}$	+3.69	+0.73
PubMed	1, $\hat{\mathbf{W}}$	+0.37	0.3, $\hat{\mathbf{W}}$	+2.03	+1.66
Computers	2, $\hat{\mathbf{W}}$	+6.20	0.2, $\mathbf{W}$	+6.05	−0.15
Photo	2, $\hat{\mathbf{W}}$	+1.98	0.3, $\hat{\mathbf{W}}$	+3.10	+1.12
Coauthor-CS	1, $\hat{\mathbf{W}}$	−0.57	0.5, $\hat{\mathbf{W}}$	+1.11	+1.68
Physics	1, $\hat{\mathbf{W}}$	+0.41	0.5, $\mathbf{W}$	+0.88	+0.47
WikiCS	2, $\hat{\mathbf{W}}$	+2.99	0.3, $\mathbf{W}$	+3.65	+0.67
roman-empire	2, $\mathbf{W}$	−3.79	0.3, $\hat{\mathbf{W}}$	+7.37	+11.15
amazon-ratings	2, $\mathbf{W}$	+1.73	0.3, $\mathbf{W}$	+2.60	+0.88
minesweeper†	1, $\mathbf{W}$	+35.44	0.7, $\mathbf{W}$	+37.79	+2.35
tolokers†	2, $\mathbf{W}$	+6.52	0.3, $\hat{\mathbf{W}}$	+7.05	+0.52
questions†	2, $\hat{\mathbf{W}}$	+6.00	0.05, $\mathbf{W}$	+6.87	+0.87
chameleon-f.	2, $\mathbf{W}$	+1.83	0.7, $\mathbf{W}$	−1.53	−3.37
squirrel-f.	2, $\mathbf{W}$	+2.14	0.05, $\mathbf{W}$	+2.11	−0.03
Actor	2, $\hat{\mathbf{W}}$	−8.05	0.9, $\mathbf{W}$	−0.35	+7.71
Round II					
Cora	2, $\hat{\mathbf{W}}$	+11.34	0.3, $\mathbf{W}$	+11.70	+0.35
CiteSeer	2, $\hat{\mathbf{W}}$	+2.89	0.5, $\hat{\mathbf{W}}$	+3.67	+0.78
PubMed	2, $\hat{\mathbf{W}}$	+0.58	0.3, $\hat{\mathbf{W}}$	+2.23	+1.65
Computers	2, $\hat{\mathbf{W}}$	+6.42	0.2, $\mathbf{W}$	+5.99	−0.42
Photo	2, $\hat{\mathbf{W}}$	+2.77	0.3, $\hat{\mathbf{W}}$	+3.61	+0.84
Coauthor-CS	1, $\hat{\mathbf{W}}$	−0.77	0.5, $\hat{\mathbf{W}}$	+1.11	+1.87
Physics	1, $\hat{\mathbf{W}}$	+0.44	0.5, $\mathbf{W}$	+0.91	+0.47
WikiCS	2, $\hat{\mathbf{W}}$	+3.75	0.3, $\mathbf{W}$	+3.77	+0.02
roman-empire	2, $\mathbf{W}$	−3.99	0.3, $\hat{\mathbf{W}}$	+7.30	+11.29
amazon-ratings	2, $\mathbf{W}$	+2.03	0.3, $\mathbf{W}$	+3.05	+1.02
minesweeper†	1, $\mathbf{W}$	+35.58	0.7, $\mathbf{W}$	+38.26	+2.68
tolokers†	2, $\mathbf{W}$	+5.54	0.2, $\hat{\mathbf{W}}$	+6.23	+0.69
questions†	4, $\hat{\mathbf{W}}$	+6.12	0.05, $\hat{\mathbf{W}}$	+6.61	+0.49
chameleon-f.	2, $\hat{\mathbf{W}}$	+0.81	0.9, $\hat{\mathbf{W}}$	+0.99	+0.18
squirrel-f.	4, $\hat{\mathbf{W}}$	+3.97	0.1, $\mathbf{W}$	+3.95	−0.02
Actor	2, $\hat{\mathbf{W}}$	−7.41	0.9, $\mathbf{W}$	−0.30	+7.11
Round III					
Penn94†	2, $\mathbf{W}$	+6.67	0.2, $\mathbf{W}$	+0.42	−6.25
Cornell5†	2, $\mathbf{W}$	+8.89	0.2, $\hat{\mathbf{W}}$	+1.91	−6.98
JohnsHopkins55†	2, $\mathbf{W}$	+14.36	0.2, $\hat{\mathbf{W}}$	+5.33	−9.03
Amherst41†	2, $\mathbf{W}$	+20.95	0.05, $\mathbf{W}$	+6.41	−14.54
Reed98†	2, $\mathbf{W}$	+10.27	0.7, $\hat{\mathbf{W}}$	−0.19	−10.47
Twitch-DE†	2, $\hat{\mathbf{W}}$	+4.66	0.2, $\hat{\mathbf{W}}$	+2.96	−1.69
Twitch-EN†	2, $\hat{\mathbf{W}}$	+2.71	0.2, $\mathbf{W}$	+2.80	+0.09
Twitch-ES†	2, $\hat{\mathbf{W}}$	+4.23	0.2, $\mathbf{W}$	+2.42	−1.81
Twitch-FR†	2, $\hat{\mathbf{W}}$	+4.52	0.1, $\hat{\mathbf{W}}$	+2.41	−2.11
Twitch-PT†	3, $\mathbf{W}$	+6.05	0.3, $\hat{\mathbf{W}}$	+2.85	−3.20
Twitch-RU†	2, $\hat{\mathbf{W}}$	+2.90	0.3, $\hat{\mathbf{W}}$	+1.77	−1.13
GitHub†	2, $\hat{\mathbf{W}}$	+1.75	0.1, $\hat{\mathbf{W}}$	+1.77	+0.02
DBLP	6, $\hat{\mathbf{W}}$	+8.18	0.05, $\hat{\mathbf{W}}$	+8.72	+0.54
Wiki	1, $\hat{\mathbf{W}}$	−4.04	0.9, $\hat{\mathbf{W}}$	+3.39	+7.43

Table 8: Band intervention on the Facebook100 graphs with the operator $\mathbf{W}$: test ROC-AUC$\times$100 of the MLP and gaps $\Delta(\omega, \omega')$ in points on the original features $\mathbf{X}$, on $\mathbf{X}^- = \mathbf{\Pi}_- \mathbf{X}$ and on $\mathbf{X}^+ = \mathbf{X} - \mathbf{X}^-$. The rows $\mathbf{X}$ use only $\mathbf{W}$ and can therefore differ from Tables 1 and 3. Shaded rows: $\mathbf{X}^-$, the band on which, under white noise, every PPR filter lowers the SNR and SGC with $K \leq 2$ raises it. Grey entries are not significant.

Graph	Features	MLP	SGC − MLP	PPR − MLP	PPR − SGC
Penn94†	$\mathbf{X}$	84.1	+6.44	+0.39	−6.05
	$\mathbf{X}^-$	77.1	+5.28	+0.00	−5.27
	$\mathbf{X}^+$	78.7	+3.10	+0.88	−2.23
Cornell5†	$\mathbf{X}$	80.7	+8.69	+1.86	−6.82
	$\mathbf{X}^-$	70.2	+4.34	−0.37	−4.71
	$\mathbf{X}^+$	76.9	+7.23	+2.31	−4.92
JohnsHopkins55†	$\mathbf{X}$	73.0	+14.51	+5.31	−9.20
	$\mathbf{X}^-$	63.1	+6.46	−0.61	−7.08
	$\mathbf{X}^+$	70.9	+11.48	+4.86	−6.62
Amherst41†	$\mathbf{X}$	68.1	+21.01	+6.58	−14.44
	$\mathbf{X}^-$	53.6	+15.48	−0.07	−15.55
	$\mathbf{X}^+$	66.3	+15.82	+4.84	−10.98
Reed98†	$\mathbf{X}$	60.6	+10.00	−1.87	−11.87
	$\mathbf{X}^-$	51.5	+10.37	−2.18	−12.56
	$\mathbf{X}^+$	55.6	+5.96	+1.23	−4.73

E ORDERS AND FLOORS OF OTHER FIXED FILTERS

Theorem 2 gives one verdict for a set of filters whose members share one score, whenever the cumulative excess has one sign. Proposition 2(a) bounds the loss of any filter whose response stays away from zero. Table 9 applies both results to common fixed filters of $\mathbf{W}$. For every value of the parameter, $g(\mu)^2$ is a non-decreasing function of the listed score on $[-1, 1]$, and the floor is $\ell = \min_{\mu\in[-1,1]} |g(\mu)|$. Both operators have their spectrum in $[-1, 1]$ (Lemma 1), so the entries hold on every graph. Every listed response has maximum 1 on $[-1, 1]$, so a floor $\ell > 0$ gives $\mathrm{SNR}(g)/\mathrm{SNR}(\mathbf{I}) \geq \ell^2$ by Proposition 2(a).

Table 9: Scores and floors of common fixed filters of $\mathbf{W}$, with $\mu \in [-1, 1]$, grouped by score. Bold: the two schemes compared in this paper.

Filter	Response $g(\mu)$	Parameter	Score	Floor ℓ
Powers (SGC)	μ^K	$K \geq 1$	$\lvert\mu\rvert$	0
PPR	$\alpha/\big(1-(1-\alpha)\mu\big)$	$\alpha \in (0, 1]$	μ	$\alpha/(2-\alpha)$
Heat kernel	$e^{-T(1-\mu)}$	$T > 0$	μ	e^{-2T}
DGC	$\big(1-\frac{T}{K}(1-\mu)\big)^K$	$0 < T < \frac{K}{2}, K \geq 1$	μ	$\big(1-\frac{2T}{K}\big)^K$
Lazy powers	$\big((1+\mu)/2\big)^K$	$K \geq 1$	μ	0
High-pass powers	$\big((1-\mu)/2\big)^K$	$K \geq 1$	$-\mu$	0
Identity	1	none	any	1

The heat kernel is one of the two diffusions of GDC (Gasteiger et al., 2019b). The DGC filter with terminal time T and K steps of size T/K (Wang et al., 2021) tends to the heat kernel with the same T as $K \to \infty$, and at $T = K/2$ it equals the lazy powers. Table 9 restricts DGC to $T < K/2$, which contains the settings released with DGC ($T/K < 0.022$). At $T = K$ the DGC filter is the power $\mathbf{W}^K$. For any other $T > K/2$ the factor $1 - \frac{T}{K}(1-\mu)$ vanishes at the interior point $\mu_0 = 1 - K/T \neq 0$, and the filter follows none of the listed scores. These diffusions share the order of PPR, so Theorem 2 with $s = \mu$ gives them the verdicts that Corollary 1(ii) and (iii) give PPR. Their floors are positive and shrink as the diffusion lengthens, as the floor of PPR does when α decreases. Lazy powers share the same order but have no floor, and high-pass powers reverse it. Powers of $\mathbf{W}$ are the only listed filters ordered by $|\mu|$. Up to this reversal, the listed filters therefore use the two orders of SGC and PPR. The floor is a separate property, and adding the identity gives any grid a member with floor 1 (Section 3.3). A set whose members do not share a score is decided filter by filter by Theorem 1. For example, S^2GC (Zhu & Koniusz, 2021) uses $g(\mu) = \alpha + (1-\alpha)K^{-1}\sum_{j=1}^{K}\mu^j$. With its default $\alpha = 0.05$, its filter for $K = 1$ ranks $\mu = -0.9$ above $\mu = -0.5$, and its filter for $K = 2$ ranks them the other way.