CALIBCF: THE GRAPH KNOWS WHERE TO CUT GRAPH-CONDITIONED SPECTRAL CUTOFFS FOR TRAINING-FREE RECOMMENDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Training-free spectral recommenders can rival learned embedding models, yet their accuracy depends on where the spectrum of the user-item graph is cut. Across five benchmarks, the best retained dimension spans 114–1,869, making a fixed rank difficult to transfer between graphs. The upper edge of the noise spectrum offers a graph-specific reference for this decision, but standard estimators are poorly suited to recommendation: user and item degrees create heterogeneous noise, while scalable methods compute only a short spectral head rather than the full bulk. We introduce CalibCF, which estimates the noise edge from the interaction graph and the head already available to a spectral recommender. User and item degrees define a separable variance profile, from which CalibCF derives a graph-conditioned template for the noise spectrum. After excluding leading signal directions, a one-parameter regression fits the template's scale to the observed squared singular values. We prove that the resulting edge estimate is consistent under a degree-corrected Bernoulli graph model and spectral regularity conditions. CalibCF then sets the operating cutoff relative to this edge and applies a continuous, closed-form shrinkage response to the retained directions. Using spectral heads covering only 4.9–10.4% of the spectrum, the template predicts the uncomputed eigenvalue distribution on four benchmarks with KS distance 0.016–0.049, a 3–12$\times$ improvement over an identically fitted Marchenko–Pastur template. CalibCF outperforms all seven compared recommenders on five benchmarks, improving NDCG@20 over the strongest baseline by up to 4.0%; 65 of 66 paired comparisons are significant after Holm correction. With a fixed operating point and no target-domain validation, it retains 98.3% of its per-graph tuned NDCG@20, compared with 96.7% for the best fixed absolute rank.

1 INTRODUCTION

Spectral collaborative filtering is one of the simplest routes to strong recommendation: extract the leading singular directions of the normalized interaction graph, weight them with a spectral filter, and rank the reconstructed scores. Without learning any embedding, such filters now match or exceed trained graph recommenders (Shen et al., 2021; Park et al., 2024; Kim et al., 2025). One decision, however, is still searched anew for every graph: *where should the filter cut the spectrum?* Keeping too many directions admits noise; keeping too few erases preference structure. The answer changes sharply across graphs. On the five benchmarks studied here, the retained dimension spans 114–1,869, and the best absolute rank of one graph loses up to 18.6% NDCG@20 when it is applied to another.

Most spectral recommenders treat this decision as part of filter design, from ideal low-pass and polynomial responses to learned coefficients (Shen et al., 2021; Kim et al., 2025; Qin et al., 2024; He et al., 2026); the scale at which the filter acts is left to per-dataset search. Random-matrix theory suggests a principled scale instead. Pure noise produces singular values only up to the upper edge of its bulk, so directions above that edge carry signal (Marčenko & Pastur, 1967; Baik et al., 2005; Gavish & Donoho, 2017). Two properties of recommendation graphs block this route. First, noise is far from homogeneous: heavy-tailed user activity and item popularity reshape the bulk, so the Marchenko–Pastur (MP) law misrepresents it. Second, only a short spectral head is computed on

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.

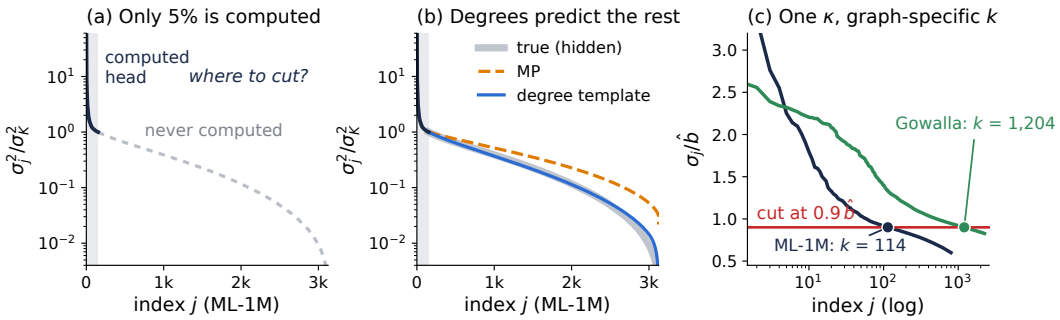

Figure 1: CalibCF on real spectra. (a) On ML-1M only the leading 4.9% of the spectrum is computed. (b) The degree template, scaled on that head, predicts the spectrum that was never computed; the MP template does not. (c) Measured against the calibrated edge, one multiplier cuts each graph at its own rank: $k = 114$ on ML-1M and $k = 1{,}204$ on Gowalla.

large graphs, whereas classical scale and rank estimators rely on the bulk that was never computed (Gavish & Donoho, 2017; Nadakuditi, 2014; Dobriban & Owen, 2019).

We propose CalibCF to estimate the noise edge from exactly the information that is available. **Our key observation is that the graph already knows the shape of its own spectrum.** The degrees that make noise heterogeneous also determine, through a separable variance profile, the shape of the noise bulk. CalibCF therefore derives this shape from user and item degrees and fits its scale to the computed spectral head. Once the graph template is normalized to a unit edge, the fit is a one-parameter regression whose coefficient is the squared edge $\hat{b}^2$, and under a degree-corrected graph model this regression provably recovers the noise edge. CalibCF then anchors a closed-form shrinkage filter at the calibrated edge (Figure 1). The uncomputed part of the spectrum is never reconstructed.

Calibrating the edge replaces the per-graph search over absolute ranks with a single multiplier κ whose meaning is the same on every graph. The calibrated coordinate is remarkably stable: the selected κ lies in $[0.85, 0.9]$ on four of five graphs while the retained dimension varies 16-fold, and without target labels one fixed multiplier retains 98.3% of the in-domain accuracy, versus 96.7% for the best fixed absolute rank. The resulting recommender outperforms every compared method, from spectral filters and EASE to MF and LightGCN, on all five benchmarks, improving NDCG@20 over truncated SVD by 2.3–7.1% and over the strongest baseline by up to 4.0%. The calibration is also robust to how much of the spectrum is computed: enlarging the head 5.3-fold moves the calibrated edge by less than 0.6%. Because neither labels nor training are involved, a complete CalibCF model is built in 9.8–26.1 seconds on one GPU, 13–331$\times$ faster than a single LightGCN training run.

Our contributions are threefold:

- **Graph-conditioned cutoff calibration.** We show that degree statistics determine the shape of the noise spectrum of recommendation graphs, and turn this into CalibCF, which calibrates a spectral cutoff from a short computed head with a single regression, without labels, training or tail reconstruction.
- **Theory of the construction.** We prove that the calibrated edge is consistent for the noise edge under a degree-corrected graph model, and that cutoff perturbations cause bounded score changes, including when the retained rank changes.
- **Evidence at the spectral and recommendation levels.** The template predicts unseen spectra 3–12$\times$ more accurately than MP; CalibCF outperforms all compared methods, from spectral filters and EASE to MF and LightGCN, on all five benchmarks; one multiplier transfers across graphs (98.3% versus 96.7% for the best fixed absolute rank); and a complete model is built in 9.8–26.1 seconds on one GPU.

2 Related Work

Collaborative filtering. Learned recommenders embed users and items, from matrix factorization with pairwise ranking losses (Koren et al., 2009; Rendle et al., 2009) and neural interaction functions

(He et al., 2017) to graph convolutions (He et al., 2020) and contrastive objectives (Wu et al., 2021; Yu et al., 2022; Cai et al., 2023). Closed-form models instead compute truncated factors or item–item weights directly (Cremonesi et al., 2010; Steck, 2019), and careful comparisons show that such simple models remain highly competitive with neural ones (Ferrari Dacrema et al., 2019). CalibCF belongs to this training-free family.

Spectral recommendation. GF-CF combines linear and ideal low-pass filters (Shen et al., 2021), SVD-GCN builds a graph recommender on a truncated SVD (Peng et al., 2022), Turbo-CF uses decomposition-free filtering (Park et al., 2024), and ChebyCF represents responses through Chebyshev interpolation (Kim et al., 2025). Diffusion and personalized graph-signal methods provide further filter constructions (Choi et al., 2023; Liu et al., 2023); SGFCF studies spectral expressiveness (Peng et al., 2024), while PolyCF and ASPIRE learn filter responses (Qin et al., 2024; He et al., 2026). These methods design the response applied to the graph and leave the scale at which it acts, such as the number of retained directions, to per-dataset search. CalibCF derives that scale from the graph and keeps a standard response, so the contribution of calibration is isolated.

Random-matrix denoising and rank selection. The MP law describes the bulk of homogeneous noise (Marčenko & Pastur, 1967), and signal directions separate from it above a sharp threshold (Baik et al., 2005; Benaych-Georges & Nadakuditi, 2012). On this basis, Gavish–Donoho derive closed-form shrinkage with an MP-median scale estimator (Gavish & Donoho, 2017); OptShrink and ScreeNOT use empirical bulk functionals (Nadakuditi, 2014; Donoho et al., 2023); parallel analysis selects the number of components from the full spectrum (Dobriban & Owen, 2019); and eOptShrink estimates an edge from leading-eigenvalue spacings before estimating a noise distribution (Su & Wu, 2025). Heteroscedastic and separable-noise models lead to distribution-aware responses (Leeb & Romanov, 2021; Zhang & Mondelli, 2024; Su, 2026), biwhitening rescales the matrix toward an MP bulk (Landa et al., 2022), and RMT cleaning has been applied to recommendation similarities (Khawar & Zhang, 2019). CalibCF is distinguished by the information it calibrates from: it keeps the recommendation operator, derives the bulk shape from the graph's variance margins, and fits only a scale to the computed head. It builds on separable Gram laws (Alt et al., 2017; Yang, 2019), degree-corrected graph models (Karrer & Newman, 2011) and regularized degree normalization (Qin & Rohe, 2013; Le et al., 2017; Hwang et al., 2019). Appendix A contrasts the estimands and inputs of the closest methods.

3 CALIBCF

A spectral cutoff is only meaningful relative to the noise level of the graph at hand, so CalibCF first estimates that level and then places the cutoff against it. The estimate uses the two sources of information that are always available: the degrees of the graph determine the *shape* of the noise spectrum (Section 3.1), and the computed head determines its *scale* (Section 3.2). Section 3.3 then turns the calibrated edge into recommendation scores. No step uses recommendation labels, and one calibration serves every operating point.

Problem setting. A deployed spectral recommender has two inputs: the fit interaction graph R and the K leading singular triples of its degree-normalized operator L (Eq.(1)), where K is a small fraction of $p = \min(m, n)$ on large graphs. We view L as a low-rank preference signal plus noise, $L = F_u \mathbb{E}[R] F_i + N$ with $N = F_u(R - \mathbb{E}R)F_i$. Pure noise produces singular values only up to its top singular value $\sigma_1(N)$, so this edge separates directions that can carry preference information from directions that noise alone explains, and a cutoff measured against it has the same meaning on every graph. Our goal is to estimate $\sigma_1(N)$ from R and the computed head alone, without recommendation labels and without the uncomputed part of the spectrum. Two properties make this hard: the entries of N have strongly heterogeneous variances, so no fixed law describes its spectrum across graphs, and the bulk statistics that classical estimators use lie in the part of the spectrum that is never computed. CalibCF resolves the first with a graph-conditioned template and the second with a scale fit that uses only computed values.

3.1 A SPECTRAL TEMPLATE FROM GRAPH STATISTICS

To know where noise ends, we first need to know what the noise spectrum of a recommendation graph looks like. Its shape is set by how unevenly interactions spread over users and items, and this unevenness is already recorded in the degrees. We therefore build the template from degree statistics alone.

Let $R \in \{0,1\}^{m\times n}$ be the fit interaction matrix, with user degrees d_u, item degrees e_i, and $N = \sum_{ui} R_{ui}$. The operator is the regularized degree normalization of the graph (Qin & Rohe, 2013; Le et al., 2017), with regularizers $\tau_u = c_\tau N/m$ and $\tau_i = c_\tau N/n$ and $c_\tau = 0.25$ throughout:

$$L = F_u R F_i, \qquad F_u = \operatorname{diag}(d+\tau_u)^{-1/2}, \qquad F_i = \operatorname{diag}(e+\tau_i)^{-1/2}. \tag{1}$$

Write $p = \min(m,n)$, $q = \max(m,n)$, $\beta = p/q$, and let $(\hat{u}_j, \sigma_j, \hat{v}_j)_{j\le K}$ be the leading $K < p$ singular triples of L, ordered by decreasing σ_j.

Degree normalization equalizes scale but not entry variances, and it is this residual heterogeneity that shapes the noise bulk. For independent Bernoulli interactions with means P and population normalizers, the normalized noise has variance $P_{ui}(1-P_{ui})/[(\bar{d}_u+\tau_u^*)(\bar{e}_i+\tau_i^*)]$. This motivates the sparse, fit-graph proxy

$$Q_{ui} = \frac{R_{ui}(1-\hat{P}_{ui})}{(d_u+\tau_u)(e_i+\tau_i)}, \qquad \hat{P}_{ui} = \min\{d_u e_i/N, 0.99\}. \tag{2}$$

Its row sums r_u, column sums c_i, and total $M = \sum Q$ define $S^0_{ui} = r_u c_i/M = \phi_u\psi_i$, a separable profile matching both margins. The separable form keeps user- and item-side variation while admitting a tractable spectral law; Appendix B.5 gives the population motivation under a degree-corrected model (Karrer & Newman, 2011).

A variance profile alone does not say where the noise ends. The separable Gram law converts it into a full spectral distribution, whose right edge is the quantity we need. For the law calculation, orient the profile as $p \times q$, swapping ϕ and ψ when $m > n$; scoring retains the original user–item orientation. The deterministic Gram-law model for centered independent entries with variance $\phi_u\psi_i$ satisfies (Alt et al., 2017; Yang, 2019)

$$\Delta(z) = \sum_{i=1}^{q} \frac{\psi_i}{1+\psi_i T(z)}, \qquad T(z) = \sum_{u=1}^{p} \frac{\phi_u}{\phi_u\Delta(z) - z}, \qquad m(z) = \frac{1}{p}\sum_{u=1}^{p} \frac{1}{\phi_u\Delta(z)-z}. \tag{3}$$

We compute the right edge directly from the real branch of this system, which avoids defining it through a smoothed density. For $T \in (-1/\max_i \psi_i, 0)$, let $z(T)$ be the unique root of the T-equation above $\max_u \phi_u\Delta(T)$; then $b_0^2 = \min_T z(T)$, found by scalar root-finding and minimization. For uniform entry variance $1/q$ it recovers the MP edge $(1+\sqrt{\beta})^2$. The CDF F_0 is obtained from $\operatorname{Im} m(x+\mathrm{i}\eta)/\pi$, truncated at b_0^2 and normalized.

3.2 CALIBRATING THE EDGE FROM THE COMPUTED HEAD

The template fixes the shape of the bulk but not its scale: the proxy is built from degrees, so the absolute noise level it implies can be off by a global factor. The computed head fixes the scale but not the shape. The two combine naturally once the template is normalized to a unit edge, because the unknown squared edge then becomes a single regression coefficient. For an excluded leading count r, let $\mathcal{J}_r$ be the zero-based window $r+5, \ldots, K - \operatorname{round}(0.15K) - 1$, which omits leading signal spikes and the least reliable end of the computed head. Define

$$h_{rj} = \frac{F_0^{-1}\big(1 - \frac{j-r+1/2}{p}\big)}{b_0^2}, \qquad \hat{b}^{\,2} = \arg\min_{x\ge 0} \sum_{j\in\mathcal{J}_r} (\sigma_{j+1}^2 - x h_{rj})^2 = \frac{h^\top\lambda}{h^\top h}, \quad \lambda_j = \sigma_{j+1}^2. \tag{4}$$

Because the number of signal spikes is itself unknown, we alternate the fit with the update $r \leftarrow \#\{j : \sigma_j > 1.02\hat{b}\}$, starting from b_0, until the count is unchanged; invalid windows and cycles are detected explicitly (Appendix C). The regression uses only computed values: the template defines a full spectral law, but CalibCF never constructs the missing empirical tail.

Algorithm 1: CalibCF. Input: fit graph R, head width K, multiplier κ, readout γ.

1. Form L by Eq.(1); compute its leading K singular triples.
2. Form Q by Eq.(2); use its margins in Eq.(3) to obtain F_0 and b_0^2.
3. Initialize $r = \#\{j : \sigma_j > 1.02b_0\}$. Alternate the fit in Eq.(4) with $r \leftarrow \#\{j : \sigma_j > 1.02\hat{b}\}$ until unchanged.
4. Set $\theta = \kappa\hat{b}$ and $w_j = \eta_\beta(\sigma_j; \theta)$ by Eq.(5).
5. Form the scores in Eq.(6), exclude observed interactions, and rank per user.

3.3 From a calibrated edge to recommendation scores

The calibrated edge gives each graph its own unit of spectral scale, and CalibCF expresses the cutoff in that unit. A multiplier κ places the operating cutoff at $\theta = \kappa\hat{b}$, giving $k = \#\{j \le K : \sigma_j > \theta\}$ retained directions. Because the cutoff is measured against the graph's own edge, a single value of κ has a comparable meaning across graphs, which an absolute rank does not.

Keeping directions all-or-nothing would make scores jump whenever the cutoff crosses a singular value. We therefore weight the retained directions continuously with the Gavish–Donoho response (Gavish & Donoho, 2017), written in factorized form as

$$\eta_\beta(\sigma;\theta) = \begin{cases} \dfrac{\sqrt{(\sigma^2-\theta^2)(\sigma^2-a_\beta\theta^2)}}{\sigma}, & \sigma > \theta, \\ 0, & \sigma \le \theta, \end{cases} \qquad a_\beta = \left(\frac{1-\sqrt{\beta}}{1+\sqrt{\beta}}\right)^2. \tag{5}$$

The response vanishes continuously at the cutoff and approaches σ far above it, so it needs nothing below the cutoff. With $w_j = \eta_\beta(\sigma_j;\theta)$, the recommendation scores are

$$\hat{S} = F_u^{-1} U_k \operatorname{diag}(w) V_k^\top D_i^{(\gamma)}, \qquad D_i^{(\gamma)} = \operatorname{diag}\left((e_i + \tau_i)^\gamma\right). \tag{6}$$

The item normalization in Eq.(1) down-weights popular items; the readout exponent γ restores a controlled amount of popularity, and $\gamma = 1/2$ undoes the normalization exactly. The user factor does not affect within-user rankings. The two operating parameters thus have distinct roles, κ for spectral retention and γ for readout, and one decomposition and one calibration serve all operating points (Algorithm 1).

4 Understanding graph-conditioned calibration

Three properties make the construction work: the graph template must change the calibration itself rather than an arbitrary scale constant, the calibrated edge must recover the true noise edge, and calibration error must not cause discontinuous score changes. We establish them in turn; proofs are in Appendices B.1, B.6 and B.2.

The template changes the calibration functional. On a fixed fitting window, write $E_h(\lambda) = h^\top\lambda/(h^\top h)$. The following result characterizes exactly when two templates merely rescale one another.

Proposition 1 (Shape dependence)**.** *For a fixed window and positive templates h, g and $c > 0$, $E_h(\lambda) = cE_g(\lambda)$ on a nonempty open set of positive, strictly ordered λ if and only if $h = g/c$. Moreover, if $\lambda = x_* h_* + \varepsilon$ with $x_* \ge 0$, then $|E_h(\lambda) - x_*| \le (\|\varepsilon\|_2 + x_*\|h_* - h\|_2)/\|h\|_2$.*

For a homogeneous profile the template is proportional to the MP template and CalibCF reduces to the corresponding MP quantile-fit rule. On all five benchmark graphs the templates are not proportional: degree conditioning changes how each observed singular value contributes to the edge. The error bound separates template mismatch from spectral noise; the next result shows that both vanish for degree-corrected graphs.

The calibrated edge is consistent. Let $N = F_u(R - \mathbb{E}R)F_i$ denote the normalized noise that L actually contains; its largest singular value $\sigma_1(N)$ is the edge a cutoff should be measured against.

Proposition 2 (Consistency)**.** *Let R follow the degree-corrected Bernoulli model of Appendix B.5, with a rank-r preference signal and mean degree $\bar{d}_n \gg \log(m+n)$, let the head grow proportionally,*

$K/p \to k_0 \in (0, 1]$, and let the noise satisfy the separable spectral regularity (A3) of Appendix B.6. Then, with probability tending to one, every exclusion count visited by the calibration is at most $r + 1$, and for each of them $\hat{b}/\sigma_1(N) \to 1$ in probability.

The proof chains Proposition 1 with the normalization transfer of Appendix B.5: empirical normalization, the low-rank signal and the fit-graph proxy margins each move the relevant spectral quantiles by a vanishing relative amount. Because h is edge-normalized, the regression also absorbs any global misspecification of the noise level, which the unscaled edge b_0 cannot.

Scores remain stable when the retained rank changes. A hard cutoff changes the support abruptly as it crosses a singular value. The continuous response in Eq.(5) bounds the resulting score perturbation without requiring the support to stay fixed.

Proposition 3 (Cutoff stability)**.** *For all $\sigma > 0$, $\theta, \theta' > 0$, and fixed singular triples and readout exponent γ,*

$$|\eta_\beta(\sigma; \theta) - \eta_\beta(\sigma; \theta')| \le \sqrt{(1 + a_\beta)|\theta^2 - \theta'^2|},$$
$$|\hat{S}_{ui}(\theta) - \hat{S}_{ui}(\theta')| \le \sqrt{d_u + \tau_u}\,(e_i + \tau_i)^\gamma \|U_{u,:}\|_2 \|V_{i,:}\|_2 \sqrt{(1 + a_\beta)|\theta^2 - \theta'^2|}.$$

The row norms refer to the common computed head, with zero weights below each cutoff. Combining the propositions at $\theta^2 = \kappa^2 E_h(\lambda)$ links calibration error directly to score error.

Why calibrate an edge instead of completing the spectrum? A distribution-dependent D-transform response requires the uncomputed singular values. Appendix B.3 constructs two binary graphs with identical degrees, leading singular triples and Frobenius energy whose full-spectrum responses nevertheless differ after any rescaling: these summaries cannot identify such a response. CalibCF sidesteps the problem. Graph margins specify a template, the head fits a single scale, and the deployed response needs nothing below its cutoff.

5 Experiments

In the experimental section, we answer four questions. **RQ1:** How accurate is CalibCF against training-free and learned recommenders? **RQ2:** Does the graph template predict the spectrum it never sees? **RQ3:** Does the calibrated multiplier transfer across graphs? **RQ4:** Which components drive accuracy, how much of the spectrum must be computed, and what does calibration cost?

Table 1: Dataset statistics. Interactions count train, validation and test events after preprocessing.

Dataset	Users	Items	Interactions	Density
ML-1M	6,034	3,125	574,376	3.046%
Gowalla	29,858	40,981	1,027,370	0.084%
Yelp2018	31,668	38,048	1,561,406	0.130%
ML-10M	69,166	8,790	5,000,415	0.822%
Pinterest	55,186	9,637	1,462,888	0.275%

5.1 Experimental setup

Datasets and protocol. We use five public benchmarks whose densities differ by more than an order of magnitude (Table 1): ML-1M and ML-10M (Harper & Konstan, 2015), Gowalla and Yelp2018 (He et al., 2020), and Pinterest (He et al., 2017). MovieLens ratings ≥ 4 are treated as positive, and MovieLens and Pinterest interactions are 5-core filtered. Every dataset is split per user into training, validation and test sets at 70%/10%/20%. All methods rank the full item catalog rather than a sampled subset (Krichene & Rendle, 2020), excluding items already seen in the fitting data, and we report Recall@20 and NDCG@20 averaged over users. Hyperparameters are selected on validation, and final models are refit on training plus validation data.

Table 2: Full-catalog Recall@20 / NDCG@20. Bold marks the best value per row. LightGCN results are averaged over the available seeds (Appendix D).

		Training-free spectral				Closed-form	Learned embeddings		
Dataset	Metric	SVD	GF-CF	Turbo-CF	ChebyCF	EASE	MF	LightGCN	CalibCF
ML-1M	R@20	0.3032	0.2828	0.2828	0.3004	0.2937	0.2440	0.2827	**0.3092**
	N@20	0.3374	0.3170	0.3143	0.3357	0.3349	0.2728	0.3136	**0.3471**
Gowalla	R@20	0.1823	0.1849	0.1830	0.1901	0.1764	0.1343	0.1817	**0.1943**
	N@20	0.1516	0.1518	0.1513	0.1584	0.1466	0.1059	0.1553	**0.1624**
Yelp2018	R@20	0.0718	0.0697	0.0693	0.0714	0.0648	0.0397	0.0624	**0.0748**
	N@20	0.0596	0.0571	0.0574	0.0593	0.0547	0.0329	0.0512	**0.0620**
ML-10M	R@20	0.3518	0.3384	0.3160	0.3654	0.3676	0.3023	0.3469	**0.3686**
	N@20	0.3361	0.3197	0.2926	0.3473	0.3554	0.2735	0.3248	**0.3555**
Pinterest	R@20	0.1706	0.1643	0.1658	0.1707	0.1628	0.1160	0.1548	**0.1739**
	N@20	0.1201	0.1151	0.1166	0.1203	0.1150	0.0791	0.1068	**0.1230**

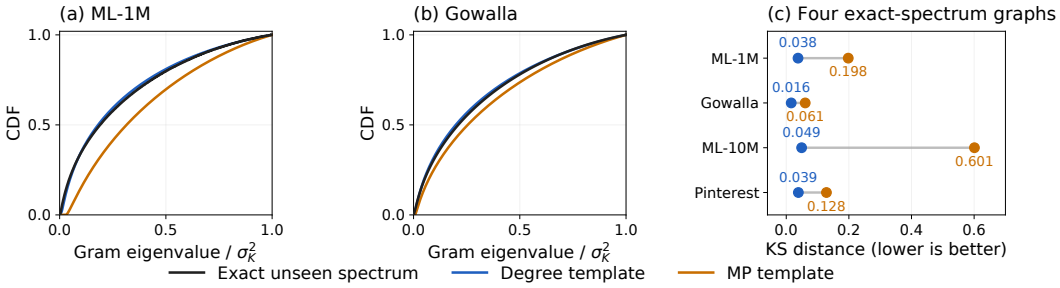

Figure 2: The graph template predicts the spectrum outside its fitting window. (a,b) Predicted and exact CDFs of uncomputed Gram eigenvalues. (c) KS distance on four graphs at the smallest evaluated head width. Both templates use the same scale-fitting procedure.

Baselines. We compare against three families. *Training-free spectral filters*: truncated SVD on the same normalized operator with its rank selected on validation (Cremonesi et al., 2010), and GF-CF, Turbo-CF and ChebyCF with their authors' implementations. *Closed-form linear model*: EASE (Steck, 2019). *Learned embeddings*: matrix factorization trained with BPR (MF) (Rendle et al., 2009) and LightGCN (He et al., 2020), the representative graph recommender against which spectral filters are routinely compared; both are retrained on all five datasets with one shared recipe and validation early stopping. CalibCF selects (κ, γ) on validation from one grid shared by all datasets, and Section 5.4 removes target-domain selection. All experiments run on a single NVIDIA RTX 4090 GPU; settings and paired statistics are in Appendix D.

5.2 RQ1: Recommendation Accuracy

Finding 1: CalibCF leads on every benchmark. CalibCF leads every row of Table 2. It improves NDCG@20 over truncated SVD by 2.3–7.1% and over the strongest external baseline on each dataset by up to 4.0%. The lead spans spectral filters, the closed-form EASE model and learned MF and LightGCN embeddings: 65 of the 66 paired baseline contrasts are significant after Holm correction. The label-free setting is already strong: with the fixed operating point $(\kappa, \gamma) = (.9, .5)$ and no target labels, CalibCF beats every tuned baseline on four of the five benchmarks. On the public Gowalla and Yelp2018 splits, CalibCF also exceeds every published training-free result we collected, including ChebyCF, PolyCF and BSPM (Table 8).

5.3 RQ2: The Graph Predicts Its Own Spectrum

Finding 2: Degree information predicts the unseen spectrum 3–12× more accurately than MP. To test the spectral model independently of recommendation metrics, we compute exact spectra for four graphs, up to Gowalla with $p = 29{,}858$. Fitting uses only the leading 4.9–10.4% of each

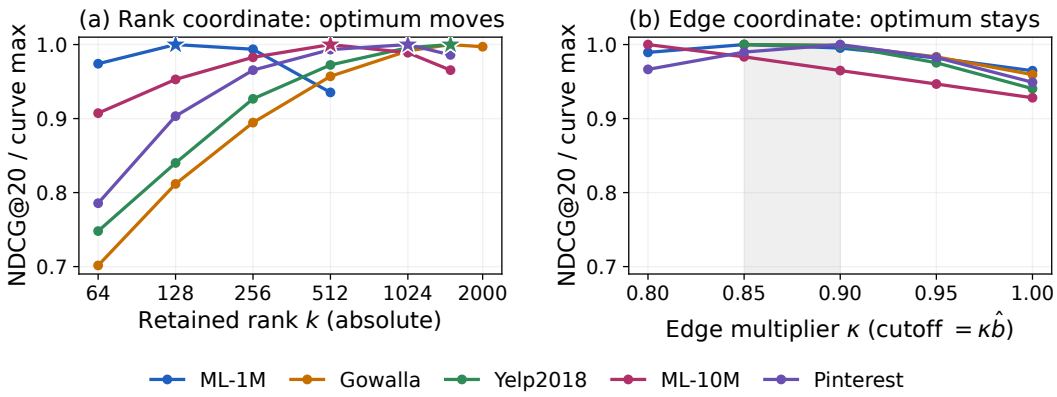

Figure 3: Where the best cutoff lies, NDCG@20 normalized by each curve's maximum ($\gamma = .5$). (a) The same response with an absolute-rank cutoff: the optimum (stars) ranges over 128–1,500 across graphs. (b) The calibrated coordinate: the optimum stays within $\kappa \in [0.85, 0.9]$ (shaded) on four graphs, although over the plotted range the retained rank changes by 3.7–8.7× on each graph.

Table 3: Operating-point transfer, NDCG@20. Fixed uses one setting for all graphs; leave-one-dataset-out (LODO) selects it on the other four datasets' validation sets. Rank uses the same response and readout with an absolute-rank cutoff; its LODO grid covers $k \leq 512$, the ranks available under every head width. The last row averages NDCG@20 relative to CalibCF in Table 2.

	CalibCF: edge multiplier κ		Rank: absolute rank k	
Dataset	fixed $(.9, .5)$	LODO (κ, γ)	fixed k=512	LODO (k, γ)
ML-1M	0.3471	0.3488 (.85,.5)	0.3255	0.3370 (256,.25)
Gowalla	0.1622	0.1490 (.9,.25)	0.1558	0.1365 (256,.25)
Yelp2018	0.0602	0.0603 (.85,.5)	0.0588	0.0560 (256,.5)
ML-10M	0.3369	0.3351 (.9,.25)	0.3554	0.3440 (256,.25)
Pinterest	0.1227	0.1215 (.85,.5)	0.1215	0.1181 (256,.5)
mean relative to Table 2	98.3%	96.5%	96.7%	92.9%

spectrum. On the remaining eigenvalues, which the fit never sees, the degree template reaches KS distance 0.016–0.049, against 0.06–0.60 for the MP template fitted identically (Figure 2). Because both templates share the same scale fit, the entire gain comes from the shape supplied by the graph. Law-completed D-transform weights also agree with exact full-spectrum weights to within 0.8%.

Controlled degree-corrected graphs, where the true noise is known, confirm Proposition 2: as the mean degree grows from 15 to 150, the relative error of the calibrated edge falls at every heterogeneity level, from 6.1% to 1.6% under the strongest. They also show where the template matters most: under strong heterogeneity the MP template places its edge inside the noise and flags 4.7–17.7 directions beyond the planted signal, whereas every direction the graph template flags lies above the top noise singular value (Appendix G).

5.4 RQ3: One multiplier across graphs

Finding 3: One edge multiplier works where one rank cannot. Figure 3 plots accuracy against the cutoff in two coordinates. Measured in absolute rank, the best cutoff moves by an order of magnitude across graphs. Measured as a multiple of the calibrated edge, it stays put: $\kappa = 0.9$ is within 0.5% of each curve's maximum on four graphs. The graph moves the cutoff, so the practitioner does not have to.

This stability carries over to unseen graphs (Table 3). Fixing $(\kappa, \gamma) = (.9, .5)$ for every dataset retains 98.3% of CalibCF's in-domain NDCG@20 on average, versus 96.7% for the best fixed absolute rank. Selecting the operating point on the other four datasets retains 96.5% versus 92.9%, with higher NDCG@20 on four of five graphs. When target validation is available, the multiplier adapts further:

Table 4: Removing one component at a time from the fixed operating point, NDCG@20. The spacing edge and hard truncation keep CalibCF's head, operator and readout.

Variant	ML-1M	Gowalla	Yelp2018	ML-10M	Pinterest
CalibCF, (κ, γ) selected on validation	0.3471	0.1624	0.0620	0.3555	0.1230
CalibCF, fixed $(\kappa, \gamma) = (.9, .5)$	0.3471	0.1622	0.0602	0.3369	0.1227
calibrated edge $\rightarrow$ eOptShrink spacing edge	0.3308	0.1159	0.0451	0.3297	0.1001
closed-form response $\rightarrow$ hard truncation	0.3324	0.1538	0.0585	0.3358	0.1147

Table 5: Calibration as the computed head grows, at the fixed operating point $(\kappa, \gamma) = (.9, .5)$. Ranges are over the evaluated head widths.

Dataset	head width K (% of spectrum)	calibrated edge $\hat{b}$	retained k	NDCG@20
ML-1M	152–800 (4.9–25.6%)	0.1286–0.1294	112–118	0.3471–0.3475
Gowalla	1,750–2,250 (5.9–7.5%)	0.3152–0.3169	1,153–1,204	0.1621–0.1622
ML-10M	500–2,000 (5.7–22.8%)	0.1411–0.1414	117–118	0.3368–0.3371
Pinterest	1,000–2,000 (10.4–20.8%)	0.2390–0.2393	739–750	0.1227

on ML-10M a deeper $\kappa = 0.7$ reaches 0.3555 NDCG@20, and the gain holds on three random splits (Figure 4).

5.5 RQ4: Components, head width and cost

Finding 4: Every component contributes. Table 4 removes one component at a time from the fixed operating point. Replacing the calibrated edge with the published leading-spacing edge estimator of eOptShrink (Su & Wu, 2025) costs 2.1–28.5% of NDCG@20, because that estimator retains far too few directions; replacing the closed-form response with hard truncation on the same support costs 0.3–6.5%. The closed-form response is also within 0.0002 of the tail-completed D-transform reference on four graphs (Appendix F), so nothing is lost by leaving the tail uncomputed. The two operating parameters play separate roles: on ML-10M the cutoff accounts for the full gain from 0.3369 to 0.3555, while on Yelp2018 the readout adds +0.0018 without changing the retained subspace (Table 16).

Finding 5: A short head suffices. CalibCF uses only the computed head, so a useful calibration should not depend on how much of the spectrum is computed. Table 5 confirms this: enlarging the head up to 5.3-fold moves the calibrated edge by less than 0.6% and NDCG@20 by at most 0.0004 on each graph in the table. Heads covering 4.9–10.4% of the spectrum are therefore enough, which keeps the decomposition, the dominant cost, small.

Finding 6: Calibration is nearly free. The graph proxy costs $O(\mathrm{nnz}\, R)$ to construct. Once the head is available, the law and edge fit take 0.39–0.63 seconds in the final configurations, and every change of operating point reuses this fit. Building a complete model from the interaction matrix and ranking the full catalog for all users takes 9.8–26.1 seconds on one GPU (Appendix E), most of it spent on the truncated SVD; a single LightGCN training run with a known number of epochs takes 128–7,558 seconds on the same GPU, 13–331$\times$ longer. The model stores two factor matrices with $(m + n)k$ entries, and exhaustive ranking costs $O(mnk)$.

6 Conclusion

CalibCF shows that a recommendation graph carries the information needed to place its own spectral cutoff: degree margins give the shape of the noise spectrum, and a short computed head gives its scale, a combination that provably recovers the noise edge under a degree-corrected model. The resulting training-free recommender outperforms all compared methods on five benchmarks, predicts unseen spectra 3–12$\times$ more accurately than an MP model, and expresses its cutoff in a coordinate that transfers across graphs. Extending graph-conditioned calibration to temporally evolving graphs and to learned spectral responses is a natural next step.

AI USAGE STATEMENT

In this work, we used generative AI tools to polish the language and check the manuscript. The authors reviewed all AI-assisted changes and take full responsibility for this submission.

REPRODUCIBILITY STATEMENT

The supplementary material contains runnable code for our proposed method that downloads the five datasets. We will publicly release the code for all experiments. Proofs and implementation details are provided in the appendix.

REFERENCES

Johannes Alt, László Erdős, and Torben Krüger. Local law for random gram matrices. *Electronic Journal of Probability*, 22:no. 25, 1–41, 2017. doi:10.1214/17-EJP42.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005. doi:10.1214/009117905000000233.

Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012. arXiv:1103.2221.

Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 56(3):2141–2161, 2020. doi: 10.1214/19-AIHP1033. arXiv:1704.02945.

Xuheng Cai, Chao Huang, Lianghao Xia, and Xubin Ren. LightGCL: Simple yet effective graph contrastive learning for recommendation. In *International Conference on Learning Representations*, 2023.

Jeongwhan Choi, Seoyoung Hong, Noseong Park, and Sung-Bae Cho. Blurring-sharpening process models for collaborative filtering. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023. arXiv:2211.09324.

Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-N recommendation tasks. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, pp. 39–46, 2010. doi:10.1145/1864708.1864721.

Edgar Dobriban and Art B. Owen. Deterministic parallel analysis: An improved method for selecting factors and principal components. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(1):163–183, 2019. doi:10.1111/rssb.12301.

David L. Donoho, Matan Gavish, and Elad Romanov. ScreeNOT: Exact MSE-optimal singular value thresholding in correlated noise. *The Annals of Statistics*, 51(1), 2023. doi:10.1214/22-AOS2232; arXiv:2009.12297.

Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pp. 101–109, 2019. doi:10.1145/3298689.3347058.

Matan Gavish and David L. Donoho. Optimal shrinkage of singular values. *IEEE Transactions on Information Theory*, 63(4):2137–2152, 2017. doi:10.1109/TIT.2017.2653801.

F. Maxwell Harper and Joseph A. Konstan. The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4):19:1–19:19, 2015. doi:10.1145/2827872.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 173–182, 2017. doi:10.1145/3038912.3052569.

Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. LightGCN: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020. arXiv:2002.02126.

Yunhang He, Cong Xu, Zhangchi Zhu, Hongzhi Yin, and Wei Zhang. ASPIRE: Make spectral graph collaborative filtering great again via adaptive filter learning. *arXiv preprint arXiv:2604.22549*, 2026.

Jong Yun Hwang, Ji Oon Lee, and Kevin Schnelli. Local law and Tracy–Widom limit for sparse sample covariance matrices. *The Annals of Applied Probability*, 29(5):3006–3036, 2019. doi:10.1214/19-AAP1472; arXiv:1806.03186.

Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011. doi:10.1103/PhysRevE.83.016107.

Farhan Khawar and Nevin L. Zhang. Cleaned similarity for better memory-based recommenders. In *Proceedings of the 42nd International ACM SIGIR Conference*, 2019. doi: 10.1145/3331184.3331310.

Chanwoo Kim, Jinkyu Sung, Yebonn Han, and Joonseok Lee. Graph spectral filtering with chebyshev interpolation for recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025. arXiv:2505.00552.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009. doi:10.1109/MC.2009.263.

Walid Krichene and Steffen Rendle. On sampled metrics for item recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1748–1757, 2020. doi:10.1145/3394486.3403226.

Boris Landa, Thomas T. C. K. Zhang, and Yuval Kluger. Biwhitening reveals the rank of a count matrix. *SIAM Journal on Mathematics of Data Science*, 2022. doi:10.1137/21M1456807; arXiv:2103.13840.

Can M. Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017. doi: 10.1002/rsa.20713. arXiv:1506.00669.

William Leeb and Elad Romanov. Optimal spectral shrinkage and PCA with heteroscedastic noise. *IEEE Transactions on Information Theory*, 67(5):3009–3037, 2021. doi:10.1109/TIT.2021.3055075; arXiv:1811.02201.

Jiahao Liu, Dongsheng Li, Hansu Gu, Tun Lu, Peng Zhang, Li Shang, and Ning Gu. Personalized graph signal processing for collaborative filtering. In *Proceedings of the ACM Web Conference 2023 (WWW)*, pp. 1264–1272, 2023. doi: 10.1145/3543507.3583466.

Albert Madansky. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, 30(3):743–746, 1959. doi: 10.1214/aoms/1177706203.

Vladimir A. Marčenko and Leonid A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967. doi:10.1070/SM1967v001n04ABEH001994.

Raj Rao Nadakuditi. OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Transactions on Information Theory*, 60(5): 3002–3018, 2014.

Jin-Duk Park, Yong-Min Shin, and Won-Yong Shin. Turbo-CF: Matrix decomposition-free graph filtering for fast recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024. arXiv:2404.14243.

Debashis Paul and Jack W. Silverstein. No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis*, 100(1): 37–57, 2009.

Shaowen Peng, Kazunari Sugiyama, and Tsunenori Mine. SVD-GCN: A simplified graph convolution paradigm for recommendation. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pp. 1625–1634, 2022. doi:10.1145/3511808.3557462.

Shaowen Peng, Xin Liu, Kazunari Sugiyama, and Tsunenori Mine. How powerful is graph filtering for recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024. arXiv:2406.08827.

Tai Qin and Karl Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Yifang Qin, Wei Ju, Xiao Luo, Yiyang Gu, Zhiping Xiao, and Ming Zhang. PolyCF: Towards the optimal spectral graph filters for collaborative filtering. *arXiv preprint arXiv:2401.12590*, 2024.

Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 452–461, 2009.

Yifei Shen, Yongji Wu, Yao Zhang, Caihua Shan, Jun Zhang, Khaled B. Letaief, and Dongsheng Li. How powerful is graph convolution for recommendation? In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, 2021. arXiv:2108.07567.

Harald Steck. Embarrassingly shallow autoencoders for sparse data. In *The World Wide Web Conference*, pp. 3251–3257, 2019. doi: 10.1145/3308558.3313710.

Pei-Chun Su. Data-driven matrix recovery via optimal shrinkage and spatially resolved singular vector denoising under high-dimensional separable noise. *arXiv preprint arXiv:2507.09013*, 2026.

Pei-Chun Su and Hau-Tieng Wu. Data-driven optimal shrinkage of singular values under high-dimensional noise with separable covariance structure with application. *Applied and Computational Harmonic Analysis*, 74:101698, 2025. doi: 10.1016/j.acha.2024.101698.

Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. Self-supervised graph learning for recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 726–735, 2021. doi:10.1145/3404835.3462862.

Fan Yang. Edge universality of separable covariance matrices. *Electronic Journal of Probability*, 24: no. 123, 1–57, 2019. doi:10.1214/19-EJP381.

Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. Are graph augmentations necessary? Simple graph contrastive learning for recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1294–1303, 2022. doi:10.1145/3477495.3531937.

Yihan Zhang and Marco Mondelli. Matrix denoising with doubly heteroscedastic noise: Fundamental limits and optimal spectral methods. *arXiv preprint arXiv:2405.13912*, 2024.

A CLOSEST METHODS

Table 6 contrasts CalibCF with the closest shrinkage and spectral methods in terms of the information each one uses to set the scale of its response. CalibCF keeps the Gavish–Donoho response and replaces its scale estimate by one that is conditioned on the graph and fitted to the computed head only.

Table 6: Comparison of calibration inputs and the quantity each method estimates.

Route	Assumption or information used	CalibCF
Gavish–Donoho (Gavish & Donoho, 2017)	MP bulk; unknown scale from the bulk median, which a short head does not contain.	Graph-conditioned bulk shape; scale from upper-bulk quantiles that are computed; same response.
OptShrink / ScreeNOT (Nadakuditi, 2014; Donoho et al., 2023)	Functionals of the empirical bulk distribution beyond the head.	One edge coordinate; the deployed response needs nothing below the cutoff.
eOptShrink (Su & Wu, 2025)	Edge from leading-eigenvalue spacings, then an estimated noise CDF.	Template from graph variance margins; edge from a regression on the upper bulk (its spacing estimator is a control in Table 17).
Biwhitening (Landa et al., 2022)	Rescales rows and columns so that the bulk becomes MP.	Keeps the recommendation operator and models its remaining heterogeneity.
Spectral CF / RMT cleaning (Khawar & Zhang, 2019; Qin et al., 2024; He et al., 2026)	Per-dataset selected or learned filters; RMT cleaning of similarities.	A graph-and-head cutoff estimator, isolated from response design and readout.

B PROOFS

B.1 THE EDGE-CALIBRATION FUNCTIONAL

Let $\lambda_j = \sigma_{j+1}^2$, $q_j = F_0^{-1}(u_j)$ and $h_j = q_j/b_0^2$ on the window $\mathcal{J}_r$. For positive h and nonnegative λ the least-squares problem in Eq.(4) is strictly convex and its unconstrained minimizer $h^\top\lambda/h^\top h$ is nonnegative, hence feasible. Fitting a scale c to the unnormalized quantiles gives $\hat{c} = q^\top\lambda/q^\top q$ and $\hat{b}^2 = b_0^2\hat{c} = h^\top\lambda/h^\top h$, so the two formulations coincide, including their exclusion updates.

Proof of Proposition 1. Write $a_h = h/(h^\top h)$ and $a_g = g/(g^\top g)$. If $E_h(\lambda) = cE_g(\lambda)$ on a nonempty open set, the linear functional $(a_h - ca_g)^\top\lambda$ vanishes on that set, hence $a_h = ca_g$; this forces $h = tg$ for some $t > 0$, and substitution gives $c = 1/t$. The converse is immediate. For the bound, $E_h(\lambda) - x_* = h^\top\{\varepsilon + x_*(h_* - h)\}/\|h\|_2^2$, and Cauchy–Schwarz with the triangle inequality gives the claim. □

For each benchmark, the degree template is not proportional to the edge-normalized MP template over the fixed fitting window, so the two templates induce distinct calibration functionals. An independent recomputation reproduces all 22 calibration iterations with relative error below 6×10^{-16}.

B.2 CUTOFF STABILITY

Proof of Proposition 3. For $\beta \in (0, 1]$ write $a = a_\beta \in [0, 1)$ and $x = \theta^2$. Factoring the radicand of Eq.(5), with its zero extension below the cutoff, $g_\sigma(x) = \eta_\beta(\sigma; \sqrt{x})^2 = (\sigma^2 - x)(\sigma^2 - ax)/\sigma^2$ for $x < \sigma^2$ and 0 otherwise. It is continuous, nonnegative, and $|g'_\sigma(x)| = |-(1 + a) + 2ax/\sigma^2| \leq$

$1 + a$ on the first interval, so it is globally $(1 + a)$-Lipschitz, including pairs that cross σ^2. Since $(\sqrt{v} - \sqrt{w})^2 \le |v - w|$ for $v, w \ge 0$, the scalar bound follows. For fixed vectors, subtract the two reconstructions and apply Cauchy–Schwarz to $\sum_j |U_{uj}V_{ij}|\,|\eta_\beta(\sigma_j;\theta) - \eta_\beta(\sigma_j;\theta')|$, then multiply by the positive user and item factors of Eq.(6). □

A sufficient condition for an unchanged top-N set follows by comparing score intervals. A numerical check over 144,000 scalar cases, including threshold crossings, confirms the scalar bound.

B.3 Response ambiguity from identical graph summaries

For $p = q$ the D-transform response reduces to $w = a/b$. Let J_d be the $d \times d$ all-ones matrix and P_{10} a cyclic permutation; define $C_A = I + P_{10} + P_{10}^2 + P_{10}^3 + P_{10}^4$, $C_B = \mathrm{diag}(J_5, J_5)$ and $R_X = \mathrm{diag}(J_8, J_7, J_6, C_X)$. Both R_X are binary 31×31 matrices with 199 edges, identical labeled degrees $(8^{[8]}, 7^{[7]}, 6^{[6]}, 5^{[10]})$, minimum degree five, identical leading three singular triples y_8, y_7, y_6 with $y_d = d/(d+\tau)$ and $\tau = 199/124$ at $c_\tau = 1/4$, identical Frobenius energy $y_8^2 + y_7^2 + y_6^2 + 50/(5+\tau)^2$, and identical variance-proxy margins. With $G_X = C_X C_X^\top/(5+\tau)^2$ and $M_X(y) = (y^2 I - G_X)^{-1}$, at $k = 2$ the excluded spectrum is y_6^2, the ten eigenvalues of G_X and eighteen zeros, so $a_X(y) = [y \,\mathrm{tr}\, M_X(y) + y/(y^2 - y_6^2) + 18/y]/31$ and $b_X(y) = [\mathrm{tr}((y^2 I + G_X)M_X(y)^2) + (y^2 + y_6^2)/(y^2 - y_6^2)^2 + 18/y^2]/31$. The weight ratios $w(y_8)/w(y_7)$ are $2.147115848\ldots$ for A and $2.010524909\ldots$ for B (verified in exact rational arithmetic): no common scale equates the two responses, so the summaries do not identify the full-spectrum response.

B.4 Finite-spectrum response enclosure

Theorem 1 (Mass–energy enclosure). *Suppose $1 \le k < K < p \le q$, the head contains exact leading singular values and $\sigma_k > \sigma_{k+1}$. Set $t = \sigma_K^2$, $m_0 = (p - K)/p$, $m_1 = (\|L\|_F^2 - \sum_{\ell \le K} \sigma_\ell^2)/p$, and for $y = \sigma_j$, $j \le k$, $f_y(x) = y/(y^2 - x)$, $h_y(x) = (y^2 + x)/(y^2 - x)^2$. For $t > 0$ and $g \in \{f_y, h_y\}$ put $J_g^- = m_0 g(m_1/m_0)$ and $J_g^+ = (m_0 - m_1/t)g(0) + (m_1/t)g(t)$ (both $m_0 g(0)$ if $t = 0$), $a_\pm = p^{-1}\sum_{\ell=k+1}^K f_y(\sigma_\ell^2) + J_{f_y}^\pm$, $b_\pm = p^{-1}\sum_{\ell=k+1}^K h_y(\sigma_\ell^2) + J_{h_y}^\pm$ and $w(a,b) = 2/\{b/a + [\beta b + (1-\beta)/y^2]/[\beta a + (1-\beta)/y]\}$. Then $w(a_-, b_+) \le w_j^{\mathrm{full}} \le w(a_+, b_-)$.*

Proof. The missing measure $\mu_{\mathrm{tail}} = p^{-1}\sum_{\ell > K} \delta_{\sigma_\ell^2}$ is supported on $[0, t]$ with mass m_0 and first moment m_1; the gap $\sigma_k > \sigma_{k+1}$ makes both kernels finite. $f_y''(x) = 2y/(y^2 - x)^3 > 0$ and $h_y''(x) = (10y^2 + 2x)/(y^2 - x)^4 > 0$, so Jensen gives $\int g\, d\mu_{\mathrm{tail}} \ge m_0 g(m_1/m_0)$ and the Edmundson–Madansky secant $g(x) \le (1 - x/t)g(0) + (x/t)g(t)$ gives the upper bound (Madansky, 1959). With $H(a,b) = b/a + (\beta b + c_0/y^2)/(\beta a + c_0/y)$, $c_0 = 1 - \beta$, one has $\partial_b H > 0$ and $\partial_a H < 0$ for $a, b > 0$, so $w = 2/H$ increases in a and decreases in b, which places w_j^{full} in the stated interval. □

The enclosure tightens as the head grows: on ML-1M the relative weight bound falls from 28.2% at $K = 152$ to 1.8% at $K = 800$. A tail-dependent response therefore depends on the head width, whereas the calibrated edge and the closed-form response do not.

B.5 Population motivation

Let independent Bernoulli entries have means $P_{ui} = \rho_n a_u b_i(\mu + f_{ui})$ with $f_{ui} = \sum_{\ell \le r} \theta_{\ell,n}(mn)^{-1/2} x_{u\ell} y_{i\ell}$, $m/n \to \gamma \in (0, \infty)$, fixed $\mu > 0$, $a_u, b_i \in [c_-, C_+]$ with limiting empirical means one, $\bar{d}_n = \rho_n \mu n \to \infty$, $\epsilon_{f,n} = \|f\|_\infty/\mu = o(1)$ and $\pi_n = \max_{ui} P_{ui} = o(1)$. With population degrees $\bar{d}_u, \bar{e}_i$, $N^* = \sum P_{ui}$ and regularizers $\tau_u^* = c_\tau N^*/m$, $\tau_i^* = c_\tau N^*/n$, the fixed-normalizer noise $F_u^*(R - P)F_i^*$ has entry variance $S_{ui} = P_{ui}(1 - P_{ui})/[(\bar{d}_u + \tau_u^*)(\bar{e}_i + \tau_i^*)]$, and for the separable benchmark $S_{ui}^0 = \rho_n \mu a_u b_i/[(\bar{d}_u + \tau_u^*)(\bar{e}_i + \tau_i^*)]$, $\max_{ui} |S_{ui}/S_{ui}^0 - 1| \le \epsilon_{f,n} + \pi_n + \epsilon_{f,n}\pi_n = o(1)$.

Proposition 4 (Uniform normalization transfer). *If in addition $\bar{d}_n/\log(m+n) \to \infty$, then with $E_u = \mathrm{diag}\,\sqrt{(\bar{d}_u + \tau_u^*)/(d_u + \tau_u)}$, $E_i = \mathrm{diag}\,\sqrt{(\bar{e}_i + \tau_i^*)/(e_i + \tau_i)}$ and $\eta_n = \sqrt{\log(m+n)/\bar{d}_n} + \log(m+n)/\bar{d}_n$, on events of probability tending to one $\|E_u - I\| + \|E_i - I\| = O_{\mathbb{P}}(\eta_n)$ and, for every X of the same shape, $F_u X F_i = E_u(F_u^* X F_i^*)E_i$ and $\|F_u X F_i - F_u^* X F_i^*\| \le O_{\mathbb{P}}(\eta_n)\|F_u^* X F_i^*\|$.*

Proof. Bounded propensities and $\|f\|_\infty/\mu = o(1)$ keep all population degrees between positive constants times $\bar{d}_n$. Bernstein's inequality with a union bound gives maximum degree deviations $O_{\mathbb{P}}(\sqrt{\bar{d}_n \log(m+n)} + \log(m+n))$ and $|N - N^*| = O_{\mathbb{P}}(\sqrt{m\bar{d}_n})$, so every relative denominator deviation is $O_{\mathbb{P}}(\eta_n)$; $x \mapsto (1+x)^{-1/2}$ is Lipschitz on $[-1/2, 1/2]$. The identity is definitional and $E_u B E_i - B = (E_u - I)B + B(E_i - I) + (E_u - I)B(E_i - I)$ gives the norm bound. □

This motivates the proxy in Eq.(2) and is the first step of the consistency proof below.

B.6 Consistency of the calibrated edge

We work in the setting of Appendix B.5 with $\bar{d}_n/\log(m+n) \to \infty$ and a fixed signal rank r. Write $X^* = F_u^*(R-P)F_i^*$ for the noise under population normalizers and $N = F_u(R-P)F_i$ for the noise contained in L. Let (ϕ^*, ψ^*) be the margins of S^0, so that $S^0_{ui} = \phi^*_u \psi^*_i$; let μ_n be their separable law from Eq.(3), b_n^2 its right edge and $\nu_n = \mu_n(b_n^2\,\cdot)$ the edge-normalized law. For a law ν on $[0,\infty)$, q_ν denotes its quantile function; $u_t = 1 - (t - 1/2)/p$.

(A3) Separable spectral regularity. (a) The Lévy distance between the empirical distribution of $\{\sigma_t(X^*)^2/b_n^2\}_{t\le p}$ and ν_n tends to zero in probability, and $\sigma_1(X^*)/b_n \to 1$ in probability. (b) For every $\delta_n \to 0$, $d_{\mathrm{L}}(\nu', \nu_n) + |b'/b_n - 1| \to 0$ uniformly over separable margins (ϕ', ψ') with $\|\phi'/\phi^* - 1\|_\infty + \|\psi'/\psi^* - 1\|_\infty \le \delta_n$, where ν' and b'^2 are their normalized law and edge.

Part (a) states that the noise follows the separable law of its variance margins with no singular values beyond its edge. This is the standard regularity of separable covariance models, established when the standardized entries are i.i.d. with a finite fourth moment (Paul & Silverstein, 2009). The non-separable part of S is uniformly negligible (Appendix B.5), and $\bar{d}_n \gg \log(m+n)$ is the regime in which the extreme eigenvalues of sparse Erdős–Rényi graphs converge to the edges of the limiting spectrum (Benaych-Georges et al., 2020). Part (b) is inherited from (a) with Gaussian entries: a relative change of separable margins is a two-sided diagonal rescaling $W \mapsto D_1 W D_2$, which multiplies every singular value by a factor between $(\|D_1^{-1}\|\|D_2^{-1}\|)^{-1}$ and $\|D_1\|\|D_2\|$.

Proof of Proposition 2. Let $r' = \mathrm{rank}(F_u P F_i) \le r+1$, since $P = \rho_n D_a(\mu \mathbf{1}\mathbf{1}^\top + f)D_b$. All statements below hold with probability tending to one.

Step 1 (normalization). $F_u = E_u F_u^*$ and $F_i = E_i F_i^*$ with E_u, E_i from Proposition 4, so $N = E_u X^* E_i$ and, for every t, $\sigma_t(N)/\sigma_t(X^*) \in [(\|E_u^{-1}\|\|E_i^{-1}\|)^{-1}, \|E_u\|\|E_i\|] = [1 - O(\eta_n), 1 + O(\eta_n)]$. With (A3a), $\sigma_1(N)/b_n \to 1$ and the normalized spectrum $\{\sigma_t(N)^2/b_n^2\}$ is within vanishing Lévy distance of ν_n.

Step 2 (signal). $L = N + F_u P F_i$, and Weyl's inequalities give $\sigma_{t+r'}(N) \le \sigma_t(L) \le \sigma_{t-r'}(N)$ for $t > r'$. Hence for every fixed $\epsilon > 0$ at most r' singular values of L exceed $(1+\epsilon)b_n$.

Step 3 (proxy margins). Uniformly in u, i, $d_u + \tau_u = (\bar{d}_u + \tau_u^*)(1 + O(\eta_n))$ by the proof of Proposition 4, and $\hat{P}_{ui} = O(\pi_n)$; thus $Q_{ui} = R_{ui} w_{ui}(1 + O(\eta_n + \pi_n))$ with $w_{ui} = [(\bar{d}_u + \tau_u^*)(\bar{e}_i + \tau_i^*)]^{-1}$. The row sum $\sum_i R_{ui} w_{ui}$ has mean $\sum_i S^0_{ui}(1 + O(\epsilon_{f,n}))$ and summands bounded by $\max_i w_{ui}$, which bounded propensities keep within a factor $O(1/\bar{d}_n)$ of that mean; Bernstein's inequality with a union bound gives relative deviations $O(\sqrt{\log(m+n)/\bar{d}_n})$ uniformly in u, and columns and the total are analogous. Both the proxy profile $r_u c_i / M$ and S^0 are products of their margins, so the proxy margins satisfy $\hat{\phi} = \alpha\phi^*$, $\hat{\psi} = \beta\psi^*$ with $\|\alpha - 1\|_\infty + \|\beta - 1\|_\infty = O(\eta_n + \pi_n + \epsilon_{f,n})$. By (A3b), $b_0/b_n \to 1$ and the normalized template law $\hat{\nu}$ satisfies $d_{\mathrm{L}}(\hat{\nu}, \nu_n) \to 0$.

Step 4 (exclusion counts). By Step 3, $1.02 b_0 \ge (1+\epsilon)b_n$ for a fixed $\epsilon > 0$, so the initial count $\#\{t : \sigma_t > 1.02 b_0\}$ is at most r' by Step 2. Step 5 shows $\hat{b}(r)/b_n \to 1$ for each of the finitely many $r \le r'$; hence every update is again at most r', and by induction so is every visited count.

Step 5 (regression). Fix $r \le r'$. The window $\mathcal{J}_r$ indexes $\lambda_t = \sigma_t(L)^2$ for $t = j+1 \in [r+6, K_b]$, so $|\mathcal{J}_r|/p \to 0.85 k_0$. Set $g_t = q_{\nu_n}(u_{t-r})$, so that $h_t = q_{\hat{\nu}}(u_{t-r})$. Proposition 1 with $h_* = g$ and $x_* = b_n^2$ gives

$$\left|\frac{\hat{b}^2}{b_n^2} - 1\right| \le \frac{\|\lambda/b_n^2 - g\|_2 + \|g - h\|_2}{\|h\|_2}.$$

Quantile functions of laws supported in $[0, 1 + o(1)]$ are monotone and bounded, so evaluating them on the grid (u_t) costs $O(1)$ in squared ℓ_2 distance and shifting indices by at most $2r' + 1$ costs $O(r')$. By Steps 1–2, $\|\lambda/b_n^2 - g\|_2^2 \le p\, W_2^2(\nu_N, \nu_n) + O(1)$, where ν_N is the normalized spectrum of N; by Step 3, $\|g - h\|_2^2 \le p\, W_2^2(\hat{\nu}, \nu_n) + O(1)$. On uniformly bounded supports Lévy convergence implies W_2 convergence, so both are $o(p)$. Finally, the window covers the top quantile levels, so $\|h\|_2^2/p \ge 0.85 k_0 \int_0^1 q_{\hat{\nu}}^2 + o(1) \ge 0.85 k_0 (\mathbb{E}_{\hat{\nu}} x)^2 + o(1)$; bounded propensities keep all variance margins within constant factors of their means, and the diagonal rescaling comparison of (A3b) then bounds $\mathbb{E}_{\nu_n} x$ away from zero. Hence $\hat{b}^2/b_n^2 \to 1$, and Step 1 gives $\hat{b}/\sigma_1(N) \to 1$. □

C Implementation

Analytic edge. For real $T \in (-1/\max_i \psi_i, 0)$, $\Delta(T) = \sum_i \psi_i/(1+\psi_i T)$ is explicit and $F_T(z) = \sum_u \phi_u/(\phi_u \Delta(T) - z)$ increases from $-\infty$ to 0^- on $z > \max_u \phi_u \Delta(T)$, so $F_T(z) = T$ has a unique root $z(T)$ bracketed by $\max_u \phi_u \Delta(T)$ and $\max_u \phi_u \Delta(T) + 2\sum_u \phi_u/|T|$. The right support edge is $\min_T z(T)$, located by a logarithmic coarse grid in $|T|$ and bounded scalar minimization; each evaluation is a bracketed root-find of cost $O(p+q)$. For entry variance $1/q$ the result equals $(1+\sqrt{\beta})^2$ to machine precision.

Law and scale fit. The density uses 600 grid points, imaginary part $\eta = 1.25 \times 10^{-4} x_{\max}$, 400 damped iterations (damping 0.5) and at most 4000 profile quantiles per side; it is truncated at the analytic edge and renormalized, and enters only through its quantiles. The fitted scale changes by less than 0.1% over a decreasing-η grid (Appendix G). With the current excluded count r, $K_b = K - \mathrm{round}(0.15K)$ and $\mathcal{J} = \{r+5, \dots, K_b - 1\}$ (zero-based), $u_j = 1 - (j - r + 0.5)/p$; a fit uses at least ten window points and a support below $K - 10$. The initial count uses the unscaled edge.

Response family. Eq.(5) is algebraically identical to the Gavish–Donoho response with its edge at θ. CalibCF uses that expression as a fixed filter family anchored at the calibrated edge. Propositions 1, 2 and 3 concern the calibration functional, its consistency and cutoff sensitivity, respectively.

Completed tail (analysis reference only). $\hat{\mu}_{k,K} = p^{-1} \sum_{j=k+1}^{K} \delta_{\sigma_j^2} + \frac{p-K}{p} \hat{\rho}_{<t}$ with $\hat{\rho}_{<t}$ the fitted density on $[0, \sigma_K^2]$ (600 quadrature points, exact mass); a moment-constrained variant tilts it exponentially to match the exact first moment. Both use training data only and are used solely to compare the law with exact spectra.

Cost. Proxy and Frobenius energy are $O(\mathrm{nnz}\, R)$; the law is $O(IGJ)$ for I iterations, G grid points and J quantiles; the edge is a few hundred root-finds of $O(p+q)$; the closed form is $O(k)$. A randomized head of width $\ell = K + s$ with t power iterations costs $O((t+1)[\mathrm{nnz}(R)\ell + (m+n)\ell^2] + \ell^3)$. After fitting, the factors $A = F_u^{-1} U_k \,\mathrm{diag}(w)$ and $B = V_k^\top D_i^{(\gamma)}$ have $(m+n)k$ entries in total, and exhaustive ranking costs $O(mnk)$ plus top-N selection.

D Datasets, baselines and statistics

Datasets and heads. Table 7 lists, for each dataset, the number of training and validation interactions used for the final fit, the head width K, the number r of leading directions excluded by the calibration, the retained dimension k at the fixed operating point $(\kappa, \gamma) = (.9, .5)$, and the rank selected for truncated SVD; the final retained dimensions are given in Appendix E. The head is computed by an exact SVD on ML-1M, by a Lanczos solver on Gowalla and Yelp2018, and by randomized SVD with eight power iterations and 30 oversampling columns on ML-10M and Pinterest.

Configurations and baselines. Every baseline uses its published settings or a configuration selected on validation NDCG@20. GF-CF uses its linear-plus-ideal response; Turbo-CF uses its published settings on the two public splits and a 144-configuration validation search on the other datasets; ChebyCF uses the authors' released coefficients; EASE selects $\lambda \in \{1, 10, 100, 500, 1000, 5000, 10000\}$ on

Table 7: Head widths and retained dimensions.

Dataset	Fit interactions	K	r	k at $(.9, .5)$	SVD rank
ML-1M	459,527	800	53	114	64
Gowalla	810,128	2,250	445	1,204	1,500
Yelp2018	1,237,259	2,000	258	1,005	400
ML-10M	4,000,429	2,000	63	117	128
Pinterest	1,170,944	2,000	177	773	256

validation. MF and LightGCN share one 64-dimensional recipe: BPR loss with uniformly sampled negatives, Adam with learning rate 10^{-3}, penalty 10^{-4}, batch 8192, and validation NDCG@20 every 20 epochs with a patience of 100 epochs, followed by a fresh train+validation refit for the selected number of epochs. LightGCN uses three propagation layers and MF none. On ML-1M and Pinterest, LightGCN results are averaged over three random seeds, and the paired comparisons use the same averages.

Table 8 compares CalibCF with the results reported by the original papers on the public splits of Gowalla and Yelp2018; CalibCF is above every reported result on both datasets.

Table 8: Reported results on the public splits of Gowalla and Yelp2018 (BSPM from its paper; PGSP as re-run by Kim et al. (2025)). SGFCF (Peng et al., 2024) uses different splits and metrics and is therefore not listed. CalibCF is the final method of Table 2.

Method	Gowalla		Yelp2018	
	R@20	N@20	R@20	N@20
ChebyCF (Kim et al., 2025)	0.1941	0.1616	–	–
PolyCF (Qin et al., 2024)	0.1934	0.1621	0.0715	0.0590
BSPM (Choi et al., 2023)	0.1920	0.1597	0.0720	0.0593
PGSP (Kim et al., 2025)	0.1917	0.1606	–	–
GF-CF (Shen et al., 2021)	0.1849	0.1518	0.0697	0.0571
LightGCN (He et al., 2020)	0.1830	0.1554	0.0649	0.0530
CalibCF (ours)	**0.1943**	**0.1624**	**0.0748**	**0.0620**

Statistics. 5,000 paired-user bootstrap replicates give percentile 95% intervals and 5,000 random sign flips give two-sided p-values (plus-one correction), Holm-corrected within each comparison family; multi-seed results are averaged as in Table 2. Additional random splits are reported in Appendix E.

E Selected configurations, additional splits, cost and statistics

This section reports the operating points selected for Table 2, results on two additional random splits, the hyperparameters selected for the closed-form and learned baselines, the computational cost of CalibCF, and the paired statistics behind the significance statements in the main text.

E.1 Selected configurations

Table 9 lists the operating point selected on validation for each dataset, the resulting retained dimension k and the size of the two factor matrices. The selected multiplier lies in $[0.85, 0.9]$ on four of the five datasets.

E.2 Additional random splits

Table 10 repeats the evaluation on two further random 70%/10%/20% splits of ML-1M, ML-10M and Pinterest, reusing the operating point selected on the main split; the edge and the factors are refitted on each split. On ML-10M the selected deeper cutoff keeps its gain over the fixed setting on both splits (Figure 4); on ML-1M and Pinterest the selected and fixed settings perform alike.

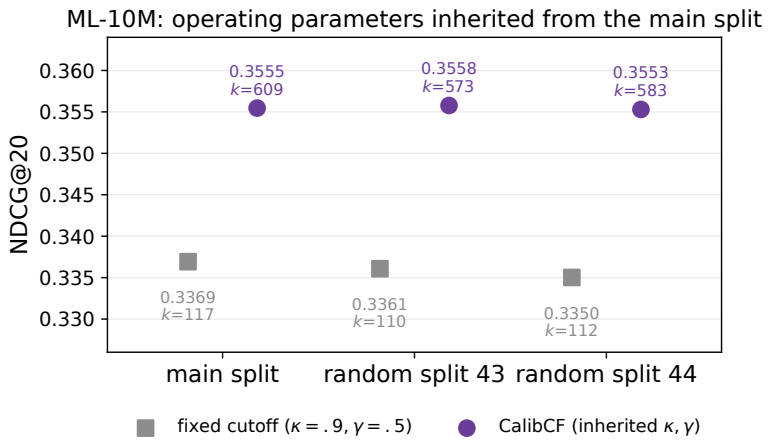

Figure 4: ML-10M across three random splits with the operating point inherited from the main split: the deeper cutoff selected there ($\kappa = 0.7$) keeps its benefit over the fixed $\kappa = .9$ on every split; the edge and factors are refitted on each split.

Table 9: Operating points selected on validation, retained dimension and factor size.

Dataset	κ	γ	k	Factor size (MiB)
ML-1M	0.9	0.5	114	4.0
Gowalla	0.85	0.5	1,869	505.1
Yelp2018	0.9	0.25	1,005	267.3
ML-10M	0.7	0.5	609	181.1
Pinterest	0.9	0.25	773	191.1

Table 10: NDCG@20 on two additional random splits.

Dataset	Split	Fixed $(\kappa, \gamma) = (.9, .5)$	Selected on the main split
ML-1M	2	0.3398	0.3398
ML-1M	3	0.3414	0.3414
ML-10M	2	0.3361	0.3558
ML-10M	3	0.3350	0.3553
Pinterest	2	0.1227	0.1227
Pinterest	3	0.1231	0.1235

E.3 Baseline hyperparameters

Table 11 lists the hyperparameters selected on validation for EASE, MF and LightGCN. On ML-1M and Pinterest, where LightGCN is trained with three seeds, its NDCG@20 varies by at most 0.0012 across seeds.

Table 11: Hyperparameters selected on validation for the closed-form and learned baselines.

Dataset	EASE λ	MF epochs	LightGCN epochs
ML-1M	500	80	380 / 400 / 360
Gowalla	100	1,020	1,240
Yelp2018	500	140	860
ML-10M	500	80	580
Pinterest	500	40	380 / 360 / 380

E.4 COMPUTATIONAL COST

Table 12 reports the median over three runs of the time needed to build a complete CalibCF model from the interaction matrix (normalization, truncated SVD, calibration and weights) and to rank the full catalog for all users; data loading and metric computation are excluded. These runs compute the head with a Lanczos solver on ML-1M and with randomized SVD on the other datasets, so their retained dimension and NDCG@20 can differ slightly from Table 2. The calibration takes 0.39–0.63 seconds on every dataset, and the SVD accounts for most of the total time.

Table 12: Computational cost of CalibCF (median of three runs; times in seconds).

Dataset	k	SVD	Calibration	Ranking	Total	Size (MiB)	NDCG@20
ML-1M	114	9.25	0.53	0.01	9.80	4	0.34712
Gowalla	1,793	25.11	0.48	0.17	26.11	485	0.16231
Yelp2018	1,033	19.85	0.49	0.14	20.68	275	0.06202
ML-10M	609	21.96	0.63	0.12	22.85	181	0.35547
Pinterest	773	14.48	0.39	0.06	15.10	191	0.12295

E.5 PAIRED STATISTICS

Tables 13 and 14 report per-user paired differences with 95% bootstrap intervals and Holm-corrected sign-flip p-values; the external-baseline family contains all 66 comparisons. All 66 differences between CalibCF and the external baselines are positive, and 65 remain significant after Holm correction. In the transfer contrasts, the transferred multiplier outperforms the transferred absolute rank on four of the five datasets, and the fixed multiplier outperforms a fixed rank of 512 on four of the five datasets; the last contrast measures the cost of transferring the operating point instead of selecting it on the target graph.

Table 13: CalibCF minus each external baseline (Holm correction over the full 66-comparison family).

		NDCG@20			Recall@20		
Dataset	comparison	difference	95% CI	p_{Holm}	difference	95% CI	p_{Holm}
ML-1M	CalibCF − SVD	+0.00976	[+0.00827, +0.01124]	0.0132	+0.00600	[+0.00406, +0.00784]	0.0132
ML-1M	CalibCF − GF-CF	+0.03008	[+0.02795, +0.03229]	0.0132	+0.02640	[+0.02358, +0.02915]	0.0132
ML-1M	CalibCF − Turbo-CF	+0.03284	[+0.03008, +0.03553]	0.0132	+0.02632	[+0.02318, +0.02952]	0.0132
ML-1M	CalibCF − ChebyCF	+0.01140	[+0.00946, +0.01332]	0.0132	+0.00873	[+0.00617, +0.01111]	0.0132
ML-1M	CalibCF − EASE	+0.01221	[+0.01011, +0.01440]	0.0132	+0.01550	[+0.01290, +0.01814]	0.0132
ML-1M	CalibCF − MF	+0.07430	[+0.07090, +0.07772]	0.0132	+0.06515	[+0.06124, +0.06906]	0.0132
ML-1M	CalibCF − LightGCN	+0.03347	[+0.03095, +0.03606]	0.0132	+0.02644	[+0.02350, +0.02946]	0.0132
Gowalla	CalibCF − SVD	+0.01080	[+0.01003, +0.01153]	0.0132	+0.01196	[+0.01079, +0.01306]	0.0132
Gowalla	CalibCF − GF-CF	+0.01054	[+0.00951, +0.01161]	0.0132	+0.00937	[+0.00790, +0.01084]	0.0132
Gowalla	CalibCF − ChebyCF	+0.00398	[+0.00312, +0.00484]	0.0132	+0.00422	[+0.00290, +0.00557]	0.0132
Gowalla	CalibCF − EASE	+0.01574	[+0.01459, +0.01687]	0.0132	+0.01790	[+0.01628, +0.01955]	0.0132
Gowalla	CalibCF − MF	+0.05648	[+0.05487, +0.05810]	0.0132	+0.06000	[+0.05797, +0.06199]	0.0132
Gowalla	CalibCF − LightGCN	+0.00709	[+0.00572, +0.00853]	0.0132	+0.01264	[+0.01083, +0.01443]	0.0132
Yelp2018	CalibCF − SVD	+0.00236	[+0.00191, +0.00282]	0.0132	+0.00300	[+0.00232, +0.00367]	0.0132
Yelp2018	CalibCF − GF-CF	+0.00490	[+0.00428, +0.00550]	0.0132	+0.00516	[+0.00435, +0.00598]	0.0132
Yelp2018	CalibCF − ChebyCF	+0.00263	[+0.00208, +0.00318]	0.0132	+0.00341	[+0.00264, +0.00415]	0.0132
Yelp2018	CalibCF − EASE	+0.00732	[+0.00653, +0.00810]	0.0132	+0.01002	[+0.00902, +0.01100]	0.0132
Yelp2018	CalibCF − MF	+0.02911	[+0.02813, +0.03009]	0.0132	+0.03516	[+0.03399, +0.03630]	0.0132
Yelp2018	CalibCF − LightGCN	+0.01081	[+0.01002, +0.01162]	0.0132	+0.01248	[+0.01142, +0.01348]	0.0132
ML-10M	CalibCF − SVD	+0.01933	[+0.01857, +0.02006]	0.0132	+0.01678	[+0.01581, +0.01776]	0.0132
ML-10M	CalibCF − GF-CF	+0.03579	[+0.03493, +0.03669]	0.0132	+0.03018	[+0.02904, +0.03126]	0.0132
ML-10M	CalibCF − Turbo-CF	+0.06291	[+0.06179, +0.06402]	0.0132	+0.05266	[+0.05128, +0.05396]	0.0132
ML-10M	CalibCF − ChebyCF	+0.00812	[+0.00753, +0.00872]	0.0132	+0.00324	[+0.00241, +0.00409]	0.0132
ML-10M	CalibCF − EASE	+0.00005	[-0.00053, +0.00063]	0.8702	+0.00099	[+0.00012, +0.00182]	0.0488
ML-10M	CalibCF − MF	+0.08193	[+0.08075, +0.08306]	0.0132	+0.06628	[+0.06491, +0.06762]	0.0132
ML-10M	CalibCF − LightGCN	+0.03062	[+0.02964, +0.03160]	0.0132	+0.02175	[+0.02052, +0.02299]	0.0132
Pinterest	CalibCF − SVD	+0.00281	[+0.00228, +0.00335]	0.0132	+0.00328	[+0.00244, +0.00413]	0.0132
Pinterest	CalibCF − GF-CF	+0.00787	[+0.00723, +0.00853]	0.0132	+0.00958	[+0.00860, +0.01057]	0.0132
Pinterest	CalibCF − Turbo-CF	+0.00639	[+0.00578, +0.00701]	0.0132	+0.00810	[+0.00717, +0.00906]	0.0132
Pinterest	CalibCF − ChebyCF	+0.00265	[+0.00213, +0.00316]	0.0132	+0.00325	[+0.00243, +0.00408]	0.0132
Pinterest	CalibCF − EASE	+0.00795	[+0.00724, +0.00866]	0.0132	+0.01108	[+0.01001, +0.01213]	0.0132
Pinterest	CalibCF − MF	+0.04388	[+0.04279, +0.04496]	0.0132	+0.05795	[+0.05655, +0.05932]	0.0132
Pinterest	CalibCF − LightGCN	+0.01615	[+0.01533, +0.01699]	0.0132	+0.01911	[+0.01802, +0.02022]	0.0132

Table 14: Operating-point transfer contrasts (Holm-corrected).

		NDCG@20			Recall@20		
Dataset	comparison	difference	95% CI	p_{Holm}	difference	95% CI	p_{Holm}
ML-1M	CalibCF (LODO) − Rank (LODO)	+0.01181	[+0.00982, +0.01383]	0.0080	+0.00481	[+0.00244, +0.00719]	0.0080
ML-1M	CalibCF ($\kappa = .9$) − Rank ($k = 512$)	+0.02166	[+0.01915, +0.02417]	0.0080	+0.01729	[+0.01424, +0.02034]	0.0080
ML-1M	CalibCF (LODO) − CalibCF (selected on target)	+0.00164	[+0.00051, +0.00274]	0.0270	+0.00141	[-0.00029, +0.00307]	0.1160
Gowalla	CalibCF (LODO) − Rank (LODO)	+0.01244	[+0.01104, +0.01377]	0.0080	+0.01983	[+0.01805, +0.02162]	0.0080
Gowalla	CalibCF ($\kappa = .9$) − Rank ($k = 512$)	+0.00637	[+0.00557, +0.00715]	0.0080	+0.00923	[+0.00798, +0.01051]	0.0080
Gowalla	CalibCF (LODO) − CalibCF (selected on target)	-0.01340	[-0.01428, -0.01254]	0.0080	-0.01009	[-0.01131, -0.00889]	0.0080
Yelp2018	CalibCF (LODO) − Rank (LODO)	+0.00431	[+0.00369, +0.00492]	0.0080	+0.00409	[+0.00325, +0.00494]	0.0080
Yelp2018	CalibCF ($\kappa = .9$) − Rank ($k = 512$)	+0.00146	[+0.00114, +0.00181]	0.0080	+0.00144	[+0.00092, +0.00199]	0.0080
Yelp2018	CalibCF (LODO) − CalibCF (selected on target)	-0.00168	[-0.00214, -0.00123]	0.0080	-0.00270	[-0.00335, -0.00204]	0.0080
ML-10M	CalibCF (LODO) − Rank (LODO)	-0.00895	[-0.00952, -0.00839]	0.0080	-0.00833	[-0.00915, -0.00751]	0.0080
ML-10M	CalibCF ($\kappa = .9$) − Rank ($k = 512$)	-0.01846	[-0.01917, -0.01776]	0.0080	-0.01523	[-0.01619, -0.01426]	0.0080
ML-10M	CalibCF (LODO) − CalibCF (selected on target)	-0.02038	[-0.02121, -0.01952]	0.0080	-0.01293	[-0.01397, -0.01187]	0.0080
Pinterest	CalibCF (LODO) − Rank (LODO)	+0.00335	[+0.00271, +0.00400]	0.0080	+0.00284	[+0.00184, +0.00383]	0.0080
Pinterest	CalibCF ($\kappa = .9$) − Rank ($k = 512$)	+0.00120	[+0.00088, +0.00151]	0.0080	+0.00142	[+0.00083, +0.00199]	0.0080
Pinterest	CalibCF (LODO) − CalibCF (selected on target)	-0.00149	[-0.00195, -0.00106]	0.0080	-0.00244	[-0.00313, -0.00168]	0.0080

F MECHANISM CONTROLS AT THE FIXED OPERATING POINT

The controls in this section fix the operating point at $(\kappa, \gamma) = (.9, .5)$, denoted CalibCF-T, and change one component at a time. Figure 5 varies the head width and compares the closed-form response with the D-transform response computed from the exact spectrum: the D-transform needs its uncomputed tail when the head is small, whereas the closed form does not. Table 15 keeps the support fixed and replaces the closed-form response by other weightings; the closed form is above every simple alternative on all five datasets and matches the tail-completed D-transform to within 0.0002 on four of them. Table 16 separates the two operating parameters, and Tables 17 and 18 replace the calibrated edge by a published edge estimator.

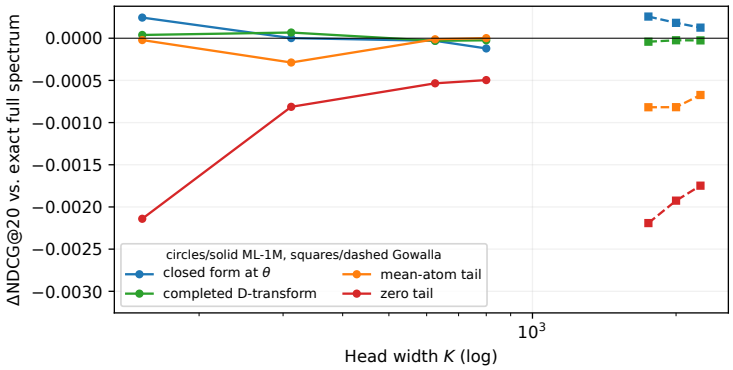

Figure 5: NDCG@20 minus that of the exact full-spectrum D-transform response against the head width on ML-1M and Gowalla: the D-transform needs a completed tail at small K, whereas the closed form at the calibrated cutoff needs no tail at all.

Edge-estimator control. Tables 17 and 18 replace the calibrated edge by the published leading-eigenvalue estimator of eOptShrink (Su & Wu, 2025, Eq. 48), which for the $p \times q$ orientation and $\lambda_j = \sigma_j^2$ reads $b_{\text{P}}^2 = \lambda_{t+1} + (\lambda_{t+1} - \lambda_{2t+1})/(2^{2/3} - 1)$ with $t = \lfloor q^c \rfloor$ and $c = \min\{1/2.01, 1/\log\log q\}$ (Spacing-P). The cutoff is placed at b_{P} or $0.9b_{\text{P}}$, with the same singular triples, normalization, response and readout as CalibCF-T. The spacing estimator retains far fewer directions and lowers both metrics on every dataset.

G ADDITIONAL ANALYSES: EXACT SPECTRA, HEAD WIDTH, SYNTHETIC GRAPHS AND SENSITIVITY

Unless stated otherwise, these analyses use the fixed operating point $(\kappa, \gamma) = (.9, .5)$, to which all ranking numbers in this section refer. For ML-10M and Pinterest, the comparisons with exact spectra

Table 15: Response controls at fixed vectors, support and readout.

Response (same support)	ML-1M	Gowalla	Yelp2018	ML-10M	Pinterest
D-transform, completed tail	0.3472	0.1620	0.0602	0.3405	0.1227
Closed form at $\kappa\hat{b}$ (CalibCF-T)	0.3471	0.1622	0.0602	0.3369	0.1227
Closed form at $\hat{b}$	0.3364	0.1558	0.0567	0.3241	0.1165
$(\sigma^2 - \theta^2)/\sigma$	0.3388	0.1597	0.0576	0.3271	0.1190
$(\sigma - \theta)_+$	0.3286	0.1574	0.0556	0.3167	0.1158
σ (hard)	0.3324	0.1538	0.0585	0.3358	0.1147
σ^2	0.3007	0.1621	0.0574	0.3004	0.1177
Support k / excluded count r	114 / 53	1204 / 445	1005 / 258	117 / 63	773 / 177

Table 16: Separating spectral retention from readout, NDCG@20. Starting from $(\kappa, \gamma) = (.9, .5)$, adjust only the cutoff, only the readout, or both (CalibCF). The final column gives retained dimensions.

Dataset	Fixed	Cutoff only	Readout only	CalibCF	k fixed / final
ML-1M	0.3471	0.3471	0.3471	0.3471	114 / 114
Gowalla	0.1622	0.1624	0.1622	0.1624	1204 / 1869
Yelp2018	0.0602	0.0603	0.0620	0.0620	1005 / 1005
ML-10M	0.3369	0.3555	0.3369	0.3555	117 / 609
Pinterest	0.1227	0.1227	0.1230	0.1230	773 / 773

Table 17: Edge-estimator control, NDCG@20 (retained dimensions).

Cutoff rule	ML-1M	Gowalla	Yelp2018	ML-10M	Pinterest
CalibCF-T	0.3471 (114)	0.1622 (1204)	0.0602 (1005)	0.3369 (117)	0.1227 (773)
Spacing-P, $\kappa = 1$	0.3172 (31)	0.1064 (46)	0.0417 (44)	0.3170 (54)	0.0938 (58)
Spacing-P, $\kappa = .9$	0.3308 (47)	0.1159 (68)	0.0451 (64)	0.3297 (85)	0.1001 (73)

Table 18: Edge-estimator control, Recall@20.

Cutoff rule	ML-1M	Gowalla	Yelp2018	ML-10M	Pinterest
CalibCF-T	0.3092	0.1939	0.0726	0.3532	0.1740
Spacing-P, $\kappa = 1$	0.2786	0.1184	0.0507	0.3342	0.1353
Spacing-P, $\kappa = .9$	0.2920	0.1314	0.0552	0.3463	0.1444

use the exact leading singular values in place of the randomized estimates (maximum differences 0.0024 and 0.0069) and keep the computed singular vectors for scoring.

G.1 Numerical convergence of the law

The bulk-quantile scale fit uses a smoothed density. At the default $\eta_{\mathrm{rel}} = 1.25 \times 10^{-4}$, the 400-iteration relative fixed-point residual is 5.3×10^{-9} (ML-1M), 5.9×10^{-6} (Gowalla), and 2.2×10^{-15} (ML-10M). Increasing to 8000 iterations changes no retained count; the largest NDCG change is 1.6×10^{-7}. Across the whole tested smoothing grid the largest 400-iteration residual is 5.8×10^{-5}.

G.2 Head width

Tables 19–23 vary the head width K at the fixed operating point. Each row gives the retained dimension k and NDCG@20 for the closed-form response of CalibCF (Closed); for the D-transform response whose uncomputed tail is completed by the degree law (DT-law), by the degree law matched to the exact first moment (DT-moment), by zeros (DT-zero) or by the mean (DT-mean); for hard truncation (Hard); and, where the full spectrum was computed, for the D-transform computed from

the exact spectrum (DT-exact). The closed form changes by at most 0.0004 NDCG@20 across head widths on every dataset. The note below each table gives the mean overlap of the top-20 lists with those of DT-exact.

Table 19: Head width on ML-1M, NDCG@20.

K	k	Closed	DT-law	DT-moment	DT-zero	DT-mean	Hard	DT-exact
152	112	0.34714	0.34694	0.34699	0.34476	0.34688	0.33273	0.34690
312	118	0.34746	0.34752	0.34747	0.34664	0.34717	0.33096	0.34745
625	115	0.34724	0.34724	0.34726	0.34673	0.34725	0.33223	0.34727
800	114	0.34712	0.34722	0.34722	0.34675	0.34724	0.33241	0.34724

Mean top-20 overlap with DT-exact (Closed / DT-law): $K = 152$: 98.0% / 99.8%; $K = 312$: 98.1% / 99.8%; $K = 625$: 98.0% / 99.9%; $K = 800$: 98.0% / 100.0%.

Table 20: Head width on Gowalla, NDCG@20.

K	k	Closed	DT-law	DT-moment	DT-zero	DT-mean	Hard	DT-exact
1750	1153	0.16206	0.16176	0.16182	0.15961	0.16098	0.15414	0.16180
2000	1180	0.16214	0.16194	0.16196	0.16004	0.16114	0.15393	0.16196
2250	1204	0.16219	0.16204	0.16207	0.16032	0.16139	0.15377	0.16207

Mean top-20 overlap with DT-exact (Closed / DT-law): $K = 1750$: 98.2% / 99.9%; $K = 2000$: 98.1% / 99.9%; $K = 2250$: 98.0% / 99.9%.

Table 21: Head width on Yelp2018, NDCG@20 (exact spectrum not computed).

K	k	Closed	DT-law	DT-moment	DT-zero	DT-mean	Hard
2000	1005	0.06022	0.06018	0.06015	0.06070	0.06047	0.05850

Table 22: Head width on ML-10M, NDCG@20.

K	k	Closed	DT-law	DT-moment	DT-zero	DT-mean	Hard	DT-exact
500	118	0.33704	0.34066	0.34064	0.34004	0.34061	0.33584	0.34065
1000	118	0.33713	0.34065	0.34064	0.34038	0.34059	0.33584	0.34065
2000	117	0.33685	0.34047	0.34048	0.34033	0.34047	0.33578	0.34047

Mean top-20 overlap with DT-exact (Closed / DT-law): $K = 500$: 90.1% / 99.9%; $K = 1000$: 90.2% / 99.9%; $K = 2000$: 90.1% / 100.0%.

Table 23: Head width on Pinterest, NDCG@20.

K	k	Closed	DT-law	DT-moment	DT-zero	DT-mean	Hard	DT-exact
1000	739	0.12271	0.12237	0.12233	0.12299	0.12275	0.11506	0.12231
2000	750	0.12270	0.12237	0.12235	0.12286	0.12254	0.11497	0.12235

Mean top-20 overlap with DT-exact (Closed / DT-law): $K = 1000$: 95.8% / 99.5%; $K = 2000$: 95.8% / 99.7%.

G.3 Prediction of the uncomputed spectrum

Tables 24–27 compare five completions of the uncomputed part of the spectrum with the exact spectrum: the degree law (law), the degree law matched to the exact first moment (moment), an MP law with the same scale fit (MP), a point mass at the mean (mean) and zeros (zero). KS and W_1/t measure the distance between the completed and exact eigenvalue distributions, with $t = \sigma_K^2$; the last two columns give the relative error of the mean eigenvalue and of the resulting D-transform weights. The two degree-law completions are closest to the exact spectrum on every dataset and head width, and they give the most accurate weights.

Table 24: Prediction of the uncomputed spectrum on ML-1M.

Dataset	K	Tail	KS	W_1/t	mean rel. err	response rel. err
ML-1M	152	law	0.038	0.85%	1.98%	0.32%
		moment	0.038	0.48%	0.00%	0.18%
		MP	0.198	9.95%	36.14%	5.10%
		mean	0.606	21.16%	0.00%	8.55%
		zero	1.000	27.54%	100.00%	21.25%
ML-1M	312	law	0.038	1.10%	2.48%	0.35%
		moment	0.039	0.53%	0.00%	0.06%
		MP	0.191	9.44%	31.37%	2.71%
		mean	0.598	22.57%	0.00%	3.14%
		zero	1.000	30.08%	100.00%	10.93%
ML-1M	625	law	0.043	0.96%	0.97%	0.09%
		moment	0.043	0.76%	0.00%	0.04%
		MP	0.195	9.87%	30.18%	1.42%
		mean	0.587	23.61%	0.00%	0.93%
		zero	1.000	32.71%	100.00%	5.24%
ML-1M	800	law	0.046	0.88%	0.09%	0.03%
		moment	0.046	0.89%	0.00%	0.03%
		MP	0.201	10.28%	30.44%	1.07%
		mean	0.583	23.89%	0.00%	0.53%
		zero	1.000	33.79%	100.00%	3.77%

Table 25: Prediction of the uncomputed spectrum on Gowalla.

Dataset	K	Tail	KS	W_1/t	mean rel. err	response rel. err
Gowalla	1750	law	0.016	0.63%	1.75%	0.23%
		moment	0.016	0.37%	0.00%	0.45%
		MP	0.061	3.85%	13.19%	3.47%
		mean	0.598	21.72%	0.00%	12.13%
		zero	1.000	29.16%	100.00%	31.81%
Gowalla	2000	law	0.017	0.72%	2.20%	0.35%
		moment	0.016	0.34%	0.00%	0.31%
		MP	0.058	3.63%	12.26%	2.87%
		mean	0.597	21.99%	0.00%	10.15%
		zero	1.000	29.64%	100.00%	27.93%
Gowalla	2250	law	0.018	0.82%	2.55%	0.45%
		moment	0.016	0.31%	0.00%	0.22%
		MP	0.056	3.47%	11.55%	2.44%
		mean	0.595	22.21%	0.00%	8.69%
		zero	1.000	30.06%	100.00%	25.01%

Table 26: Prediction of the uncomputed spectrum on ML-10M.

Dataset	K	Tail	KS	W_1/t	mean rel. err	response rel. err
ML-10M	500	law	0.049	1.06%	1.77%	0.20%
		moment	0.050	0.78%	0.00%	0.04%
		MP	0.601	32.52%	119.33%	7.79%
		mean	0.605	22.11%	0.00%	2.33%
		zero	1.000	27.26%	100.00%	8.40%
ML-10M	1000	law	0.052	1.33%	1.81%	0.17%
		moment	0.053	1.02%	0.00%	0.07%
		MP	0.591	31.95%	107.48%	5.00%
		mean	0.602	23.75%	0.00%	1.18%
		zero	1.000	29.72%	100.00%	5.48%
ML-10M	2000	law	0.059	1.21%	1.37%	0.02%
		moment	0.058	1.33%	0.00%	0.03%
		MP	0.613	34.08%	109.43%	2.88%
		mean	0.601	24.24%	0.00%	0.39%
		zero	1.000	31.15%	100.00%	2.79%

Table 27: Prediction of the uncomputed spectrum on Pinterest.

Dataset	K	Tail	KS	W_1/t	mean rel. err	response rel. err
Pinterest	1000	law	0.039	2.05%	2.14%	0.80%
		moment	0.032	1.18%	0.00%	0.11%
		MP	0.128	5.15%	10.06%	1.80%
		mean	0.513	20.61%	0.00%	10.31%
		zero	1.000	51.22%	100.00%	30.03%
Pinterest	2000	law	0.040	2.29%	1.82%	0.42%
		moment	0.036	1.55%	0.00%	0.15%
		MP	0.140	5.36%	9.81%	0.93%
		mean	0.503	20.94%	0.00%	2.77%
		zero	1.000	54.62%	100.00%	12.49%

G.4 Synthetic degree-corrected Bernoulli graphs

Edge errors are relative to the largest singular value of the true normalized noise N. The mean error of the degree template is positive in every cell and shrinks with the mean degree at every heterogeneity level, as Proposition 2 predicts. In every run the detection threshold $1.02\hat{b}$ of the degree template exceeds $\sigma_1(N)$, so by Weyl's inequality the directions it flags never outnumber the rank of the normalized mean matrix $F_u P F_i$, which clipping of heavy-tailed propensities raises beyond the planted rank. Under moderate and strong heterogeneity the MP error is negative, placing its edge inside the noise; under weak heterogeneity the profile is nearly homogeneous and both templates are accurate. Table 28 reports means and standard deviations over three random graphs per setting; "spurious" counts the directions flagged beyond the planted signal.

Table 28: Synthetic degree-corrected Bernoulli graphs.

		Edge error		Bulk KS		Spurious directions	
Het. σ	Mean degree	Degree	MP	Degree	MP	Degree	MP
0.4	15	+2.6%±0.3	+1.3%±0.3	0.021	0.009	0.0	0.0
0.4	50	+1.2%±0.1	+0.6%±0.1	0.009	0.004	0.0	0.0
0.4	150	+0.7%±0.2	+0.5%±0.2	0.006	0.005	0.0	0.0
0.8	15	+3.4%±0.3	-1.0%±0.2	0.032	0.029	0.0	0.0
0.8	50	+1.6%±0.0	-0.9%±0.0	0.011	0.023	0.0	0.0
0.8	150	+0.2%±0.3	-1.2%±0.3	0.008	0.013	1.0	1.0
1.2	15	+6.1%±0.3	-6.3%±0.3	0.060	0.109	0.0	17.7
1.2	50	+2.3%±0.1	-5.2%±0.4	0.019	0.070	1.0	13.3
1.2	150	+1.6%±0.4	-2.7%±0.6	0.012	0.049	2.0	4.7

G.5 Sensitivity

Tables 29–31 vary one constant at a time around the defaults $\kappa = 0.9$, $c_\tau = 0.25$, the cap 0.99 on $\hat{P}$ and $\eta_{\mathrm{rel}} = 1.25 \times 10^{-4}$, and report NDCG@20 with the retained dimension in parentheses. Across the smoothing grid and the variants of the variance proxy, NDCG@20 changes by at most 0.0004.

Table 29: Sensitivity to the edge multiplier κ, NDCG@20 (retained k).

Dataset	0.85	0.88	0.9	0.92	0.95
ML-1M	0.3488 (177)	0.3480 (137)	0.3471 (114)	0.3461 (96)	0.3424 (77)
Gowalla	0.1624 (1,869)	0.1629 (1,445)	0.1622 (1,204)	0.1613 (1,000)	0.1597 (768)
Yelp2018	0.0603 (1,769)	0.0605 (1,284)	0.0602 (1,005)	0.0599 (775)	0.0588 (519)
ML-10M	0.3434 (170)	0.3395 (134)	0.3369 (117)	0.3342 (103)	0.3306 (88)
Pinterest	0.1215 (1,304)	0.1224 (975)	0.1227 (773)	0.1221 (598)	0.1206 (392)

Table 30: Sensitivity to the regularization constant c_τ (top) and to the variance-proxy correction $(1 - \hat{P})$ (bottom), NDCG@20 (retained k).

Dataset	$c_\tau = 0.1$	0.25	0.5
ML-1M	0.3386 (87)	0.3471 (114)	0.3490 (119)
Gowalla	0.1608 (1,172)	0.1622 (1,204)	0.1616 (1,090)

Dataset	no correction	cap 0.90	cap 0.95	cap 0.99	no cap
ML-1M	0.3467 (106)	0.3471 (114)	0.3471 (114)	0.3471 (114)	0.3471 (114)
Gowalla	0.1622 (1,204)	0.1622 (1,204)	0.1622 (1,204)	0.1622 (1,204)	0.1622 (1,204)

Table 31: Sensitivity to the law smoothing η_{rel}, NDCG@20 (retained k).

Dataset	5×10^{-4}	2.5×10^{-4}	1.25×10^{-4}	6.25×10^{-5}
ML-1M	0.3473 (116)	0.3472 (114)	0.3471 (114)	0.3471 (114)
Gowalla	0.1622 (1,217)	0.1622 (1,210)	0.1622 (1,204)	0.1622 (1,203)
Yelp2018	0.0602 (1,010)	0.0602 (1,005)	0.0602 (1,005)	0.0602 (1,005)
ML-10M	0.3373 (119)	0.3370 (118)	0.3369 (117)	0.3369 (117)
Pinterest	0.1227 (777)	0.1227 (774)	0.1227 (773)	0.1227 (772)