SCoGF: Spectral Consensus Graph Fusion for Scalable Incomplete Multi-View Clustering

Anonymous authors
Paper under double-blind review

Abstract

Spectral multi-view clustering under view-level missingness faces an unresolved tension: scalable methods rely on kernel evaluations that are undefined when a sample lacks a view, while incomplete-view methods either impute missing views and propagate recovery error, or couple all N samples in every iteration at cubic cost. We propose SCoGF (Spectral Consensus Graph Fusion), a single-pass method in which the missing masks shape the graph construction directly rather than entering through any later correction. SCoGF learns per-view anchor graphs from simplex-constrained reconstruction weights on observed anchors only, fuses them by a uniform masked average that is the exact minimizer of the masked aggregation subproblem, and embeds all samples through a randomized SVD with a missing-aware cross-anchor extension. Deterministic k-means anchor selection makes the output reproducible given the masks. The pipeline is imputation-free, non-iterative, free of kernel bandwidths and learned view weights, and near-linear in N. Theoretically, we prove that the masked average is statistically sound with variance decaying in the view count, the anchor embedding is stable under explicit budgets on the anchor and view counts, and the resulting partition admits an end-to-end normalized-cut consistency bound. Empirically, across multi-view benchmarks under both symmetric and view-dependent missingness, SCoGF is the only compared method that never collapses on feature-heterogeneous data, while running orders of magnitude faster than iterative competitors and requiring no per-rate tuning.

1 Introduction

Multi-view clustering partitions data by exploiting the consensus and complementarity of several feature representations of the same samples (Xu et al., 2013; Li et al., 2022). Spectral approaches are particularly attractive because they recover non-convex cluster geometry through the spectrum of a graph Laplacian (Ng et al., 2002; von Luxburg, 2007). Two practical obstacles, however, continue to limit their deployment. The first is *incompleteness*: owing to sensor failures, privacy constraints, or collection costs, not every view is observed for every sample (Li et al., 2014; Liu et al., 2019). We focus on view-level missingness, where the entire feature vector of a view is absent for a subset of samples. The second is *scalability*: spectral clustering requires the eigendecomposition of an $N \times N$ graph Laplacian, costing $O(N^3)$ time and $O(N^2)$ memory (Fowlkes et al., 2004; Williams & Seeger, 2001). Both issues are increasingly common in the same applications, since views produced by pretrained encoders are cheap to obtain, but availability is dictated by privacy, sensing cost, and modality drop-out.

Each obstacle has been studied extensively in isolation, yet the standard remedies are mutually incompatible. Scalable spectral clustering is built on anchor (landmark) graphs and Nyström-type extensions, whose elementary operation, a kernel evaluation between a sample and a landmark, is undefined whenever the sample lacks the corresponding view (Fowlkes et al., 2004; Cai & Chen, 2015). Incomplete multi-view methods, conversely, either *impute* the missing views before clustering (Lin et al., 2021; 2023; Li et al., 2023), which couples recovery quality to a training procedure and propagates imputation error into fusion as the missing rate grows, or learn over all N samples in every iteration without imputation (Li et al., 2014; Shao et al., 2015; Hu & Chen, 2018; Wen et al., 2020), which forfeits scalability. A third, often overlooked failure mode is *dataset sensitivity*:

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.

methods that excel on one benchmark family can collapse to random-guess accuracy on feature-heterogeneous benchmarks whose views differ sharply in dimensionality and scale (Section 6.2).

These failures trace to three recurring mechanisms. Alignment-based methods estimate cross-view correspondences from samples observed in several views at once, but the fraction of such samples decays as $(1-\eta)^V$ and vanishes quickly as the view count or missing rate grows, leaving the alignment underdetermined. Alternating solvers propagate mask-induced errors across iterations and can settle on degenerate solutions, and their per-iteration cost is what couples all N samples in the first place. And zero- or model-filled entries distort the scale and neighborhood structure of heterogeneous descriptors, so the graph built on imputed values no longer reflects the data. We refer to these mechanisms as *joint visibility*, *iterative fusion*, and *imputed geometry*.

We propose SCoGF (Spectral Consensus Graph Fusion), a single algorithm designed to avoid all three mechanisms. The key idea is to let the missing masks shape the graph construction directly, rather than correcting for missingness after the fact. Each view builds its anchor graph only on the anchors it observes, so an incomplete view contributes proportionally less mass to the consensus instead of corrupting it. Pairwise co-observation replaces joint visibility, since an anchor pair needs only one view that sees both endpoints, an event whose probability grows exponentially in the view count (Theorem 5.1). The per-view graphs are fused by a uniform masked average, the closed-form minimizer of the aggregation objective, so fusion is a single closed-form pass with no iterative solver. Because nothing is ever filled in, the graph geometry is computed exclusively from observed entries. The consensus graph is turned into a partition by a randomized truncated SVD with a missing-aware cross-anchor extension, and anchors are selected deterministically by k-means over the most-observed samples, so the anchor set is reproducible given the masks.

This design optimizes the worst case rather than the peak. In deployment, the missing rate and the view family are rarely known in advance, and clustering offers no labels for per-dataset validation, so a method is used at whatever configuration it ships with. The relevant figure of merit is then not peak accuracy after tuning on a favorable benchmark, but the accuracy retained when the benchmark, the missing rate, and the hyperparameters are all unfavorable. Our evaluation therefore reports worst-case behavior alongside averages (Section 6.2). Our contributions are as follows:

- **A mask-native spectral pipeline with a closed-form fusion.** Each view builds its anchor graph only on the anchors it observes, and missing anchors enter as zero rows and columns rather than as imputed values. The pipeline requires only pairwise co-observation instead of jointly visible samples, fuses the per-view graphs by a uniform masked average that is the exact minimizer of the masked aggregation problem (Proposition A.1), and is imputation-free. It is near-linear in N, uses deterministic anchor selection given the masks, and exposes a single hyperparameter.
- **End-to-end theoretical guarantee.** We establish statistical soundness of the uniform masked average (unsupported-pair coverage improving exponentially in the view count and $O(1/V)$ entrywise variance under MCAR, Theorem 5.1); anchor spectral stability with explicit budgets in the anchor count, the view count, and the eigengap (Theorem 5.2); and normalized-cut consistency of the final partition, with the total error decomposing into an anchor-subproblem term and a cross-anchor extension term (Theorem 5.3).
- **Empirical study with an applicability analysis.** Across five benchmarks covering feature-homogeneous and feature-heterogeneous view families, two missingness protocols, and missing rates up to 70%, we report where each method class applies and where it breaks. Within our comparison, SCoGF attains the best average rank (2.04) and the highest worst-case cell (0.390 ACC, with next best 0.361) across the tested benchmarks, missing rates, and missingness protocols, while running 2–9$\times$ faster than the strongest scalable iterative competitors on the largest benchmark and one to two orders of magnitude faster than classical iterative methods.

2 Related Work

Incomplete multi-view clustering. Existing methods fall into two families (Wen et al., 2020; Wang et al., 2019). *Imputation-based* methods first recover the missing views and then cluster: COMPLETER (Lin et al., 2021) recovers missing views through contrastive prediction, DCP (Lin et al., 2023) refines this with dual contrastive prediction, and ProImp (Li et al., 2023) imputes via learned prototypes. These methods exploit cross-view consistency to infer missing information, but suffer

from error propagation that worsens with the missing rate. *Imputation-free* methods work directly with the available data: PVC (Li et al., 2014) learns a common subspace from partially observed views, MIC (Shao et al., 2015) uses weighted nonnegative matrix factorization, DAIMC (Hu & Chen, 2018) aligns latent factors with explicit missing masks, AWSR (Feng et al., 2024) adopts adaptive weighted self-representation, and Wen et al. (2020) perform incomplete spectral clustering with adaptive graph learning. The graph-based members operate on full $N \times N$ Laplacians whose eigendecomposition is cubic in N; the subspace- and NMF-based members still couple all N samples in every iteration. The anchor-based RISE (Wang et al., 2025b) and JVILG (Wang et al., 2025a) are scalable, but inherit the kernel-evaluation problem or learn view weights that, in our protocol, can collapse on feature-heterogeneous data (Section 6.2). A parallel line of deep methods (Xu et al., 2022; Bai et al., 2024; Dai et al., 2025; Wu et al., 2025) also avoids imputation by aggregating the available views in representation space. SCoGF shares their imputation-free philosophy but operates on per-view anchor graphs in the shallow spectral setting, characterizes its aggregation as the exact minimizer of a masked problem, and backs it with statistical and spectral-stability guarantees.

Anchor graphs and scalable spectral clustering. Anchor graphs reduce the cost of spectral clustering to near-linear in N by representing each sample through a small anchor set (Cai & Chen, 2015; Fowlkes et al., 2004). Our anchor-graph construction descends from locally linear reconstruction weights (Roweis & Saul, 2000) and adaptive-neighbor graph learning (Nie et al., 2014), transplanted to the anchor level under view-level missingness. Large-scale multi-view bipartite-graph methods (Li et al., 2015; 2022), LMVSC (Kang et al., 2020), and Procrustes alignment (Nie et al., 2018) assume fully observed views: the affinity between a sample and an anchor is undefined when the sample lacks the view, and affinities are typically computed from fixed kernels with no local adaptivity. SCoGF removes both restrictions. The closest scalable incomplete-view competitor, IMVC-CBG (Wang et al., 2022), fuses fixed anchor graphs but provides no end-to-end guarantees. Nyström and randomized SVD (Williams & Seeger, 2001; Fowlkes et al., 2004; Gittens & Mahoney, 2016; Halko et al., 2011) approximate spectral objects of large graphs in near-linear time but assume fully observed anchor graphs; SCoGF integrates a randomized eigensolver into an incomplete multi-view pipeline whose guarantees are preserved under masking.

Table 3 (Appendix B) summarizes how SCoGF sits relative to representative methods: none of those examined handles incomplete views, scales, tolerates mis-set hyperparameters, and offers guarantees at once. The fragility behind the robustness column is quantified by the tuning sweep of Section 6.1.

3 Problem Formulation

Setup and notation. Bold uppercase letters denote matrices, bold lowercase vectors, calligraphic letters sets. For a matrix $\mathbf{A}$, $\lambda_i(\mathbf{A})$ is the i-th largest eigenvalue; $\odot$ and $\oslash$ denote element-wise product and division. Throughout, N is the sample count, V the view count, m the anchor count, K the cluster count, and d_v the dimension of view v; $\eta_v \in [0, 1)$ is the missing rate of view v (and η the shared rate under symmetric missingness), and $\bar{\rho} := \frac{1}{V}\sum_v (1-\eta_v)^2$ is the aggregate observation level that recurs in the theory (under symmetric MCAR, $\bar{\rho} = (1-\eta)^2$). For a matrix $\mathbf{A}$ with rows $\mathbf{a}_i$, $\mathrm{rowNorm}(\mathbf{A})$ divides each row by its ℓ_1 norm and $\mathrm{rowL2}(\mathbf{A})$ divides each row by its ℓ_2 norm.

Multi-view data with view-level missingness. Consider $\mathbf{X}^{(v)} \in \mathbb{R}^{N \times d_v}$, $v = 1, \ldots, V$. For each sample i and view v, either the entire row $\mathbf{X}^{(v)}_{i,:}$ is observed or none of it is. The availability pattern is encoded by $\mathbf{o}^{(v)} \in \{0, 1\}^N$ with $o^{(v)}_i = 1$ if sample i is observed in view v, and every sample retains at least one observed view. The constructions below never evaluate any view on entries it did not observe; hence no imputation of missing rows is required at any stage.

Masks and anchor graphs. Let $\mathcal{A} = \{a_1, \ldots, a_m\}$, $m \ll N$, be an anchor set (Section 4), and let $\mathbf{S}^{(v)} \in \mathbb{R}^{m \times m}$ be the anchor graph of view v constructed in a mask-native manner (Section 4): by construction, $\mathbf{S}^{(v)}$ has zero rows and columns at anchors the view does not observe. The anchor-level mask of view v is $\mathbf{m}^{(v)} \in \{0, 1\}^m$ with $m^{(v)}_i = 1$ if anchor a_i is observed in view v.

Aggregation semantics. The primitive underlying SCoGF is the aggregation functional

$$\Psi(\mathbf{G}) = \sum_{v=1}^{V} \left\| \mathbf{S}^{(v)} - \mathbf{G} \right\|_F^2, \tag{1}$$

in which masking enters through the graphs themselves: since $\mathbf{S}^{(v)}$ is zero on every pair the view cannot support, an incomplete view contributes less total mass automatically, and pairs observed by no view contribute only the penalty on $\mathbf{G}$ itself. Elementwise, the minimizer of equation 1 at pair (i, j) is the average $\frac{1}{V}\sum_v S_{ij}^{(v)}$. This average form is unbiased only up to the pair-independent global scale $\bar{\rho}$ (Theorem 5.1(b)), which the degree normalization of Section 4 annihilates.

4 THE SCOGF ALGORITHM

The algorithm is a single-pass, three-stage pipeline consisting of mask-native anchor graphs (Stage 1), closed-form aggregation (Stage 2), and consensus spectral embedding with a missing-aware extension (Stage 3). Algorithm 1 summarizes the complete procedure.

Deterministic anchor selection. Random anchor draws inject avoidable variance into every downstream stage, and under heavy missingness can land on samples whose views are almost entirely unobserved. SCoGF instead selects anchors deterministically: the candidate pool consists of the samples with the most observed views, with ties broken by a fixed permutation, and mini-batch k-means with a fixed seed maps the pool to m anchors (Algorithm 1). Each sample is represented by its concatenated standardized views with unobserved entries set to zero, which, after standardization, equals column-mean imputation; this fill-in is used *only* for anchor selection and never enters the graph construction. Given the missing masks, the selected anchor set is deterministic; the randomized SVD and k-means are the only stochastic steps that remain, and both admit fixed seeds, which simplifies reporting over repeated mask draws.

Stage 1: mask-native learned anchor graphs. For each view v, let $\mathcal{O}^{(v)} = \{i \in \mathcal{A} : o_i^{(v)} = 1\}$ be the observed anchors, with standardized features collected in $\mathbf{P}^{(v)}$. Each observed anchor $i \in \mathcal{O}^{(v)}$ is represented by its r nearest observed neighbors $\mathcal{N}_r(i) \subset \mathcal{O}^{(v)} \setminus \{i\}$ through simplex-constrained least squares with a ridge term:

$$\mathbf{z}_i^{(v)} = \arg\min_{\mathbf{z}\geq\mathbf{0},\ \mathbf{1}^\top\mathbf{z}=1} \left\|\mathbf{p}_i^{(v)} - \left(\mathbf{P}^{(v)}_{\mathcal{N}_r(i)}\right)^\top \mathbf{z}\right\|_2^2 + \lambda_z\|\mathbf{z}\|_2^2, \tag{2}$$

solved in closed form subject to the equality constraint, followed by an active-set refinement that zeroes negative coordinates and renormalizes. The ridge weight is scaled by the mean diagonal of the local Gram matrix so that the penalty is invariant to the per-view feature scale. The per-view anchor graph is $\mathbf{A}^{(v)} = (\mathbf{Z}^{(v)} + (\mathbf{Z}^{(v)})^\top)/2$ on $\mathcal{O}^{(v)} \times \mathcal{O}^{(v)}$, with rows and columns of missing anchors set to zero, and $\mathbf{S}^{(v)}$ its row-normalized counterpart with self-loops $\mathrm{diag}(\mathbf{S}^{(v)}) = \mathbf{m}^{(v)}$ for numerical stability. Unlike kernel anchor graphs, equation 2 carries no bandwidth parameter and adapts the weight profile to the local geometry of each view; unlike imputation-based pipelines, it never evaluates a view on an anchor the view does not observe. Both neighborhood search and reconstruction are confined to $\mathcal{O}^{(v)}$.

Stage 2: closed-form masked aggregation. The closed-form minimizer of equation 1 (Proposition A.1 in Appendix A) is the uniform masked average $\mathbf{G}^\star = \frac{1}{V}\sum_{v=1}^{V}\mathbf{S}^{(v)}$, followed by symmetrization $\mathbf{G}^\star \leftarrow (\mathbf{G}^\star + \mathbf{G}^{\star\top})/2$ and zeroing the diagonal. Because equation 1 is elementwise separable, symmetrizing and zeroing the diagonal after minimization coincides with minimizing equation 1 over symmetric $\mathbf{G}$ with $\mathrm{diag}(\mathbf{G}) = \mathbf{0}$, so the produced graph is still an exact constrained minimizer. Every anchor pair co-observed by at least one view accumulates exactly those views; pairs observed by no view are zero. The computation is a single $O(m^2V)$ pass with no iterative solver.

Stage 3: consensus embedding and missing-aware extension. Let $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{G}^\star\mathbf{D}^{-1/2}$ be the normalized adjacency of the consensus graph. The embedding matrix $\mathbf{U}_K \in \mathbb{R}^{m\times K}$ collects the second through $(K{+}1)$-th left singular vectors of $\mathbf{W}$, computed from the top-$(K{+}1)$ by a randomized truncated SVD (Halko et al., 2011) with the leading component dropped. Since $\mathbf{W}$ is symmetric, its singular vectors coincide with its eigenvectors up to sign and are taken in decreasing algebraic order; the drop-first rule is justified in the block-structured regime of the theory by Proposition A.4. The embedding is extended to all N samples through a cross-anchor matrix $\mathbf{C} \in \mathbb{R}^{N\times m}$ built from the Stage-1 reconstruction weights: for each sample i and each view v in which it is observed, i is reconstructed from its r nearest observed anchors via the simplex weights of equation 2, the weights are accumulated uniformly across its observed views, and row i is renormalized by the number of

its available views. The full-sample embedding is

$$\mathbf{F} = \mathrm{rowL2}\big(\mathrm{rowNorm}(\mathbf{C})\,\mathbf{U}_K\big), \tag{3}$$

and the final partition is obtained by k-means on $\mathbf{F}$.

Algorithm 1 SCoGF (Spectral Consensus Graph Fusion) Algorithm

Require: Incomplete views $\{\mathbf{X}^{(v)} \in \mathbb{R}^{N\times d_v}\}_{v=1}^{V}$; cluster number K; anchor number m; neighbor size r; ridge λ_z

Ensure: Cluster labels $\mathbf{y} \in \{1,\dots,K\}^N$

Anchor selection (deterministic)

pool $\leftarrow$ top-$\min(N, \max(3m, \lfloor N/5 \rfloor))$ samples by observed-view count, ties broken by a fixed permutation

$\mathcal{A} \leftarrow$ pool samples nearest to the m centroids of fixed-seed mini-batch k-means on the pooled, standardized, zero-filled features

Stage 1: per-view mask-native anchor graphs

for $v = 1$ to V **do**

  $\mathbf{m}^{(v)} \leftarrow$ anchor observability; $\mathbf{P}^{(v)} \leftarrow$ standardized observed-anchor features

  for each observed anchor i **do**

    solve equation 2 for $\mathbf{z}_i^{(v)}$ over the r nearest observed anchors (closed-form equality solve + active-set refinement)

  end for

  $\mathbf{A}^{(v)} \leftarrow (\mathbf{Z}^{(v)} + (\mathbf{Z}^{(v)})^\top)/2$ on observed anchors; $\mathbf{S}^{(v)} \leftarrow$ row-normalized $\mathbf{A}^{(v)}$; $\mathrm{diag}(\mathbf{S}^{(v)}) \leftarrow \mathbf{m}^{(v)}$ ▷ missing anchors remain zero rows/columns in $\mathbf{S}^{(v)}$

end for

Stage 2: closed-form masked aggregation

$\mathbf{G}^\star \leftarrow \frac{1}{V}\sum_v \mathbf{S}^{(v)}$; $\mathbf{G}^\star \leftarrow (\mathbf{G}^\star + \mathbf{G}^{\star\top})/2$; $\mathrm{diag}(\mathbf{G}^\star) \leftarrow 0$ ▷ exact minimizer of equation 1; no inversion, no iteration

Stage 3: consensus spectral embedding and extension

$\mathbf{W} \leftarrow \mathbf{D}^{-1/2}\mathbf{G}^\star\mathbf{D}^{-1/2}$; $\mathbf{U}_K \leftarrow$ second through $(K{+}1)$-th left singular vectors of $\mathbf{W}$ by randomized SVD (Halko et al., 2011) ▷ drop the leading component

build $\mathbf{C} \in \mathbb{R}^{N\times m}$ from Stage-1 weights with uniform view accumulation, renormalize per sample by its available-view count

$\mathbf{F} \leftarrow \mathrm{rowL2}(\mathrm{rowNorm}(\mathbf{C})\mathbf{U}_K)$; $\mathbf{y} \leftarrow k$-means$(\mathbf{F}, K)$

return $\mathbf{y}$

In all experiments m follows the size rule $m = \min(500, N)$ and $\lambda_z = 0.01$ is fixed globally, so r is the only tuned hyperparameter.

Complexity. Stage 1 costs $O(m^2 d_v)$ for neighbor search and $O(mr^2 d_v + mr^3)$ for the closed-form ridge solves, per view. Stage 2 costs a single $O(m^2 V)$ pass. Stage 3 costs $O(m^2 K)$ for the randomized truncated SVD, $O(Nm\bar{d}V)$ for building $\mathbf{C}$ (with $\bar{d} = \max_v d_v$), $O(NmK)$ for the embedding, and $O(NK^2 I_{\mathrm{km}})$ for k-means. The total, $O\big(N(mK + m\bar{d}V + rV) + V(m^2 d_v + mr^2 d_v + mr^3) + m^2 V + m^2 K\big)$, is linear in N with no iterative solver anywhere in the pipeline. Memory is $O(m^2 + Nr)$ with sparse $\mathbf{C}$, versus $O(N^2)$ for full-graph methods. SCoGF exposes a single hyperparameter, the neighbor size r; the ridge weight is globally fixed at $\lambda_z = 0.01$, and m follows the size rule stated in Section 6.1.

5 Theoretical Analysis

We establish three results: statistical soundness of the masked average (Theorem 5.1), anchor spectral stability (Theorem 5.2), and an end-to-end clustering guarantee (Theorem 5.3). Throughout, ι denotes the failure probability in concentration statements, $\Delta_{\mathrm{sep}} > 0$ the support separation, D the cluster-diameter bound, $\bar{d} = \max_v d_v$ (as in Section 4), and B a uniform bound on graph entries. All proofs are deferred to Appendix A.

Assumption 1 (Data generation and within-cluster geometry)**.** (i) For each view v, samples are drawn i.i.d. from a mixture of K probability densities $\{p_k\}_{k=1}^{K}$ on $\mathbb{R}^{d_v}$, each p_k continuous and supported on a connected compact set $\mathcal{X}_k$. (ii) The supports are separated: $\min_{k\neq\ell} \inf_{x\in\mathcal{X}_k, y\in\mathcal{X}_\ell} \|x -$

$y\|_2 \geq \Delta_{\text{sep}} > 0$. (iii) Each $\mathcal{X}_k$ is regular: diameter $\mathsf{D}_k \leq \mathsf{D}$, and for every $\varepsilon \leq \Delta_{\text{sep}}/2$ it admits a partition into at most $(C_{\text{cov}}\mathsf{D}_k/\varepsilon)^{d_v}$ cells, each of diameter at most ε and Lebesgue measure at least $(\varepsilon/C_{\text{cov}})^{d_v}$. (iv) The cluster densities are bounded: $0 < p_{\min} \leq p_k(x) \leq p_{\max}$. (v) There is a deterministic pairwise coverage floor $\rho_0 > 0$ on the realized masks: writing $\Omega^{(v)} = \{(i,j) : o_i^{(v)} = o_j^{(v)} = 1\}$, $\min_v |\Omega^{(v)}|/N^2 \geq \rho_0$. Under symmetric MCAR the floor $\rho_0 = (1-\eta)^2/2$ holds with probability at least $1 - e^{-cN}$; under view-dependent missingness, $\rho_0 = \min_v(1-\eta_v)^2/2$.

Assumption 2 (Missingness mechanism)**.** Conditional on the data, view indicators are independent Bernoulli variables with $\mathbb{P}(o_i^{(v)} = 0) = \eta_v$, independent across views and samples.

Assumption 3 (Anchor coverage and clusterability)**.** (a) The anchor set is modeled as a uniform draw for the coverage bounds used in the proofs. This idealizes the implementation's deterministic k-means selection (Section 4), which Appendix F shows dominates uniform draws empirically, with the gap growing in the sample count and the missing rate. (b) The target $\mathbf{W}^\star$ has eigengap $\delta_K = \lambda_K(\mathbf{W}^\star) - \lambda_{K+1}(\mathbf{W}^\star) > 0$. The implemented drop-first convention (Section 4) is analyzed without any extra gap assumption by Proposition A.4. (c) The full-sample consensus graph $\widetilde{\mathbf{G}}$ has a normalized-Laplacian eigengap $\delta_{\text{Lap}} = \lambda_{K+1}(\mathbf{L}) - \lambda_K(\mathbf{L}) > 0$, with Laplacian eigenvalues taken in ascending order.

Theorem 5.1 (Statistical soundness of the uniform masked average)**.** *Under Assumption 2:*

(a) Coverage. *Any anchor pair (i,j), $i \neq j$, is co-observed by at least one view with probability $P_{\text{obs}} = 1 - \prod_{v=1}^{V}(2\eta_v - \eta_v^2)$; the fraction of pairs supported by no view decays exponentially in V, and with one complete view coverage is exactly* 1.

(b) Mean. *Assume the conditional mean is view-homogeneous, i.e., $\mathbb{E}[S_{ij}^{(v)} \mid i,j \text{ observed in } v] = s_{ij}^\star$ for all v. Then $\mathbb{E}[\mathbf{G}_{ij}^\star] = \bar{\rho}\, s_{ij}^\star$ with the pair-independent scale $\bar{\rho} = \frac{1}{V}\sum_v (1-\eta_v)^2$. Without view homogeneity, the same statement holds with $s_{ij}^\star$ replaced by the mask-weighted view average, and the residual enters the bias term of Theorem 5.2.*

(c) Variance. *If $\mathbb{E}\big[(S_{ij}^{(v)})^2 \,\big|\, i,j \text{ observed in } v\big] \leq \sigma^2 + B^2$, and the per-view entries are conditionally independent across views given the masks, then $\mathrm{Var}(\mathbf{G}_{ij}^\star) \leq (\sigma^2 + B^2)\,\bar{\rho}/V = O\big((\sigma^2+B^2)/V\big)$.*

Theorem 5.2 (Anchor spectral stability)**.** *Let $\mathbf{S}^\star \in [0,B]^{m\times m}$ be the population counterpart of the consensus graph on the anchor set, with entries $s_{ij}^\star = \mathbb{E}[S_{ij}^{(v)} \mid i,j \text{ observed in } v]$, row means $s_i := \frac{1}{m}\sum_j s_{ij}^\star$, and $s_{\min} := \min_i s_i > 0$; let $\mathbf{W}^\star = (\mathbf{D}^\star)^{-1/2}\bar{\rho}\,\mathbf{S}^\star(\mathbf{D}^\star)^{-1/2}$ with $\mathbf{D}^\star = \mathrm{diag}(\bar{\rho}\, m s_i)$. Let $\widehat{\mathbf{W}}$ be the normalized adjacency of $\mathbf{G}^\star$, and $\mathbf{U}_K, \mathbf{U}_K^\star$ the top-K eigenvector matrices of $\widehat{\mathbf{W}}$ and $\mathbf{W}^\star$. Under Assumptions 1–3, there exist constants $c_1, c_2 > 0$ such that for every $\epsilon \in (0, \delta_K/2)$, whenever*

$$V \geq \frac{c_1 B^4 K \log(4m/\iota)}{\epsilon^2 \delta_K^2\, \bar{\rho}\, s_{\min}^4}, \qquad m \geq c_2\, \rho_0^{-1}\, r \log(4m/\iota)\Big(\frac{\sqrt{K}\, B^2 \mathsf{D}}{\lambda_z\, \epsilon\, \delta_K\, \bar{\rho}\, s_{\min}^2}\Big)^{\bar{d}}, \tag{4}$$

with probability at least $1-\iota$, $\big\|\mathbf{U}_K\mathbf{U}_K^\top - \mathbf{U}_K^\star(\mathbf{U}_K^\star)^\top\big\|_F \leq \epsilon$. The implemented drop-first convention is covered by Proposition A.4.

The budgets identify the scaling variables of the problem: the view budget grows as $1/\bar{\rho}\,((1-\eta)^{-2}$ under symmetric missingness), the anchor budget as ρ_0^{-1}, and both depend only polynomially on K, B, $1/\delta_K$, and r (the $r^{1/2+1/\bar{d}}$ factor from the weight-stability layer is absorbed by the stated linear power). At practical feature dimensions the power of $\bar{d}$ makes the constants loose, so we read equation 4 as an existential scaling statement rather than a numeric sample-size prediction.

Two lemmas underpin the stability analysis, a structural event (anchor occupancy, neighborhood purity, block-diagonal structure, and a per-block spectral gap) and the Lipschitz stability of the Stage-1 weight map in its local data (Appendix A, Lemmas A.2 and A.3). The cross-anchor extension contributes an additional error decomposing into a smoothing term and an anchor-subproblem term (Lemma A.5, Appendix A). Proposition A.4 shows the drop-first rule is safe exactly where the top-K argument degenerates: the dropped direction is a solver-random, delocalized mixture of the K cluster directions, and every cluster retains a $1 - O(K^{-1}\log K)$ fraction of its signal.

Theorem 5.3 (End-to-end clustering consistency). *Let $\widehat{\mathbf{G}}$ be the extended consensus graph of Lemma A.5, $\widetilde{\mathbf{G}}$ the full-sample optimum with minimum degree $d_{\min}$, and $\widehat{\mathcal{C}}$ the partition returned by spectral clustering on $\widehat{\mathbf{G}}$ with a $(1+\varepsilon_0)$-approximate k-means. The analysis idealizes Stage 3 as spectral clustering performed directly on $\widehat{\mathbf{G}}$ (the standard anchor-graph extension idealization), so the bound concerns the partition of the extended graph rather than the realized output of the cross-anchor pipeline. Under Assumptions 1–3, with probability at least $1-\iota$,*

$$\mathrm{NCut}_K(\widehat{\mathcal{C}};\widetilde{\mathbf{G}}) \le (1+\varepsilon_0)\,\mathrm{NCut}_K(\mathcal{C}^\star;\widetilde{\mathbf{G}}) + O\!\left(\frac{K\,C_L\,\|\mathbf{E}\|_F}{d_{\min}}\right) + O\!\left(\frac{K^2\,C_L^2\,\|\mathbf{E}\|_F^2}{d_{\min}^2\,\delta_{\mathrm{Lap}}^2}\right), \tag{5}$$

where $\mathcal{C}^\star$ is the optimal K-way partition of $\widetilde{\mathbf{G}}$, $\mathbf{E}=\widehat{\mathbf{G}}-\widetilde{\mathbf{G}}$ with $\|\mathbf{E}\|_F$ decomposed by Lemma A.5, and C_L is the Lipschitz constant of the normalized Laplacian map.

6 Experiments

The empirical study covers effectiveness, sensitivity, efficiency, and applicability boundaries.

6.1 Setup

Datasets. Five multi-view benchmarks spanning two view families. The *feature-heterogeneous* family, whose views are hand-crafted descriptors with sharply different dimensionalities and scales, contains MSRC-v5 (Winn et al., 2005) (N=210, V=5, K=7), 100Leaves (Mallah et al., 2013) (N=1,600, V=3, K=100), mfeat (Dua & Graff, 2019) (N=2,000, V=6, K=10), and ALOI (Geusebroek et al., 2005) (N=10,800, V=4, K=100). The *feature-homogeneous* family, whose views are comparable, contains Caltech101-20 (Fei-Fei et al., 2007) (N=2,386, V=6, K=20).

Missing-view generation. Following the standard protocol (Liu et al., 2019; Wang et al., 2022), we remove view instances uniformly at random under two schemes: *symmetric* missingness, where all views share a common rate $\eta \in \{0.0, 0.3, 0.5, 0.7\}$; and *view-dependent* missingness, where a rate template $((0.1, 0.1, 0.1, 0.3, 0.5, 0.7)$ for V=6, with the natural sub-templates $(0.1, 0.1, 0.3, 0.5, 0.7)$ for V=5, $(0.1, 0.1, 0.5, 0.7)$ for V=4, $(0.1, 0.3, 0.7)$ for V=3) is randomly permuted across views in each trial. Whenever a sample would lose all of its views under the draw, exactly one view, chosen uniformly at random, is restored, so every sample retains at least one observed view. Missing masks and algorithmic randomness are drawn from paired seeds so that every compared method faces identical masks in each trial.

Baselines. Sixteen methods in three groups. (i) Simple imputation baselines: ZeroFill, Concat, BestSV, KernelAdd. (ii) Classical multi-view methods (complete-data methods applied after column-mean imputation): AMGL (Nie et al., 2016), MKKM with incomplete kernels (Liu et al., 2020), LMSC (Zhang et al., 2017), MCAN (Nie et al., 2017), RMSC (Xia et al., 2014). (iii) Incomplete-view and scalable methods: PVC (Li et al., 2014), DAIMC (Hu & Chen, 2018), LMVSC (Kang et al., 2020), IMVC-CBG (Wang et al., 2022), AWSR (Feng et al., 2024), and the two most recent scalable competitors RISE (Wang et al., 2025b) and JVILG (Wang et al., 2025a). Which methods run on which dataset follows an explicit three-tier complexity protocol, mirroring what each method can actually execute: (a) datasets with $N < 2{,}000$ (MSRC-v5, 100Leaves) run *all* baselines, including the cubic-cost methods RMSC and AWSR; (b) datasets with $2{,}000 \le N < 5{,}000$ (mfeat, Caltech101-20) run all methods except the cubic pair; (c) datasets with $N \ge 5{,}000$ (ALOI) run only the sub-quadratic group: PVC, DAIMC, LMVSC, IMVC-CBG, RISE, JVILG, and SCoGF.

Metrics and implementation. We report clustering accuracy (ACC) and normalized mutual information (NMI), averaged over 30 trials per configuration on every dataset; higher is better, and the top two entries per column are marked in red. SCoGF exposes a single tuned hyperparameter, the neighbor number r of the learned anchor graph; the anchor number follows the size rule $m = \min(500, N)$ and $\lambda_z = 0.01$ is fixed globally. The same size rule is applied *uniformly to every anchor-based method* (SCoGF, RISE, JVILG, LMVSC, IMVC-CBG), so no competitor is disadvantaged by a smaller anchor budget.

Hyperparameter protocol. Every tunable hyperparameter of every method is selected *once per dataset* at a fixed tuning rate ($\eta = 0.5$, symmetric): each grid configuration is run for 3 trials, the

configuration with the best mean ACC is selected, and the selected configuration is then frozen and used at all missing rates and under both missingness schemes. No method's parameters are re-tuned per missing rate, and the missing-rate dimension is never seen during selection. Following common practice in clustering evaluation, model selection uses ground-truth labels (an oracle-tuning protocol) applied identically to every compared method. Although this selection sees ground-truth labels, SCoGF's accuracy is nearly flat across its small grid (Section 6.3), so its reported results do not depend on the oracle; the oracle budget matters mainly for competitors whose performance hinges on precise tuning. The anchor number m is never tuned and follows the uniform size rule stated above. The full grids and selected configurations are reported in Appendices D and G; methods not listed there expose no tuned parameter under this protocol.

6.2 Main results: effectiveness across datasets and missing rates

Table 1 reports ACC at the two highest symmetric missing rates, the regime that separates the methods; full ACC/NMI results at all four rates and under view-dependent missingness over 30 trials are in Appendix C. Throughout, the best entry per column is in bold red and the second best in red, and methods excluded by the complexity protocol (Section 6.1) are marked "–".

Table 1: ACC at symmetric $\eta \in \{0.5, 0.7\}$ (means over 30 trials).

	MSRC-v5		100Leaves		mfeat		Caltech101-20		ALOI	
Method	η=.5	η=.7	η=.5	η=.7	η=.5	η=.7	η=.5	η=.7	η=.5	η=.7
AMGL	0.469	0.327	0.400	0.340	0.524	0.327	0.242	0.186	–	–
AWSR	0.407	0.275	0.381	0.321	–	–	–	–	–	–
BestSV	0.429	0.343	0.336	0.256	0.403	0.297	0.337	0.337	–	–
Concat	0.489	0.365	0.391	0.332	0.574	0.346	0.257	0.205	–	–
DAIMC	0.436	0.286	0.196	0.203	0.231	0.140	0.441	0.423	0.084	0.120
IMVC-CBG	0.416	0.334	0.290	0.263	0.453	0.306	0.372	0.298	0.270	0.194
JVILG	0.155	0.223	0.310	0.333	0.538	0.194	**0.628**	**0.577**	0.403	0.365
KernelAdd	0.471	0.339	0.405	0.344	0.544	0.319	0.252	0.194	–	–
LMSC	0.329	0.276	0.366	0.306	0.402	0.263	0.240	0.175	–	–
LMVSC	0.166	0.165	0.153	0.096	0.206	0.103	0.352	0.339	0.127	0.053
MCAN	0.505	0.363	0.400	0.345	0.600	0.437	0.304	0.238	–	–
MKKM	0.318	0.265	0.380	0.324	0.327	0.214	0.230	0.180	–	–
PVC	0.608	0.442	0.255	0.230	0.656	0.412	0.348	0.317	0.246	0.172
RISE	0.568	0.449	0.435	0.361	0.763	0.549	0.483	0.371	0.477	0.411
RMSC	0.405	0.314	0.343	0.303	–	–	–	–	–	–
ZeroFill	0.491	0.362	0.390	0.327	0.571	0.364	0.262	0.205	–	–
SCoGF (ours)	**0.733**	**0.524**	**0.533**	**0.390**	**0.831**	**0.803**	0.461	0.392	**0.696**	**0.552**

Findings. On MSRC-v5, where $m{=}N$ and the anchor machinery buys nothing on complete data, SCoGF leads at every $\eta \geq 0.3$, while JVILG returns trivial or near-trivial solutions at every missing rate, collapsing exactly to the single-cluster solution ($0.143{=}1/7$) on complete data. On 100Leaves, SCoGF leads in ACC at every $\eta \geq 0.3$ but with the thinnest margins of any dataset: its worst cell across the study (0.390 at $\eta{=}0.7$) occurs here, JVILG takes the NMI lead at $\eta{=}0.7$, and with only three views the masked average has little to aggregate, making this the clearest boundary case of SCoGF's design (Table 2). On mfeat, full-graph methods lead on complete data but degrade quickly; SCoGF passes all baselines at $\eta{=}0.5$ and widens the margin at $\eta{=}0.7$, and under view-dependent missingness RISE is strongest and SCoGF second in both metrics. On Caltech101-20, the feature-homogeneous benchmark, JVILG dominates throughout, since its learned view weights exploit cross-view structure that a uniform masked average does not model, while SCoGF is mid-pack and degrades the least among the anchor-based competitors. On ALOI, the largest benchmark, only the sub-quadratic group runs; SCoGF leads at every missing rate by a wide margin, while DAIMC and LMVSC degenerate at $\eta \geq 0.3$.

Worst-case behavior. Table 2 summarizes the worst-case and average behavior of each method across all dataset–protocol cells. SCoGF attains the best average rank (2.04) of all seventeen methods, and its worst cell (0.390, 100Leaves at $\eta{=}0.7$) stays well above the failure modes of otherwise-strong competitors: JVILG's worst cell is 0.143 (the trivial $1/K$ solution on MSRC-v5, reached on complete data), LMVSC's 0.053, and DAIMC's 0.054, despite an identical oracle-tuning budget.

RISE and JVILG are the strongest competitors overall, with learned embeddings and view weights that exploit the geometry well where it is favorable; the same machinery is also their failure mode, since learned view weights can concentrate on a view that missingness has emptied and iterative solvers can settle on degenerate solutions, as in JVILG's collapse on MSRC-v5, RISE's degeneration off its narrow plateau (Appendix D), and the order-of-magnitude losses of DAIMC and LMVSC at $\eta \geq 0.3$. SCoGF has no learned weights and no iterative solver, so these failure modes do not arise; it gives up a few points of peak accuracy on easy, complete, homogeneous data in exchange for effectiveness that survives everywhere we tested.

Table 2: Applicability summary: worst-case ACC and average rank across all dataset $\times$ missingness cells in which the method participated (Table 1 and Tables 4–8). Ranks are computed per cell among the participating methods, with view-dependent missingness counted as one cell per dataset.

	AMGL	AWSR	BestSV	Concat	DAIMC	IMVC-CBG	JVILG	KernelAdd	LMSC	LMVSC	MCAN	MKKM	PVC	RISE	RMSC	ZeroFill	SCoGF (ours)
Worst ACC	0.186	0.275	0.256	0.205	0.054	0.194	0.143	0.194	0.175	0.053	0.238	0.180	0.172	0.361	0.303	0.205	**0.390**
Avg. rank	8.72	11.00	10.03	6.33	9.40	9.52	9.40	7.40	12.25	10.28	4.12	12.85	7.42	3.52	10.70	6.47	**2.04**

6.3 Sensitivity

The tuning sweep behind the protocol of Section 6.1 is reported in full in Appendix D (Tables 9–13) and substantiates the robustness column of Table 3. Across the grid $r \in \{5, 10, 15\}$, SCoGF's ACC varies by at most 0.03 on every dataset; on MSRC-v5 and ALOI the largest r is best, consistent with denser anchor graphs helping when N is small or K is large. Competitors behave differently: RISE drops from 0.62 to 0.22 on MSRC-v5 when β moves from 100 to 1, JVILG returns trivial or near-trivial solutions on MSRC-v5 across its four configurations, DAIMC fails to converge on ALOI for two of its three α values, and IMVC-CBG loses over 40% of its MSRC-v5 accuracy between λ=0.1 and λ=10, so a mis-set parameter costs SCoGF a few points but can cost several competitors the entire result.

6.4 Efficiency

Table 17 (Appendix H) reports wall-clock timings at symmetric η=0.5. SCoGF sits in the lowest cost bracket among methods that remain accurate at high missing rates: 13.9s on ALOI versus 105.8s for LMVSC and 33.5s for RISE; on the smaller benchmarks the gap widens to one to two orders of magnitude against the classical methods (e.g., 46.7s for RMSC on 100Leaves versus 1.2s for SCoGF), and the cubic pair cannot run on ALOI at all. Only PVC and IMVC-CBG are comparably cheap, and both are far less accurate.

These timings match the complexity analysis of Section 4: under the size rule the anchor count is constant for $N \geq 500$, so all super-linear work operates on the fixed-size anchor system and the total cost reduces to $O\big(N(K + \bar{d}V)\big)$ with memory $O(m^2 + Nr)$ instead of $O(N^2)$. The non-iterative design has two further consequences: the runtime is nearly invariant to the missing rate, since the same code path with the same m and r runs at every η; and the quoted timings are the total cost, whereas a tuned competitor must multiply its per-run cost by the number of grid points searched.

7 Conclusion

We presented SCoGF, a single-pass, imputation-free algorithm for incomplete multi-view spectral clustering built on mask-native constructions: deterministic k-means anchor selection, learned simplex anchor graphs on observed anchors only, a closed-form uniform masked aggregation (Proposition A.1), and a randomized spectral embedding with a missing-aware cross-anchor extension. We proved statistical soundness of the masked average, anchor spectral stability with explicit budgets, and an end-to-end normalized-cut consistency bound. Within our comparison, SCoGF attains the best average rank and the highest worst-case cell across the tested benchmarks and missing-rate range, including feature-heterogeneous benchmarks on which recent competitors degrade to random-guess accuracy, while running 2–9$\times$ faster than the strongest scalable iterative competitors on the largest benchmark and one to two orders of magnitude faster than classical iterative methods, and requiring no per-rate tuning, with accuracy insensitive to its single hyperparameter. Limitations and future directions are discussed in Appendix E.

AI USE STATEMENT

We used generative AI tools to aid and polish the writing, and to search for and identify related work. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the final paper.

ETHICS STATEMENT

This work uses only publicly available benchmark datasets and does not involve human subjects, personally identifiable information, or deployed models that raise fairness, privacy, or safety concerns. The authors declare no conflicts of interest.

REPRODUCIBILITY STATEMENT

The method is fully specified in Algorithm 1, all assumptions and complete proofs are given in the appendix, and the datasets, missingness protocols, baselines, hyperparameter grids, and metrics are described in Section 6.1 and the appendix. Source code will be released upon acceptance.

REFERENCES

Sen Bai, Qinghai Zheng, Xiaojuan Ren, and Jihua Zhu. Graph-guided imputation-free incomplete multi-view clustering. *Expert Systems with Applications*, 258:125165, 2024.

Deng Cai and Xiaofei Chen. Large scale spectral clustering via landmark-based sparse representation. *IEEE Transactions on Cybernetics*, 45(8):1669–1680, 2015.

Yuzhuo Dai, Jiaqi Jin, Zhibin Dong, Siwei Wang, Xinwang Liu, En Zhu, Xihong Yang, Xinbiao Gan, and Yu Feng. Imputation-free and alignment-free: Incomplete multi-view clustering driven by consensus semantic learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

Dheeru Dua and Casey Graff. UCI machine learning repository. `http://archive.ics.uci.edu/ml`, 2019. Multiple Features data set.

Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.

Lishan Feng, Guoxu Zhou, and Jingya Chang. An adaptive weighted self-representation method for incomplete multi-view clustering. *International Journal of Machine Learning and Cybernetics*, 15:4521–4536, 2024. Also arXiv:2302.05563, 2023.

Charless Fowlkes, Serge Belongie, Fan Chung, and Jitendra Malik. Spectral grouping using the Nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225, 2004.

Jan-Mark Geusebroek, Gertjan Burghouts, and Arnold Smeulders. The Amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.

Alex Gittens and Michael W. Mahoney. Revisiting the Nyström method for improved large-scale machine learning. *Journal of Machine Learning Research*, 17(1):3977–4041, 2016.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53 (2):217–288, 2011.

Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 2262–2268, 2018.

Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 4412–4419, 2020.

Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 3911–3919, 2023.

Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, pp. 1968–1974, 2014.

Xiuyi Li, He Zhang, Rong Wang, and Feiping Nie. Multi-view clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):330–344, 2022.

Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-scale multi-view spectral clustering via bipartite graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, pp. 2750–2756, 2015.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. COMPLETER: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11174–11183, 2021.

Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2023.

Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, and Wen Gao. Late fusion incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2410–2423, 2019.

Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, En Zhu, Tongliang Liu, Marius Kloft, Dinggang Shen, Jianping Yin, and Wen Gao. Multiple kernel k-means with incomplete kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(5):1191–1204, 2020.

Charbel Mallah, James Cope, and James Orwell. Plant leaf classification using probabilistic integration of shape, texture and margin features. In *Proceedings of the International Conference on Signal Processing, Pattern Recognition and Applications*, 2013.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, volume 14, pp. 849–856, 2002.

Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 977–986, 2014.

Feiping Nie, Jing Li, and Xuelong Li. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 1881–1887, 2016.

Feiping Nie, Guohao Cai, and Xuelong Li. Multi-view clustering and semi-supervised classification with adaptive neighbours. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Feiping Nie, Lai Tian, and Xuelong Li. Multiview clustering via adaptively weighted procrustes. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2022–2030, 2018.

Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

Weixiang Shao, Lifang He, and Philip S. Yu. Multiple incomplete views clustering via weighted nonnegative matrix factorization with $\ell_{2,1}$ regularization. In *Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*, pp. 318–334, 2015.

Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

Hao Wang, Linlin Zong, Bing Liu, Yan Yang, and Wei Zhou. Spectral perturbation meets incomplete multi-view data. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 3677–3683, 2019.

Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9776–9785, 2022.

Xinxin Wang, Yongshan Zhang, and Yicong Zhou. Learn multi-task anchor: Joint view imputation and label generation for incomplete multi-view clustering. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 6451–6459, 2025a. doi: 10.24963/ijcai.2025/718.

Xinxin Wang, Yongshan Zhang, and Yicong Zhou. Highly efficient rotation-invariant spectral embedding for scalable incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21312–21320, 2025b. doi: 10.1609/aaai.v39i20.35431.

Jie Wen, Yong Xu, and Hong Liu. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Transactions on Cybernetics*, 50(4):1418–1429, 2020.

Christopher K. I. Williams and Matthias Seeger. Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems*, volume 13, pp. 682–688, 2001.

John Winn, Antonio Criminisi, and Thomas Minka. Object categorization by learned universal visual dictionary. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pp. 1800–1807, 2005.

Benyu Wu, Wei Du, Jun Wang, and Guoxian Yu. Imputation-free incomplete multi-view clustering via knowledge distillation. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 6570–6578, 2025.

Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, pp. 2149–2155, 2014.

Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning, 2013. arXiv preprint arXiv:1304.5634.

Jie Xu, Chao Li, Yazhao Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 4455–4463, 2022.

Stella X. Yu and Jianbo Shi. Multiclass spectral clustering. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)*, volume 1, pp. 313–319, 2003. doi: 10.1109/ICCV.2003.1238361.

Changqing Zhang, Qinghua Hu, Huazhu Fu, Pengfei Zhu, and Xiaochun Cao. Latent multi-view subspace clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4279–4287, 2017.

A PROOFS

A.1 NOTATION AND DEFINITIONS USED IN THE PROOFS

We write $\pi_1, \ldots, \pi_K$ for the mixing proportions of the mixture of Assumption 1(i) and $\pi_{\min} := \min_k \pi_k$; $M_{ij}^{(v)} := o_i^{(v)} o_j^{(v)} \in \{0, 1\}$ for the indicator that i and j are co-observed in view v;

$n_v := \sum_i o_i^{(v)}$ for the observed-sample count of view v; and c for a universal positive constant that may change from line to line. The normalized cut of a K-way partition $\mathcal{C} = \{\mathcal{C}_k\}_{k=1}^K$ of a graph with degrees d_i is $\mathrm{NCut}_K(\mathcal{C}) = \sum_{k=1}^K \mathrm{cut}(\mathcal{C}_k, \bar{\mathcal{C}}_k)/\mathrm{vol}(\mathcal{C}_k)$, where $\mathrm{vol}(\mathcal{C}_k) = \sum_{i\in\mathcal{C}_k} d_i$.

A.2 Auxiliary results

Proposition A.1 (Closed-form masked aggregation)**.** *The unique minimizer of the masked aggregation problem equation 1 is the uniform masked average*

$$\mathbf{G}^\star = \frac{1}{V}\sum_{v=1}^{V} \mathbf{S}^{(v)}, \tag{6}$$

followed by symmetrization and diagonal zeroing: each anchor pair accumulates exactly the views that observe both endpoints, and pairs observed by no view are zero. Symmetrizing and zeroing the diagonal after minimization coincides with minimizing Eq. equation 1 over symmetric $\mathbf{G}$ with $\mathrm{diag}(\mathbf{G}) = \mathbf{0}$, so the produced graph is still an exact constrained minimizer.

Proof of Proposition A.1. The functional $\Psi(\mathbf{G}) = \sum_v \|\mathbf{S}^{(v)} - \mathbf{G}\|_F^2$ of Eq. equation 1 is separable across entries and, for each pair (i, j), is a sum of strictly convex quadratics. Setting the derivative to zero,

$$\frac{\partial \Psi}{\partial G_{ij}} = 2\sum_v \left(G_{ij} - S_{ij}^{(v)}\right) = 0 \quad \Longrightarrow \quad G_{ij}^\star = \frac{1}{V}\sum_v S_{ij}^{(v)},$$

and strict convexity makes this the unique minimizer. Since $\mathbf{S}^{(v)}$ is zero at every pair view v cannot support (missing anchors remain zero rows/columns), the sum accumulates exactly the co-observing views, and pairs supported by no view are zero. For the constrained variant, minimize Ψ directly over symmetric $\mathbf{G}$: the objective couples each unordered pair $\{i, j\}$ through $\sum_v \left[(S_{ij}^{(v)} - G_{ij})^2 + (S_{ji}^{(v)} - G_{ij})^2\right]$, whose unique minimizer is $G_{ij} = \frac{1}{2V}\sum_v \left(S_{ij}^{(v)} + S_{ji}^{(v)}\right)$, exactly the symmetrized average. Restricting further to $\mathrm{diag}(\mathbf{G}) = \mathbf{0}$ leaves the off-diagonal entries untouched by the same entrywise separability, so symmetrizing and zeroing the diagonal after unconstrained minimization coincides with constrained minimization over symmetric zero-diagonal graphs. □

Lemma A.2 (Structural event)**.** *Under Assumptions 1 and 3(a), there is a constant $C_2 > 0$ absorbing $\left(4C_{\mathrm{cov}}^2 \mathsf{D}/\Delta_{\mathrm{sep}}\right)^{\bar{d}} r\, p_{\max}/(p_{\min}\pi_{\min}\rho_0)$ such that, whenever $m \geq C_2 K \log(4KV/\iota)$, an event of probability at least $1 - \iota/2$ holds on which, simultaneously for all k, v:*

(i) Occupancy. $m_k^{(v)} \geq \frac{1}{2}\, m\, \pi_k \sqrt{\rho_0}$.

(ii) Purity. The r nearest observed neighbors of every observed anchor in view v all belong to the cluster of that anchor.

(iii) Block structure. Every Stage-1 graph $\mathbf{S}^{(v)}$ and the anchor consensus graph $\mathbf{G}^\star$ are block-diagonal with respect to the cluster partition.

(iv) Gap. Each diagonal block of the anchor consensus graph (before the Stage-2 diagonal zeroing) is connected and carries a positive diagonal, hence is primitive, so $\delta_{\mathrm{bd}} := \min_k \left(1 - \lambda_2(\mathbf{W}_k)\right) > 0$, where $\mathbf{W}_k$ are the normalized blocks. (This is a per-block spectral margin; it is distinct from the cross-view eigengap δ_K of Assumption 3(b).)

Proof of Lemma A.2.

(i) Occupancy. Each anchor is uniform over $\{1, \ldots, N\}$, hence lands in cluster k with probability π_k; by Assumption 1(v) the realized observed fraction of view v satisfies $n_v/N \geq \sqrt{\rho_0}$ (since $|\Omega^{(v)}|/N^2 = (n_v/N)^2 \geq \rho_0$), independently of the anchor draw. Each anchor therefore lands in cluster k and is observed in view v with probability at least $\pi_k\sqrt{\rho_0}$. The count $m_k^{(v)}$ is a sum of m indicators sampled without replacement; Serfling's inequality with mean $\mu = m\pi_k\sqrt{\rho_0}$ gives $\mathbb{P}[m_k^{(v)} < \mu/2] \leq e^{-\mu/8}$.

(ii) Purity. Fix k, v and set $\varepsilon = \Delta_{\mathrm{sep}}/2$. By Assumption 1(iii), $\mathcal{X}_k$ partitions into $N_{\mathrm{cell}} \le (2C_{\mathrm{cov}}\mathsf{D}_k/\Delta_{\mathrm{sep}})^{d_v}$ cells of diameter $\le \Delta_{\mathrm{sep}}/2$ and measure $\ge (\Delta_{\mathrm{sep}}/(2C_{\mathrm{cov}}))^{d_v}$. An anchor of cluster k falls in any given cell with probability at least $q = p_{\min}(\Delta_{\mathrm{sep}}/(2C_{\mathrm{cov}}))^{d_v}$, so cell occupancy dominates a $\mathrm{Binomial}(m_k^{(v)}, q)$; by Chernoff, $\mathbb{P}[\text{cell} < r] \le \exp\big(-(m_k^{(v)}q - r)^2/(2m_k^{(v)}q)\big)$, which is at most $\iota/(4KVN_{\mathrm{cell}})$ whenever $m_k^{(v)} \ge 8\big(4C_{\mathrm{cov}}^2\mathsf{D}_k/\Delta_{\mathrm{sep}}\big)^{d_v} p_{\min}^{-1}\big[r + \log\big(8KV(2C_{\mathrm{cov}}\mathsf{D}_k/\Delta_{\mathrm{sep}})^{d_v}/\iota\big)\big]$. Combining with (i) and absorbing the factors into C_2 gives the stated condition on m. Union-bounding over cells, clusters, and views: every anchor has at least r same-cluster observed anchors within distance $\Delta_{\mathrm{sep}}/2$, while every cross-cluster observed anchor is at distance $\ge \Delta_{\mathrm{sep}}$ (Assumption 1(ii)). Ties occur with probability zero, so the r nearest observed neighbors are all same-cluster.

(iii) Block structure. On the purity event, the candidate set $\mathcal{N}_r(i)$ of every anchor contains only same-cluster anchors, so the Stage-1 weight vector $\mathbf{z}_i$ is supported within the cluster of i; symmetrization, row normalization, and the zeroing of missing-anchor rows all preserve zero blocks. The Stage-2 aggregation is element-wise (a uniform average of the $\mathbf{S}^{(v)}$ followed by symmetrization and diagonal zeroing), so the consensus graph $\mathbf{G}^\star$ inherits the same zero cross-blocks.

(iv) Gap. On the block-structure event, consider the consensus graph before the Stage-2 diagonal zeroing, with normalized adjacency whose blocks are $\mathbf{W}_k = \mathbf{D}_k^{-1/2}\mathbf{G}_k\mathbf{D}_k^{-1/2}$. On the occupancy and purity event, every anchor has $r \ge 1$ same-cluster neighbors within distance $\Delta_{\mathrm{sep}}/2$, and chaining across the $O((\mathsf{D}/\Delta_{\mathrm{sep}})^{\bar{d}})$ cells connects each block. Moreover, the Stage-1 self-loops $\mathrm{diag}(\mathbf{S}^{(v)}) = \mathbf{m}^{(v)}$ (Section 4) contribute positive diagonal mass to every observed anchor, which persists through the masked average and the symmetrization, so each diagonal block has a positive diagonal entry at every anchor observed by at least one view. A connected graph with a positive diagonal entry is aperiodic, hence primitive. (The Stage-2 diagonal zeroing removes these self-loops from the realized graph but perturbs $\mathbf{W}$ by at most $B/(m\bar{\rho}s_{\min})$ in operator norm, which is absorbed in Layer 4 of the proof of Theorem 5.2.) By Perron–Frobenius, 1 is a simple eigenvalue of $\mathbf{W}_k$ with eigenvector $\mathbf{D}_k^{1/2}\mathbf{1}$ and all other eigenvalues are strictly below 1 in absolute value. The K largest eigenvalues of $\mathbf{W}$ are exactly the leading eigenvalues of the K blocks, giving $\delta_{\mathrm{bd}} = \min_k(1 - \lambda_2(\mathbf{W}_k)) > 0$. □

Lemma A.3 (Stability of Stage-1 weights)**.** *Suppose the standardized features satisfy $\|\mathbf{p}\|_2 \le c_p$ and $\|\mathbf{P}\|_2 \le c_P$ for constants c_p, c_P. Note that $\mathbf{P}$ is the local neighbor matrix $\mathbf{P}^{(v)}_{\mathcal{N}_r(i)} \in \mathbb{R}^{r\times d_v}$: under z-score standardization of features supported on the compact sets of Assumption 1, the entries remain bounded by a constant, so $\|\mathbf{P}\|_2 \le \|\mathbf{P}\|_F = O(\sqrt{r\,d_v})$. Thus c_p, c_P are constants of the pipeline, depending only on r and d_v and not on m or N; in the budget of Theorem 5.2 they are absorbed into the constants of β_e, and the explicit $\sqrt{r}$ factor there tracks the neighborhood perturbation $\|\mathbf{P} - \mathbf{P}'\|_F \le 2\sqrt{r}\,\varepsilon_r$ itself. Then the solution map $(\mathbf{p}, \mathbf{P}) \mapsto \mathbf{z}^\star(\mathbf{p}, \mathbf{P})$ of the Stage-1 problem (Eq. equation 2) over the simplex $\mathcal{S} = \{\mathbf{z} \ge \mathbf{0},\ \mathbf{1}^\top\mathbf{z} = 1\}$ is Lipschitz with constant $C(c_p, c_P)/\lambda_z$:*

$$\big\|\mathbf{z}^\star(\mathbf{p}, \mathbf{P}) - \mathbf{z}^\star(\mathbf{p}', \mathbf{P}')\big\|_2 \le \frac{c_p + 2c_P}{\lambda_z}\big(\|\mathbf{p} - \mathbf{p}'\|_2 + \|\mathbf{P} - \mathbf{P}'\|_2\big).$$

Proof of Lemma A.3. Write $f(\mathbf{z}; \mathbf{p}, \mathbf{P}) = \|\mathbf{p} - \mathbf{P}\mathbf{z}\|_2^2 + \lambda_z\|\mathbf{z}\|_2^2$, with gradient $\nabla_{\mathbf{z}} f = 2\mathbf{P}^\top(\mathbf{P}\mathbf{z} - \mathbf{p}) + 2\lambda_z\mathbf{z}$. The Hessian $2(\mathbf{P}^\top\mathbf{P} + \lambda_z\mathbf{I}) \succeq 2\lambda_z\mathbf{I}$, so $\nabla_{\mathbf{z}} f$ is $2\lambda_z$-strongly monotone in $\mathbf{z}$. Let $\mathbf{z}, \mathbf{z}'$ be the minimizers over the simplex $\mathcal{S}$ for data $(\mathbf{p}, \mathbf{P})$ and $(\mathbf{p}', \mathbf{P}')$. The variational inequalities of optimality, $\langle \nabla f(\mathbf{z}; \mathbf{p}, \mathbf{P}), \mathbf{z}' - \mathbf{z}\rangle \ge 0$ and $\langle \nabla f(\mathbf{z}'; \mathbf{p}', \mathbf{P}'), \mathbf{z} - \mathbf{z}'\rangle \ge 0$, give, respectively,

$$\big\langle \nabla f(\mathbf{z}; \mathbf{p}, \mathbf{P}),\ \mathbf{z} - \mathbf{z}'\big\rangle \le 0 \quad \text{and} \quad \big\langle \nabla f(\mathbf{z}'; \mathbf{p}', \mathbf{P}'),\ \mathbf{z} - \mathbf{z}'\big\rangle \ge 0.$$

Subtracting the gradient at the same data $(\mathbf{p}, \mathbf{P})$ from both sides of the second inequality,

$$\big\langle \nabla f(\mathbf{z}; \mathbf{p}, \mathbf{P}) - \nabla f(\mathbf{z}'; \mathbf{p}, \mathbf{P}),\ \mathbf{z} - \mathbf{z}'\big\rangle \le \big\langle \nabla f(\mathbf{z}'; \mathbf{p}', \mathbf{P}') - \nabla f(\mathbf{z}'; \mathbf{p}, \mathbf{P}),\ \mathbf{z} - \mathbf{z}'\big\rangle.$$

The left side is at least $2\lambda_z\|\mathbf{z} - \mathbf{z}'\|_2^2$ by strong monotonicity at fixed data; the right side is at most $\|\nabla f(\mathbf{z}'; \mathbf{p}', \mathbf{P}') - \nabla f(\mathbf{z}'; \mathbf{p}, \mathbf{P})\|_2\,\|\mathbf{z} - \mathbf{z}'\|_2$ by Cauchy–Schwarz. Dividing by $\|\mathbf{z} - \mathbf{z}'\|_2$,

$$2\lambda_z\|\mathbf{z} - \mathbf{z}'\|_2 \le \big\|\nabla f(\mathbf{z}'; \mathbf{p}', \mathbf{P}') - \nabla f(\mathbf{z}'; \mathbf{p}, \mathbf{P})\big\|_2.$$

For the gradient difference at $\mathbf{z}'$, using $\|\mathbf{p}\|_2, \|\mathbf{p}'\|_2 \le c_p$, $\|\mathbf{P}\|_2, \|\mathbf{P}'\|_2 \le c_P$, and $\|\mathbf{z}'\|_2 \le \|\mathbf{z}'\|_1 = 1$,

$$\begin{aligned}
&\big\|\mathbf{P}^\top(\mathbf{P}\mathbf{z}' - \mathbf{p}) - \mathbf{P}'^\top(\mathbf{P}'\mathbf{z}' - \mathbf{p}')\big\|_2 \\
&\quad\le \|(\mathbf{P}-\mathbf{P}')^\top(\mathbf{P}\mathbf{z}' - \mathbf{p})\|_2 + \|\mathbf{P}'^\top\big((\mathbf{P}-\mathbf{P}')\mathbf{z}' - (\mathbf{p}-\mathbf{p}')\big)\|_2 \\
&\quad\le (c_P + c_p)\,\|\mathbf{P}-\mathbf{P}'\|_2 + c_P\big(\|\mathbf{P}-\mathbf{P}'\|_2\,\|\mathbf{z}'\|_2 + \|\mathbf{p}-\mathbf{p}'\|_2\big) \\
&\quad\le c_P\,\|\mathbf{p}-\mathbf{p}'\|_2 + (c_p + 2c_P)\,\|\mathbf{P}-\mathbf{P}'\|_2.
\end{aligned}$$

The bound above applies to the half-gradient $\mathbf{P}^\top(\mathbf{P}\mathbf{z} - \mathbf{p}) + \lambda_z\mathbf{z}$; the full gradient carries an additional factor 2, so

$$\|\mathbf{z} - \mathbf{z}'\|_2 \le \frac{c_P}{\lambda_z}\,\|\mathbf{p}-\mathbf{p}'\|_2 + \frac{c_p + 2c_P}{\lambda_z}\,\|\mathbf{P}-\mathbf{P}'\|_2 \le \frac{c_p + 2c_P}{\lambda_z}\big(\|\mathbf{p}-\mathbf{p}'\|_2 + \|\mathbf{P}-\mathbf{P}'\|_2\big).$$

□

A.3 Proof of Theorem 5.1

Proof of Theorem 5.1. (a) Under view-level MCAR (Assumption 2), anchors i, j are co-observed by view v with probability $(1-\eta_v)^2$, hence fail to be co-observed with probability $1-(1-\eta_v)^2 = 2\eta_v - \eta_v^2$; independence across views gives $P_{\text{obs}} = 1 - \prod_v (2\eta_v - \eta_v^2)$. Each factor equals $2\eta_v - \eta_v^2 = 1-(1-\eta_v)^2 < 1$ whenever $\eta_v < 1$, so the unsupported fraction decays exponentially in V at rate $1-(1-\eta_{\max})^2$ per view (slow at high missing rates: at $\eta_{\max} = 0.7$ the base is 0.91, but exponential throughout the experimental range); a single complete view ($\eta_v = 0$) annihilates the product.

(b) Masks are independent of the graph entries given the data, so, under the view-homogeneous conditional mean of the theorem statement,

$$\mathbb{E}\big[\mathbf{G}^\star_{ij}\big] = \frac{1}{V}\sum_v \mathbb{E}\big[M^{(v)}_{ij} S^{(v)}_{ij}\big] = \frac{1}{V}\sum_v (1-\eta_v)^2\, s^\star_{ij} = \bar\rho\, s^\star_{ij},$$

with $\bar\rho = \frac{1}{V}\sum_v (1-\eta_v)^2$ independent of the pair (i,j). If the conditional mean varies across views, define $\tilde s_{ij} := \frac{1}{V\bar\rho}\sum_v (1-\eta_v)^2\,\mathbb{E}[S^{(v)}_{ij} \mid i,j \text{ observed in } v]$; then $\mathbb{E}[\mathbf{G}^\star_{ij}] = \bar\rho\,\tilde s_{ij}$ and the deviation $\tilde s_{ij} - s^\star_{ij}$ from any single population target enters the anchor-sampling bias term of Theorem 5.2. Multiplication of $\mathbf{G}^\star$ by any positive global constant leaves the normalized adjacency $\mathbf{W} = \mathbf{D}^{-1/2}\mathbf{G}^\star\mathbf{D}^{-1/2}$ invariant (the factor cancels between $\mathbf{G}^\star$ and $\mathbf{D}$), hence leaves the embedding invariant.

(c) Given the masks, the per-view entries are conditionally independent across views. Writing $q_v = (1-\eta_v)^2$ and using $\mathrm{Var}\big(M^{(v)}_{ij} S^{(v)}_{ij}\big) \le \mathbb{E}\big[M^{(v)}_{ij}(S^{(v)}_{ij})^2\big] \le q_v\,\mathbb{E}\big[(S^{(v)}_{ij})^2 \,\big|\, i,j \text{ observed in } v\big] \le q_v(\sigma^2 + B^2)$,

$$\mathrm{Var}\big(\mathbf{G}^\star_{ij}\big) = \frac{1}{V^2}\sum_v \mathrm{Var}\big(M^{(v)}_{ij} S^{(v)}_{ij}\big) \le \frac{(\sigma^2+B^2)\bar\rho}{V} = O\big((\sigma^2+B^2)/V\big).$$

□

A.4 Proof of Theorem 5.2

Proof of Theorem 5.2. Write $\widehat{\mathbf{G}} = \mathbf{G}^\star = \frac{1}{V}\sum_v \mathbf{S}^{(v)}$ and let $\mathbf{E} := \widehat{\mathbf{G}} - \bar\rho\mathbf{S}^\star$. We bound $\|\widehat{\mathbf{W}} - \mathbf{W}^\star\|_2$ in four layers, then apply Davis–Kahan (Layer 5).

Layer 1 (mask noise, entrywise Bernstein). Fix the anchor set. Each entry $\widehat G_{ij} = \frac{1}{V}\sum_v M^{(v)}_{ij} S^{(v)}_{ij}$ is an average of V terms $X_v := M^{(v)}_{ij} S^{(v)}_{ij} \in [0, B]$ that are independent across v (the masks are independent by Assumption 2, and the $\mathbf{S}^{(v)}$ are deterministic given the anchor features) with $\mathbb{E}X_v = q_v S^{(v)}_{ij}$ and $\mathrm{Var}(X_v) \le B^2 q_v$. Strictly, the realized $\mathbf{S}^{(v)}$ depend on the anchor-level masks, since neighborhood search is confined to observed anchors (Section 4); we therefore read the display conditionally on the anchor features and on the structural event of Lemma A.2, with the randomness of X_v ascribed to the indicators $M^{(v)}_{ij}$, and the residual mask–graph coupling controlled on the

structural event and absorbed into the bias term β_e of Layer 2. Bernstein's inequality for the average, followed by a union bound over the m^2 entries, gives with probability at least $1-\iota/4$,

$$\max_{ij}\left|\widehat{G}_{ij}-A_{ij}\right|\le\omega_1,\qquad \omega_1:=\sqrt{\frac{4B^2\bar{\rho}\log(4m^2/\iota)}{V}}+\frac{4B\log(4m^2/\iota)}{3V},$$

where $A_{ij}:=\frac{1}{V}\sum_v q_v S_{ij}^{(v)}$ is the mask-average of the realized graphs.

Layer 2 (anchor-sampling bias). On the structural event (Lemma A.2), the realized r-NN neighborhoods of every anchor are same-cluster and within radius $\varepsilon_r=O\big(\mathsf{D}(r\log(4m/\iota)/(\rho_0 m))^{1/\bar{d}}\big)$. Replacing the population neighborhood by the realized one perturbs $\mathbf{p}$ by at most $2\varepsilon_r$ and the local matrix $\mathbf{P}\in\mathbb{R}^{r\times\bar{d}_v}$ by at most $2\sqrt{r}\,\varepsilon_r$ in Frobenius norm (each of the r rows shifts by $O(\varepsilon_r)$), and Lemma A.3 (with $\|\mathbf{P}-\mathbf{P}'\|_2\le\|\mathbf{P}-\mathbf{P}'\|_F$) converts the combined perturbation into an entrywise weight perturbation of at most $\big((c_p+2c_P)/\lambda_z\big)\cdot O(\sqrt{r}\,\varepsilon_r)$; symmetrization and row normalization (degrees bounded away from zero on the occupancy event) preserve the order. With probability at least $1-\iota/4$,

$$\max_{ij}\left|A_{ij}-\bar{\rho}s^\star_{ij}\right|\le\beta_e,\qquad \beta_e=O\!\left(\frac{B\mathsf{D}\sqrt{r}}{\lambda_z}\Big(\frac{r\log(4m/\iota)}{\rho_0 m}\Big)^{1/\bar{d}}\right).$$

Layer 3 (degree concentration). Let $\widehat{d}_i=\sum_j\widehat{G}_{ij}$ and $d_i^\star=\bar{\rho}\,m s_i\ge m\bar{\rho}s_{\min}$. By Layers 1–2, $|\widehat{d}_i-d_i^\star|\le m(\omega_1+\beta_e)$ for all i; whenever $\omega_1+\beta_e\le\bar{\rho}s_{\min}/2$ (implied by Eq. equation 4), $\widehat{d}_i\in(1\pm\frac{1}{2})d_i^\star$.

Layer 4 (normalized perturbation). Decompose $\widehat{W}_{ij}-W^\star_{ij}=\frac{\widehat{G}_{ij}-\bar{\rho}s^\star_{ij}}{\sqrt{\widehat{d}_i\widehat{d}_j}}+\bar{\rho}s^\star_{ij}\big(\frac{1}{\sqrt{\widehat{d}_i\widehat{d}_j}}-\frac{1}{\sqrt{d_i^\star d_j^\star}}\big)$. By Layer 3, the first term is at most $\frac{\omega_1+\beta_e}{m\bar{\rho}s_{\min}/2}$ and the second is $O\big(\frac{B}{ms_{\min}}\cdot\frac{\omega_1+\beta_e}{\bar{\rho}s_{\min}}\big)$. Since $s_{\min}$ is a row-*mean* floor while B bounds the entries themselves, $s_{\min}\le B$, so $1/s_{\min}\le B/s^2_{\min}$ and both terms are of the same order, giving

$$\max_{ij}\left|\widehat{W}_{ij}-W^\star_{ij}\right|=O\!\left(\frac{B\,(\omega_1+\beta_e)}{m\,\bar{\rho}\,s^2_{\min}}\right),$$
$$\left\|\widehat{\mathbf{W}}-\mathbf{W}^\star\right\|_2\le m\max_{ij}\left|\widehat{W}_{ij}-W^\star_{ij}\right|=O\!\left(\frac{B\,(\omega_1+\beta_e)}{\bar{\rho}\,s^2_{\min}}\right).$$

(The diagonal zeroing of $\mathbf{G}^\star$ changes $\mathbf{W}$ by at most $B/(m\bar{\rho}s_{\min})$ in operator norm and is absorbed.)

Layer 5 (Davis–Kahan). The first condition of Eq. equation 4 enforces $B\omega_1/(\bar{\rho}s^2_{\min})\le\frac{\epsilon\delta_K}{8\sqrt{2K}}$ and the second enforces $B\beta_e/(\bar{\rho}s^2_{\min})\le\frac{\epsilon\delta_K}{8\sqrt{2K}}$ (after solving the $\bar{d}$-th power). The $\sqrt{r}$ factor of β_e raises the power of r in the solved anchor budget to $r^{1/2+1/\bar{d}}$; since $r^{1/2+1/\bar{d}}\le r$ for $\bar{d}\ge 2$, the stated budget in Eq. equation 4 dominates and suffices as written. Hence $\|\widehat{\mathbf{W}}-\mathbf{W}^\star\|_2\le\epsilon\delta_K/(2\sqrt{2K})\le\delta_K/2$. The Davis–Kahan $\sin\Theta$ theorem in projector form yields

$$\left\|\mathbf{U}_K\mathbf{U}_K^\top-\mathbf{U}^\star_K(\mathbf{U}^\star_K)^\top\right\|_F\le\frac{\sqrt{2K}\,\|\widehat{\mathbf{W}}-\mathbf{W}^\star\|_2}{\delta_K-\|\widehat{\mathbf{W}}-\mathbf{W}^\star\|_2}\le\epsilon.$$

A union bound over Layers 1–2 and the structural event gives the stated probability $1-\iota$. The implemented drop-first convention (Section 4) is analyzed in Proposition A.4. □

A.5 Safety of the drop-first rule: statement and proof of Proposition A.4

Theorem 5.2 is stated for the top-K embedding, whose gap $\delta_K=\lambda_K(\mathbf{W}^\star)-\lambda_{K+1}(\mathbf{W}^\star)$ equals the per-block margin $\delta_{\mathrm{bd}}>0$ on the structural event of Lemma A.2. The implementation instead drops the leading component and retains the second through $(K+1)$-th directions (Section 4). On the structural event this is exactly the degenerate case: $\lambda_1(\mathbf{W})=\cdots=\lambda_K(\mathbf{W})=1$, no gap separates the leading direction from the rest of the Perron subspace, and any gap assumption at λ_2 is incompatible with the block structure the theory is built on. The proposition below analyzes the drop-first rule directly in this regime, in two layers: the *outer* layer identifies the Perron subspace as a

whole through the gap $\lambda_K - \lambda_{K+1} = \delta_{\mathrm{bd}}$, to which Davis–Kahan applies without any internal basis identification, and the *inner* layer controls the orientation within the degenerate Perron subspace through the randomness of the solver's sketch. The relevant property of the dropped direction is not constancy but *delocalization* across the K cluster directions.

Proposition A.4 (Safety of the drop-first rule under block structure)**.** *Work on the structural event of Lemma A.2, on which the realized normalized adjacency* $\mathbf{W}$ *of the consensus graph is block diagonal with* K *primitive blocks. Let* $\mathbf{u}_1, \dots, \mathbf{u}_K$ *be the block-Perron directions, orthonormal with disjoint supports,* $\mathcal{P} = \mathrm{span}\{\mathbf{u}_k\}$ *the top-*K *eigenspace, and* $\lambda_{K+1}(\mathbf{W}) \le 1 - \delta_{\mathrm{bd}} < 1$*. Condition on the structural event and take the randomness to be the Gaussian sketch of the randomized SVD (Halko et al., 2011); model the orientation of the returned basis within the degenerate subspace* $\mathcal{P}$ *as Haar-random (the standard idealization of sketch-based eigensolvers at exact degeneracy; the conclusion holds for any returned basis whose leading vector is delocalized). Then, with probability at least* $1 - \iota$ *over the sketch:*

(i) the dropped leading direction $\mathbf{v}_1 = \sum_k \alpha_k \mathbf{u}_k$ *satisfies* $\max_k \alpha_k^2 = O\big(K^{-1}\log(K/\iota)\big)$*;*

(ii) the cluster representatives in the retained Perron coordinates, $\mathbf{b}_k^{(-1)} := \mathbf{b}_k - \alpha_k \boldsymbol{\alpha}$*, where* $\mathbf{b}_k$ *is the coordinate vector of* $\mathbf{u}_k$ *in the returned basis, satisfy* $\|\mathbf{b}_k^{(-1)}\|_2^2 = 1 - \alpha_k^2 \ge 1 - O\big(K^{-1}\log(K/\iota)\big)$ *and* $\|\mathbf{b}_k^{(-1)} - \mathbf{b}_l^{(-1)}\|_2^2 \ge 2 - O\big(K^{-1}\log(K/\iota)\big)$ *for all* $k \ne l$*;*

(iii) the additionally retained $(K{+}1)$*-th direction* $\mathbf{v}_{K+1}$ *is supported within a single block* k_0 *(or within the tied blocks if* λ_{K+1} *is not simple). Consequently, after* rowL2 *normalization, every row in a block* $k \ne k_0$ *maps exactly to* $\widetilde{\mathbf{b}}_k := \mathbf{b}_k^{(-1)}/\|\mathbf{b}_k^{(-1)}\|_2$ *with zero within-cluster spread, and the* K *cluster directions remain at pairwise distance* $\sqrt{2}\,\big(1 - O(K^{-1}\log(K/\iota))\big)$*; rows in block* k_0 *are confined to the ray* $\big\{\mathrm{normalize}\big(\widetilde{\mathbf{b}}_{k_0} + t\,\mathbf{e}_{K+1}\big) : t \ge 0\big\}$*, whose axis* $\mathbf{e}_{K+1}$ *is exactly orthogonal to every other cluster direction. Writing* $t_i := v_{K+1}(i)/u_{k_0}(i)$ *with* $u_{k_0}(i) = \sqrt{d_i/\mathrm{vol}_{k_0}}$*, a row of block* k_0 *can be misassigned only when* t_i *exceeds a constant threshold; under approximate degree regularity, Markov's inequality bounds the fraction of such rows in block* k_0 *by* $O(1/\tau^2)$ *at any threshold* τ*, while the remaining* $K - 1$ *blocks are recovered exactly.*

Proof of Proposition A.4. On the structural event, $\mathbf{W}$ is block diagonal with K primitive blocks (Lemma A.2(iii)–(iv)), so 1 is an eigenvalue of multiplicity exactly K with eigenspace $\mathcal{P}$; the $\mathbf{u}_k$ are orthonormal by the disjointness of their supports; and every remaining eigenvalue is at most $1 - \delta_{\mathrm{bd}}$, so the gap $\lambda_K - \lambda_{K+1} = \delta_{\mathrm{bd}}$ identifies $\mathcal{P}$ as a whole.

(i) Under the Haar model, $\boldsymbol{\alpha}$ is uniform on $\mathbb{S}^{K-1}$. Each coordinate satisfies $\alpha_k^2 \sim \mathrm{Beta}\big(\frac{1}{2}, \frac{K-1}{2}\big)$, and the standard concentration inequality for uniform vectors on the sphere gives $\mathbb{P}\big(\alpha_k^2 > c\log(K/\iota)/K\big) \le \iota/K$ for an absolute constant c; a union bound over k yields (i).

(ii) Let $\mathbf{V}$ be the orthogonal matrix whose columns are the returned basis vectors expressed in $\{\mathbf{u}_k\}$, so that $\mathbf{b}_k = \mathbf{V}^\top \mathbf{e}_k$ and the first coordinates across clusters form the first row $\boldsymbol{\alpha}^\top$ of $\mathbf{V}$. Dropping the first coordinate gives $\mathbf{b}_k^{(-1)} = \mathbf{b}_k - \alpha_k \boldsymbol{\alpha}$, hence $\|\mathbf{b}_k^{(-1)}\|_2^2 = \|\mathbf{b}_k\|_2^2 - \alpha_k^2 = 1 - \alpha_k^2$ and, since $\|\mathbf{b}_k - \mathbf{b}_l\|_2^2 = 2$,

$$\|\mathbf{b}_k^{(-1)} - \mathbf{b}_l^{(-1)}\|_2^2 = 2 - (\alpha_k - \alpha_l)^2 \ge 2 - 4\max_k \alpha_k^2,$$

and (i) gives (ii).

(iii) Eigenvectors of a block-diagonal matrix associated with simple eigenvalues are supported within single blocks, so $\mathbf{v}_{K+1}$ is supported within the block k_0 attaining λ_{K+1} (or within the tied blocks). The embedding row of an anchor i in cluster k decomposes exactly as

$$\mathrm{row}_i = u_k(i)\,\mathbf{b}_k^{(-1)} + v_{K+1}(i)\,\mathbf{e}_{K+1}, \qquad u_k(i) = \sqrt{d_i/\mathrm{vol}_k},$$

because every retained Perron coordinate of row i carries the same factor $u_k(i)$. For $k \ne k_0$ the second term vanishes and $u_k(i) > 0$, so rowL2 normalization maps every such row exactly to $\widetilde{\mathbf{b}}_k$; the pairwise distances of these cluster points are $\sqrt{2}\,\big(1 - O(K^{-1}\log(K/\iota))\big)$ by (ii). For $k = k_0$,

normalization removes the common factor $u_{k_0}(i)$ and the row lands on the ray $\mathrm{normalize}\big(\widetilde{\mathbf{b}}_{k_0} + t_i\,\mathbf{e}_{K+1}\big)$ with $t_i = v_{K+1}(i)/u_{k_0}(i)$. The axis $\mathbf{e}_{K+1}$ is exactly orthogonal to every $\widetilde{\mathbf{b}}_l$, $l \neq k_0$, whose $(K+1)$-th coordinate is zero, and the angle between $\widetilde{\mathbf{b}}_{k_0}$ and any $\widetilde{\mathbf{b}}_l$ is $\pi/2 - O(K^{-1}\log(K/\iota))$, since $\mathbf{b}_k^{(-1)\top}\mathbf{b}_l^{(-1)} = -\alpha_k\alpha_l$ by the orthonormality of the $\mathbf{b}_k$. Misassignment of row i to another cluster therefore requires the direction $\widetilde{\mathbf{b}}_{k_0} + t_i\mathbf{e}_{K+1}$ to cross the angular bisector of two cluster directions, i.e., requires t_i to exceed the constant threshold $\tan\big(\pi/4 - O(K^{-1}\log(K/\iota))\big) = \Theta(1)$. Finally, $\sum_{i\in k_0} t_i^2 = \mathrm{vol}_{k_0}\sum_{i\in k_0} v_{K+1}(i)^2/d_i = O(m_{k_0})$, using $\|\mathbf{v}_{K+1}\|_2 = 1$ and approximate degree regularity $d_i = \Theta(\mathrm{vol}_{k_0}/m_{k_0})$ over the block, as certified by the degree concentration of Theorem 5.2 (Layer 3 of its proof) whenever the row-mean ratio $s_{\max}/s_{\min}$ is bounded. Markov's inequality then gives $\#\{i \in k_0 : t_i > \tau\} = O(m_{k_0}/\tau^2)$ for every $\tau > 0$, i.e., at most an $O(1/\tau^2)$-fraction of the rows of block k_0 cross any threshold τ, while every row of the remaining $K-1$ blocks is assigned exactly. □

Three comments are in order. (a) *Source of randomness.* The Haar model is justified by the implementation itself: the randomized SVD forms its basis from a Gaussian sketch, so at exact degeneracy the returned orientation within $\mathcal{P}$ is a function of the sketch alone and is rotationally invariant in distribution. For the implemented randomized solver the Haar model is the exact description at exact degeneracy; for a deterministic solver the returned basis is arbitrary but fixed, and the conclusion holds whenever that basis is delocalized. (b) *Finite-sample splitting.* Off the structural event, the K-fold root splits by $O(\|\widehat{\mathbf{W}} - \mathbf{W}^\star\|_2)$, biasing the basis toward particular directions; since the bias is of the same order as the perturbation controlled in Theorem 5.2, the delocalization conclusion survives with an additional remainder of that order. (c) *The connected regime and small K.* When the consensus graph is connected and $\lambda_1 = 1$ is simple, the leading eigenvector is $\mathbf{D}^{1/2}\mathbf{1}/\|\mathbf{D}^{1/2}\mathbf{1}\|$—constant only on degree-regular graphs—and the drop-first rule reduces to the classical near-constant-direction heuristic; the proposition covers the opposite, degenerate regime in which that heuristic has no gap to appeal to. The bound of (i) is qualitative for small K (MSRC-v5, K=7) and strong for large K (100Leaves and ALOI, K=100), in line with the empirical observation that SCoGF is most stable where K is large.

A.6 Statement and proof of Lemma A.5

Lemma A.5 (Cross-anchor extension error decomposition)**.** *Let $\widetilde{\mathbf{G}} \in \mathbb{R}^{N\times N}$ be the full-sample consensus graph with entries in $[0, B]$, $\widetilde{\mathbf{G}}_{\mathcal{A}}$ its restriction to the anchor set, and $\widehat{\mathbf{G}} := \mathrm{sym}\big(\widetilde{\mathbf{C}}\,\mathbf{G}^\star\widetilde{\mathbf{C}}^\top\big)$ its extension to all samples, where $\widetilde{\mathbf{C}} := \mathrm{rowNorm}(\mathbf{C})$ is the row-normalized cross-anchor matrix of Section 4. Define $\mathbf{G}^{\mathrm{ca}} := \mathrm{sym}\big(\widetilde{\mathbf{C}}\,\widetilde{\mathbf{G}}_{\mathcal{A}}\,\widetilde{\mathbf{C}}^\top\big)$ and let $\kappa_{\mathbf{C}} := \max_j \sum_i \widetilde{\mathbf{C}}_{ij}$ be the maximal anchor column mass. Then*

$$\big\|\widetilde{\mathbf{G}} - \widehat{\mathbf{G}}\big\|_F \le \underbrace{\big\|\widetilde{\mathbf{G}} - \mathbf{G}^{\mathrm{ca}}\big\|_F}_{\text{extension (smoothing) error}} + \underbrace{\kappa_{\mathbf{C}}\,\big\|\widetilde{\mathbf{G}}_{\mathcal{A}} - \mathbf{G}^\star\big\|_F}_{\text{anchor-subproblem error}}. \tag{7}$$

Moreover: (i) under uniform-coverage anchor sets, $\kappa_{\mathbf{C}} = O(N/m)$ with probability at least $1-\iota/4$; (ii) if every sample has at least r observed anchor neighbors in each of its observed views, an event that holds with probability at least $1-\iota/4$ whenever $m \ge c\,(r + \log(N/\iota))/(\pi_{\min}\rho_0)$, then $\|\widetilde{\mathbf{G}} - \mathbf{G}^{\mathrm{ca}}\|_F = O\big(NB(rV\beta_e/w_{\min} + \varepsilon_r)\big)$, where β_e is the entrywise Stage-1 weight bias, ε_r the r-NN radius (Layer 2 of the proof of Theorem 5.2), and $w_{\min} = \min_i |\{v : o_i^{(v)} = 1\}|/V > 0$ the row-renormalization floor; (iii) the anchor-subproblem error is controlled by Theorem 5.2.

Proof of Lemma A.5.

The decomposition in Lemma A.5 is the triangle inequality applied to $\widetilde{\mathbf{G}} - \widehat{\mathbf{G}} = (\widetilde{\mathbf{G}} - \mathbf{G}^{\mathrm{ca}}) + (\mathbf{G}^{\mathrm{ca}} - \widehat{\mathbf{G}})$, with

$$\|\mathbf{G}^{\mathrm{ca}} - \widehat{\mathbf{G}}\|_F \le \|\widetilde{\mathbf{C}}\|_2^2\,\|\widetilde{\mathbf{G}}_{\mathcal{A}} - \mathbf{G}^\star\|_F \le \kappa_{\mathbf{C}}\,\|\widetilde{\mathbf{G}}_{\mathcal{A}} - \mathbf{G}^\star\|_F,$$

using submultiplicativity and $\|\widetilde{\mathbf{C}}\|_2^2 \le \|\widetilde{\mathbf{C}}\|_1\|\widetilde{\mathbf{C}}\|_\infty = \kappa_{\mathbf{C}}$ (row sums of $\widetilde{\mathbf{C}}$ equal 1 by row normalization).

(i) Row normalization makes every row of $\widetilde{\mathbf{C}}$ sum to one, so the total mass of $\widetilde{\mathbf{C}}$ is N, distributed over m anchor columns. Under uniform anchor sampling each anchor receives expected column mass N/m; the row contributions are independent across samples, and Bernstein's inequality with a union bound over the m columns gives $\kappa_{\mathbf{C}} = O\big(N/m + \sqrt{N\log(m/\iota)/m}\big) = O(N/m)$ with probability at least $1 - \iota/4$.

(ii) On the per-sample occupancy event, every row of $\mathbf{C}$ has its full complement of r observed neighbors per observed view. Entry (i, j) of $\mathbf{G}^{\mathrm{ca}}$ reads $G^{\mathrm{ca}}_{ij} = \sum_{a,b\in\mathcal{A}} \widetilde{\mathbf{C}}_{ia}\widetilde{\mathbf{C}}_{jb}\,\widetilde{G}_{ab}$, a weighted local average of $\widetilde{\mathbf{G}}$ over the anchor neighborhoods of i and j. For (i, j) in the same cluster, the structural event (Lemma A.2) restricts both neighborhoods to same-cluster anchors; the r-NN radii are at most ε_r, so $|\widetilde{G}_{ab} - \widetilde{G}_{ij}|$ is of order $B\varepsilon_r$; and the realized weights deviate from their population counterparts by at most β_e per entry (Lemma A.3), which row normalization divides by $w_i \geq w_{\min}$ (the map $(\mathbf{x}, w) \mapsto \mathbf{x}/w$ is $1/w_{\min}$-Lipschitz when $w \geq w_{\min}$). With at most rV nonzero weights per row, the entrywise error is $\big|G^{\mathrm{ca}}_{ij} - \widetilde{G}_{ij}\big| \leq \frac{2B\,rV\,\beta_e}{w_{\min}} + O(B\varepsilon_r)$, and for (i, j) in different clusters both sides are zero on the structural event. Taking the Frobenius norm of the resulting entrywise perturbation over the N^2 entries gives $\|\widetilde{\mathbf{G}} - \mathbf{G}^{\mathrm{ca}}\|_F = O\big(NB(rV\beta_e/w_{\min} + \varepsilon_r)\big)$.

(iii) The anchor-subproblem error $\|\widetilde{\mathbf{G}}_{\mathcal{A}} - \mathbf{G}^\star\|_F$ is controlled by Theorem 5.2: the mask-noise layer ω_1 and the anchor-sampling bias layer β_e together bound $\|\mathbf{G}^\star - \bar{\rho}\mathbf{S}^\star\|_F \leq m(\omega_1 + \beta_e)$ on the structural event, and $\widetilde{\mathbf{G}}_{\mathcal{A}} = \bar{\rho}\mathbf{S}^\star$ by definition of the population counterpart. □

A.7 Proof of Theorem 5.3

Proof of Theorem 5.3.

Write $\mathbf{L} = \mathbf{L}(\widetilde{\mathbf{G}})$ and $\widehat{\mathbf{L}} = \mathbf{L}(\widehat{\mathbf{G}})$, and $\mathbf{E} = \widehat{\mathbf{G}} - \widetilde{\mathbf{G}}$.

Step 1 (Laplacian perturbation). For normalized Laplacians, $\|\widehat{\mathbf{L}} - \mathbf{L}\|_2 \leq C_L\|\mathbf{E}\|_2/d_{\min}$, with C_L absorbing the bounded-degree ratios (von Luxburg et al., 2008).

Step 2 (eigenspace perturbation). By Assumption 3(c) and Davis–Kahan, whenever $C_L\|\mathbf{E}\|_2/(d_{\min}\delta_{\mathrm{Lap}}) \leq 1/2$,

$$\|\sin\Theta(\widehat{\mathbf{U}}_K, \mathbf{U}_K^{\mathrm{full}})\|_F \leq \frac{\sqrt{2K}\,\|\widehat{\mathbf{L}} - \mathbf{L}\|_2}{\delta_{\mathrm{Lap}}} \leq \frac{\sqrt{2K}\,C_L\|\mathbf{E}\|_2}{d_{\min}\delta_{\mathrm{Lap}}}.$$

Step 3 (from embedding to partition). Let $\widehat{\mathbf{H}}$ and $\mathbf{H}^\star$ be the *normalized* indicator matrices of $\widehat{\mathcal{C}}$ and $\mathcal{C}^\star$, with columns $H_{\cdot k} = \mathbf{1}_{\mathcal{C}_k}/\sqrt{\mathrm{vol}(\mathcal{C}_k)}$, so that $\mathbf{H}^\top\mathbf{H} = \mathbf{I}$ and $\|\mathbf{H}\|_F^2 = K$; this normalization is the one for which $\mathrm{NCut}_K(\mathcal{C}; \mathbf{A}) = \mathrm{Tr}(\mathbf{H}^\top\mathbf{L}(\mathbf{A})\mathbf{H})$ holds. Spectral clustering runs k-means on the normalized rows of $\widehat{\mathbf{U}}_K$; we adopt the standard approximate-k-means perturbation reduction (Ng et al., 2002; Yu & Shi, 2003), under which a $(1 + \varepsilon_0)$-approximate k-means solution satisfies $\mathrm{Tr}(\widehat{\mathbf{H}}^\top\widehat{\mathbf{L}}\widehat{\mathbf{H}}) \leq (1 + \varepsilon_0)\min_{\mathbf{H}}\mathrm{Tr}(\mathbf{H}^\top\widehat{\mathbf{L}}\mathbf{H}) + O(K\|\sin\Theta\|_F^2)$, where the second term is the perturbation of the k-means objective induced by replacing the population embedding with the realized one. Using the NCut identity above and $|\mathrm{Tr}(\mathbf{H}^\top(\widehat{\mathbf{L}} - \mathbf{L})\mathbf{H})| \leq K\|\widehat{\mathbf{L}} - \mathbf{L}\|_2$ for every normalized indicator $\mathbf{H}$ (since $\|\mathbf{H}\|_F^2 = K$), switch from $\widehat{\mathbf{L}}$ to $\mathbf{L}$:

$$\mathrm{NCut}_K(\widehat{\mathcal{C}}; \widetilde{\mathbf{G}}) = \mathrm{Tr}(\widehat{\mathbf{H}}^\top\mathbf{L}\widehat{\mathbf{H}}) \leq \mathrm{Tr}(\widehat{\mathbf{H}}^\top\widehat{\mathbf{L}}\widehat{\mathbf{H}}) + K\|\widehat{\mathbf{L}} - \mathbf{L}\|_2,$$

and bound the first term against $\mathbf{H}^\star$ before switching back:

$$\begin{aligned}\mathrm{Tr}(\widehat{\mathbf{H}}^\top\widehat{\mathbf{L}}\widehat{\mathbf{H}}) &\leq (1 + \varepsilon_0)\mathrm{Tr}((\mathbf{H}^\star)^\top\widehat{\mathbf{L}}\mathbf{H}^\star) + O(K\|\sin\Theta\|_F^2)\\ &\leq (1 + \varepsilon_0)\mathrm{NCut}_K(\mathcal{C}^\star; \widetilde{\mathbf{G}}) + (1 + \varepsilon_0)K\|\widehat{\mathbf{L}} - \mathbf{L}\|_2 + O(K\|\sin\Theta\|_F^2).\end{aligned}$$

Combining and substituting Steps 1–2 gives the two-term bound of Eq. equation 5, using $\|\mathbf{E}\|_2 \leq \|\mathbf{E}\|_F$; the squared $\sin\Theta$ term contributes the factor $\|\sin\Theta\|_F^2 \leq 2K\,C_L^2\|\mathbf{E}\|_2^2/(d_{\min}^2\delta_{\mathrm{Lap}}^2)$, hence the K^2 in the quadratic term of Eq. equation 5. Under the second small-error condition $K\,C_L\|\mathbf{E}\|_F/(d_{\min}\delta_{\mathrm{Lap}}^2) \leq 1$, the quadratic term is at most the linear one and the bound reduces to its linear form.

Step 4 (error budget). Lemma A.5 gives $\|\mathbf{E}\|_F = O\big(NB(rV\beta_e/w_{\min} + \varepsilon_r)\big) + \kappa_{\mathbf{C}}\|\widetilde{\mathbf{G}}_{\mathcal{A}} - \mathbf{G}^\star\|_F$, where the anchor-level error is controlled by Theorem 5.2 (the mask-noise and bias layers, with $\|\widetilde{\mathbf{G}}_{\mathcal{A}} - \mathbf{G}^\star\|_F \le m(\omega_1 + \beta_e)$). Since the aggregation is a single closed-form pass with no iterative solver, there is no optimization-error layer. As $m \to N$, the anchor set exhausts the sample and the extension and bias remainders vanish; the mask-noise layer ω_1 does not vanish with m and decays only in V, so the bound approaches the optimal normalized cut up to the factor $(1 + \varepsilon_0)$ and the residual mask noise, in line with the consistency framework of von Luxburg et al. (2008). □

B Positioning Table

Table 3: Positioning of SCoGF relative to representative methods. "Robust to hyperparams" reports whether accuracy degrades gracefully under mis-set hyperparameter values, with no collapse observed in the tuning sweep of Section 6.1, followed by the number of hyperparameters tuned per dataset under that protocol in parentheses; entries with (0) run with fixed defaults and take no tuned parameter. The anchor number m follows the uniform size rule $m = \min(500, N)$ for every anchor-based method (SCoGF, RISE, JVILG, LMVSC, IMVC-CBG) and is never tuned, so it is excluded from all counts.

Method	Incomplete views	Scalability	Robust to hyperparams (# tuned)	Theory
PVC (Li et al., 2014)	✓	×	✓(0)	×
DAIMC (Hu & Chen, 2018)	✓	×	× (1)	×
AWSR (Feng et al., 2024)	✓	×	× (2)	×
RISE (Wang et al., 2025b)	✓	✓	× (1)	×
JVILG (Wang et al., 2025a)	✓	✓	× (2)	×
RMSC (Xia et al., 2014)	×	×	× (1)	×
AMGL (Nie et al., 2016)	×	×	✓(0)	×
MCAN (Nie et al., 2017)	×	×	✓(1)	×
LMVSC (Kang et al., 2020)	×	✓	× (1)	×
IMVC-CBG (Wang et al., 2022)	✓	✓	× (1)	×
SCoGF (ours)	✓	✓	✓(1)	✓

C Full Results

Tables 4–8 report the complete results behind Table 1, one table per dataset ordered by sample size from small to large; within each table, baselines are listed alphabetically and our method last. The tables cover ACC and NMI at all four symmetric missing rates and under view-dependent missingness. Entries are means over 30 trials with standard deviations as superscripts, the best entry per column is in bold red and the second best in red, and methods excluded by the complexity protocol (Section 6.1) are omitted from the table rather than marked.

D Tuning Sweep Results

Tables 9–13 report the complete tuning grids behind the protocol of Section 6.1, one table per dataset ordered by sample size. Every configuration is evaluated at symmetric η=0.5 over 3 runs, and the configuration with the best mean ACC (in bold) is frozen and applied to all missing rates in the main experiments. Throughout the sweep tables, only methods with a non-degenerate grid are shown, the anchor number follows the uniform size rule $m = \min(500, N)$ of Section 6.1 and, together with internal constants, is omitted from the config strings, and parameter-free methods (ZeroFill, Concat, BestSV, KernelAdd, AMGL, MKKM, PVC) take no part in the sweep. The tables give direct evidence for the robustness claims of Sections 2 and 6.3. SCoGF's accuracy varies by at most 0.03 across its entire grid on every dataset. By contrast, RISE loses two thirds of its MSRC-v5 accuracy when β leaves its plateau, JVILG returns trivial or near-trivial solutions on MSRC-v5

Table 4: ACC and NMI on MSRC-v5 (reporting conventions as in Appendix C).

	η=0		η=0.3		η=0.5		η=0.7		view-dep.	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
AMGL	$0.644^{\pm 0.007}$	$0.577^{\pm 0.009}$	$0.606^{\pm 0.048}$	$0.513^{\pm 0.040}$	$0.469^{\pm 0.052}$	$0.348^{\pm 0.046}$	$0.327^{\pm 0.030}$	$0.194^{\pm 0.030}$	$0.643^{\pm 0.069}$	$0.550^{\pm 0.061}$
AWSR	$0.547^{\pm 0.023}$	$0.530^{\pm 0.033}$	$0.500^{\pm 0.054}$	$0.380^{\pm 0.052}$	$0.407^{\pm 0.052}$	$0.267^{\pm 0.055}$	$0.275^{\pm 0.026}$	$0.122^{\pm 0.028}$	$0.491^{\pm 0.052}$	$0.386^{\pm 0.066}$
BestSV	$0.712^{\pm 0.025}$	$0.582^{\pm 0.021}$	$0.536^{\pm 0.037}$	$0.426^{\pm 0.034}$	$0.429^{\pm 0.026}$	$0.342^{\pm 0.026}$	$0.343^{\pm 0.027}$	$0.267^{\pm 0.024}$	$0.609^{\pm 0.055}$	$0.493^{\pm 0.035}$
Concat	**0.843**$^{\pm 0.011}$	**0.737**$^{\pm 0.011}$	$0.700^{\pm 0.051}$	$0.578^{\pm 0.046}$	$0.489^{\pm 0.043}$	$0.357^{\pm 0.048}$	$0.365^{\pm 0.033}$	$0.239^{\pm 0.044}$	$0.719^{\pm 0.060}$	$0.598^{\pm 0.062}$
DAIMC	$0.705^{\pm 0.057}$	$0.621^{\pm 0.045}$	$0.632^{\pm 0.069}$	$0.539^{\pm 0.062}$	$0.436^{\pm 0.074}$	$0.334^{\pm 0.077}$	$0.286^{\pm 0.064}$	$0.159^{\pm 0.067}$	$0.626^{\pm 0.091}$	$0.535^{\pm 0.088}$
IMVC-CBG	$0.397^{\pm 0.019}$	$0.288^{\pm 0.019}$	$0.427^{\pm 0.044}$	$0.304^{\pm 0.040}$	$0.416^{\pm 0.043}$	$0.284^{\pm 0.041}$	$0.334^{\pm 0.034}$	$0.214^{\pm 0.030}$	$0.358^{\pm 0.071}$	$0.236^{\pm 0.059}$
JVILG	$0.143^{\pm 0.000}$	$0.000^{\pm 0.000}$	$0.150^{\pm 0.017}$	$0.009^{\pm 0.018}$	$0.155^{\pm 0.028}$	$0.015^{\pm 0.032}$	$0.223^{\pm 0.039}$	$0.086^{\pm 0.043}$	$0.148^{\pm 0.012}$	$0.005^{\pm 0.011}$
KernelAdd	$0.734^{\pm 0.031}$	$0.671^{\pm 0.029}$	$0.614^{\pm 0.042}$	$0.521^{\pm 0.038}$	$0.471^{\pm 0.055}$	$0.348^{\pm 0.046}$	$0.339^{\pm 0.036}$	$0.206^{\pm 0.030}$	$0.680^{\pm 0.058}$	$0.576^{\pm 0.056}$
LMSC	$0.609^{\pm 0.066}$	$0.457^{\pm 0.046}$	$0.421^{\pm 0.057}$	$0.256^{\pm 0.051}$	$0.329^{\pm 0.031}$	$0.161^{\pm 0.028}$	$0.276^{\pm 0.022}$	$0.112^{\pm 0.022}$	$0.467^{\pm 0.066}$	$0.303^{\pm 0.054}$
LMVSC	$0.253^{\pm 0.021}$	$0.088^{\pm 0.017}$	$0.172^{\pm 0.007}$	$0.065^{\pm 0.008}$	$0.166^{\pm 0.005}$	$0.055^{\pm 0.003}$	$0.165^{\pm 0.003}$	$0.054^{\pm 0.003}$	$0.168^{\pm 0.006}$	$0.059^{\pm 0.008}$
MCAN	$0.779^{\pm 0.004}$	$0.714^{\pm 0.004}$	$0.705^{\pm 0.050}$	$0.609^{\pm 0.050}$	$0.505^{\pm 0.067}$	$0.420^{\pm 0.053}$	$0.363^{\pm 0.038}$	$0.273^{\pm 0.035}$	$0.730^{\pm 0.052}$	$0.642^{\pm 0.051}$
MKKM	$0.642^{\pm 0.008}$	$0.584^{\pm 0.014}$	$0.406^{\pm 0.045}$	$0.268^{\pm 0.048}$	$0.318^{\pm 0.031}$	$0.167^{\pm 0.030}$	$0.265^{\pm 0.018}$	$0.105^{\pm 0.026}$	$0.390^{\pm 0.049}$	$0.267^{\pm 0.060}$
PVC	$0.763^{\pm 0.053}$	$0.676^{\pm 0.038}$	$0.723^{\pm 0.061}$	$0.606^{\pm 0.046}$	$0.608^{\pm 0.074}$	$0.496^{\pm 0.061}$	$0.442^{\pm 0.057}$	$0.313^{\pm 0.049}$	$0.727^{\pm 0.070}$	$0.617^{\pm 0.056}$
RISE	$0.698^{\pm 0.053}$	$0.639^{\pm 0.036}$	$0.636^{\pm 0.047}$	$0.579^{\pm 0.032}$	$0.568^{\pm 0.053}$	$0.487^{\pm 0.051}$	$0.449^{\pm 0.047}$	$0.352^{\pm 0.051}$	$0.636^{\pm 0.058}$	$0.570^{\pm 0.044}$
RMSC	$0.627^{\pm 0.015}$	$0.532^{\pm 0.013}$	$0.545^{\pm 0.048}$	$0.423^{\pm 0.052}$	$0.405^{\pm 0.041}$	$0.278^{\pm 0.045}$	$0.314^{\pm 0.045}$	$0.176^{\pm 0.046}$	$0.503^{\pm 0.077}$	$0.387^{\pm 0.088}$
ZeroFill	$0.819^{\pm 0.015}$	$0.699^{\pm 0.013}$	$0.713^{\pm 0.043}$	$0.587^{\pm 0.037}$	$0.491^{\pm 0.059}$	$0.364^{\pm 0.052}$	$0.362^{\pm 0.034}$	$0.236^{\pm 0.030}$	$0.717^{\pm 0.066}$	$0.597^{\pm 0.061}$
SCoGF (ours)	$0.765^{\pm 0.002}$	$0.718^{\pm 0.002}$	**0.758**$^{\pm 0.015}$	**0.699**$^{\pm 0.023}$	**0.733**$^{\pm 0.021}$	**0.637**$^{\pm 0.029}$	**0.524**$^{\pm 0.054}$	**0.410**$^{\pm 0.048}$	**0.767**$^{\pm 0.021}$	**0.700**$^{\pm 0.028}$

Table 5: ACC and NMI on 100Leaves (reporting conventions as in Appendix C).

	η=0		η=0.3		η=0.5		η=0.7		view-dep.	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
AMGL	$0.855^{\pm 0.013}$	$0.934^{\pm 0.004}$	$0.535^{\pm 0.015}$	$0.746^{\pm 0.006}$	$0.400^{\pm 0.011}$	$0.678^{\pm 0.006}$	$0.340^{\pm 0.009}$	$0.646^{\pm 0.005}$	$0.498^{\pm 0.030}$	$0.733^{\pm 0.014}$
AWSR	$0.807^{\pm 0.015}$	$0.911^{\pm 0.006}$	$0.504^{\pm 0.018}$	$0.708^{\pm 0.010}$	$0.381^{\pm 0.010}$	$0.641^{\pm 0.008}$	$0.321^{\pm 0.007}$	$0.608^{\pm 0.006}$	$0.468^{\pm 0.052}$	$0.695^{\pm 0.030}$
BestSV	$0.608^{\pm 0.013}$	$0.795^{\pm 0.004}$	$0.431^{\pm 0.011}$	$0.623^{\pm 0.007}$	$0.336^{\pm 0.011}$	$0.528^{\pm 0.008}$	$0.256^{\pm 0.006}$	$0.442^{\pm 0.008}$	$0.494^{\pm 0.050}$	$0.703^{\pm 0.037}$
Concat	$0.913^{\pm 0.013}$	$0.965^{\pm 0.004}$	$0.549^{\pm 0.017}$	$0.750^{\pm 0.008}$	$0.391^{\pm 0.008}$	$0.668^{\pm 0.004}$	$0.332^{\pm 0.010}$	$0.638^{\pm 0.005}$	$0.521^{\pm 0.035}$	$0.743^{\pm 0.018}$
DAIMC	$0.769^{\pm 0.020}$	$0.914^{\pm 0.008}$	$0.222^{\pm 0.023}$	$0.527^{\pm 0.023}$	$0.196^{\pm 0.012}$	$0.474^{\pm 0.019}$	$0.203^{\pm 0.011}$	$0.484^{\pm 0.016}$	$0.247^{\pm 0.082}$	$0.561^{\pm 0.066}$
IMVC-CBG	$0.425^{\pm 0.012}$	$0.748^{\pm 0.005}$	$0.344^{\pm 0.014}$	$0.627^{\pm 0.009}$	$0.290^{\pm 0.011}$	$0.575^{\pm 0.007}$	$0.263^{\pm 0.010}$	$0.559^{\pm 0.008}$	$0.278^{\pm 0.036}$	$0.550^{\pm 0.032}$
JVILG	$0.316^{\pm 0.000}$	$0.672^{\pm 0.000}$	$0.311^{\pm 0.026}$	$0.659^{\pm 0.029}$	$0.310^{\pm 0.028}$	$0.662^{\pm 0.030}$	$0.333^{\pm 0.037}$	**0.683**$^{\pm 0.035}$	$0.222^{\pm 0.031}$	$0.543^{\pm 0.034}$
KernelAdd	$0.847^{\pm 0.014}$	$0.929^{\pm 0.005}$	$0.532^{\pm 0.015}$	$0.744^{\pm 0.008}$	$0.405^{\pm 0.012}$	$0.679^{\pm 0.007}$	$0.344^{\pm 0.010}$	$0.648^{\pm 0.005}$	$0.528^{\pm 0.048}$	$0.746^{\pm 0.025}$
LMSC	$0.816^{\pm 0.012}$	$0.907^{\pm 0.004}$	$0.489^{\pm 0.014}$	$0.707^{\pm 0.008}$	$0.366^{\pm 0.012}$	$0.639^{\pm 0.005}$	$0.306^{\pm 0.009}$	$0.604^{\pm 0.006}$	$0.434^{\pm 0.056}$	$0.680^{\pm 0.032}$
LMVSC	$0.542^{\pm 0.022}$	$0.812^{\pm 0.012}$	$0.268^{\pm 0.019}$	$0.499^{\pm 0.023}$	$0.153^{\pm 0.009}$	$0.332^{\pm 0.021}$	$0.096^{\pm 0.009}$	$0.230^{\pm 0.019}$	$0.254^{\pm 0.059}$	$0.516^{\pm 0.069}$
MCAN	**0.960**$^{\pm 0.004}$	**0.983**$^{\pm 0.001}$	$0.565^{\pm 0.017}$	$0.759^{\pm 0.006}$	$0.400^{\pm 0.011}$	$0.682^{\pm 0.006}$	$0.345^{\pm 0.011}$	$0.659^{\pm 0.005}$	$0.554^{\pm 0.018}$	$0.766^{\pm 0.010}$
MKKM	$0.854^{\pm 0.015}$	$0.935^{\pm 0.004}$	$0.523^{\pm 0.013}$	$0.742^{\pm 0.007}$	$0.380^{\pm 0.011}$	$0.669^{\pm 0.007}$	$0.324^{\pm 0.009}$	$0.641^{\pm 0.006}$	$0.464^{\pm 0.040}$	$0.717^{\pm 0.019}$
PVC	$0.781^{\pm 0.016}$	$0.918^{\pm 0.006}$	$0.379^{\pm 0.018}$	$0.647^{\pm 0.012}$	$0.255^{\pm 0.012}$	$0.546^{\pm 0.011}$	$0.230^{\pm 0.011}$	$0.526^{\pm 0.012}$	$0.374^{\pm 0.049}$	$0.657^{\pm 0.030}$
RISE	$0.793^{\pm 0.018}$	$0.919^{\pm 0.007}$	$0.551^{\pm 0.014}$	$0.726^{\pm 0.009}$	$0.435^{\pm 0.013}$	$0.635^{\pm 0.012}$	$0.361^{\pm 0.013}$	$0.587^{\pm 0.012}$	$0.480^{\pm 0.025}$	$0.684^{\pm 0.018}$
RMSC	$0.861^{\pm 0.012}$	$0.937^{\pm 0.004}$	$0.492^{\pm 0.018}$	$0.715^{\pm 0.008}$	$0.343^{\pm 0.010}$	$0.638^{\pm 0.006}$	$0.303^{\pm 0.008}$	$0.617^{\pm 0.006}$	$0.397^{\pm 0.046}$	$0.670^{\pm 0.025}$
ZeroFill	$0.907^{\pm 0.011}$	$0.964^{\pm 0.004}$	$0.552^{\pm 0.016}$	$0.751^{\pm 0.007}$	$0.390^{\pm 0.010}$	$0.668^{\pm 0.005}$	$0.327^{\pm 0.010}$	$0.634^{\pm 0.005}$	$0.519^{\pm 0.036}$	$0.743^{\pm 0.019}$
SCoGF (ours)	$0.868^{\pm 0.011}$	$0.941^{\pm 0.004}$	**0.676**$^{\pm 0.019}$	**0.820**$^{\pm 0.008}$	**0.533**$^{\pm 0.016}$	**0.736**$^{\pm 0.009}$	**0.390**$^{\pm 0.014}$	$0.663^{\pm 0.007}$	**0.636**$^{\pm 0.030}$	**0.802**$^{\pm 0.016}$

across its four configurations, DAIMC fails on ALOI for two of its three α values, and IMVC-CBG loses over 40% of its MSRC-v5 accuracy across a two-decade change of its graph parameter λ.

E Limitations and Future Work

The stability theory idealizes anchors as uniform draws (Appendix F); a dedicated analysis through coherence conditions is open. The budgets of Theorem 5.2 are existential rather than operational at practical feature dimensions, and the analysis assumes MCAR missingness, so MNAR mechanisms are an important direction. The consistency bound degrades with $1/(d_{\min}^2 \delta_{\mathrm{Lap}}^2)$ and is loose at extreme missing rates; explicit sample-complexity bounds relating N, m, and the missing rate are a natural next step. Incorporating pretrained deep features as inputs is a promising extension.

F Anchor Selection Ablation

Assumption 3(a) covers uniform anchor draws, while the implementation selects anchors deterministically by k-means (Section 4). To close this gap empirically, we rerun SCoGF on every dataset under the protocol of Section 6.1 with the anchor rule as the only variable, comparing the deterministic k-means rule against uniform random draws. Tables 14 and 15 report mean$\pm$std ACC and NMI; both rules use the same 30 mask draws ($m = \min(500, N)$ and the frozen r as in the main experiments), so the trials are paired across anchor rules. On MSRC-v5, $m = N = 210$, so the random rule draws no subsample at all and selects the entire sample (hence its zero variance at η=0,

Table 6: ACC and NMI on mfeat (reporting conventions as in Appendix C).

	η=0		η=0.3		η=0.5		η=0.7		view-dep.	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
AMGL	$0.951^{\pm 0.002}$	$0.896^{\pm 0.003}$	$0.722^{\pm 0.010}$	$0.682^{\pm 0.010}$	$0.524^{\pm 0.013}$	$0.536^{\pm 0.015}$	$0.327^{\pm 0.022}$	$0.294^{\pm 0.021}$	$0.748^{\pm 0.058}$	$0.710^{\pm 0.040}$
BestSV	$0.754^{\pm 0.012}$	$0.697^{\pm 0.010}$	$0.541^{\pm 0.022}$	$0.488^{\pm 0.017}$	$0.403^{\pm 0.015}$	$0.369^{\pm 0.016}$	$0.297^{\pm 0.014}$	$0.272^{\pm 0.015}$	$0.645^{\pm 0.041}$	$0.589^{\pm 0.032}$
Concat	$0.802^{\pm 0.016}$	$0.743^{\pm 0.011}$	$0.772^{\pm 0.021}$	$0.689^{\pm 0.017}$	$0.574^{\pm 0.049}$	$0.522^{\pm 0.030}$	$0.346^{\pm 0.027}$	$0.320^{\pm 0.025}$	$0.790^{\pm 0.047}$	$0.710^{\pm 0.038}$
DAIMC	$0.790^{\pm 0.060}$	$0.749^{\pm 0.033}$	$0.637^{\pm 0.100}$	$0.664^{\pm 0.044}$	$0.231^{\pm 0.137}$	$0.213^{\pm 0.190}$	$0.140^{\pm 0.012}$	$0.077^{\pm 0.015}$	$0.743^{\pm 0.065}$	$0.708^{\pm 0.036}$
IMVC-CBG	$0.803^{\pm 0.052}$	$0.752^{\pm 0.011}$	$0.605^{\pm 0.033}$	$0.504^{\pm 0.018}$	$0.453^{\pm 0.034}$	$0.378^{\pm 0.016}$	$0.306^{\pm 0.014}$	$0.252^{\pm 0.013}$	$0.560^{\pm 0.195}$	$0.497^{\pm 0.168}$
JVILG	$0.693^{\pm 0.000}$	$0.794^{\pm 0.000}$	$0.607^{\pm 0.087}$	$0.715^{\pm 0.060}$	$0.538^{\pm 0.079}$	$0.656^{\pm 0.072}$	$0.194^{\pm 0.090}$	$0.256^{\pm 0.217}$	$0.605^{\pm 0.106}$	$0.720^{\pm 0.074}$
KernelAdd	$0.822^{\pm 0.000}$	$0.821^{\pm 0.000}$	$0.791^{\pm 0.027}$	$0.730^{\pm 0.014}$	$0.544^{\pm 0.022}$	$0.544^{\pm 0.019}$	$0.319^{\pm 0.024}$	$0.281^{\pm 0.026}$	$0.845^{\pm 0.051}$	$0.784^{\pm 0.039}$
LMSC	$0.708^{\pm 0.053}$	$0.618^{\pm 0.033}$	$0.587^{\pm 0.052}$	$0.454^{\pm 0.041}$	$0.402^{\pm 0.044}$	$0.276^{\pm 0.025}$	$0.263^{\pm 0.020}$	$0.152^{\pm 0.018}$	$0.618^{\pm 0.079}$	$0.505^{\pm 0.061}$
LMVSC	$0.833^{\pm 0.006}$	$0.899^{\pm 0.009}$	$0.729^{\pm 0.049}$	$0.752^{\pm 0.022}$	$0.206^{\pm 0.056}$	$0.216^{\pm 0.100}$	$0.103^{\pm 0.001}$	$0.010^{\pm 0.001}$	$0.802^{\pm 0.032}$	$0.826^{\pm 0.023}$
MCAN	**0.975**$^{\pm 0.000}$	**0.945**$^{\pm 0.000}$	$0.784^{\pm 0.045}$	$0.744^{\pm 0.018}$	$0.600^{\pm 0.029}$	$0.625^{\pm 0.021}$	$0.437^{\pm 0.034}$	$0.439^{\pm 0.032}$	$0.826^{\pm 0.097}$	$0.823^{\pm 0.046}$
MKKM	$0.648^{\pm 0.010}$	$0.646^{\pm 0.006}$	$0.471^{\pm 0.007}$	$0.459^{\pm 0.006}$	$0.327^{\pm 0.006}$	$0.283^{\pm 0.007}$	$0.214^{\pm 0.016}$	$0.144^{\pm 0.018}$	$0.463^{\pm 0.135}$	$0.455^{\pm 0.148}$
PVC	$0.754^{\pm 0.061}$	$0.719^{\pm 0.030}$	$0.734^{\pm 0.051}$	$0.687^{\pm 0.030}$	$0.656^{\pm 0.064}$	$0.576^{\pm 0.035}$	$0.412^{\pm 0.041}$	$0.329^{\pm 0.040}$	$0.718^{\pm 0.061}$	$0.679^{\pm 0.026}$
RISE	$0.954^{\pm 0.034}$	$0.917^{\pm 0.014}$	**0.921**$^{\pm 0.032}$	**0.858**$^{\pm 0.017}$	$0.763^{\pm 0.042}$	$0.686^{\pm 0.021}$	$0.549^{\pm 0.040}$	$0.466^{\pm 0.025}$	**0.913**$^{\pm 0.053}$	**0.863**$^{\pm 0.036}$
ZeroFill	$0.813^{\pm 0.030}$	$0.758^{\pm 0.017}$	$0.766^{\pm 0.037}$	$0.689^{\pm 0.024}$	$0.571^{\pm 0.047}$	$0.523^{\pm 0.030}$	$0.364^{\pm 0.026}$	$0.330^{\pm 0.025}$	$0.776^{\pm 0.053}$	$0.706^{\pm 0.038}$
SCoGF (ours)	$0.830^{\pm 0.022}$	$0.865^{\pm 0.003}$	$0.846^{\pm 0.024}$	$0.849^{\pm 0.011}$	**0.831**$^{\pm 0.024}$	**0.808**$^{\pm 0.013}$	**0.803**$^{\pm 0.040}$	**0.721**$^{\pm 0.015}$	$0.848^{\pm 0.019}$	$0.847^{\pm 0.014}$

Table 7: ACC and NMI on Caltech101-20 (reporting conventions as in Appendix C).

	η=0		η=0.3		η=0.5		η=0.7		view-dep.	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
AMGL	$0.362^{\pm 0.015}$	$0.521^{\pm 0.006}$	$0.283^{\pm 0.012}$	$0.396^{\pm 0.012}$	$0.242^{\pm 0.010}$	$0.315^{\pm 0.011}$	$0.186^{\pm 0.011}$	$0.226^{\pm 0.009}$	$0.310^{\pm 0.017}$	$0.425^{\pm 0.011}$
BestSV	$0.418^{\pm 0.014}$	$0.587^{\pm 0.025}$	$0.339^{\pm 0.010}$	$0.442^{\pm 0.016}$	$0.337^{\pm 0.011}$	$0.352^{\pm 0.015}$	$0.337^{\pm 0.009}$	$0.261^{\pm 0.012}$	$0.368^{\pm 0.018}$	$0.485^{\pm 0.078}$
Concat	$0.411^{\pm 0.028}$	$0.596^{\pm 0.013}$	$0.334^{\pm 0.020}$	$0.491^{\pm 0.012}$	$0.257^{\pm 0.016}$	$0.388^{\pm 0.013}$	$0.205^{\pm 0.007}$	$0.301^{\pm 0.009}$	$0.339^{\pm 0.051}$	$0.502^{\pm 0.042}$
DAIMC	$0.464^{\pm 0.021}$	$0.613^{\pm 0.015}$	$0.458^{\pm 0.015}$	$0.590^{\pm 0.013}$	$0.441^{\pm 0.022}$	$0.530^{\pm 0.017}$	$0.423^{\pm 0.023}$	$0.369^{\pm 0.019}$	$0.462^{\pm 0.020}$	$0.583^{\pm 0.033}$
IMVC-CBG	$0.481^{\pm 0.033}$	$0.427^{\pm 0.030}$	$0.454^{\pm 0.028}$	$0.364^{\pm 0.020}$	$0.372^{\pm 0.023}$	$0.292^{\pm 0.012}$	$0.298^{\pm 0.025}$	$0.226^{\pm 0.009}$	$0.438^{\pm 0.055}$	$0.357^{\pm 0.055}$
JVILG	**0.750**$^{\pm 0.000}$	**0.820**$^{\pm 0.000}$	**0.660**$^{\pm 0.077}$	**0.766**$^{\pm 0.047}$	**0.628**$^{\pm 0.066}$	**0.726**$^{\pm 0.034}$	**0.577**$^{\pm 0.053}$	**0.634**$^{\pm 0.053}$	**0.606**$^{\pm 0.073}$	**0.706**$^{\pm 0.084}$
KernelAdd	$0.365^{\pm 0.017}$	$0.526^{\pm 0.006}$	$0.298^{\pm 0.015}$	$0.414^{\pm 0.010}$	$0.252^{\pm 0.014}$	$0.314^{\pm 0.014}$	$0.194^{\pm 0.011}$	$0.235^{\pm 0.008}$	$0.325^{\pm 0.019}$	$0.452^{\pm 0.018}$
LMSC	$0.390^{\pm 0.014}$	$0.508^{\pm 0.013}$	$0.319^{\pm 0.015}$	$0.372^{\pm 0.013}$	$0.240^{\pm 0.015}$	$0.275^{\pm 0.013}$	$0.175^{\pm 0.008}$	$0.201^{\pm 0.009}$	$0.320^{\pm 0.048}$	$0.400^{\pm 0.050}$
LMVSC	$0.583^{\pm 0.022}$	$0.658^{\pm 0.023}$	$0.576^{\pm 0.057}$	$0.491^{\pm 0.025}$	$0.352^{\pm 0.033}$	$0.047^{\pm 0.046}$	$0.339^{\pm 0.001}$	$0.022^{\pm 0.002}$	$0.563^{\pm 0.047}$	$0.546^{\pm 0.032}$
MCAN	$0.506^{\pm 0.019}$	$0.589^{\pm 0.010}$	$0.392^{\pm 0.026}$	$0.461^{\pm 0.012}$	$0.304^{\pm 0.019}$	$0.368^{\pm 0.010}$	$0.238^{\pm 0.019}$	$0.306^{\pm 0.011}$	$0.430^{\pm 0.032}$	$0.500^{\pm 0.021}$
MKKM	$0.317^{\pm 0.008}$	$0.494^{\pm 0.006}$	$0.266^{\pm 0.013}$	$0.362^{\pm 0.008}$	$0.230^{\pm 0.009}$	$0.286^{\pm 0.009}$	$0.180^{\pm 0.009}$	$0.226^{\pm 0.009}$	$0.269^{\pm 0.023}$	$0.366^{\pm 0.028}$
PVC	$0.336^{\pm 0.017}$	$0.516^{\pm 0.011}$	$0.344^{\pm 0.020}$	$0.487^{\pm 0.014}$	$0.348^{\pm 0.035}$	$0.425^{\pm 0.017}$	$0.317^{\pm 0.031}$	$0.311^{\pm 0.017}$	$0.344^{\pm 0.028}$	$0.494^{\pm 0.025}$
RISE	$0.572^{\pm 0.007}$	$0.642^{\pm 0.014}$	$0.553^{\pm 0.026}$	$0.565^{\pm 0.014}$	$0.483^{\pm 0.038}$	$0.460^{\pm 0.012}$	$0.371^{\pm 0.040}$	$0.308^{\pm 0.013}$	$0.557^{\pm 0.030}$	$0.585^{\pm 0.020}$
ZeroFill	$0.395^{\pm 0.022}$	$0.582^{\pm 0.010}$	$0.331^{\pm 0.020}$	$0.486^{\pm 0.011}$	$0.262^{\pm 0.015}$	$0.385^{\pm 0.012}$	$0.205^{\pm 0.008}$	$0.299^{\pm 0.010}$	$0.340^{\pm 0.048}$	$0.505^{\pm 0.039}$
SCoGF (ours)	$0.490^{\pm 0.024}$	$0.619^{\pm 0.007}$	$0.480^{\pm 0.041}$	$0.580^{\pm 0.016}$	$0.461^{\pm 0.036}$	$0.540^{\pm 0.013}$	$0.392^{\pm 0.035}$	$0.469^{\pm 0.013}$	$0.494^{\pm 0.051}$	$0.586^{\pm 0.021}$

where the pipeline is fully deterministic); the only residual difference from the k-means rule is that the latter uses learned anchor centers rather than the sample points themselves.

The k-means rule matches uniform draws at low missing rates and pulls ahead as the sample count and the missing rate grow, with the largest gap exactly where robustness matters most (ALOI at η=0.7: +0.19 ACC). On MSRC-v5 no anchor subsampling takes place (m=N), and the two rules agree to within trial noise (≤ 0.02 ACC), confirming that the gap on the larger datasets is attributable to anchor placement rather than the rule itself.

G Selected Hyperparameters

Table 16 lists the configuration selected by the protocol of Section 6.1 for every tuned method on every dataset. The anchor number is unified at $m = \min(500, N)$ for all anchor-based methods (SCoGF, RISE, JVILG, LMVSC, IMVC-CBG) and is therefore not listed; "–" marks dataset–method pairs excluded by the complexity protocol.

AMGL, MKKM, PVC, ZeroFill, Concat, BestSV, and KernelAdd are parameter-free under our protocol and are omitted from the table. LMVSC uses $\alpha = 0.1$ throughout; with m fixed by the size rule, α is its only tuned constant.

H Runtime Results

Table 8: ACC and NMI on ALOI (reporting conventions as in Appendix C).

	η=0		η=0.3		η=0.5		η=0.7		view-dep.	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
DAIMC	$0.447^{\pm 0.015}$	$0.726^{\pm 0.008}$	$0.054^{\pm 0.006}$	$0.171^{\pm 0.034}$	$0.084^{\pm 0.006}$	$0.235^{\pm 0.016}$	$0.120^{\pm 0.015}$	$0.303^{\pm 0.039}$	$0.092^{\pm 0.043}$	$0.282^{\pm 0.108}$
IMVC-CBG	$0.569^{\pm 0.014}$	$0.779^{\pm 0.006}$	$0.374^{\pm 0.013}$	$0.576^{\pm 0.009}$	$0.270^{\pm 0.010}$	$0.471^{\pm 0.006}$	$0.194^{\pm 0.007}$	$0.394^{\pm 0.009}$	$0.284^{\pm 0.055}$	$0.523^{\pm 0.072}$
JVILG	$0.377^{\pm 0.000}$	$0.701^{\pm 0.000}$	$0.383^{\pm 0.022}$	$0.725^{\pm 0.009}$	$0.403^{\pm 0.018}$	$0.740^{\pm 0.011}$	$0.365^{\pm 0.031}$	$\mathbf{0.690}^{\pm 0.025}$	$0.333^{\pm 0.039}$	$0.655^{\pm 0.038}$
LMVSC	$0.312^{\pm 0.012}$	$0.624^{\pm 0.015}$	$0.198^{\pm 0.011}$	$0.436^{\pm 0.018}$	$0.127^{\pm 0.009}$	$0.293^{\pm 0.026}$	$0.053^{\pm 0.009}$	$0.123^{\pm 0.032}$	$0.198^{\pm 0.034}$	$0.465^{\pm 0.029}$
PVC	$0.514^{\pm 0.018}$	$0.762^{\pm 0.008}$	$0.361^{\pm 0.017}$	$0.604^{\pm 0.008}$	$0.246^{\pm 0.012}$	$0.495^{\pm 0.009}$	$0.172^{\pm 0.006}$	$0.420^{\pm 0.007}$	$0.367^{\pm 0.030}$	$0.622^{\pm 0.022}$
RISE	$0.652^{\pm 0.014}$	$0.799^{\pm 0.007}$	$0.555^{\pm 0.013}$	$0.697^{\pm 0.007}$	$0.477^{\pm 0.013}$	$0.622^{\pm 0.007}$	$0.411^{\pm 0.013}$	$0.567^{\pm 0.010}$	$0.533^{\pm 0.035}$	$0.690^{\pm 0.020}$
SCoGF (ours)	$\mathbf{0.842}^{\pm 0.011}$	$\mathbf{0.889}^{\pm 0.003}$	$\mathbf{0.788}^{\pm 0.016}$	$\mathbf{0.844}^{\pm 0.005}$	$\mathbf{0.696}^{\pm 0.017}$	$\mathbf{0.767}^{\pm 0.007}$	$\mathbf{0.552}^{\pm 0.016}$	$0.668^{\pm 0.007}$	$\mathbf{0.767}^{\pm 0.019}$	$\mathbf{0.830}^{\pm 0.009}$

Table 9: Tuning sweep on MSRC-v5 (symmetric η=0.5; conventions as in Section D).

Method	Config	ACC
AWSR	γ=128, λ=0.1	$\mathbf{0.444}^{\pm 0.091}$
	γ=128, λ=0.3	$0.444^{\pm 0.091}$
	γ=8, λ=0.1	$0.317^{\pm 0.046}$
	γ=8, λ=0.3	$0.297^{\pm 0.026}$
DAIMC	α=0.1, β=0.1	$\mathbf{0.479}^{\pm 0.127}$
	α=1, β=0.1	$0.478^{\pm 0.027}$
	α=10, β=0.1	$0.441^{\pm 0.064}$
IMVC-CBG	λ=0.1	$\mathbf{0.444}^{\pm 0.043}$
	λ=1	$0.281^{\pm 0.025}$
	λ=10	$0.256^{\pm 0.027}$
JVILG	β=1, p=0.5	$0.143^{\pm 0.000}$
	β=1, p=1	$\mathbf{0.162}^{\pm 0.033}$
	β=10, p=0.5	$0.143^{\pm 0.000}$
	β=10, p=1	$0.151^{\pm 0.014}$
LMSC	λ=0.01	$0.378^{\pm 0.052}$
	λ=0.1	$0.379^{\pm 0.029}$
	λ=1	$\mathbf{0.383}^{\pm 0.033}$
LMVSC	α=0.1	$\mathbf{0.175}^{\pm 0.003}$
MCAN	k=10	$0.452^{\pm 0.050}$
	k=15	$0.470^{\pm 0.026}$
	k=20	$\mathbf{0.475}^{\pm 0.030}$
RISE	β=1	$0.217^{\pm 0.012}$
	β=10	$0.551^{\pm 0.076}$
	β=100	$\mathbf{0.619}^{\pm 0.021}$
RMSC	λ_e=0.01	$0.187^{\pm 0.017}$
	λ_e=0.1	$0.313^{\pm 0.011}$
	λ_e=1	$\mathbf{0.414}^{\pm 0.058}$
SCoGF	r=10	$0.714^{\pm 0.034}$
	r=15	$\mathbf{0.732}^{\pm 0.007}$
	r=5	$0.703^{\pm 0.065}$

Table 10: Tuning sweep on 100Leaves (symmetric η=0.5; conventions as in Section D).

Method	Config	ACC
AWSR	γ=128, λ=0.1	**0.390**$^{\pm 0.016}$
	γ=128, λ=0.3	0.389$^{\pm 0.018}$
	γ=8, λ=0.1	0.354$^{\pm 0.008}$
	γ=8, λ=0.3	0.357$^{\pm 0.015}$
DAIMC	α=0.1, β=0.1	0.194$^{\pm 0.022}$
	α=1, β=0.1	**0.207**$^{\pm 0.003}$
	α=10, β=0.1	0.194$^{\pm 0.014}$
IMVC-CBG	λ=0.1	0.237$^{\pm 0.009}$
	λ=1	**0.288**$^{\pm 0.021}$
	λ=10	0.255$^{\pm 0.010}$
JVILG	β=1, p=0.5	**0.308**$^{\pm 0.011}$
	β=1, p=1	0.171$^{\pm 0.003}$
	β=10, p=0.5	0.281$^{\pm 0.018}$
	β=10, p=1	0.149$^{\pm 0.010}$
LMSC	λ=0.01	0.372$^{\pm 0.005}$
	λ=0.1	**0.378**$^{\pm 0.004}$
	λ=1	0.372$^{\pm 0.003}$
LMVSC	α=0.1	**0.154**$^{\pm 0.010}$
MCAN	k=10	**0.395**$^{\pm 0.014}$
	k=15	0.392$^{\pm 0.016}$
	k=20	0.386$^{\pm 0.009}$
RISE	β=1	0.091$^{\pm 0.008}$
	β=10	0.390$^{\pm 0.020}$
	β=100	**0.442**$^{\pm 0.008}$
RMSC	λ_e=0.01	0.270$^{\pm 0.009}$
	λ_e=0.1	**0.347**$^{\pm 0.007}$
	λ_e=1	0.340$^{\pm 0.016}$
SCoGF	r=10	0.521$^{\pm 0.009}$
	r=15	0.510$^{\pm 0.017}$
	r=5	**0.536**$^{\pm 0.003}$

Table 11: Tuning sweep on mfeat (symmetric η=0.5; conventions as in Section D).

Method	Config	ACC
DAIMC	α=0.1, β=0.1	**0.307**$^{\pm 0.174}$
	α=1, β=0.1	0.109$^{\pm 0.000}$
	α=10, β=0.1	0.109$^{\pm 0.000}$
IMVC-CBG	λ=0.1	0.399$^{\pm 0.023}$
	λ=1	0.427$^{\pm 0.023}$
	λ=10	**0.490**$^{\pm 0.007}$
JVILG	β=1, p=0.5	0.514$^{\pm 0.199}$
	β=1, p=1	**0.643**$^{\pm 0.080}$
	β=10, p=0.5	0.531$^{\pm 0.301}$
	β=10, p=1	0.568$^{\pm 0.091}$
LMSC	λ=0.01	**0.395**$^{\pm 0.036}$
	λ=0.1	0.386$^{\pm 0.005}$
	λ=1	0.389$^{\pm 0.006}$
LMVSC	α=0.1	**0.193**$^{\pm 0.009}$
MCAN	k=10	0.549$^{\pm 0.014}$
	k=15	0.577$^{\pm 0.025}$
	k=20	**0.592**$^{\pm 0.013}$
RISE	β=1	0.138$^{\pm 0.032}$
	β=10	0.667$^{\pm 0.030}$
	β=100	**0.677**$^{\pm 0.029}$
SCoGF	r=10	0.844$^{\pm 0.012}$
	r=15	0.848$^{\pm 0.013}$
	r=5	**0.851**$^{\pm 0.010}$

Table 12: Tuning sweep on Caltech101-20 (symmetric η=0.5; conventions as in Section D).

Method	Config	ACC
DAIMC	α=0.1, β=0.1	0.427$^{\pm 0.004}$
	α=1, β=0.1	0.424$^{\pm 0.011}$
	α=10, β=0.1	**0.458**$^{\pm 0.030}$
IMVC-CBG	λ=0.1	**0.362**$^{\pm 0.024}$
	λ=1	0.283$^{\pm 0.028}$
	λ=10	0.336$^{\pm 0.021}$
JVILG	β=1, p=0.5	0.653$^{\pm 0.177}$
	β=1, p=1	**0.681**$^{\pm 0.043}$
	β=10, p=0.5	0.621$^{\pm 0.061}$
	β=10, p=1	0.626$^{\pm 0.070}$
LMSC	λ=0.01	0.234$^{\pm 0.007}$
	λ=0.1	0.232$^{\pm 0.002}$
	λ=1	**0.235**$^{\pm 0.001}$
LMVSC	α=0.1	**0.388**$^{\pm 0.079}$
MCAN	k=10	**0.299**$^{\pm 0.014}$
	k=15	0.297$^{\pm 0.010}$
	k=20	0.281$^{\pm 0.009}$
RISE	β=1	0.332$^{\pm 0.030}$
	β=10	**0.472**$^{\pm 0.009}$
	β=100	0.457$^{\pm 0.008}$
SCoGF	r=10	**0.472**$^{\pm 0.036}$
	r=15	0.465$^{\pm 0.051}$
	r=5	0.464$^{\pm 0.046}$

Table 13: Tuning sweep on ALOI (symmetric η=0.5; conventions as in Section D).

Method	Config	ACC
DAIMC	α=0.1, β=0.1	**0.090**$^{\pm 0.009}$
	α=1, β=0.1	0.000$^{\pm 0.000}$
	α=10, β=0.1	0.000$^{\pm 0.000}$
IMVC-CBG	λ=0.1	0.150$^{\pm 0.004}$
	λ=1	0.185$^{\pm 0.004}$
	λ=10	**0.251**$^{\pm 0.007}$
JVILG	β=1, p=0.5	0.189$^{\pm 0.028}$
	β=1, p=1	0.399$^{\pm 0.010}$
	β=10, p=0.5	0.204$^{\pm 0.014}$
	β=10, p=1	**0.413**$^{\pm 0.033}$
LMVSC	α=0.1	**0.130**$^{\pm 0.005}$
RISE	β=1	0.054$^{\pm 0.006}$
	β=10	0.448$^{\pm 0.015}$
	β=100	**0.486**$^{\pm 0.012}$
SCoGF	r=10	0.677$^{\pm 0.008}$
	r=15	**0.685**$^{\pm 0.029}$
	r=5	0.658$^{\pm 0.019}$

Table 14: Deterministic k-means anchors versus uniform random anchors, mean$\pm$std ACC over 30 paired trials under the protocol of Section 6.1; "view-dep." is the asymmetric missingness setting of the main experiments. NMI is reported in Table 15.

Dataset	Anchor rule	η=0	η=0.3	η=0.5	η=0.7	view-dep.
MSRC-v5	k-means	0.765$^{\pm 0.002}$	0.758$^{\pm 0.015}$	0.733$^{\pm 0.021}$	0.524$^{\pm 0.054}$	0.767$^{\pm 0.021}$
	random	0.767$^{\pm 0.000}$	0.771$^{\pm 0.016}$	0.740$^{\pm 0.020}$	0.538$^{\pm 0.062}$	0.772$^{\pm 0.021}$
100Leaves	k-means	0.868$^{\pm 0.011}$	0.676$^{\pm 0.019}$	0.533$^{\pm 0.016}$	0.390$^{\pm 0.014}$	0.636$^{\pm 0.030}$
	random	0.845$^{\pm 0.016}$	0.643$^{\pm 0.015}$	0.487$^{\pm 0.017}$	0.354$^{\pm 0.010}$	0.597$^{\pm 0.030}$
mfeat	k-means	0.830$^{\pm 0.022}$	0.846$^{\pm 0.024}$	0.831$^{\pm 0.024}$	0.803$^{\pm 0.040}$	0.848$^{\pm 0.019}$
	random	0.854$^{\pm 0.023}$	0.842$^{\pm 0.023}$	0.828$^{\pm 0.026}$	0.778$^{\pm 0.045}$	0.841$^{\pm 0.024}$
Caltech101-20	k-means	0.490$^{\pm 0.024}$	0.480$^{\pm 0.041}$	0.461$^{\pm 0.036}$	0.392$^{\pm 0.035}$	0.494$^{\pm 0.051}$
	random	0.489$^{\pm 0.045}$	0.451$^{\pm 0.040}$	0.411$^{\pm 0.027}$	0.361$^{\pm 0.030}$	0.458$^{\pm 0.045}$
ALOI	k-means	0.842$^{\pm 0.011}$	0.788$^{\pm 0.016}$	0.696$^{\pm 0.017}$	0.552$^{\pm 0.016}$	0.767$^{\pm 0.019}$
	random	0.834$^{\pm 0.010}$	0.738$^{\pm 0.016}$	0.581$^{\pm 0.017}$	0.360$^{\pm 0.015}$	0.719$^{\pm 0.021}$

Table 15: Same comparison as Table 14, mean$\pm$std NMI over the same 30 paired trials.

Dataset	Anchor rule	η=0	η=0.3	η=0.5	η=0.7	view-dep.
MSRC-v5	k-means	0.718$^{\pm 0.002}$	0.699$^{\pm 0.023}$	0.637$^{\pm 0.029}$	0.410$^{\pm 0.048}$	0.700$^{\pm 0.028}$
	random	0.719$^{\pm 0.000}$	0.717$^{\pm 0.017}$	0.654$^{\pm 0.033}$	0.425$^{\pm 0.061}$	0.707$^{\pm 0.029}$
100Leaves	k-means	0.941$^{\pm 0.004}$	0.820$^{\pm 0.008}$	0.736$^{\pm 0.009}$	0.663$^{\pm 0.007}$	0.802$^{\pm 0.016}$
	random	0.927$^{\pm 0.008}$	0.800$^{\pm 0.008}$	0.712$^{\pm 0.008}$	0.646$^{\pm 0.006}$	0.780$^{\pm 0.015}$
mfeat	k-means	0.865$^{\pm 0.003}$	0.849$^{\pm 0.011}$	0.808$^{\pm 0.013}$	0.721$^{\pm 0.015}$	0.847$^{\pm 0.014}$
	random	0.862$^{\pm 0.012}$	0.840$^{\pm 0.010}$	0.800$^{\pm 0.014}$	0.695$^{\pm 0.019}$	0.842$^{\pm 0.011}$
Caltech101-20	k-means	0.619$^{\pm 0.007}$	0.580$^{\pm 0.016}$	0.540$^{\pm 0.013}$	0.469$^{\pm 0.013}$	0.586$^{\pm 0.021}$
	random	0.593$^{\pm 0.015}$	0.568$^{\pm 0.013}$	0.525$^{\pm 0.013}$	0.429$^{\pm 0.016}$	0.572$^{\pm 0.020}$
ALOI	k-means	0.889$^{\pm 0.003}$	0.844$^{\pm 0.005}$	0.767$^{\pm 0.007}$	0.668$^{\pm 0.007}$	0.830$^{\pm 0.009}$
	random	0.883$^{\pm 0.004}$	0.802$^{\pm 0.008}$	0.696$^{\pm 0.008}$	0.572$^{\pm 0.007}$	0.798$^{\pm 0.011}$

Table 16: Selected hyperparameters per dataset (tuned at η=0.5, frozen across missing rates).

Method	MSRC-v5	100Leaves	mfeat	Caltech101-20	ALOI
SCoGF (r)	15	5	5	10	15
RISE (β)	100	100	100	10	100
JVILG (β, p)	1, 1	1, 0.5	1, 1	1, 1	10, 1
LMVSC (α)	0.1	0.1	0.1	0.1	0.1
DAIMC (α)	0.1	1.0	0.1	10	0.1
IMVC-CBG (λ)	0.1	1.0	10	0.1	10
MCAN (knn)	20	10	20	10	–
LMSC (λ)	1.0	0.1	0.01	1.0	–
RMSC (λ_e)	1.0	0.1	–	–	–
AWSR (γ, λ)	128, 0.1	128, 0.1	–	–	–

Table 17: Average wall-clock runtime in seconds per trial (symmetric $\eta = 0.5$), datasets ordered by sample size. "–" marks cells excluded by the complexity protocol.

Method	MSRC-v5	100Leaves	mfeat	Caltech101-20	ALOI
AMGL	0.13	1.59	1.75	3.03	–
AWSR	0.49	7.93	–	–	–
BestSV	0.33	63.92	2.01	3.55	–
Concat	0.06	0.70	0.35	0.88	–
DAIMC	0.52	23.00	1.09	7.00	123.86
IMVC-CBG	2.02	3.84	12.39	14.17	12.01
JVILG	0.35	8.13	4.92	7.73	40.13
KernelAdd	0.10	1.00	1.33	2.36	–
LMSC	0.26	4.26	3.05	10.18	–
LMVSC	1.11	12.73	30.63	44.23	105.76
MCAN	0.23	14.65	22.56	35.15	–
MKKM	0.12	8.97	12.89	14.22	–
PVC	0.19	1.37	1.25	6.95	14.47
RISE	0.55	9.20	4.10	8.62	33.52
RMSC	0.62	46.73	–	–	–
ZeroFill	0.05	0.64	0.26	0.65	–
SCoGF (ours)	0.81	1.23	2.18	8.19	13.90