IS A SMALL MATRIX EIGENDECOMPOSITION SUFFICIENT FOR SPECTRAL CLUSTERING?

Anonymous authors
Paper under double-blind review

ABSTRACT

Spectral clustering is effective but expensive at large scales because its standard formulation constructs an $n \times n$ affinity matrix and performs eigendecomposition. A central challenge is to retain informative cluster representations while reducing the matrix size to depend only on the number of clusters k. We address this challenge from a distributional perspective: each representative summarizes a distribution rather than an individual point. D-SPEC constructs an $n \times k$ point–distribution bipartite graph and reduces eigendecomposition to a $k \times k$ matrix. We establish exact partition recovery under explicit affinity-profile conditions, perturbation-to-clustering error guarantees, and finite-partition bounds for adaptively selected representatives. Extensive experiments on synthetic and real-world datasets demonstrate the effectiveness and scalability of this distributional construction.

1 INTRODUCTION

Clustering is an unsupervised learning method that does not require labeled data, so it plays an important role in many fields where labeling is difficult. Spectral clustering (von Luxburg, 2007; Shi & Malik, 2000), as one of the most widely used clustering algorithms, has solid theory and good clustering performance, and has been applied to many fields, such as image segmentation (Li et al., 2012), cell analysis (Zu et al., 2023), and multi-view clustering (Lu et al., 2023; Tang et al., 2023).

Spectral clustering has two key steps: constructing a similarity matrix and eigendecomposition with time complexities of $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$ (n is the number of points), respectively (von Luxburg, 2007; Li et al., 2022). The expensive time complexity limits the application of spectral clustering in large-scale data (Li et al., 2022; Huang et al., 2020; Xie et al., 2023; Macgregor, 2023; Zhang et al., 2026). In order to apply the superior clustering performance of spectral clustering to large-scale data, many efficient spectral clustering algorithms have been proposed in recent years. In order to reduce the high complexity of similarity matrix construction, the main idea is to construct a sparse graph, which not only reduces the time and memory of graph construction but also speeds up the subsequent eigendecomposition process (Spielman & Srivastava, 2011; Saturluri et al., 2011). Nyström approximation is a simple method to construct a sparse graph (Fowlkes et al., 2004; Musco & Musco, 2017; Yang et al., 2012; Chen & Cai, 2011), but the clustering performance is greatly affected by the sampling points, so the landmark method based on k-means is proposed as an improvement (Bouneffouf & Birol, 2015; Rafailidis et al., 2017; Huang et al., 2020; Li et al., 2022; Xie et al., 2023). These methods reduce the time of graph construction from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$. To reduce the time complexity of eigendecomposition, using transfer cut (Li et al., 2012; Huang et al., 2020; Li et al., 2022) on the constructed $n \times p$ sparse graph can achieve eigendecomposition with $\mathcal{O}(n)$ (Huang et al., 2020; Li et al., 2022) or $\mathcal{O}(p^3)$ (p is the number of landmarks) (Xie et al., 2023) time complexity.

These methods accelerate eigendecomposition by replacing a large graph with a smaller representation, but a limited point-landmark budget may fail to retain the cluster structure needed for clustering. *Can a compact distributional representation preserve this structure while accelerating spectral clustering?*

Landmark constructions typically use many more representative vertices than clusters to maintain clustering quality. This motivates a second question: *Can a distributional representation support effective spectral clustering with only k representative vertices and a $k \times k$ reduced eigenproblem?*

A distributional perspective provides a principled way to address these questions. D-SPEC summarizes each candidate cluster as a distribution and constructs an $n \times k$ graph. This perspective

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

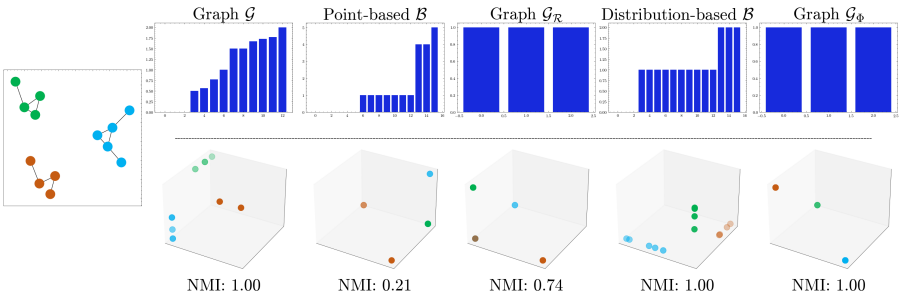


Figure 1: An illustrative comparison of distribution-based and point-based constructions. The distributional representation retains the three-cluster structure more accurately. The ideal graph has three connected components; the corresponding disconnected bipartite construction has three zero normalized-Laplacian eigenvalues, paired with unit eigenvalues of its reduced normalized affinity.

decouples the number of representative vertices from the number of points contributing to them. Our analysis proves exact partition recovery under explicit affinity-profile conditions, while experiments examine the representation beyond this idealized setting[1]. Figure 1 illustrates the original graph, point-based and distribution-based constructions, their spectra and embeddings, and the Normalized Mutual Information (NMI) (Strehl & Ghosh, 2002) obtained by k-means (Wu et al., 2008).

Using this idea, we propose distribution-based spectral clustering (D-SPEC). We evaluate it on synthetic and real datasets ranging from 3 hundred to 20 million points. These experiments support the effectiveness and scalability of distribution representatives with a $k \times k$ reduced eigenproblem.

We summarize our contributions below:

- Introducing a distributional perspective on scalable spectral clustering: representative vertices summarize distributions rather than individual landmarks, enabling an $n \times k$ point–distribution graph.
- Developing D-SPEC, whose bipartite spectral problem reduces exactly to a $k \times k$ matrix while each representative aggregates information from multiple points.
- Establishing explicit conditions for exact partition recovery, perturbation-to-clustering error guarantees, and finite-partition concentration valid for adaptively selected distribution representatives.
- Demonstrating strong clustering performance and scalability through extensive experiments on synthetic and real-world datasets.

2 Related Work

In this section, we provide a concise overview of the related work. Spectral clustering aims to partition the data points into k clusters using the spectrum of the graph Laplacian (von Luxburg, 2007). Spectral clustering first constructs a similarity matrix between points. After calculating the Laplacian matrix, it performs eigendecomposition to map the points to the $\mathbb{R}^k$ space. Finally, k-means is employed in the $\mathbb{R}^k$ space to complete the clustering. Although spectral clustering has good clustering performance and a theoretical basis, its high time complexity limits its extension to large-scale data. In recent years, many algorithms have been developed to accelerate it.

Nyström approximation-based spectral clustering (Fowlkes et al., 2004; Musco & Musco, 2017; Yang et al., 2012; Bouneffouf & Birol, 2015) first randomly selects a small subset to construct a similarity sub-matrix, which can efficiently construct the similarity matrix and eigendecomposition, but the clustering performance is greatly affected by the sampling points. The landmark-based methods employ k-means to improve the cluster performance (Chen & Cai, 2011; Rafailidis et al., 2017; Li et al., 2012; Huang et al., 2020; Li et al., 2022; Xie et al., 2023).

[1] See the reduced point-landmark-budget comparison in Appendix M.

In addition, some methods approximate the similarity of the original graph by random feature mapping, thereby efficiently constructing the similarity matrix and accelerating the eigendecomposition (Wu et al., 2018; He et al., 2019). Some methods calculate the approximate eigenvector by the power method without eigendecomposition (Macgregor, 2023).

3 Algorithm Description and Analysis

We now introduce D-SPEC[2], which constructs a bipartite graph between n points and k distributions and solves a $k \times k$ reduced eigenproblem. We analyze exact partition recovery, perturbation stability, and finite-partition approximation under explicit conditions.

3.1 Distribution-based Spectral Clustering

Given a dataset $\mathbb{X} \in \mathbb{R}^d = \{x_1, \ldots, x_n\}$, Spectral clustering (SC) constructs a fully connected undirected graph $\mathcal{G} = \{\mathbb{X}, \boldsymbol{W}\}$ with affinity matrix $\boldsymbol{W}$, where the element w_{ij} of $\boldsymbol{W}$ indicates the similarity between points x_i and x_j. The Laplacian matrix of $\mathcal{G}$ is $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D}$ is formed from row sums $d_{ii} = \sum_j w_{ij} > 0$ of the affinity being normalized; if self-loops are removed, they are removed before computing these sums. After eigendecomposition of the normalized Laplacian $\boldsymbol{N} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}}$, the eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ corresponding to the smallest k eigenvalues $\lambda_1, \ldots, \lambda_k$ are taken as the new feature $\mathbb{X}_{spec} \in \mathbb{R}^k$, Finally, clustering is completed using k-means on $\mathbb{X}_{spec} \in \mathbb{R}^k$.

Instead of a $n \times n$ fully connected graph, we construct an $n \times k$ bipartite graph $\mathcal{B} = \left\{\mathbb{X}, \mathbb{K}, \boldsymbol{W}^{\Phi}\right\}$ with similarity $\boldsymbol{W}^{\Phi}$ ($n \times k$), where the element w_{ij}^{Φ} represents the similarity between node x_i and subgraph $\mathcal{G}_j$ of $\mathcal{G}$, which is measured by the similarity between point x_i and distribution $P(\mathbb{C}_j)$, where $\mathbb{C}_j = \{x_1^j, \ldots, x_{nk}^j\}$ ($\bigcup_j \mathbb{C}_j = \mathbb{X}$) indicates the j-th cluster. The $P(\mathbb{C}_j)$ is obtained in two steps: i) acquire cluster $\mathbb{C}_j$, ii) compute $P(\mathbb{C}_j)$ based on $\mathbb{C}_j$.

Assumption 3.1. Let $\mathcal{G} = \{\mathbb{X}, \boldsymbol{W}\}$ be the fully connected undirected graph formed by $\mathbb{X}$ with affinity matrix $\boldsymbol{W}$, then $\hat{w}(x, y) > \hat{w}(x, z), \forall x \in \mathbb{C}_i, z \in \mathbb{C}_j, j \neq i, i \in \{1, \ldots, k\}$. Where $\hat{w}(x, y)$ is the similarity between point $x \in \mathbb{C}_i$ and its nearest neighbor $y \in \mathbb{C}_i$, and $\hat{w}(x, z)$ is the similarity between point x and z in other $\mathbb{C}_j (j \neq i)$.

Acquiring candidate distributions: D-SPEC uses the kernel feature map to compute affinities on a sample of p points. Assumption 3.1 motivates thresholding by favoring local within-cluster connections. Recovery with a single global threshold also requires that the sampled within-cluster graphs remain connected after removing between-cluster edges; Lemma C.1 gives the precise condition. The implementation retains off-diagonal affinities at least $\vartheta = \tau \max_{u,v} s_{uv}$ and takes the k largest connected components as candidate representatives. The maximum includes diagonal entries, while self-loops do not affect component membership. Corollary C.10 accounts for this relative threshold and its random scaling.

Assumption 3.1 may fail at noisy or boundary points; existing empirical diagnostics are reported in Appendix B. Sampling controls graph-construction cost, while aggregation limits representative contamination under the appendix conditions. Lemma C.11 quantifies sampling coverage, coverage alone is distinct from threshold connectivity. If fewer than k components are found (threshold is too small), the recovery guarantees for k representatives do not apply.

Computing $P(\mathbb{C}_j)$: After obtaining the approximate cluster $\mathbb{G}_j$ of $\mathbb{C}_j$, D-SPEC employs kernel mean embedding to estimate $\mathcal{G}_j$:

$$\hat{\Phi}(\mathcal{G}_j) \approx \hat{\Phi}(\mathcal{G}_j^b) = \frac{\sum_{x \in \mathcal{G}_j^b} \Phi(x)}{|\mathcal{G}_j^b|} = \frac{\sum_{x \in \mathbb{G}_j} \Phi(x)}{|\mathbb{G}_j|} \tag{1}$$

Here $\Phi(x)$ is the feature map in the RKHS. Kernel mean affinities can also be evaluated through kernel values, including for Gaussian kernels. We use the finite-dimensional feature map of the Isolation Distribution Kernel (IDK) (Ting et al., 2023b) to aggregate distributions explicitly and

[2]The code and datasets are available at `https://anonymous.4open.science/r/D-SPEC/`.

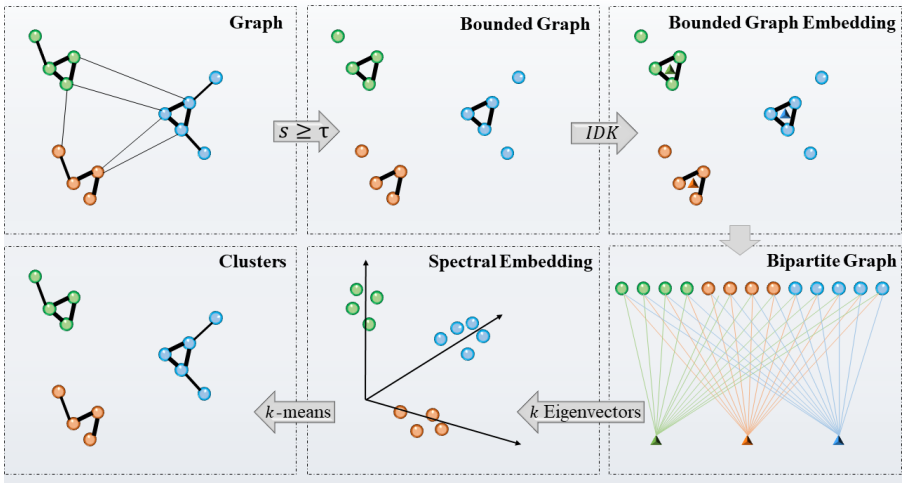


Figure 2: Illustration of D-SPEC. A Bounded graph is obtained by the threshold τ, and then the distribution of each subgraph is obtained using Equation 1. A bipartite graph of nodes and distributions is constructed, and then eigendecomposition is performed to map the data to $\mathbb{R}^k$. Finally, the k-means clustering is employed in $\mathbb{R}^k$.

compute $w_{ij}^{\Phi} = \langle \Phi(x_i), \hat{\Phi}(\mathcal{G}_j) \rangle$ efficiently. Theorem C.9 quantifies the approximation even when representatives are selected from the same random features.

Algorithm 1 D-SPEC: Distribution-based Spectral Clustering

Require: Data $\mathbb{X}$, clusters k, threshold τ, sample size p, kernel parameters (ψ, t)
Ensure: Clustering result
1: Sample $\min(p, n)$ points and construct their graphs
2: Threshold the sampled graph and find its connected components
3: Select the k largest components as $\mathcal{G}_1^b, \ldots, \mathcal{G}_k^b$
4: Compute mapping $\hat{\Phi}(\mathcal{G}_j)$ using Equation 1
5: Construct bipartite graph $\mathcal{B}$ between $\mathbb{X}$ and $\hat{\Phi}(\mathcal{G}_j)$
6: Perform transfer cut to obtain k-dimensional spectral embedding
7: Apply k-means clustering to the spectral embedding to obtain the clustering result

Write $\boldsymbol{W} = \boldsymbol{W}^{\Phi} \in \mathbb{R}_{\geq 0}^{n \times k}$. The full bipartite adjacency, in point-first order, is $\boldsymbol{B}_{\text{full}} = \begin{bmatrix} 0 & \boldsymbol{W} \\ \boldsymbol{W}^{\top} & 0 \end{bmatrix}$. Its degree blocks are $D_x = \operatorname{diag}(\boldsymbol{W}\mathbf{1}_k)$ and $D_y = \operatorname{diag}(\boldsymbol{W}^{\top}\mathbf{1}_n)$. Transfer cut (Li et al., 2012; Huang et al., 2020; Li et al., 2022) uses the reduced affinity $\boldsymbol{W}_{\mathbb{K}} = \boldsymbol{W}^{\top} D_x^{-1} \boldsymbol{W}$, whose degree matrix is D_y. It replaces a dense $(n+k)^2$ eigenproblem by a $k \times k$ one. Forming this matrix costs $\mathcal{O}(nk^2)$ and its dense eigendecomposition costs $\mathcal{O}(k^3)$. Algorithm 1 summarizes the procedure.

3.2 Distributional Recovery and Perturbation Guarantees

The distributional construction represents the data by an $n \times k$ point–distribution affinity matrix and reduces the spectral problem to a $k \times k$ matrix. In this subsection, write $B = \boldsymbol{W}^{\Phi}$ for the point–distribution affinity, not the full bipartite adjacency. Define $D_x = \operatorname{diag}(B\mathbf{1}_k)$, $D_y = \operatorname{diag}(B^{\top}\mathbf{1}_n)$, and $T = D_x^{-1/2} B D_y^{-1/2}$, with positive degrees. The reduced affinity is $B^{\top} D_x^{-1} B$, its degree matrix is D_y, and its normalized Laplacian is $I_k - T^{\top}T$. Forming the reduced matrix costs $\mathcal{O}(nk^2)$; its dense eigendecomposition costs $\mathcal{O}(k^3)$. Appendix C.1 establishes the exact transfer-cut relation to the same bipartite operator.

Theorem 3.2 (Exact partition recovery). *Let H be the membership matrix of k nonempty clusters. If $B = \operatorname{diag}(a)HC$, with $a_i > 0$ and a nonnegative $k \times k$ matrix C having no zero row or column and no two rows that are positive scalar multiples, the mathematical row-normalized transfer embedding*

takes exactly one distinct value per cluster. Every global optimum of k-means with k nonempty clusters recovers the partition. If $C = I_k$, the embedding is HV for an orthogonal matrix V, and the normalized bipartite Laplacian has exactly k zero eigenvalues.

Theorem 3.2 guarantees preservation of cluster membership under explicit distributional affinity conditions. The general profile model permits nonzero between-cluster affinities; the diagonal-profile special case gives a one-hot representation up to an orthogonal change of basis. Appendix C.1 separates the sampled graph's threshold-connectivity condition from the affinity-profile condition and gives the complete recovery proof. The numerical denominator offset is omitted in this mathematical statement.

Theorem 3.3 (Eigenvalue perturbation certificate)**.** *Let B and $\widetilde{B} = B + E$ satisfy the positivity and relative-degree conditions of Theorem C.3. For the $n \times n$ point-side normalized Laplacians $L_d = I_n - TT^\top$ and $\widetilde{L}_d$, define $\varepsilon_d = \min\left\{1, 2\left(\frac{1+q_E}{\sqrt{(1-\alpha_x)(1-\alpha_y)}} - 1\right)\right\}$, where α_x, α_y, q_E are given in equation 7. Then $\max_{1\le i\le n} |\lambda_i(\widetilde{L}_d) - \lambda_i(L_d)| \le \|\widetilde{L}_d - L_d\|_2 \le \varepsilon_d$. For SC on the same points, let ε_s be the normalized-affinity perturbation bound defined in Appendix C.1. If $\varepsilon_d \le \varepsilon_s$, the D-SPEC eigenvalue perturbation certificate is no larger than the SC certificate.*

Theorem 3.4 (Spectral-subspace perturbation certificate)**.** *Assume $n > k$. For $\ell \in \{d, s\}$, let U_ℓ and $\widetilde{U}_\ell$ be orthonormal bases for the bottom-k spectral subspaces of the clean and perturbed $n \times n$ normalized Laplacians. Let $\delta_\ell = \lambda_{k+1}(L_\ell) - \lambda_k(L_\ell)$, and assume $\|\widetilde{L}_\ell - L_\ell\|_2 \le \varepsilon_\ell < \delta_\ell$. Then $\|\sin\Theta(U_\ell, \widetilde{U}_\ell)\|_2 \le \frac{\varepsilon_\ell}{\delta_\ell - \varepsilon_\ell}$. If $\varepsilon_d/\delta_d \le \varepsilon_s/\delta_s$, the D-SPEC subspace perturbation certificate is no larger than the SC certificate.*

Theorems 3.3–3.4 identify a regime of tighter robustness guarantees through explicit affinity perturbations, degree changes, and spectral gaps. For kernel mean representatives, replacing a fraction ρ_j of the samples in representative j changes each associated affinity by at most $K_{\max}\rho_j$ when $0 \le \kappa \le K_{\max}$ and the fitted kernel is fixed. The appendix connects this averaging bound to ε_d, and separately proves stability of the actual row-normalized embedding in Proposition C.6. Theorem C.7 further converts an embedding bound ε_Y and the achieved k-means objective J into at most $4(\varepsilon_Y + \sqrt{J})^2/\Delta^2$ mislabeled points, after a label permutation, when this quantity is below the smallest true cluster size; Δ is the ideal center separation. The analysis thus connects the embedding to the returned clustering. Theorem C.9 and Corollary C.10 additionally control finite-partition approximation and relative-threshold component recovery, including adaptive selection of representatives. The SC comparison concerns certified upper bounds; experiments report realized behavior.

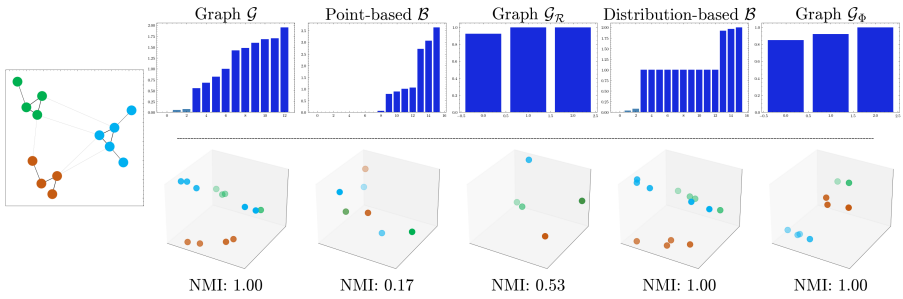


Figure 3: An example with noisy edges (We do eigendecomposition on the normalized Laplacian matrix of graph $\mathcal{G}$, $\mathcal{B}$, and do eigendecomposition on the normalized adjacency matrix of graph $\mathcal{G}_R$, $\mathcal{G}_\Phi$). The performance of point-based methods decreased compared to Figure 1, while that of distribution-based methods did not.

In Figure 3, the distributional bipartite construction is less affected by added cross-cluster edges than the point-based construction in this example.

4 EXPERIMENTS

In this section, we empirically study whether the eigendecomposition of only a $k \times k$ matrix can achieve effective clustering. We demonstrate the superiority of D-SPEC through the following four comparisons: (a) Performance on the benchmark datasets. (b) Scalability on large-scale datasets. (c) Performance on the fundamental limitations of spectral clustering. (d) A comparison of the ensemble version of D-SPEC and U-SPEC.

We compare the proposed D-SPEC with: SC (Shi & Malik, 2000): The original spectral clustering (Ncut) is used as our baseline. U-SPEC (Huang et al., 2020): Ultra-Scalable Spectral Clustering, A hybrid landmark selection method that combines random initialization of candidate samples with k-means to determine cluster centroids as representatives, followed by the computation of approximate K-nearest neighbor representatives. DNCSC (Li et al., 2022): Divide-and-conquer spectral clustering, a divide-and-conquer-based landmark selection method to generate high-quality landmarks. FastSC (Macgregor, 2023): Fast spectral clustering method using the power method to calculate approximate eigenvectors. GBSC (Xie et al., 2023): Spectral clustering algorithm based on granular-ball, which generates p granular-balls from the original data, and performs spectral clustering only on the p granular-balls. Further experiments, such as comparisons with methods not shown in Table 1 and ablation analyses, are presented in the appendix.

We use Normalized Mutual Information (NMI) (Strehl & Ghosh, 2002), Adjusted Rand Index (ARI) (Hubert & Arabie, 1985; Gates & Ahn, 2017) and F-measure (van Rijsbergen, 1979) as evaluation metrics. The experiments are executed on a Linux machine with 1TB RAM and an AMD 128-core CPU, with each core running at 2 GHz.

4.1 EXPERIMENTS ON BENCHMARK DATASETS

We use fifteen datasets, including the datasets used in the U-SPEC and DNSCS papers. The characteristics of the datasets we used are shown in the Appendix. The results in terms of NMI are shown in Table 1. The results in terms of ARI and F-measure are shown in Appendix F. SC requires a significant amount of time for clustering, exceeding two days for datasets larger than MNIST in size. GBSC is not able to handle datasets larger than mnist due to the memory consumption. Our D-SPEC method achieves the best scores on most of the fifteen benchmark datasets.

Table 1: Average NMI scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	n	d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
spiral	312	2	3	**1.000**	**1.000**	**1.000**	0.693	0.009	**1.000**
4C	1000	2	4	**1.000**	**1.000**	**1.000**	0.726	0.528	**1.000**
AC	1004	2	2	**1.000**	**1.000**	**1.000**	0.340	0.610	**1.000**
RingG	1536	2	4	0.794	0.845	0.761	0.779	0.694	**0.987**
complex9	3031	2	9	**1.000**	0.971	0.951	0.810	0.662	**1.000**
cure-t2-4k	4200	2	7	0.843	0.886	0.872	0.810	0.772	**0.951**
landsat	2000	36	6	0.281	0.668	0.647	**0.740**	0.646	0.647
spambase	4601	57	2	0.011	0.013	0.033	0.020	0.162	**0.166**
waveform3	5000	21	3	0.371	0.370	0.369	**0.605**	0.370	0.406
pendigits	10992	16	10	0.641	0.826	0.813	0.523	0.596	**0.847**
usps	11000	256	10	0.676	0.654	0.652	0.564	0.338	**0.778**
letters	20000	16	26	0.278	0.455	0.437	0.308	O/M	**0.478**
mnist	70000	784	10	**0.766**	0.699	0.736	0.651	O/M	0.746
skin	245057	3	2	N/A	0.025	0.508	0.001	O/M	**0.767**
covertype	581012	54	7	N/A	0.212	0.086	**0.695**	O/M	0.218
Avg.score				0.666	0.642	0.658	0.551	0.490	**0.733**
Avg.rank				3.567	2.833	3.433	4.067	5.400	**1.700**

* O/M indicates the out-of-memory error.
* N/A indicates that no clustering results were obtained within two days.

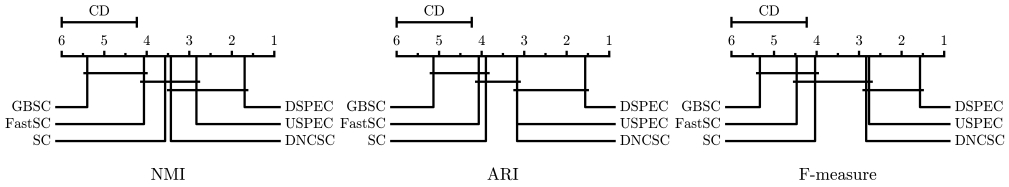


Figure 4: The Nemenyi significance test results at 0.1 significance level.

The results of the Nemenyi significance test (Nemenyi, 1963) are shown in Figure 4. D-SPEC attains the strongest aggregate ranking across the three metrics and is the only method showing a statistically significant improvement over SC in this comparison. This finding is consistent with aggregated distributional representatives and threshold-based suppression of weak connections.

To examine empirical robustness alongside the perturbation analysis, we introduced uniform noise $U(0, 1)$ with ratios ranging from 0 to 0.2. We performed this on the smallest suitable 4C and the largest usps datasets, among the datasets where all comparison algorithms could be successfully executed. The results are illustrated in Figure 5.

On 4C, all algorithms generally show a trend of worsening performance as the ratio increases. When the noise ratio reaches 0.2, only DNCSC and D-SPEC achieve an NMI greater than 0.9, while the NMI of SC is 0.52. On usps, only U-SPEC, DNCSC, and D-SPEC are almost unaffected by noise. This may be because the clustering results of U-SPEC and DNCSC are not very high in the absence of noise. They are all around 0.65, much lower than 0.78 of D-SPEC. For this reason, we compare them on the second largest dataset pendigits, because they all perform well (>0.8) and have closed results when there is no noise in this data. As the ratio of noise increases, the NMI of U-SPEC decreases from 0.83 to 0.79, the NMI of DNCSC decreases from 0.81 to 0.76, and only the NMI of D-SPEC remains almost unchanged.

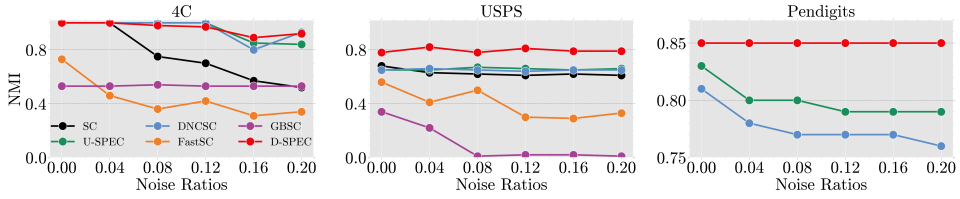


Figure 5: Comparison of robustness to noise.

4.2 Comparison on Large-scale Datasets

In this subsection, we summarize the time complexity of D-SPEC and compare D-SPEC with other algorithms on large-scale datasets.

With a fixed feature-map configuration, forming the affinity graph on p sampled points and extracting components costs $\mathcal{O}(p^2)$, while point–distribution affinities cost $\mathcal{O}(nk)$. Transfer cut costs $\mathcal{O}(nk^2 + k^3)$, including a $\mathcal{O}(k^3)$ dense eigendecomposition. These terms in Table 2 exclude feature-map construction and final k-means. Explicitly, for feature dimension q, dense affinity evaluation costs $\mathcal{O}(p^2q + nkq)$ and representative averaging costs $\mathcal{O}(pq)$; sparse IDK blocks can reduce this cost. With k-dimensional embeddings, I Lloyd iterations cost $\mathcal{O}(Ink^2)$ per restart. Linear scaling with n treats the kernel configuration, p, k, and iteration budget as fixed.

In order to demonstrate that D-SPEC can achieve efficient and effective clustering for large-scale data by only performing eigendecomposition on a $k \times k$ matrix, we selected the five large-scale datasets used in U-SPEC, with points ranging from 1 million to 20 million. We compared the NMI and runtime of these algorithms on these large-scale data, as shown in the Figure 6, and the results of ARI and F-measure are in the Appendix H.

The gray bar in the figure indicates that the algorithm cannot handle the dataset. SC and GBSC failed on all five datasets. FastSC can run on the first two datasets, but the clustering effect is very poor

Table 2: Comparison of the computational complexity.

	SC	USPEC	DNCSC	FastSC	GBSC	D-SPEC
LS	NA	$\mathcal{O}(p^2)$	$\mathcal{O}(n\alpha)$	NA	$\mathcal{O}(n \log n)$	$\mathcal{O}(p^2)$
Sco	$\mathcal{O}(n^2)$	$\mathcal{O}(np^{\frac{1}{2}})$	$\mathcal{O}(nK)$	$\mathcal{O}(n^2)$	$\mathcal{O}(p^2)$	$\mathcal{O}(nk)$
ED	$\mathcal{O}(n^3)$	$\mathcal{O}(nK(K+k)+p^3)$	$\mathcal{O}(nK(K+k)+p^3)$	$\mathcal{O}(\frac{nK}{\epsilon^2})$	$\mathcal{O}(p^3)$	$\mathcal{O}(nk^2+k^3)$

* LS: Landmark selection. Sco: Similarity construction. ED: reduced-matrix formation and spectral decomposition.
* n: number of points. k: number of clusters. p: number of landmards. K: number of nearest neighbors.

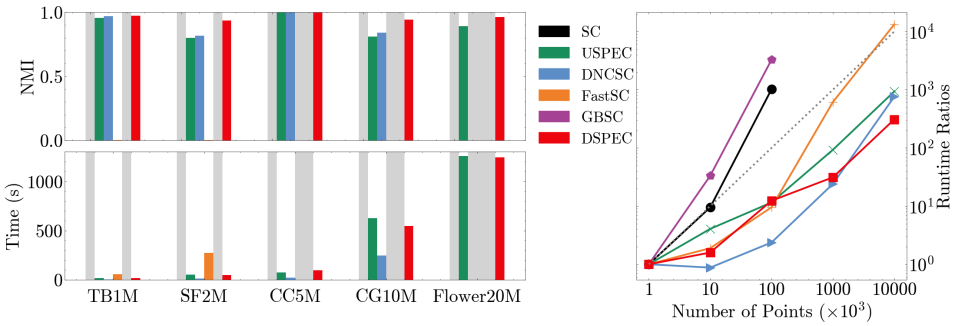


Figure 6: Results in terms of NMI and runtime in 5 large-scale datasets (left). And scale-up test on CG10M dataset (right).

because it uses approximate eigenvectors. DNCSC overflowed the memory on the largest dataset. Only USPEC and D-SPEC can handle all datasets, and D-SPEC achieves the best clustering results.

Since the running time of an algorithm is not only related to the number of points in the dataset, but also to the distribution, dimension and the number of clusters of the data, in order to compare the running time of D-SPEC and other algorithms that only change with the size of the dataset, we randomly sampled 1k, 10k, 100k, 1M, and 10M points from CG-10M dataset, and each algorithm was experimented with the same parameters on all datasets. The results are shown in the Figure 6 (right), where the vertical axis (logarithmized) is the ratio of the running time on different datasets to the running time on the 1k dataset. The results are consistent with Table 2. D-SPEC has a very low time complexity, especially on 1M and 10M datasets.

In short, experiments on large-scale datasets show that the D-SPEC algorithm, which only performs eigendecomposition on the $k \times k$ matrix, is effective and efficient.

4.3 Fundamental Limitations of Spectral Clustering

Nadler & Galun (2006) pointed out that spectral clustering has some fundamental limitations. We compare it with the three example datasets given by (Nadler & Galun, 2006). The experimental results of these three datasets are shown in Figure 7.

The first row of the figure illustrates the dataset, while the second row presents the NMI scores of various clustering algorithms. In the first dataset, which is composed of three clusters with an equal point ratio of 1:1:1 and a density ratio of 1:8:8, all algorithms perform well except for FastSC and GBSC. However, when the point ratio is altered to 8:1:1 (second dataset), the NMI scores of all algorithms except D-SPEC decline below 0.8. For the third dataset, consisting of a Gaussian distribution and a uniform distribution, the NMI of all algorithms except U-SPEC and D-SPEC is poor and below 0.6.

D-SPEC is the only algorithm that maintains good performance on all three datasets. There are two main reasons. (i) D-SPEC first extracts a bounded graph, which removes some noise edges, such as those caused by cluster overlap. (ii) D-SPEC constructs a bipartite graph based on distribution, which can preserve cluster information.

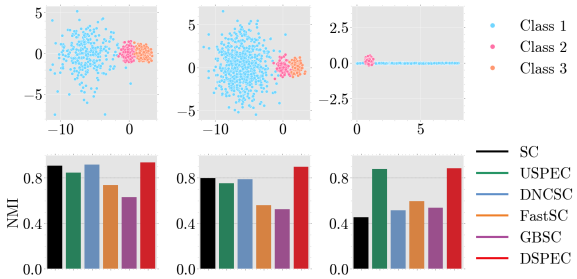


Figure 7: Comparison of algorithms for addressing the fundamental limitations of spectral clustering.

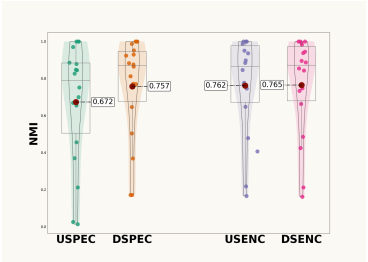


Figure 8: Comparison between U-SPEC and D-SPEC and their ensemble versions.

For example, the two clusters in the second dataset are not ignored (which would be ignored by other clustering algorithms because each cluster has very few points), and the uniform distribution line cluster in the third dataset spans a large area, but their distribution information is preserved, thus achieving more effective spectral clustering. In a nutshell, D-SPEC provides a possible way to overcome the basic limitations of spectral clustering.

4.4 Ensemble Distribution-based Spectral Clustering

Ensemble learning is often used to combine multiple base models to improve the performance of the base model. For example, the U-SENC (Huang et al., 2020), proposed as the ensemble version of U-SPEC, significantly improved the clustering quality of the U-SPEC. We compared D-SPEC with U-SENC, the results are shown in Figure 8. Even compared with U-SENC, the overall performance of our proposed D-SPEC algorithm is better.

After we performed the same ensemble method on D-SPEC, we found that the improvement in clustering performance was very small (as shown in the Figure 8). The reason is that ensemble learning often works on weak base models, while D-SPEC has good clustering performance. In addition, ensemble learning requires a large diversity. Since U-SPEC is point-based, the diversity is large, while the diversity based on distribution is small.

The tested ensemble configuration yields only limited improvement for D-SPEC. Developing distributional ensembles with greater diversity remains an open question. U-SENC first uses U-SPEC to obtain k' clusters, where $k' \in \{20, \ldots, 60\}$, and repeats this m times. It constructs an $n \times \hat{k}$ graph, with $\hat{k} = \sum_{i=1}^{m} k'_i$, before a final spectral clustering step. We apply the same procedure to D-SPEC, denoted D-SENC (DSENC in the figure); displayed values are average NMI scores.

We conducted an extensive series of additional experiments to further validate the performance of D-SPEC. These include comparative analyses against coreset-based spectral clustering Jourdan et al. (2025) and deep spectral clustering Meng et al. (2025), as well as ablation studies examining the impact of varying the size p and employing different sampling strategies. Furthermore, we evaluated D-SPEC on the downstream task of image segmentation. Detailed results for these experiments are provided in the Appendix.

5 Conclusion

We introduced a distributional perspective on scalable spectral clustering, replacing individual landmark vertices by distribution representatives in an $n \times k$ bipartite graph. Its spectral problem reduces exactly to a $k \times k$ matrix. Our theory establishes exact partition recovery under explicit profile conditions, quantifies how affinity perturbations propagate to normalized operators and returned labels, and provides finite-partition guarantees compatible with adaptive representative selection. Existing experiments demonstrate strong clustering performance and scalability on the evaluated datasets. These results support distributional representation as a useful direction for efficient spectral clustering.

AI USE STATEMENT

Generative AI assisted with language editing, literature and reference checks, and the checking of theoretical statements and proofs. This introduced no new experimental measurements. The authors retain responsibility for the final manuscript, including verification of all AI-assisted content.

ETHICS STATEMENT

This work concerns unsupervised clustering of benchmark feature matrices. No new human-subject data were collected. Clustering can amplify representation bias, so deployment in consequential settings requires dataset-specific fairness, privacy, and error analysis.

REFERENCES

E. Alpaydin and Fevzi Alimoglu. Pen-Based Recognition of Handwritten Digits. UCI Machine Learning Repository, 1996. URL https://archive.ics.uci.edu/dataset/81/pen+based+recognition+of+handwritten+digits.

Rajen Bhatt and Abhinav Dhall. Skin Segmentation. UCI Machine Learning Repository, 2009. URL https://archive.ics.uci.edu/dataset/229/skin+segmentation.

Jock A Blackard and Denis J Dean. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture*, 24(3):131–151, 1999. doi: 10.1016/S0168-1699(99)00046-0. URL https://doi.org/10.1016/S0168-1699(99)00046-0.

Djallel Bouneffouf and Inanc Birol. Sampling with minimum sum of squared similarities for Nyström-based large scale spectral clustering. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 2313–2319, 2015. URL https://www.ijcai.org/Proceedings/15/Papers/327.pdf.

Deng Cai, Xiaofei He, and Jiawei Han. Speed up kernel discriminant analysis. *The VLDB Journal*, 20(1):21–33, 2011a. doi: 10.1007/s00778-010-0189-3. URL https://doi.org/10.1007/s00778-010-0189-3.

Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1548–1560, 2011b. doi: 10.1109/TPAMI.2010.231. URL https://doi.org/10.1109/TPAMI.2010.231.

Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27, 1974. doi: 10.1080/03610927408827101. URL https://doi.org/10.1080/03610927408827101.

Hong Chang and Dit-Yan Yeung. Robust path-based spectral clustering. *Pattern Recognition*, 41(1):191–203, 2008. doi: 10.1016/j.patcog.2007.04.010. URL https://doi.org/10.1016/j.patcog.2007.04.010.

Xinlei Chen and Deng Cai. Large scale spectral clustering with landmark-based representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 25, pp. 313–318, 2011. doi: 10.1609/aaai.v25i1.7900. URL https://doi.org/10.1609/aaai.v25i1.7900.

Dongdong Cheng, Xiaocui Jiang, Shuyin Xia, Guoyin Wang, Jinlong Huang, Sulan Zhang, and Yi Wang. Fast spectral clustering via pseudo-label-based granular-ball division for large-scale data. *IEEE Transactions on Knowledge and Data Engineering*, 38(5):2807–2817, 2026. doi: 10.1109/TKDE.2026.3672824. URL https://doi.org/10.1109/TKDE.2026.3672824.

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. doi: 10.1109/TPAMI.2004.1262185. URL https://doi.org/10.1109/TPAMI.2004.1262185.

Peter W Frey and David J Slate. Letter recognition using Holland-style adaptive classifiers. *Machine Learning*, 6(2):161–182, 1991. doi: 10.1023/A:1022606404104. URL https://doi.org/10.1023/A:1022606404104.

Alexander J Gates and Yong-Yeol Ahn. The impact of random models on clustering similarity. *Journal of Machine Learning Research*, 18(87):1–28, 2017. URL https://www.jmlr.org/papers/v18/17-039.html.

Li He, Nilanjan Ray, Yisheng Guan, and Hong Zhang. Fast large-scale spectral clustering via explicit feature mapping. *IEEE Transactions on Cybernetics*, 49(3):1058–1071, 2019. doi: 10.1109/TCYB.2018.2794998. URL https://doi.org/10.1109/TCYB.2018.2794998.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830. URL https://doi.org/10.1080/01621459.1963.10500830.

Mark Hopkins, Erik Reeber, George Forman, and Jaap Suermondt. Spambase. UCI Machine Learning Repository, 1999. URL https://archive.ics.uci.edu/dataset/94/spambase.

Dong Huang, Chang-Dong Wang, Jian-Sheng Wu, Jian-Huang Lai, and Chee-Keong Kwoh. Ultra-scalable spectral clustering and ensemble clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1212–1226, 2020. doi: 10.1109/TKDE.2019.2903410. URL https://doi.org/10.1109/TKDE.2019.2903410.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985. doi: 10.1007/BF01908075. URL https://link.springer.com/article/10.1007/BF01908075.

Ben Jourdan, Gregory Schwartzman, Peter Macgregor, and He Sun. Coreset spectral clustering. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/8400e654f8fb7154a262295dcc2bd348-Abstract-Conference.html.

Félix Laplante and Christophe Ambroise. Spectral bridges. *Computo*, 2024. doi: 10.57750/1gr8-bk61. URL https://computo-journal.org/published-202412-ambroise-spectral/.

Hongmin Li, Xiucai Ye, Akira Imakura, and Tetsuya Sakurai. Divide-and-conquer based large-scale spectral clustering. *Neurocomputing*, 501:664–678, 2022. doi: 10.1016/j.neucom.2022.06.006. URL https://doi.org/10.1016/j.neucom.2022.06.006.

Zhenguo Li, Xiao-Ming Wu, and Shih-Fu Chang. Segmentation using superpixels: A bipartite graph partitioning approach. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 789–796. IEEE, 2012. doi: 10.1109/CVPR.2012.6247750. URL https://doi.org/10.1109/CVPR.2012.6247750.

Zhoumin Lu, Feiping Nie, Rong Wang, and Xuelong Li. A differentiable perspective for multi-view spectral clustering with flexible extension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7087–7098, 2023. doi: 10.1109/TPAMI.2022.3224978. URL https://doi.org/10.1109/TPAMI.2022.3224978.

Peter Macgregor. Fast and simple spectral clustering in theory and practice. In *Advances in Neural Information Processing Systems*, volume 36, pp. 34410–34425. Curran Associates, Inc., 2023. doi: 10.52202/075280-1492. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/6c5b82193c5d8e6aa5806239676ddc97-Abstract-Conference.html.

Dmitrii Marin, Meng Tang, Ismail Ben Ayed, and Yuri Boykov. Kernel clustering: Density biases and solutions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):136–147, 2019. doi: 10.1109/TPAMI.2017.2780166. URL https://doi.org/10.1109/TPAMI.2017.2780166.

Hua Meng, Yueyi Zhang, and Zhiguo Long. Deep spectral clustering by integrating local structure and prior information. *Knowledge-Based Systems*, 308:112743, 2025. doi: 10.1016/j.knosys.2024.112743. URL https://doi.org/10.1016/j.knosys.2024.112743.

Cameron Musco and Christopher Musco. Recursive sampling for the Nyström method. In *Advances in Neural Information Processing Systems*, volume 30, pp. 3833–3845, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/a03fa30821986dff10fc66647c84c9c3-Abstract.html.

Boaz Nadler and Meirav Galun. Fundamental limitations of spectral clustering. In *Advances in Neural Information Processing Systems*, volume 19, pp. 1017–1024. MIT Press, 2006. URL https://proceedings.neurips.cc/paper_files/paper/2006/hash/bdb6920adcd0457aa17b53b22963dad9-Abstract.html.

Peter Bjorn Nemenyi. *Distribution-free multiple comparisons*. PhD thesis, Princeton University, 1963. URL https://www.genealogy.math.ndsu.nodak.edu/id.php?id=29470.

Feiping Nie, Chaodie Liu, Rong Wang, and Xuelong Li. A novel and effective method to directly solve spectral clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46 (12):10863–10875, 2024. doi: 10.1109/TPAMI.2024.3447287. URL https://doi.org/10.1109/TPAMI.2024.3447287.

Shenfei Pei, Huijuan Dong, Nianci Guan, Zhongqi Lin, Feiping Nie, Xudong Jiang, and Zengwei Zheng. A greedy strategy for graph cut. *IEEE Transactions on Image Processing*, 35:2224–2234, 2026. doi: 10.1109/TIP.2026.3661874. URL https://doi.org/10.1109/TIP.2026.3661874.

Dimitrios Rafailidis, Eleni Constantinou, and Yannis Manolopoulos. Landmark selection for spectral clustering based on Weighted PageRank. *Future Generation Computer Systems*, 68:465–472, 2017. doi: 10.1016/j.future.2016.03.006. URL https://doi.org/10.1016/j.future.2016.03.006.

Bishwajit Saha, Dmitry Krotov, Mohammed J Zaki, and Parikshit Ram. End-to-end differentiable clustering with associative memories. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 29649–29670. PMLR, 2023. URL https://proceedings.mlr.press/v202/saha23a.html.

S. Salvador and P. Chan. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *16th IEEE International Conference on Tools with Artificial Intelligence*, pp. 576–584, 2004. doi: 10.1109/ICTAI.2004.50. URL https://doi.org/10.1109/ICTAI.2004.50.

Venu Satuluri, Srinivasan Parthasarathy, and Yiye Ruan. Local graph sparsification for scalable clustering. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, pp. 721–732, 2011. doi: 10.1145/1989323.1989399. URL https://doi.org/10.1145/1989323.1989399.

Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. doi: 10.1109/34.868688. URL https://doi.org/10.1109/34.868688.

Daniel A Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011. doi: 10.1137/080734029. URL https://doi.org/10.1137/080734029.

Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(Dec):583–617, 2002. URL https://www.jmlr.org/papers/v3/strehl02a.html.

Chang Tang, Zhenglai Li, Jun Wang, Xinwang Liu, Wei Zhang, and En Zhu. Unified one-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):6449–6460, 2023. doi: 10.1109/TKDE.2022.3172687. URL https://doi.org/10.1109/TKDE.2022.3172687.

Kai Ming Ting, Jonathan R. Wells, and Ye Zhu. Point-set kernel clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):5147–5158, 2023a. doi: 10.1109/TKDE.2022.3144914. URL https://doi.org/10.1109/TKDE.2022.3144914.

Kai Ming Ting, Bi-Cun Xu, Takashi Washio, and Zhi-Hua Zhou. Isolation distributional kernel: A new tool for point and group anomaly detections. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):2697–2710, 2023b. doi: 10.1109/TKDE.2021.3120277. URL https://doi.org/10.1109/TKDE.2021.3120277.

Toon Van Craenendonck and Hendrik Blockeel. Constraint-based clustering selection. *Machine Learning*, 106(9–10):1497–1521, 2017. doi: 10.1007/s10994-017-5643-7. URL https://doi.org/10.1007/s10994-017-5643-7.

C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 2nd edition, 1979. ISBN 978-0-408-70929-3. URL https://www.dcs.gla.ac.uk/Keith/Preface.html.

Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4): 395–416, 2007. doi: 10.1007/s11222-007-9033-z. URL https://doi.org/10.1007/s11222-007-9033-z.

Yizhang Wang, Jiaxin Qian, Muhammad Hassan, Xinyu Zhang, Tao Zhang, Chao Yang, Xingxing Zhou, and Fengjin Jia. Density peak clustering algorithms: A review on the decade 2014–2023. *Expert Systems with Applications*, 238:121860, 2024. doi: 10.1016/j.eswa.2023.121860. URL https://doi.org/10.1016/j.eswa.2023.121860.

Hermann Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912. doi: 10.1007/BF01456804. URL https://doi.org/10.1007/BF01456804.

Lingfei Wu, Pin-Yu Chen, Ian En-Hsu Yen, Fangli Xu, Yinglong Xia, and Charu Aggarwal. Scalable spectral clustering using random binning features. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2506–2515, 2018. doi: 10.1145/3219819.3220090. URL https://doi.org/10.1145/3219819.3220090.

Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2008. doi: 10.1007/s10115-007-0114-2. URL https://doi.org/10.1007/s10115-007-0114-2.

Jiang Xie, Weiyu Kong, Shuyin Xia, Guoyin Wang, and Xinbo Gao. An efficient spectral clustering algorithm based on granular-ball. *IEEE Transactions on Knowledge and Data Engineering*, 35(9): 9743–9753, 2023. doi: 10.1109/TKDE.2023.3249475. URL https://doi.org/10.1109/TKDE.2023.3249475.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 478–487. PMLR, 2016. URL https://proceedings.mlr.press/v48/xieb16.html.

Tianbao Yang, Yu-Feng Li, Mehrdad Mahdavi, Rong Jin, and Zhi-Hua Zhou. Nyström method vs random Fourier features: A theoretical and empirical comparison. In *Advances in Neural Information Processing Systems*, volume 25, 2012. URL https://proceedings.neurips.cc/paper/2012/hash/621bf66ddb7c962aa0d22ac97d69b793-Abstract.html.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008. URL https://doi.org/10.1093/biomet/asv008.

Hang Zhang, Kai Ming Ting, and Ye Zhu. Kernel-bounded clustering: Achieving the objective of spectral clustering without eigendecomposition. *Artificial Intelligence*, 350:104440, 2026. doi: 10.1016/j.artint.2025.104440. URL https://doi.org/10.1016/j.artint.2025.104440.

Songpeng Zu, Yang Eric Li, Kangli Wang, Ethan J. Armand, Sainath Mamde, Maria Luisa Amaral, Yuelai Wang, Andre Chu, Yang Xie, Michael Miller, Jie Xu, Zhaoning Wang, Kai Zhang, Bojing Jia, Xiaomeng Hou, Lin Lin, Qian Yang, Seoyeon Lee, Bin Li, Samantha Kuan, Hanqing Liu, Jingtian Zhou, Antonio Pinto-Duarte, Jacinta Lucero, Julia Osteen, Michael Nunn, Kimberly A. Smith, Bosiljka Tasic, Zizhen Yao, Hongkui Zeng, Zihan Wang, Jingbo Shang, M. Margarita Behrens, Joseph R. Ecker, Allen Wang, Sebastian Preissl, and Bing Ren. Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature*, 624(7991):378–389, 2023. doi: 10.1038/s41586-023-06824-9. URL https://doi.org/10.1038/s41586-023-06824-9.

Wendi Zuo and Xinmin Hou. An improved probability propagation algorithm for density peak clustering based on natural nearest neighborhood. *Array*, 15:100232, 2022. doi: 10.1016/j.array.2022.100232. URL https://doi.org/10.1016/j.array.2022.100232.

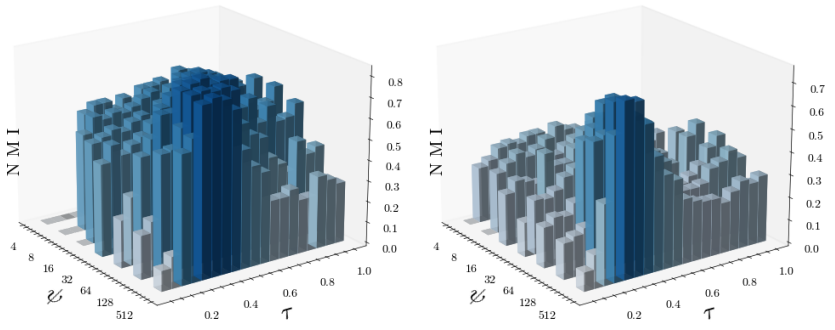
ORGANIZATION OF THE APPENDIX

This appendix provides supplementary materials, detailed proofs, and extensive experimental results to support the findings presented in the main paper. The content is organized as follows:

- **Limitation:** Section A discusses the limitations of D-SPEC.
- **Theoretical Proofs:** Section B reports empirical diagnostics of Assumption 3.1. Section C proves recovery, perturbation, clustering-error, finite-partition, and sampling results. Section D reports additional spectral diagnostics.
- **Experiments Details** Section E shows the datasets we used and parameter settings. Section F reports the complete results in terms of ARI and F-measure. Section G analyses the results on the 3 datasets where D-SPEC performs worse than other methods. Section H presents results on large-scale datasets. Section I reports the results on the fundamental limitations of spectral clustering.
- **Comparisons with State-of-the-Arts:** We provide detailed comparisons with some variations of the methods in the main text J, Deep Spectral Clustering (Section K) and Coreset Spectral Clustering (Section L), along with an analysis of point-based versus distribution-based methods (Section M).
- **Ablation Studies and Parameter Analysis:** Sections N–Q discuss the ablation studies regarding the size p, bounded graphs, and sampling strategies.
- **Applications:** Section R demonstrates the application of D-SPEC in image segmentation.

A LIMITATION OF D-SPEC

D-SPEC has two parameters, ψ (parameter of IDK) and τ (parameter of D-SPEC). Since D-SPEC is a distribution-based method, its parameters are sensitive to the distribution of data. We show the impact of these two parameters on the Pendigits, and Mnist datasets, the results are shown in the Figure 9. Like most clustering algorithms, such as DBSCAN, DP, and Spectral clustering (the parameters of the Gaussian kernel), it is a limitation of D-SPEC that a suitable parameter must be selected to maximize the performance of the algorithm. D-SPEC typically performs well only when ψ is in the range from 64 to 512 and τ is in the range from 0.2 to 0.5.



(a) Pendigits (b) Mnist

Figure 9: The sensitivity analysis of parameters ψ and τ of D-SPEC.

B ASSUMPTION 3.1 IN BENCHMARK DATASET

Assumption 3.1 is a latent local separability condition motivating graph construction without supplying labels to the algorithm. Exact recovery additionally uses the threshold-connectivity and affinity-profile conditions in Appendix C. The existing diagnostics on seven real datasets measure

Table 3: Performance comparison across benchmark datasets.

Dataset	Strict Assump. 3.1	Cross exceed	KME
Landsat	0.871	0.020%	0.932
Spambase	0.843	0.036%	0.944
Waveform3	0.739	0.031%	0.933
Pendigits	0.989	0.005%	0.990
USPS	0.939	0.013%	0.970
Letters	0.829	0.049%	0.926
MNIST	0.926	0.017%	0.969

local separation: *Cross exceed* is the fraction of cross-cluster comparisons exceeding the nearest within-cluster similarity, and *KME* is post-hoc class-mean separability.

The seven-dataset macro strict rate is 0.877 (shown in Table 3). Allowing equality in the discrete similarity raises it to 0.903. More importantly, the nearest same-cluster neighbor outranks a random cross-cluster point with probability 0.994, while only 0.024% of cross comparisons are true exceedances. Thus, failures of the all-cross quantifier arise from isolated extreme points or ties, not a reversal of the neighborhood structure. Aggregation then raises macro cluster separability to 0.952, supporting the distributional design. These diagnostics concern local separation and aggregation; they do not alone establish single-threshold connectivity on every dataset.

C Proof of the Theorems

C.1 Operators, recovery, and perturbation guarantees

We establish threshold connectivity, exact recovery for weighted distributional profiles with potentially nonzero between-cluster affinities, and explicit eigenvalue, subspace, and row-normalized embedding bounds. We then connect these bounds to clustering error and finite-partition approximation. Comparison with SC is conditional on the stated perturbation and spectral-gap inequalities.

Notation and normalization. In this supplement, $B \in \mathbb{R}^{n\times k}_{\geq 0}$ denotes the point–distribution affinity matrix, rather than the full bipartite adjacency matrix. Set

$$d = B\mathbf{1}_k, \quad c = B^\top \mathbf{1}_n, \quad D_x = \mathrm{diag}(d), \quad D_y = \mathrm{diag}(c), \quad T = D_x^{-1/2} B D_y^{-1/2}. \tag{2}$$

All degrees used in inverses are assumed positive. Degree matrices are always formed from the row sums of the affinity matrix being normalized; if self-loops are removed, they are removed before computing these sums. The reduced affinity is $W_y = B^\top D_x^{-1} B$, and $W_y \mathbf{1}_k = B^\top D_x^{-1} d = c$. Its degree matrix is therefore D_y. We use the three explicitly distinguished symmetric operators

$$L_y = I_k - T^\top T, \qquad L_d = I_n - TT^\top, \qquad L_{\mathcal{B}} = I_{n+k} - \begin{bmatrix} 0 & T \\ T^\top & 0 \end{bmatrix}. \tag{3}$$

The point-side affinity associated with L_d is $W_x = B D_y^{-1} B^\top$, whose row-sum degree matrix is D_x. Eigenvalues of a symmetric Laplacian are ordered increasingly throughout. The identity $L = I - A$ pairs an eigenvalue λ of L with the eigenvalue $1 - \lambda$ of A associated with the same eigenvector; it does not pair arbitrarily ordered eigenvalues.

Exactness of transfer cut. If $T^\top T v = \sigma^2 v$, $\|v\|_2 = 1$, and $\sigma > 0$, define $u = Tv/\sigma$. Then $\|u\|_2 = 1$ and

$$\begin{bmatrix} 0 & T \\ T^\top & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \sigma \begin{bmatrix} u \\ v \end{bmatrix}.$$

The eigenvalues are $1 - \sigma^2$ for L_y, L_d and $\gamma = 1 - \sigma$ for $L_{\mathcal{B}}$, with $1 - \sigma^2 = \gamma(2 - \gamma)$. The $k \times k$ problem exactly recovers these modes of the same bipartite operator, not the spectrum of an arbitrary original point graph. Forming the reduced matrix costs $\mathcal{O}(nk^2)$; its dense eigendecomposition costs $\mathcal{O}(k^3)$.

Lemma C.1 (A threshold condition for component recovery)**.** *Let a finite sample $\mathcal{Z}$ be partitioned into $k \geq 2$ nonempty sets $\mathcal{Z}_1, \ldots, \mathcal{Z}_k$, and let κ be the symmetric nonnegative affinity actually used to construct the sampled graph. Define*

$$b_* = \max_{r \neq s} \max_{x \in \mathcal{Z}_r,\, y \in \mathcal{Z}_s} \kappa(x, y), \qquad \beta_* = \min_{r} \max_{\mathcal{T} \in \mathrm{ST}(\mathcal{Z}_r)} \min_{\{x,y\} \in E(\mathcal{T})} \kappa(x, y),$$

where $\mathrm{ST}(\mathcal{Z}_r)$ is the set of spanning trees on $\mathcal{Z}_r$, and the connectivity level of a singleton is $+\infty$. For the graph retaining $x \neq y$ when $\kappa(x, y) \geq \vartheta$, the components are exactly the prescribed sets if and only if

$$b_* < \vartheta \leq \beta_*. \tag{4}$$

Consequently such a threshold exists if and only if $b_* < \beta_*$.

Proof. The inequality $b_* < \vartheta$ removes every between-set edge. If $\vartheta \leq \beta_*$, each nonsingleton set has a spanning tree all of whose edges are retained, so each induced graph remains connected. Conversely, exact component recovery requires every between-set affinity to be less than ϑ. Each recovered component contains a spanning tree of retained edges, whose smallest affinity is at least ϑ. This yields both inequalities in equation 4. □

Assumption 3.1 expresses a local nearest-neighbor separation property. Lemma C.1 states the additional connectivity condition needed for recovery by a single global threshold; the local property alone does not imply this condition. For sampled data, the condition concerns the sampled affinity, not merely the full-data graph. If additional components remain, the largest-k rule selects one pure representative $\mathbb{G}_r \subseteq \mathbb{C}_r$ per cluster provided these k representatives are nonempty components and

$$\min_{r} |\mathbb{G}_r| > \max_{\mathcal{Q} \notin \{\mathbb{G}_1, \ldots, \mathbb{G}_k\}} |\mathcal{Q}|,$$

where the maximum is zero when no additional component exists. This follows directly by ordering component sizes. Neither purity nor this size condition is implied by the nearest-neighbor comparison alone.

Theorem C.2 (Exact recovery from distributional affinity profiles)**.** *Let $H \in \{0, 1\}^{n \times k}$ be the membership matrix of k nonempty clusters. Suppose the actual point–distribution affinity satisfies*

$$B = \mathrm{diag}(a) H C, \qquad a_i > 0, \qquad C \in \mathbb{R}^{k \times k}_{\geq 0}. \tag{5}$$

Assume C has no zero row or column and no two rows are positive scalar multiples of one another. Let $V \in \mathbb{R}^{k \times k}$ contain all k orthonormal eigenvectors of $T^\top T$. The idealized transfer embedding of D-SPEC, with the numerical denominator offset omitted, is

$$Y = \mathrm{rownorm}(D_x^{-1} B D_y^{-1/2} V).$$

Its rows are constant within each cluster and take exactly k distinct values. Any global optimum of k-means with k nonempty clusters therefore recovers the partition up to permutation.

If $C = I_k$, then $Y = HV$, $W_y = D_y$, and $T^\top T = I_k$. The reduced Laplacian is zero, and the normalized bipartite Laplacian has exactly k zero eigenvalues. This special case is the precise zero-eigenvalue setting of Theorem 3.2.

Proof. The conditions on C imply positive row sums $q = C\mathbf{1}_k$ and no zero column. Since every cluster is nonempty, all row and column degrees of B are positive. If $x_i \in \mathbb{C}_r$, then $d_i = a_i q_r$ and

$$(D_x^{-1} B D_y^{-1/2} V)_{i,:} = q_r^{-1} e_r^\top C D_y^{-1/2} V.$$

Positive row scaling disappears under row normalization, and multiplication by the orthogonal matrix V preserves row norms. Hence

$$Y = H\, \mathrm{rownorm}(C D_y^{-1/2}) V. \tag{6}$$

Right multiplication by the invertible positive diagonal matrix $D_y^{-1/2}$ preserves whether two rows are positive scalar multiples. The assumed distinct row directions of C therefore give distinct normalized rows of $C D_y^{-1/2}$. Assigning each of these k distinct values to its own cluster gives zero k-means objective. A zero-objective cluster cannot contain two different row values. With exactly k nonempty

clusters, every global minimizer must therefore recover the stated partition. This is a property of the objective's global optimum, not a guarantee that an arbitrary Lloyd initialization finds it.

For $C = I_k$, put $s_r = \sum_{i:x_i \in \mathbb{C}_r} a_i$. Then $D_x = \mathrm{diag}(a)$ and $D_y = \mathrm{diag}(s)$, so

$$W_y = B^\top D_x^{-1} B = H^\top \mathrm{diag}(a) H = D_y, \qquad T^\top T = I_k.$$

Equation equation 6 reduces to $Y = HV$. Each distribution vertex is joined with positive weight to exactly the points of its own cluster, giving exactly k bipartite components. Alternatively, the k singular values of T are all one, so the normalized bipartite adjacency has eigenvalue one with multiplicity k; its Laplacian has zero with the same multiplicity. □

The matrix V may be any valid orthogonal eigenbasis. Thus literal one-hot rows are a choice of basis, whereas recovery up to a common orthogonal transformation is basis-independent. Moreover, the general profile model equation 5 permits nonzero between-cluster affinities while retaining exact partition recovery; it need not have k zero eigenvalues. Nonsingularity of C is sufficient but not necessary: the recovery argument also permits rank-deficient profiles with distinct row directions. It uses all k right eigenvectors and no division by singular values. A positive spectral gap is a separate condition used only in the subspace stability result below.

For the kernel mean construction,

$$B_{ir} = \frac{1}{|\mathbb{G}_r|} \sum_{z \in \mathbb{G}_r} \kappa(x_i, z).$$

The diagonal-profile case follows if these affinities vanish for every incorrect representative and are positive for the correct one. Purity of $\mathbb{G}_r$ alone does not imply that every point has positive affinity to it. Likewise, deleting edges in the sampled threshold graph does not set their values to zero in the unthresholded kernel used in B. Representative coverage, threshold connectivity, and distribution-profile structure are therefore separate conditions. The recovery conclusion preserves cluster membership, not all edge weights of the original graph.

Theorem C.3 (Perturbation of the normalized distributional operator)**.** *Let $B, \widetilde{B} = B + E \in \mathbb{R}^{n \times k}_{\geq 0}$ have the same query points and matched distribution columns. Assume the degrees in equation 2 are positive. Define perturbed degrees and operators by the same formulas, and set*

$$\begin{aligned} \alpha_x &= \max_i \frac{|(E\mathbf{1}_k)_i|}{d_i} < 1, & \alpha_y &= \max_j \frac{|(E^\top \mathbf{1}_n)_j|}{c_j} < 1, \\ q_E &= \|D_x^{-1/2} E D_y^{-1/2}\|_2, & g &= [(1-\alpha_x)(1-\alpha_y)]^{-1/2}, \\ \eta &= g(1+q_E) - 1, & \varepsilon_d &= \min\{1, 2\eta\}. \end{aligned} \tag{7}$$

Then

$$\|\widetilde{L}_d - L_d\|_2 \leq \varepsilon_d, \qquad \max_{1 \leq i \leq n} |\lambda_i(\widetilde{L}_d) - \lambda_i(L_d)| \leq \varepsilon_d. \tag{8}$$

Both Laplacians here are $n \times n$ point-side operators.

Proof. The degree conditions ensure positive perturbed degrees. Put

$$Q = D_x^{-1/2} E D_y^{-1/2}, \quad R = \widetilde{D}_x^{-1/2} D_x^{1/2}, \quad S = D_y^{1/2} \widetilde{D}_y^{-1/2}.$$

Then $\widetilde{T} = R(T+Q)S$. With $r = (1-\alpha_x)^{-1/2}$ and $s = (1-\alpha_y)^{-1/2}$, diagonal scaling gives

$$\|R\|_2 \leq r, \quad \|S\|_2 \leq s, \quad \|R - I\|_2 \leq r - 1, \quad \|S - I\|_2 \leq s - 1.$$

For any $z \in \mathbb{R}^k$, weighted Cauchy–Schwarz yields

$$\|Tz\|_2^2 = \sum_i \frac{(\sum_j B_{ij} z_j / \sqrt{c_j})^2}{d_i} \leq \sum_{i,j} B_{ij} \frac{z_j^2}{c_j} = \|z\|_2^2.$$

Hence $\|T\|_2 \leq 1$, and the same argument gives $\|\widetilde{T}\|_2 \leq 1$. Expanding the difference gives

$$\begin{aligned} \|\widetilde{T} - T\|_2 &\leq \|RQS\|_2 + \|(R-I)TS\|_2 + \|T(S-I)\|_2 \\ &\leq rsq_E + (r-1)s + (s-1) = g(1+q_E) - 1 = \eta. \end{aligned}$$

Consequently,

$$\|\widetilde{L}_d - L_d\|_2 \le (\|\widetilde{T}\|_2 + \|T\|_2)\|\widetilde{T} - T\|_2 \le 2\eta.$$

Also $0 \preceq TT^\top, \widetilde{T}\widetilde{T}^\top \preceq I_n$, so their difference has operator norm at most one. This proves the first bound; Weyl's inequality (Weyl, 1912) proves the second. □

Corollary C.4 (Contamination of kernel mean representatives)**.** *Fix the fitted kernel, the query points, and a correspondence between the distribution columns. Assume* $0 \le \kappa(x, z) \le K_{\max}$ *and*

$$B_{ij} = \mathbb{E}_{P_j}\kappa(x_i, Z), \qquad \widetilde{P}_j = (1-\rho_j)P_j + \rho_j Q_j, \qquad 0 \le \rho_j \le 1.$$

Then $|E_{ij}| \le K_{\max}\rho_j$*. In particular,*

$$\begin{aligned}
\alpha_x \le \bar{\alpha}_x &:= \frac{K_{\max}\sum_j \rho_j}{\min_i d_i}, \\
\alpha_y \le \bar{\alpha}_y &:= nK_{\max}\max_j \frac{\rho_j}{c_j}, \\
q_E \le \bar{q}_E &:= K_{\max}\sqrt{\left(\sum_i d_i^{-1}\right)\left(\sum_j \rho_j^2/c_j\right)}.
\end{aligned} \tag{9}$$

If $\bar{\alpha}_x, \bar{\alpha}_y < 1$*, substituting these upper bounds into equation 7 gives a valid spectral perturbation certificate.*

Proof. The mixture identity gives $E_{ij} = \rho_j(\mathbb{E}_{Q_j}\kappa(x_i, Z) - \mathbb{E}_{P_j}\kappa(x_i, Z))$. Both expectations lie in $[0, K_{\max}]$, proving the entrywise bound. Summing absolute values gives the row- and column-degree bounds. Finally,

$$q_E^2 \le \|D_x^{-1/2} E D_y^{-1/2}\|_F^2 = \sum_{i,j}\frac{E_{ij}^2}{d_i c_j} \le K_{\max}^2\left(\sum_i d_i^{-1}\right)\left(\sum_j \rho_j^2/c_j\right).$$

The expression in equation 7 is nondecreasing in these three arguments, justifying substitution. □

For an empirical representative of size m_j, replacing r_j samples gives the same entrywise bound with $\rho_j = r_j/m_j$: write the difference of the two means as m_j^{-1} times the sum of the r_j changed kernel values, each of magnitude at most $K_{\max}$. This is a deterministic averaging guarantee. It does not require all individual points to satisfy Assumption 3.1. The fixed-kernel and matched-column conditions are essential: if thresholding or resampling changes representatives discontinuously, Theorem C.3 still applies to the resulting E, but the contamination bound requires a separate justification for that E.

Theorem C.5 (Spectral-subspace stability)**.** *Assume* $n > k$*. Let* $U, \widetilde{U} \in \mathbb{R}^{n\times k}$ *contain orthonormal eigenvectors for the* k *smallest eigenvalues of* $L_d, \widetilde{L}_d$*. Under Theorem C.3, suppose*

$$\delta_d = \lambda_{k+1}(L_d) - \lambda_k(L_d) > \varepsilon_d.$$

Then

$$\|\sin\Theta(U, \widetilde{U})\|_2 \le \frac{\varepsilon_d}{\delta_d - \varepsilon_d}. \tag{10}$$

If $\varepsilon_d \le \delta_d/2$*, the bound is at most* $2\varepsilon_d/\delta_d$*. If* T *has full column rank, then* $\delta_d = \sigma_k(T)^2$*. In the diagonal-profile case of Theorem C.2,* $\delta_d = 1$*.*

Proof. Write $F = \widetilde{L}_d - L_d$. Let $U_\perp$ contain the remaining clean eigenvectors, with eigenvalue matrix $\Lambda_\perp$, and let $\widetilde{\Lambda}$ contain the selected perturbed eigenvalues. The eigenvalue equations imply, for $X = U_\perp^\top \widetilde{U}$,

$$\Lambda_\perp X - X\widetilde{\Lambda} = -U_\perp^\top F\widetilde{U}.$$

Weyl's inequality gives $\lambda_{\max}(\widetilde{\Lambda}) \le \lambda_k(L_d) + \varepsilon_d$. Set $h = \delta_d - \varepsilon_d > 0$ and $\zeta = \lambda_{\max}(\widetilde{\Lambda})$. Then

$$M = \Lambda_\perp - \zeta I \succeq hI, \qquad N = \zeta I - \widetilde{\Lambda} \succeq 0.$$

The Sylvester equation has the convergent integral solution

$$X = -\int_0^\infty e^{-tM}(U_\perp^\top F\widetilde{U})e^{-tN}\,dt.$$

Indeed, differentiating the integrand verifies the equation, and the boundary term at infinity vanishes because $M \succeq hI$. Therefore

$$\|X\|_2 \le \int_0^\infty e^{-th}\|F\|_2\,dt \le \varepsilon_d/h.$$

Since $\|U_\perp^\top \widetilde{U}\|_2 = \|\sin\Theta(U,\widetilde{U})\|_2$, this proves equation 10. Finally, full column rank gives k positive eigenvalues $\sigma_1(T)^2,\dots,\sigma_k(T)^2$ of $TT^\top$ and $n-k$ zeros. Thus $\lambda_{k+1}(L_d) = 1$ and $\lambda_k(L_d) = 1-\sigma_k(T)^2$. The special case follows from $T^\top T = I_k$. □

This is a Davis–Kahan-type bound with the spectral separation made explicit; see also Yu et al. (2015) for general variants. The denominator accounts for the displacement of the selected perturbed eigenvalues; the simpler clean-gap bound follows under the stated half-gap condition.

Conditional comparison with SC. For symmetric nonnegative SC affinities $W, \widetilde{W} = W + F_s$ on the same n points, set $d_s = W\mathbf{1}_n > 0$, $D_s = \mathrm{diag}(d_s)$, and

$$L_s = I_n - D_s^{-1/2} W D_s^{-1/2}, \qquad \alpha_s = \max_i \frac{|(F_s\mathbf{1}_n)_i|}{(d_s)_i} < 1,$$
$$q_s = \|D_s^{-1/2} F_s D_s^{-1/2}\|_2, \qquad \varepsilon_s = \min\left\{2, \frac{q_s+\alpha_s}{1-\alpha_s}\right\}.$$

Then $\|\widetilde{L}_s - L_s\|_2 \le \varepsilon_s$. To see this, set $A = D_s^{-1/2} W D_s^{-1/2}$, $Q_s = D_s^{-1/2} F_s D_s^{-1/2}$, and $R_s = \widetilde{D}_s^{-1/2} D_s^{1/2}$. The perturbed normalized affinity is $R_s(A+Q_s)R_s$, while $\|A\|_2 \le 1$. With $r_s = (1-\alpha_s)^{-1/2}$, expansion as above gives

$$\|\widetilde{L}_s - L_s\|_2 \le r_s^2 q_s + (r_s-1)r_s + (r_s-1) = \frac{q_s+\alpha_s}{1-\alpha_s}.$$

The trivial bound two follows because both normalized affinities have operator norm at most one.

Let $\delta_s = \lambda_{k+1}(L_s) - \lambda_k(L_s)$ and suppose $\varepsilon_d < \delta_d$ and $\varepsilon_s < \delta_s$. If

$$\varepsilon_d \le \varepsilon_s, \qquad \frac{\varepsilon_d}{\delta_d} \le \frac{\varepsilon_s}{\delta_s}, \tag{11}$$

then the D-SPEC certificates are no larger than the SC certificates: Weyl gives the eigenvalue comparison, and monotonicity of $r/(1-r)$ on $[0,1)$ gives

$$\frac{\varepsilon_d}{\delta_d - \varepsilon_d} \le \frac{\varepsilon_s}{\delta_s - \varepsilon_s}.$$

Condition equation 11 depends on affinity perturbations, degrees, and clean gaps, with D-SPEC terms bounded by Corollary C.4. It compares certificates, not realized errors for every possible perturbation.

Proposition C.6 (Stability of the implemented row-normalized embedding)**.** *Under Theorem C.3, let $V, \widetilde{V}$ contain all k orthonormal right-side eigenvectors, and define*

$$R_B = BD_y^{-1/2}, \qquad Z = \mathrm{rownorm}(R_B), \qquad Y = ZV,$$

with perturbed counterparts. Write $h_i = (R_B)_{i,:}$, $e_i = E_{i,:}$, and $s = (1-\alpha_y)^{-1/2}$. Then

$$\min_{O^\top O = I_k} \|\widetilde{Y}O - Y\|_F \le \left(\sum_{i=1}^n b_i^2\right)^{1/2}, \quad b_i = \min\left\{2,\ 2\left[(s-1) + s\frac{\|e_i D_y^{-1/2}\|_2}{\|h_i\|_2}\right]\right\}. \tag{12}$$

This statement concerns the mathematical row normalization; the small denominator offset used solely for numerical stabilization is omitted.

Proof. All rows of B and $\widetilde{B}$ are nonzero. Positive multiplication by D_x^{-1} disappears under row normalization, and orthogonal right multiplication preserves row norms. The implemented embedding is therefore exactly $Y = \mathrm{rownorm}(BD_y^{-1/2})V$. Let $S = D_y^{1/2}\widetilde{D}_y^{-1/2}$ as in the perturbation proof. Then $\widetilde{h}_i = (h_i + e_i D_y^{-1/2})S$, so

$$\|\widetilde{h}_i - h_i\|_2 \le (s-1)\|h_i\|_2 + s\|e_i D_y^{-1/2}\|_2.$$

For nonzero vectors u, v, the triangle and reverse-triangle inequalities give

$$\left\| \frac{u}{\|u\|_2} - \frac{v}{\|v\|_2} \right\|_2 \le \min\left\{2, \frac{2\|u-v\|_2}{\|v\|_2}\right\}.$$

Apply this row by row and take $O = \widetilde{V}^\top V$. Then $\widetilde{Y}O - Y = (\widetilde{Z} - Z)V$; orthogonal invariance of the Frobenius norm proves equation 12. □

Unlike the eigenspace bound equation 10, equation 12 directly handles row normalization and right-eigenbasis rotation. Neither stability bound alone guarantees unchanged k-means assignments without an additional cluster-separation condition. The next subsection supplies this missing link for the labels actually returned by k-means.

Numerical normalization offset. The denominator offset can also be included explicitly. Let $\widetilde{a}_i = (\widetilde{D}_x^{-1}\widetilde{B}\widetilde{D}_y^{-1/2}\widetilde{V})_{i,:}$ and $r_i = \|\widetilde{a}_i\|_2 > 0$. With offset $\zeta \ge 0$, the implemented row is $\widetilde{a}_i/(r_i+\zeta)$ instead of $\widetilde{a}_i/r_i$, whose difference has norm $\zeta/(r_i+\zeta)$. Thus for the offset-normalized output $\widetilde{Y}_\zeta$,

$$\min_{O^\top O = I_k} \|\widetilde{Y}_\zeta O - Y\|_F \le \left(\sum_i b_i^2\right)^{1/2} + \left(\sum_i \left(\frac{\zeta}{r_i+\zeta}\right)^2\right)^{1/2}. \tag{13}$$

This follows from the triangle inequality and orthogonal invariance; the second term can be included in ε_Y below when using the objective of the offset-normalized implementation.

C.2 From embedding stability to clustering error

Theorem C.7 (A certificate for the returned clustering)**.** *Let $Y = HM \in \mathbb{R}^{n\times k}$ be an ideal embedding with $k \ge 2$ nonempty clusters and distinct row centers $\mu_1, \dots, \mu_k$ of M. Write z_i for the true cluster of point i, and define*

$$\Delta = \min_{r\ne s} \|\mu_r - \mu_s\|_2 > 0, \qquad n_{\min} = \min_r |\mathbb{C}_r|.$$

Suppose $\min_{O^\top O = I_k} \|\widetilde{Y}O - Y\|_F \le \varepsilon_Y$. For any returned labels $\widehat{z}_i \in \{1,\dots,k\}$ and centers $\widehat{\mu}_1, \dots, \widehat{\mu}_k$, let

$$J = \sum_i \|\widetilde{Y}_{i,:} - \widehat{\mu}_{\widehat{z}_i}\|_2^2, \qquad A = \frac{4(\varepsilon_Y + \sqrt{J})^2}{\Delta^2}.$$

If $A < n_{\min}$, a permutation π of the cluster labels satisfies

$$|\{i : \pi(\widehat{z}_i) \ne z_i\}| \le A. \tag{14}$$

Thus the misclustering rate is at most A/n. This statement uses the achieved objective, not an assumption of globally optimal k-means.

Proof. The minimum over the compact orthogonal group is attained. Choose an alignment O achieving the embedding bound and put $X = \widetilde{Y}O$, $\widehat{Z}_{i,:} = \widehat{\mu}_{\widehat{z}_i}O$. Rotating the points and centers preserves their assignments and objective. Hence

$$\|\widehat{Z} - Y\|_F \le \|\widehat{Z} - X\|_F + \|X - Y\|_F \le \sqrt{J} + \varepsilon_Y.$$

Let $\mathcal{E} = \{i : \|\widehat{Z}_{i,:} - \mu_{z_i}\|_2 \ge \Delta/2\}$. Summing squared errors gives $|\mathcal{E}| \le 4\|\widehat{Z} - Y\|_F^2/\Delta^2 \le A$. Since $A < n_{\min}$, every true cluster has a point outside $\mathcal{E}$. Two such points from different true clusters cannot share a returned label: their common fitted center would imply

$$\|\mu_{z_i} - \mu_{z_j}\|_2 \le \|\mu_{z_i} - \widehat{Z}_{i,:}\|_2 + \|\widehat{Z}_{j,:} - \mu_{z_j}\|_2 < \Delta,$$

contradicting separation. Each true cluster therefore uses at least one distinct returned label among points outside $\mathcal{E}$. There are only k labels, so these correspondences form a bijection. Under its permutation, all misclassified points belong to $\mathcal{E}$, proving equation 14. □

Corollary C.8 (Approximate global k-means)**.** *In Theorem C.7, suppose $J \le \xi\,\mathrm{OPT}_k(\widetilde{Y})$ for $\xi \ge 1$. If $4(1+\sqrt{\xi})^2\varepsilon_Y^2/\Delta^2 < n_{\min}$, then*

$$\min_{\pi} \frac{|\{i : \pi(\widehat{z}_i) \neq z_i\}|}{n} \le \frac{4(1+\sqrt{\xi})^2\varepsilon_Y^2}{n\Delta^2}. \tag{15}$$

A right-hand side strictly below $1/n$ guarantees exact recovery.

Proof. For the aligned embedding X, the ideal assignments and centers $Y = HM$ are a feasible k-means candidate. Orthogonal invariance gives $\mathrm{OPT}_k(\widetilde{Y}) = \mathrm{OPT}_k(X) \le \|X - Y\|_F^2 \le \varepsilon_Y^2$. Thus $\sqrt{J} \le \sqrt{\xi}\varepsilon_Y$, and substitution into Theorem C.7 proves the bound. The last statement uses the integer-valued number of errors. □

For the distributional profile model, the ideal centers are the rows of $M = \mathrm{rownorm}(CD_y^{-1/2})V$, and Proposition C.6 supplies $\varepsilon_Y = (\sum_i b_i^2)^{1/2}$, with the additional term in equation 13 if a denominator offset is used. In the diagonal-profile case $C = I_k$, $\Delta = \sqrt{2}$. A global k-means solution then has misclustering rate at most $8\varepsilon_Y^2/n$ whenever $8\varepsilon_Y^2 < n_{\min}$. An arbitrary Lloyd run need not have a known approximation factor ξ; its achieved residual J can instead be used in Theorem C.7. The ideal separation and perturbation assumptions remain necessary for either certificate.

C.3 Finite-partition guarantees for adaptive representatives

The following result concerns the random feature construction, not an i.i.d. sampling model for the selected connected components. In the Isolation Kernel implementation (Ting et al., 2023b), each partition contributes a binary feature block with at most one nonzero entry per point; points outside its retained regions have a zero block. Scaling the concatenated features by $1/\sqrt{t}$ makes their inner product an average of t co-membership indicators.

Theorem C.9 (Uniform finite-partition approximation)**.** *Condition on the fixed data $\mathbb{X} = \{x_1, \dots, x_n\}$, a fixed reference set, and a fixed ψ. Suppose the t partitions are independently generated with the same conditional law. Let*

$$h_\ell(x,z) = \sum_{r=1}^{\psi} \phi_{\ell r}(x)\phi_{\ell r}(z) \in \{0,1\}, \qquad \kappa_t(x,z) = \frac{1}{t}\sum_{\ell=1}^{t} h_\ell(x,z),$$

and let $\kappa_\infty(x,z)$ be the conditional expectation of $h_1(x,z)$. For $0 < \delta < 1$, set

$$\epsilon_t = \sqrt{\frac{\log(n(n+1)/\delta)}{2t}}.$$

With conditional probability at least $1-\delta$,

$$\max_{i,j} |\kappa_t(x_i,x_j) - \kappa_\infty(x_i,x_j)| \le \epsilon_t. \tag{16}$$

On this event, simultaneously for all probability vectors u, w on the data, including those selected using these same partitions,

$$|u^\top(K_t - K_\infty)w| \le \epsilon_t, \qquad K_{t,ij} = \kappa_t(x_i,x_j). \tag{17}$$

Proof. For a fixed pair, Hoeffding's inequality (Hoeffding, 1963) for independent variables in $[0,1]$ gives

$$\Pr\{|\kappa_t(x_i,x_j) - \kappa_\infty(x_i,x_j)| > \epsilon\} \le 2e^{-2t\epsilon^2}.$$

A union bound over the $n(n+1)/2$ unordered pairs, including diagonals, gives failure probability at most $n(n+1)e^{-2t\epsilon^2}$. Substitution of ϵ_t proves equation 16. Once that event holds, nonnegativity and unit sums imply

$$|u^\top(K_t - K_\infty)w| \le \sum_{i,j} u_i w_j \epsilon_t = \epsilon_t.$$

This deterministic inequality holds for every u, w on the same event, so they may be selected adaptively from K_t. □

Taking u to select one query point and w to be uniform on a nonempty representative $\mathbb{G}_j$ gives

$$\left| \frac{1}{|\mathbb{G}_j|} \sum_{z \in \mathbb{G}_j} [\kappa_t(x_i, z) - \kappa_\infty(x_i, z)] \right| \le \epsilon_t.$$

Thus there is no additional union bound over exponentially many possible representatives. This compares the *same selected set* under two kernels; it does not assume that a selected component is an unbiased i.i.d. sample from a population cluster. If ψ is chosen from a fixed finite candidate set Ψ, a further union bound replaces $n(n+1)$ by $|\Psi|n(n+1)$ and permits adaptive choice among these candidates.

Corollary C.10 (Stability of the implemented relative threshold)**.** *On event equation 16, consider a sampled set $\mathcal{Z} \subseteq \mathbb{X}$ with prescribed nonempty parts $\mathcal{Z}_1, \dots, \mathcal{Z}_k$. Let $b_{\mathcal{Z}}$ and $\beta_{\mathcal{Z}}$ be the between-part maximum and within-part connectivity level of Lemma C.1, computed using κ_∞. Set*

$$M_{\mathcal{Z}} = \max_{x,z \in \mathcal{Z}} \kappa_\infty(x, z), \qquad M_{t,\mathcal{Z}} = \max_{x,z \in \mathcal{Z}} \kappa_t(x, z).$$

For $0 \le \tau \le 1$, retain distinct points when $\kappa_t(x, z) \ge \vartheta_t = \tau M_{t,\mathcal{Z}}$. If

$$b_{\mathcal{Z}} + (1+\tau)\epsilon_t < \tau M_{\mathcal{Z}} \le \beta_{\mathcal{Z}} - (1+\tau)\epsilon_t, \tag{18}$$

then the components are exactly $\mathcal{Z}_1, \dots, \mathcal{Z}_k$.

Proof. Uniform approximation gives $|M_{t,\mathcal{Z}} - M_{\mathcal{Z}}| \le \epsilon_t$. Every between-part affinity satisfies

$$\kappa_t(x, z) \le b_{\mathcal{Z}} + \epsilon_t < \tau(M_{\mathcal{Z}} - \epsilon_t) \le \tau M_{t,\mathcal{Z}},$$

so no such edge survives. For each nonsingleton part, choose a maximum-bottleneck spanning tree under κ_∞. Every tree edge satisfies

$$\kappa_t(x, z) \ge \beta_{\mathcal{Z}} - \epsilon_t \ge \tau(M_{\mathcal{Z}} + \epsilon_t) \ge \tau M_{t,\mathcal{Z}}.$$

Each part therefore remains connected. The implication is deterministic on equation 16, including for adaptively chosen samples. □

The factor $1 + \tau$ accounts for the random scaling in the actual threshold $\tau \max K_t$. Its diagonal entries need not be one, since a point may fall outside a partition's retained regions. For a fixed absolute threshold ϑ, the sufficient condition simplifies to $b_{\mathcal{Z}} + \epsilon_t < \vartheta \le \beta_{\mathcal{Z}} - \epsilon_t$. In particular, with margin $\gamma_{\mathcal{Z}} = \min\{\tau M_{\mathcal{Z}} - b_{\mathcal{Z}}, \beta_{\mathcal{Z}} - \tau M_{\mathcal{Z}}\} > 0$, the sufficient requirement is

$$t > \frac{(1+\tau)^2}{2\gamma_{\mathcal{Z}}^2} \log \frac{n(n+1)}{\delta}.$$

This is a finite-partition sufficient bound, not a universal guarantee for any particular fixed choice of t.

Propagation to normalized operators and labels. For any selected k nonempty representatives, define B_t, B_∞ by averaging their respective kernels over the same sets. Then $|(B_t - B_\infty)_{ij}| \le \epsilon_t$ on the uniform event. With degrees and h_i formed from B_∞, the proofs above give

$$\alpha_x \le \frac{k\epsilon_t}{\min_i d_i}, \quad \alpha_y \le \frac{n\epsilon_t}{\min_j c_j}, \quad q_E \le \epsilon_t \sqrt{\Big(\sum_i d_i^{-1}\Big)\Big(\sum_j c_j^{-1}\Big)}.$$

Also $\|E_{i,:} D_y^{-1/2}\|_2 \le \epsilon_t \sqrt{\sum_j c_j^{-1}}$. When the relative-degree bounds are below one, these expressions can be substituted into Theorem C.3 and Proposition C.6. If the same representatives give an ideal profile model with separated centers, the resulting ε_Y further yields the label bound of Theorem C.7. These are implications on the common probability-$1 - \delta$ event; profile structure and center separation are explicit additional conditions, not consequences of concentration alone.

C.4 SAMPLING COVERAGE AND COMPONENT RECOVERY

Lemma C.11 (Cluster coverage under sampling without replacement)**.** *Let the k true clusters have sizes $n_1, \dots, n_k > 0$, and sample p points uniformly without replacement, with $1 \le p \le n$. For $\pi_{\min} = \min_r n_r/n$,*

$$\Pr\{\exists r : \mathcal{Z} \cap \mathbb{C}_r = \varnothing\} \le \sum_{r=1}^{k} \frac{\binom{n-n_r}{p}}{\binom{n}{p}} \le k e^{-p\pi_{\min}}, \tag{19}$$

where a binomial numerator is zero if $p > n - n_r$.

Proof. Exactly $\binom{n-n_r}{p}$ of the $\binom{n}{p}$ possible samples avoid cluster r. For $p \le n - n_r$, the ratio is

$$\prod_{\ell=0}^{p-1} \left(1 - \frac{n_r}{n-\ell}\right) \le (1 - n_r/n)^p \le e^{-pn_r/n}.$$

A union bound proves equation 19. □

Coverage alone does not guarantee sampled connectivity: sampling can remove bridge points. The next result makes a stronger, explicit sufficient separation assumption to give an unconditional probability bound over the stated sampling and partition randomness.

Corollary C.12 (An explicit sampling-and-partition guarantee)**.** *Under Theorem C.9 and Lemma C.11, fix $k \ge 2$ and $0 \le \tau \le 1$. All probabilities are conditional on the fixed data, reference set, and ψ of Theorem C.9; under this same conditioning, $\mathcal{Z}$ is a uniformly sampled p-element subset. Define the full-data quantities under κ_∞ by*

$$b = \max_{r \ne s} \max_{x \in \mathbb{C}_r, z \in \mathbb{C}_s} \kappa_\infty(x, z), \qquad a = \min_{r:|\mathbb{C}_r| \ge 2} \min_{x \ne z \in \mathbb{C}_r} \kappa_\infty(x, z),$$
$$q_- = \min_i \kappa_\infty(x_i, x_i), \quad q_+ = \max_i \kappa_\infty(x_i, x_i).$$

Set $a = +\infty$ if every cluster is a singleton. If

$$b + (1+\tau)\epsilon_t < \tau q_-, \qquad \tau q_+ \le a - (1+\tau)\epsilon_t,$$

the sampled relative-threshold graph has exactly the nonempty parts $\mathcal{Z} \cap \mathbb{C}_1, \dots, \mathcal{Z} \cap \mathbb{C}_k$ as its k connected components with probability at least $1 - \delta - \sum_r \binom{n-n_r}{p} / \binom{n}{p} \ge 1 - \delta - ke^{-p\pi_{\min}}$.

Proof. For each partition, $h_\ell(x, z) \le h_\ell(x, x)$, so for every nonempty sample $q_- \le M_{\mathcal{Z}} \le q_+$. If all true clusters are covered, their sampled between-part maximum is at most b, and each sampled nonsingleton part has a spanning tree with every edge of weight at least a. Thus the stated inequalities imply equation 18. On the uniform event, Corollary C.10 gives component recovery. A union bound for the failure of coverage and of the uniform event gives the probability claim; independence of these two events is not required. □

This last corollary is a sufficient regime with uniform within-cluster separation, not a replacement for the weaker spanning-tree condition of Lemma C.1. It guarantees the sampled components; recovery of all n labels additionally uses the distributional profiles and clustering-error conditions established above.

D COMPARISON OF EIGENVALUES BETWEEN D-SPEC AND U-SPEC

Table 14 reports the previously computed eigenvalue and eigenvector diagnostics for D-SPEC and U-SPEC under the stated noise settings. These original diagnostics are retained without redefinition; they differ from the basis-invariant subspace certificates above and do not directly verify the SC certificate comparison.

E PARAMETER SETTING

Datasets: spiral Chang & Yeung (2008), 4C Zuo & Hou (2022), AC Marin et al. (2019), RingG Ting et al. (2023a), complex9 Salvador & Chan (2004), cure-t2-4k Van Craenendonck & Blockeel (2017), landsat Wang et al. (2024), spambase Hopkins et al. (1999), waveform3 Wang et al. (2024), pendigits Alpaydin & Alimoglu (1996), usps Cai et al. (2011b), letters Frey & Slate (1991), mnist Cai et al. (2011a), skin Bhatt & Dhall (2009), covertype Blackard & Dean (1999).

In order to reduce the contingency introduced by hyper-parameter settings, we tried multiple hyper-parameters for each algorithm and reported the best results Xie et al. (2016); Saha et al. (2023); Nie et al. (2024); Tang et al. (2023). The hyperparameters of the algorithms used in the experiments are shown in Table 4.

Table 4: Parameter search ranges.

Algorithm	Parameter search ranges
SC	$\sigma \in \{2^q \mid q = -5, \ldots, 5\}$
U-SPEC	$K \in \{2, 3, 5, 7, 10, 15, 20, 30, 45, 60, 80\}$
DNCSC	$K \in \{2, 3, 5, 7, 10, 15, 20, 30, 45, 60, 80\}$
GBSC	$\sigma \in \{2^q \mid q = -5, \ldots, 5\}$
CSC	$\sigma \in \{2^q \mid q = -5, \ldots, 5\}$
D-SPEC	$\psi \in \{4, 8, 16, 32, 64, 128, 256, 512\}$ $\tau \in \{0.2, 0.25, \ldots, 0.8\}$

The inability to automatically determine parameters is a limitation of our algorithm, and we leave this as future work. Since NMI requires labels, in practice, one approach is to use label-independent metrics such as CH (Calinski-Harabasz)Caliński & Harabasz (1974) to select hyperparameters. Figure 10 shows the results of CH and NMI. CH and NMI have similar distributions: parameters with higher CH also tend to have higher NMI, and parameters with higher NMI also tend to have higher CH, especially when $\psi \geq 32$. Therefore, in practice, we can use unsupervised methods such as CH to select parameters. We also noticed that D-SPEC has better results in the band area where ψ ranges from 32 to 512 and τ ranges from 0.3-0.5 (ψ=32) to 0.1-0.3 (ψ=512).

F MAIN RESULTS IN TERMS OF ARI AND F-MEASURE

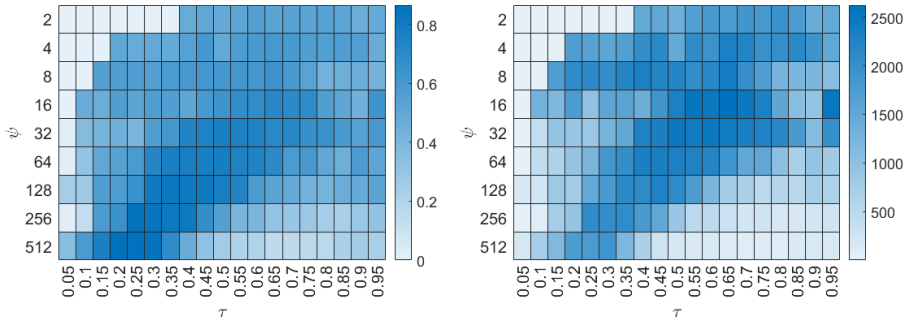
The results in terms of ARI are shown in Table 5, And the results in terms of F-measure are shown in Table 6.

G RESULTS ON LANDSAT, WAVEFORM3, AND COVERTYPE

D-SPEC is best or tied on 11/15 datasets and has an average rank 1.700. On the seven real datasets Landsat–MNIST, its macro-NMI is 0.581, above SC 0.432, U-SPEC 0.526, DNCSC 0.527, and FastSC 0.487.

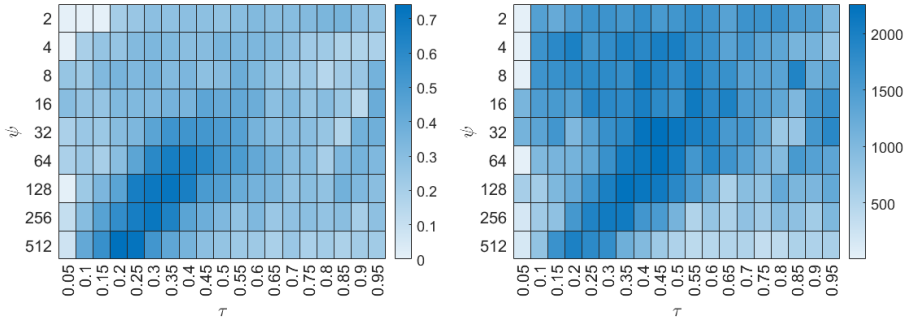
On Landsat, D-SPEC 0.647 is close to U-SPEC 0.668. On Waveform3, D-SPEC 0.406 exceeds SC/U-SPEC/DNCSC/GBSC (0.369–0.371) and is second only to FastSC 0.605. On Covertype, D-SPEC 0.218 exceeds U-SPEC 0.212 and DNCSC 0.086, again trailing FastSC. A method winning these three rows does not erase D-SPEC's overall rank and its $k \times k$ eigenproblem.

Waveform3 has the lowest strict pointwise rate 0.739 and edge homophily 0.773, while cluster imbalance affects Covertype's uniform sample. Our method is the best on Covertype, except for FastSC. FastSC performs better on this dataset because it constructs an n^2 graph, but this approach will not be suitable for larger datasets, such as the five largest datasets, where it will either be ineffective or take a long time.



(a) NMI on Pendigits

(b) CH on Pendigits



(c) NMI on Mnist

(d) CH on Mnist

Figure 10: The heatmap of CH and NMI on Pendigits and Mnist.

Table 5: Average ARI scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
spiral	312	2	3	**1.000**	**1.000**	**1.000**	0.612	0.004	**1.000**
4C	1000	2	4	**1.000**	**1.000**	**1.000**	0.562	0.399	**1.000**
AC	1004	2	2	**1.000**	**1.000**	**1.000**	0.202	0.656	**1.000**
RingG	1536	2	3	0.571	0.776	0.522	0.662	0.582	**0.987**
complex9	3031	2	9	**1.000**	0.932	0.896	0.691	0.391	**1.000**
cure-t2-4k	4200	2	7	0.889	0.920	0.869	0.707	0.620	**0.951**
landsat	2000	36	6	0.088	0.597	0.581	0.527	0.591	**0.598**
spambase	4601	57	2	0.005	0.001	0.019	0.029	0.160	**0.250**
waveform3	5000	21	3	0.253	0.252	0.267	**0.495**	0.252	0.377
pendigits	10992	16	10	0.573	0.724	0.680	0.482	0.425	**0.787**
usps	11000	256	10	0.455	0.505	0.510	0.325	0.189	**0.676**
letters	20000	16	26	0.020	0.179	0.187	0.045	O/M	**0.224**
mnist	70000	784	10	0.604	0.611	0.658	0.464	O/M	**0.677**
skin	245057	3	2	N/A	0.001	0.565	0.020	O/M	**0.878**
covertype	581012	54	7	N/A	0.089	0.008	**0.566**	O/M	0.167
Avg.score				0.574	0.572	0.584	0.426	0.388	**0.705**
Avg.rank				3.900	3.167	3.167	4.067	5.133	**1.567**

H Results of the Five Largest Datasets

The results in terms of NMI, ARI, F-measure, and runtime on the five largest datasets are shown in Table 7, Table 8, Table 9, Table 10, respectively.

Table 6: Average F-measure scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
spiral	312	2	3	**1.000**	**1.000**	**1.000**	0.667	0.394	**1.000**
4C	1000	2	4	**1.000**	**1.000**	**1.000**	0.550	0.440	**1.000**
AC	1004	2	2	**1.000**	**1.000**	**1.000**	0.657	0.909	**1.000**
RingG	1536	2	3	0.666	0.728	0.604	0.636	0.633	**0.987**
complex9	3031	2	9	**1.000**	0.939	0.903	0.587	0.522	**1.000**
cure-t2-4k	4200	2	7	0.772	0.811	0.832	0.579	0.759	**0.941**
landsat	2000	36	6	0.222	0.717	0.708	0.519	0.723	**0.735**
spambase	4601	57	2	0.374	0.391	0.428	0.377	0.320	**0.721**
waveform3	5000	21	3	0.509	0.517	0.535	0.525	0.510	**0.721**
pendigits	10992	16	10	0.692	0.819	0.758	0.553	0.611	**0.863**
usps	11000	256	10	0.534	0.583	0.590	0.433	0.135	**0.756**
letters	20000	16	26	0.183	0.341	0.335	0.189	O/M	**0.380**
mnist	70000	784	10	0.623	0.711	**0.750**	0.539	O/M	0.747
skin	245057	3	2	N/A	0.442	0.808	0.519	O/M	**0.953**
covertype	581012	54	7	N/A	0.267	0.164	**0.531**	O/M	0.262
Avg.score				0.660	0.684	0.694	0.524	0.541	**0.804**
Avg.rank				4.033	2.767	2.833	4.467	5.333	**1.567**

Table 7: Average NMI scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
TB1M	1M	2	2	N/A	0.957	0.970	0.001	O/M	**0.974**
SF2M	2M	2	4	N/A	0.799	0.818	0.001	O/M	**0.936**
CC5M	5M	2	3	N/A	**0.999**	0.998	N/A	O/M	**0.999**
CG10M	10M	2	11	N/A	0.809	0.841	N/A	O/M	**0.942**
Flower20M	20M	2	13	N/A	0.892	O/M	N/A	O/M	**0.963**

Table 8: Average ARI scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
TB1M	1M	2	2	N/A	0.981	0.988	0.001	O/M	**0.989**
SF2M	2M	2	4	N/A	0.748	0.903	0.001	O/M	**0.966**
CC5M	5M	2	3	N/A	**1.000**	0.999	N/A	O/M	0.999
CG10M	10M	2	11	N/A	0.525	0.913	N/A	O/M	**0.959**
Flower20M	20M	2	13	N/A	0.811	O/M	N/A	O/M	**0.966**

Table 9: Average F-measure scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
TB1M	1M	2	2	N/A	0.995	**0.997**	0.509	O/M	**0.997**
SF2M	2M	2	4	N/A	0.735	0.739	0.243	O/M	**0.966**
CC5M	5M	2	3	N/A	**1.000**	0.999	N/A	O/M	**1.000**
CG10M	10M	2	11	N/A	0.586	0.802	N/A	O/M	**0.974**
Flower20M	20M	2	13	N/A	0.836	O/M	N/A	O/M	**0.974**

Table 10: Runtime (seconds) over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
TB1M	1M	2	2	N/A	17.512	**5.179**	57.982	O/M	15.920
SF2M	2M	2	4	N/A	53.107	**13.201**	274.534	O/M	49.310
SF5M	5M	2	3	N/A	74.690	**20.174**	N/A	O/M	98.572
SF10M	10M	2	11	N/A	625.271	**245.621**	N/A	O/M	545.864
SF20M	20M	2	13	N/A	1258.085	O/M	N/A	O/M	**1245.407**

I Fundamental Limitations of Spectral Clustering

The results in terms of NMI, ARI, and F-measure on the three datasets are shown in Table 11, Table 12, and Table 13, respectively.

Table 11: Average NMI scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
DSSS	900	2	3	0.909	0.848	0.917	0.737	0.630	**0.937**
DSDS	900	2	3	0.797	0.752	0.787	0.559	0.525	**0.899**
OGOL	1400	2	2	0.454	0.878	0.515	0.595	0.538	**0.885**

Table 12: Average ARI scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
DSSS	900	2	3	0.937	0.819	0.945	0.604	0.479	**0.966**
DSDS	900	2	3	0.908	0.814	0.899	0.540	0.348	**0.973**
OGOL	1400	2	2	0.409	0.922	0.497	0.597	0.530	**0.936**

Table 13: Average F-measure scores over 10 runs. The best score in each dataset is highlighted in bold.

dataset	#n	#d	k	SC	U-SPEC	DNCSC	FastSC	GBSC	D-SPEC
DSSS	900	2	3	0.978	0.850	0.981	0.608	0.524	**0.987**
DSDS	900	2	3	0.547	0.540	0.597	0.632	0.601	**0.964**
OGOL	1400	2	2	0.813	0.980	0.849	0.737	0.861	**0.981**

J Comparison with Variations of the Methods in Main Text

In the tables in the main text, we present a variety of representative methods for reducing time complexity. Here, we compare some other recent methods: Spectral Bridges Laplante & Ambroise (2024), GGC Pei et al. (2026), and FSC-PLGB Cheng et al. (2026). They share the same intuition as the main text comparison. For example, Spectral Bridges and GBSC are similar in that they divide the space and then cluster it by region. FSC-PLGB is also based on Granular-Ball. The results in Table 15 indicate that the performance of these methods is inferior to D-SPEC.

K Comparison with Deep Spectral Clustering

We compare with the recently deep spectral clustering DSC-LSP Meng et al. (2025), the results in terms of NMI are shown in Table 16. When we run DSC-LSP on the dataset we used, we get a very

Table 14: The differences in eigenvalues and the distances between eigenvectors

ratio	**eigenvalues**						**eigenvectors**					
	4C		usps		pendigit		4C		usps		pendigit	
	U	D	U	D	U	D	U	D	U	D	U	D
0.04	0.049	0.005	0.635	0.027	0.022	0.008	1.622	1.446	1.891	1.772	1.825	0.681
0.08	0.040	0.007	0.634	0.250	0.020	0.003	1.723	1.224	1.723	1.129	1.654	0.636
0.12	0.069	0.016	0.629	0.113	0.023	0.007	1.906	1.605	1.815	0.913	1.834	1.397
0.16	0.079	0.029	0.631	0.047	0.029	0.005	1.795	1.503	1.875	1.281	1.952	1.548
0.20	0.079	0.022	0.633	0.166	0.033	0.007	1.958	1.423	1.853	1.351	1.800	1.377

D: D-SPEC. U: U-SPEC.

Table 15: Comparison across benchmark datasets.

NMI	Landsat	Spambase	Waveform3	Pendigits	USPS	Letters	MNISTAve
D-SPEC	0.647	**0.166**	**0.406**	**0.847**	**0.778**	**0.478**	0.746
U-SPEC	**0.668**	0.013	0.370	0.826	0.654	0.455	0.699
GGC	0.642	0.033	0.367	**0.847**	0.755	0.423	**0.862**
Spectral Bridges	0.635	0.076	0.391	0.801	0.708	0.461	0.788
FSC-PLGB	0.646	0.026	0.368	0.817	0.656	0.452	0.673

low NMI. We speculate this may be because we did not set the appropriate hyperparameters (there are 7 hyperparameters in DSC-LSP). Therefore, we compare it on the dataset used in the DSC-LSP paper and use the NMI reported in their paper. We find that DSC-LSP has better results for image data (such as mnist, fashion-mnist) because it can get better data representation, and spectral clustering based on this representation will be better, but for tabular data (such as mfeat-kar, isolet1234, and pendigits), DSC-LSP is not as good as D-SPEC.

Table 16: Results in terms of NMI scores compared with DSC-LSP.

Dataset	#n	#d	k	D-SPEC	DSC-LSP
mnist	70000	784	10	0.746	**0.919**
mnist-test	10000	784	10	0.773	**0.784**
fashion-mnist	70000	784	10	0.594	**0.693**
cnae	1080	856	9	0.597	**0.599**
msra25-uni	1799	256	12	**0.713**	0.704
mfeat-kar	2000	64	10	**0.818**	0.726
isolet1234	6238	617	26	**0.762**	0.714
pendigits	10992	16	10	**0.847**	0.786

L Comparison with Coreset Spectral Clustering

Recently, a coreset-based spectral clustering algorithm, CSC, was proposed Jourdan et al. (2025). CSC first obtains the coreset in the kernel space and then uses spectral clustering to cluster these coresets. After obtaining the k centers of the coreset, all points are assigned to the nearest center to complete the clustering. CSC just replaces the step of using k-means to obtain k centers in the coreset-based kernel k-means with using spectral clustering to obtain k centers. which can be replaced by any other algorithm (such as GMM, etc.) In addition, using spectral clustering to replace k-means will increase the time complexity from $O(n)$ to $O(n^2)$. We compare the methods of D-SPEC and CSC as shown in Table 17. The coreset ratio was set to 0.1 on all datasets. D-SPEC outperforms CSC on all datasets used.

Table 17: Results compared with CSC.

dataset	#n	#d	k	NMI		ARI		F1	
				D-SPEC	CSC	D-SPEC	CSC	D-SPEC	CSC
spiral	312	2	3	1.000	0.156	1.000	0.055	1.000	0.488
4C	1000	2	4	1.000	0.231	1.000	0.174	1.000	0.444
AC	1004	2	2	1.000	0.814	1.000	0.903	1.000	0.970
RingG	1536	2	3	0.987	0.596	0.987	0.569	0.987	0.607
complex9	3031	2	9	1.000	0.683	0.999	0.416	0.999	0.437
cure-t2-4k	4200	2	7	0.951	0.801	0.951	0.721	0.941	0.775
landsat	2000	36	6	0.647	0.506	0.598	0.369	0.735	0.623
spambase	4601	57	2	0.166	0.040	0.250	0.077	0.721	0.607
waveform3	5000	21	3	0.406	0.305	0.377	0.265	0.721	0.571
pendigits	10992	16	10	0.847	0.748	0.787	0.636	0.863	0.802
usps	11000	256	10	0.778	0.453	0.676	0.293	0.756	0.368
letters	20000	16	26	0.478	0.350	0.224	0.124	0.380	0.278
mnist	70000	784	10	0.747	0.420	0.677	0.187	0.747	0.260
skin	245057	3	2	0.767	NA	0.878	NA	0.953	NA
covertype	581012	54	7	0.218	NA	0.167	NA	0.262	NA

M Comparison of Distributions and Points

Point-landmark representations can lose accuracy when the representative budget is sharply reduced. Table 18 compares U-SPEC using $p = k^2$ point landmarks with D-SPEC's k distribution vertices on selected datasets with $k > 5$. The average NMI of U-SPEC decreases from 0.671 to 0.577, while D-SPEC achieves 0.708 with an $n \times k$ graph.

Table 18: A comparison of USPEC and USPEC$_k$ (setting $p = k \times k$) with D-SPEC.

	NMI			ARI			F1		
dataset	D-SPEC	USPEC	USPEC$_k$	D-SPEC	USPEC	USPEC$_k$	D-SPEC	USPEC	USPEC$_k$
complex9	1.000	0.971	0.770	0.999	0.932	0.530	0.999	0.939	0.620
curet24k	0.951	0.886	0.763	0.951	0.920	0.587	0.941	0.811	0.686
landsat	0.647	0.668	0.618	0.598	0.598	0.521	0.735	0.717	0.715
pendigits	0.847	0.826	0.774	0.787	0.724	0.670	0.863	0.819	0.792
usps	0.778	0.654	0.527	0.676	0.505	0.381	0.756	0.583	0.519
letters	0.478	0.455	0.445	0.224	0.179	0.177	0.380	0.341	0.348
mnist	0.747	0.699	0.573	0.677	0.611	0.460	0.747	0.711	0.602
covertype	0.218	0.212	0.144	0.167	0.089	0.034	0.262	0.267	0.211
Ave.Score	**0.708**	0.671	0.577	**0.635**	0.570	0.420	**0.710**	0.648	0.562

We compare the U-SPEC and D-SPEC on four high-dimensional datasets as shown in Table 19.

Table 19: Comparison of D-SPEC and U-SPEC on high-dimensional datasets.

		NMI			ARI		
Dataset	$n \times d, k$	D-SPEC	U-SPEC	Δ	D-SPEC	U-SPEC	Δ
UCI HAR	$10{,}299 \times 561, 6$	**.708**	.674	**+.034**	**.598**	.574	**+.024**
Abdominal CT	$58{,}830 \times 784, 11$	**.492**	.447	**+.045**	**.245**	.233	**+.012**
Gas Sensor Drift	$13{,}910 \times 128, 6$	**.578**	.008	**+.570**	**.475**	0.001	**+.474**
Sensorless Drive	$58{,}509 \times 48, 11$	**.421**	.338	**+.083**	**.233**	.033	**+.200**

D-SPEC improves NMI/ARI on all four; Abdominal CT combines 58,830 samples with 784 dimensions. These cross-modal gains are consistent with within-class subdistributions being better retained by distribution vertices than point landmarks.

N Comparison of D-SPEC under Different p

In all our experiments, we set p to 10000. Our results on the MNIST dataset, where p ranges from 1000 to 20000, are shown in Table 20: when $p = 1000$, the clustering results are lower than when $p = 10000$, but when $p \geq 2000$, the clustering results do not change much.

Table 20: A comparison of D-SPEC under different p.

p	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000	20000
NMI	0.696	0.737	0.748	0.749	0.747	0.747	0.747	0.757	0.750	0.747	0.769
ARI	0.582	0.619	0.679	0.643	0.661	0.661	0.659	0.649	0.652	0.677	0.657
F1	0.690	0.732	0.746	0.740	0.751	0.751	0.735	0.736	0.752	0.747	0.758

O Ablation of D-SPEC under Different Bounded Graphs

The performance of distribution-based bipartite graphs is significantly better than point-based methods. Table 21 compares the results of bounded graphs obtained using D-SPEC and k-means spectral clustering (SC). D-SPEC performs best, while k-means performs worst. This indicates that both bounded graphs and distribution-based bipartite graphs contribute to achieving good clustering results with eigenvalue decomposition only on $k \times k$ matrices.

Table 21: Comparison of D-SPEC under different bounded graphs.

	NMI			ARI			F1		
dataset	D-SPEC	k-means	SC	D-SPEC	k-means	SC	D-SPEC	k-means	SC
landsat	0.647	0.616	0.646	0.598	0.559	0.568	0.735	0.722	0.721
spambase	0.166	0.070	0.031	0.250	0.095	0.020	0.721	0.648	0.477
waveform3	0.406	0.368	0.361	0.377	0.256	0.293	0.721	0.609	0.645
pendigits	0.847	0.691	0.664	0.787	0.567	0.527	0.863	0.686	0.643
usps	0.778	0.510	0.658	0.676	0.313	0.514	0.756	0.502	0.577
letters	0.478	0.387	0.446	0.224	0.166	0.195	0.380	0.299	0.340
mnist	0.747	0.533	0.588	0.677	0.322	0.496	0.747	0.483	0.646
skin	0.767	0.278	0.580	0.878	0.229	0.642	0.953	0.663	0.877
covertype	0.218	0.124	0.172	0.167	0.067	0.116	0.262	0.218	0.230

P Ablation of D-SPEC between points-based and distribution-based bipartite graph based on bounded graph

We compared our D-SPEC with DS_b, which directly constructs a point-based bipartite graph from bounded graphs (sampling 100 points due to their large size) to construct a $100 \times n$ bipartite graph. DS_k, we directly compute the mean of the bounded graphs to obtain k centroids (rather than k distributions), and subsequently construct a bipartite graph involving these k centroids and the n data points.

The results shown in Table 22 demonstrate that our proposed bounded graph and the bipartite graph constructed based on distributions are effective.

Table 22: Comparison of D-SPEC between points-based and distribution-based bipartite graph based on bounded graph

dataset	NMI			ARI			F-measure		
	D-SPEC	DS_b	DS_k	D-SPEC	DS_b	DS_k	D-SPEC	DS_b	DS_k
landsat	0.647	0.632	0.632	0.598	0.572	0.572	0.735	0.740	0.736
spambase	0.166	0.257	0.236	0.250	0.353	0.324	0.721	0.787	0.770
waveform3	0.406	0.361	0.367	0.377	0.308	0.310	0.721	0.580	0.579
pendigits	0.847	0.674	0.691	0.787	0.538	0.575	0.863	0.704	0.736
usps	0.778	0.418	0.507	0.676	0.267	0.385	0.756	0.469	0.578
letters	0.478	0.328	0.408	0.224	0.108	0.164	0.380	0.238	0.312
mnist	0.747	0.458	0.527	0.677	0.340	0.431	0.747	0.527	0.591
skin	0.767	0.619	0.690	0.878	0.695	0.771	0.953	0.896	0.922
covertype	0.218	0.175	0.177	0.167	0.095	0.095	0.262	0.257	0.272

Q Ablation of Different Sampling Methods

Across all experiments, we sample p (set $p = 10000$) points by uniformly sampling from each dataset. Our empirical results demonstrate that D-SPEC, even with this straightforward sampling strategy, achieves performance superior to state-of-the-art algorithms. We further compared random uniform sampling with hybrid representative selection (used in USPEC and DNCSC) on three large-scale real-world datasets. As shown in Table 23, hybrid representative selection (D-SPEC$_{HRS}$) yields additional performance gains for D-SPEC in the mnist and skin datasets.

Table 23: Comparison of different sampling methods.

	NMI		ARI		F1	
dataset	D-SPEC	D-SPEC$_{HRS}$	D-SPEC	D-SPEC$_{HRS}$	D-SPEC	D-SPEC$_{HRS}$
mnist	0.747	0.770	0.677	0.692	0.747	0.792
skin	0.767	0.810	0.878	0.892	0.953	0.963
covertype	0.218	0.203	0.167	0.140	0.262	0.270

Table 24 reports results for varying values of p; clustering performance exhibits minimal variation beyond $p > 2000$, consistent with the behavior observed under random uniform sampling.

Table 24: Comparison of different sampling methods at different p values.

p	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
D-SPEC	0.696	0.737	0.748	0.749	0.747	0.747	0.747	0.757	0.750	0.747
D-SPEC$_{HRS}$	0.736	0.745	0.773	0.766	0.771	0.771	0.779	0.769	0.771	0.770

R Image Segmentation

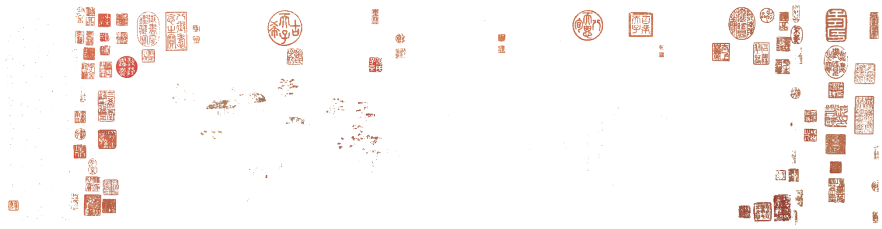
We show an example that uses D-SPEC to cluster images for image segmentation. We first convert the images (shown in Figure 11) from sRGB values to CIE 1976 Lab* values. (`https://en.wikipedia.org/wiki/CIELAB_color_space`). Then we cluster the images in LAB space. The segmented results of D-SPEC are shown in Figure 12. The results of U-SPEC and k-means are shown in Figure 13 and Figure 14 respectively. D-SPEC completely separates the rice paper, seal, and painting. U-SPEC separates the painting but not the rice paper and seal, while k-means fails to separate even the painting.



Figure 11: Zhao Mengfu's Autumn Colors.



(a) Cluster 1

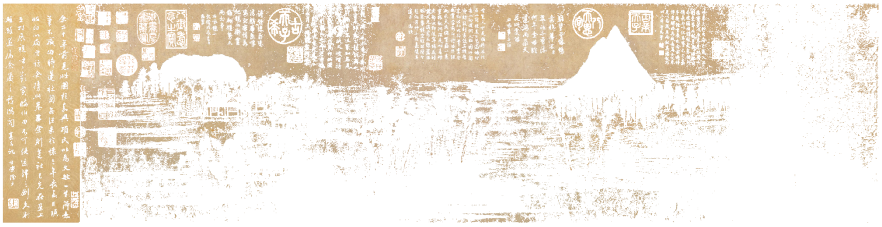


(b) Cluster 2

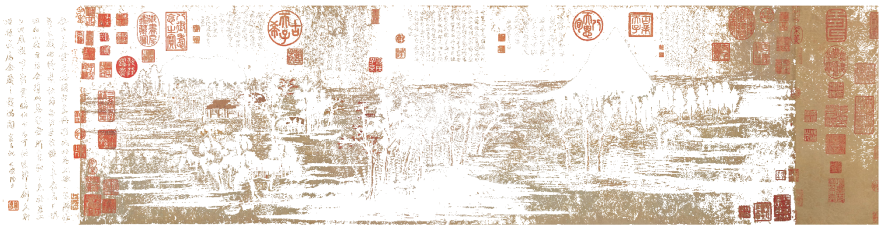


(c) Cluster 3

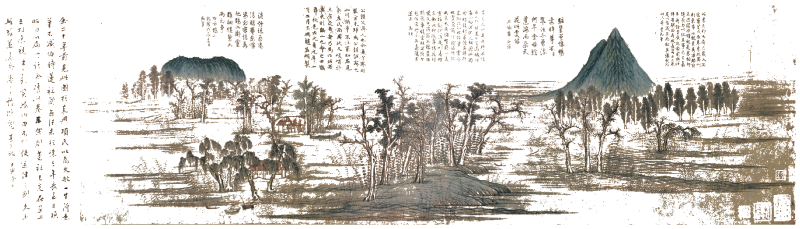
Figure 12: Image Segmentation of D-SPEC.



(a) Cluster 1

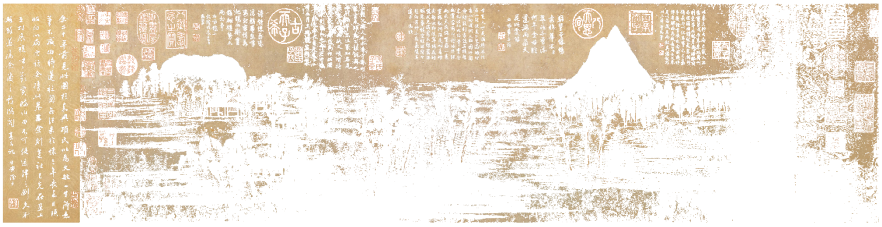


(b) Cluster 2

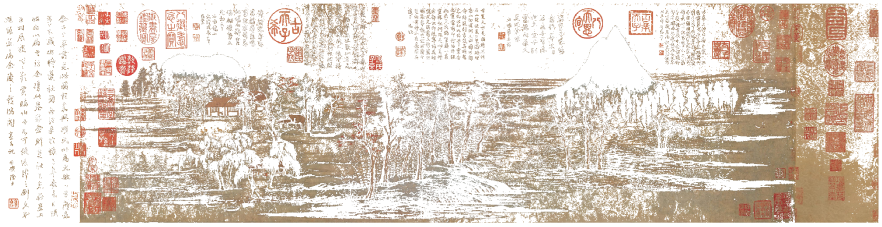


(c) Cluster 3

Figure 13: Image Segmentation of U-SPEC.



(a) Cluster 1



(b) Cluster 2



(c) Cluster 3

Figure 14: Image Segmentation of k-means.