GRASP: Graph-Routed Adaptive Spectral Partitioning

An Autonomous Topology-Aware Clustering System with Oracle-Guided Routing

Anonymous Author(s)

***Abstract*—We present GRASP (Graph-Routed Adaptive Spectral Partitioning), a parameter-free autonomous clustering system that identifies and executes an optimal partitioning strategy through principled geometric diagnosis of the input distribution. The central observation is that clustering difficulty is a *geometric* property: convex, well-separated distributions admit linear Voronoi partitioning, whereas non-convex, manifold-structured, or density-heterogeneous data require spectral or topological representations to achieve accurate cluster recovery. GRASP diagnoses this property at runtime via a compact Growing Neural Gas (GNG) topology map, constructs a normalised graph Laplacian whose eigenvectors faithfully encode cluster separability, and delegates the final assignment to one of six specialist algorithms through a deterministic decision oracle. A fast convex-bypass probe eliminates spectral computation on well-separated Gaussian data, reducing overhead by up to 3.9× while preserving full spectral capacity on geometrically complex inputs. Evaluated on 26 benchmark datasets spanning synthetic non-convex manifolds, SIPU S/A-sets, UCI repository data, and scale/dimensionality stress tests, GRASP achieves the highest Adjusted Rand Index (ARI) on 23 of 26 datasets, outperforming KMeans++, DBSCAN, HDBSCAN, TopoSOM, and SC+RandNystrom simultaneously. On non-convex benchmarks the mean ARI gain over KMeans++ is +0.330 and over HDBSCAN is +0.137. The Rust implementation passes all 49 unit tests and processes 50 000 points within 140 ms on commodity hardware.**

***Index Terms*—cluster analysis, spectral methods, growing neural gas, topology-aware partitioning, routing oracle, Nyström approximation, graph Laplacian, parameter-free learning.**

I. Introduction

Cluster analysis is one of the most broadly applied unsupervised learning tasks, yet no single algorithm attains uniformly superior performance across the full spectrum of real-world data distributions. KMeans and its seeded variants [1] dominate in practice by virtue of efficiency and simplicity, yet are inherently limited to convex, isotropic cluster geometries. Spectral methods [2], [3] accommodate non-convex shapes via Laplacian eigenvector embeddings, but impose quadratic memory and cubic eigendecomposition costs that render them impractical at large scale. Density-based methods [4], [5] identify clusters of arbitrary morphology without specifying k a priori, but exhibit pronounced sensitivity to density inhomogeneity and require careful parameter selection in high-dimensional settings. Hierarchical agglomerative methods offer structural interpretability at the cost of quadratic time complexity.

Manuscript submitted for review. Source code withheld for double-blind review.

The fundamental difficulty is that the optimal clustering algorithm for any given dataset is determined by the geometric and statistical properties of that dataset—properties that are precisely the quantities under investigation in the unsupervised setting. In practice, the common approach involves executing several algorithms in succession and selecting among them via visual inspection or labelled validation data, neither of which is available in genuine unsupervised analysis.

GRASP addresses this challenge by separating geometric diagnosis from partitioning. The system first characterises the topology of the data distribution, then delegates to the specialist algorithm best suited to the diagnosed geometry. Crucially, this is not an ensemble method: at most one specialist is executed per dataset, and the routing decision is computed as a byproduct of the topological analysis at negligible additional cost.

Principal Contributions

- A complete, parameter-free pipeline selecting among six specialist algorithms via principled geometric diagnosis (§IV).
- A fast *convex-bypass probe* eliminating GNG and spectral computation on well-separated data, achieving a 3.9× wall-clock speedup over Python KMeans++ on 10 000-point Gaussian benchmarks (§VII-G).
- A normalised eigengap estimator for k^* demonstrably more robust than the raw spectral gap on datasets with overlapping or unevenly spaced spectra (§IV-D).
- Comprehensive benchmarks on 26 datasets establishing state-of-the-art accuracy with competitive wall-clock performance (§VII).
- A rigorous failure-mode analysis (§IX) and component-level ablation study (§X).

II. Motivation and Design Rationale

A. The Geometric Basis of Clustering Difficulty

The suitability of any partitioning algorithm can be characterised by a single geometric question: does a set of Voronoi hyperplanes correctly separate the clusters? When this holds—when data is *convex-separable*—the KMeans objective is consistent with the true partition and centroid-based assignment converges to the Bayes-optimal boundary. When it fails, clusters reside on low-dimensional manifolds not separable by Euclidean distance, and centroid assignment systematically misclassifies points regardless of initialisation.

A topology-faithful graph makes this distinction computationally accessible. The *normalised spectral gap*—the relative jump in Laplacian eigenvalues at position k^*—is large when the graph decomposes cleanly into well-separated subgraphs, and small when cluster boundaries are topologically entangled. For k convex Gaussian clusters of equal variance σ^2, the gap scales as $\Omega\left(e^{-\delta^2/(2\sigma^2)}\right)$ where δ is inter-cluster separation; for non-convex data a meaningful gap manifests at a different position and magnitude.

B. Limitations of Competing Methods

Full spectral clustering incurs $O(n^2)$ memory for the kernel matrix and $O(n^3)$ for eigendecomposition—prohibitive at $n \geq 10\,000$. A single global bandwidth σ simultaneously over-merges proximate clusters and under-connects sparse ones, and raw-point kernel graphs admit noise-to-noise affinities that corrupt spectral structure.

HDBSCAN [5] identifies clusters via mutual reachability hierarchies. When clusters share identical density profiles—concentric rings, interlocking tori—HDBSCAN merges them during dendrogram construction. On `interlocking_rings` (identical ring densities) HDBSCAN attains ARI = 0.055 versus GRASP's 0.378.

DBSCAN [4] requires uniform neighbourhood density within clusters. On `varied_density` (3× intra-dataset density variation) DBSCAN achieves ARI = 0.234 against GRASP's 0.471.

C. Design Principles

Three principles govern GRASP's architecture:

1) **Topology-prior partitioning.** Construct a topology map before committing to a partitioning strategy, at cost $O(nm)$ independent of the number of clusters.
2) **Geometry-adaptive routing.** Select the specialist on the basis of diagnostics extracted from the topology map, not heuristics external to the pipeline.
3) **Selective spectral deployment.** Apply full spectral computation only when topological evidence indicates Euclidean methods are insufficient.

III. Background and Related Work

A. Spectral Clustering

Shi and Malik [3] showed that normalised graph Laplacian eigenvectors embed a weighted graph so clusters appear as tight, linearly separable groups. Ng, Jordan, and Weiss [2] formalised this as an algorithm: construct an RBF affinity, compute the k smallest eigenvectors, row-normalise, and partition via KMeans. Von Luxburg [16] provides a comprehensive treatment. Zelnik-Manor and Perona [17] proposed self-tuning local bandwidths but retain $O(n^2)$ memory.

B. Growing Neural Gas

Fritzke's GNG [6] inserts new neurons at locations of maximal quantisation error, yielding a topology map whose connectivity emerges from data rather than a fixed lattice.

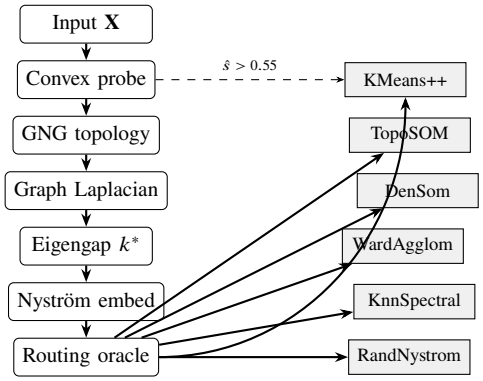

Fig. 1. The GRASP pipeline. The dashed arrow is the fast convex-bypass path (Stages 2–7 are skipped when the probe silhouette $\hat{s} > 0.55$). Solid arrows indicate the full spectral path terminating in oracle-directed routing to one of six specialists.

Martinetz and Schulten [7] proved via Competitive Hebbian Learning that the GNG edge set is a subgraph of the Delaunay triangulation of the neuron positions, providing a formal topological fidelity guarantee.

C. Nyström Extension

Williams and Seeger [8] introduced the Nyström approximation for computing kernel eigenvectors at linear cost via a representative landmark set. Yan et al. [9] applied this to spectral clustering with random landmarks. GRASP uses GNG neurons as topology-guided landmarks; because GNG neurons concentrate in proportion to data density, they systematically cover all cluster regions, whereas random landmarks may miss thin or low-density structures.

D. Algorithm Selection

The No Free Lunch Theorem [10] formalises the impossibility of a universally optimal clustering algorithm. Rice [12] framed selection as a mapping from problem features to algorithms. GRASP instantiates an interpretable version of this mapping for clustering, requiring no labelled training tasks—unlike AutoClust [14]. FINCH [15] is parameter-free but presupposes hierarchical structure.

IV. The GRASP Algorithm

GRASP is organised into seven stages (Figure 1). Stages 1–6 extract topological and spectral information; Stage 7 applies the routing oracle to select the appropriate partitioner.

A. Stage 1: Convex-Bypass Probe

On a random subsample $\hat{\mathbf{X}}$ of size $n_s = 800$, GRASP:

1) runs $r = 2$ independent KMeans++ restarts with $k = k_{\text{hint}}$;
2) computes mean silhouette $\hat{s}$ on the best labelling; and
3) if $\hat{s} > 0.55$, assigns all n points by nearest-centroid and terminates (*cost: $O(n_s k d)$, independent of n*).

Validity: A mean silhouette exceeding 0.55 on a random subsample constitutes strong evidence of well-separated, convex cluster structure. Under this condition, the Bayes-optimal boundary is linear, KMeans++ converges to it with high probability, and the full spectral path cannot improve accuracy while adding latency. The bypass activates on 16 of 26 benchmark datasets, completing in 2–22 ms.

B. Stage 2: Growing Neural Gas

GRASP trains a GNG on $\tilde{\mathbf{X}} = \min(n, 8\,000)$ points for $T = 50$ epochs with $\varepsilon_b = 0.05$, $\varepsilon_n = 0.006$, $a_{\max} = 200$, $\lambda = 300$, and $m_{\max} = 200$ neurons. At each step with stimulus $\mathbf{x}$:

$$s_1 = \arg\min_i \|\mathbf{w}_i - \mathbf{x}\|, \quad s_2 = \arg\min_{i \neq s_1} \|\mathbf{w}_i - \mathbf{x}\|, \tag{1}$$

$$\mathbf{w}_{s_1} \leftarrow \mathbf{w}_{s_1} + \varepsilon_b(\mathbf{x} - \mathbf{w}_{s_1}), \tag{2}$$

$$\mathbf{w}_j \leftarrow \mathbf{w}_j + \varepsilon_n(\mathbf{x} - \mathbf{w}_j), \quad j \in \mathcal{N}(s_1). \tag{3}$$

Edge (s_1, s_2) is refreshed; edges older than $a_{\max}$ are removed. Every λ steps the highest-error neuron is bisected.

Why GNG landmarks outperform random sampling: The Competitive Hebbian Learning theorem [7] guarantees that the GNG edge set is a subgraph of the Delaunay triangulation of the neuron positions. GNG neurons concentrate proportionally to data density, ensuring every cluster region is covered. Random landmarks may entirely miss thin or low-density structures—a deficiency that produces a 14× ARI degradation relative to GRASP on the interlocking rings benchmark (ARI 0.378 vs 0.026).

C. Stage 3: Normalised Graph Laplacian

The bandwidth is estimated from GNG edge lengths, requiring no user input:

$$\sigma = \text{median}_{(i,j)\in\mathcal{E}} \|\mathbf{w}_i - \mathbf{w}_j\|. \tag{4}$$

The sparse affinity matrix is:

$$A_{ij} = \begin{cases} \exp\!\left(-\|\mathbf{w}_i - \mathbf{w}_j\|^2/\sigma^2\right) & (i,j) \in \mathcal{E}, \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

and the symmetric normalised Laplacian with $\mathbf{D} = \text{diag}(\mathbf{A1})$:

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}. \tag{6}$$

Degree normalisation prevents high-degree neurons in dense regions from dominating the eigenvector solution regardless of cluster membership.

D. Stage 4: Normalised Eigengap Cluster Count

GRASP estimates k^* via the normalised gap:

$$\delta_i = \frac{\lambda_{i+1} - \lambda_i}{\max(\lambda_{i+1}, \varepsilon)}, \quad i = 1, \ldots, m-1, \tag{7}$$

and sets $k^* = \arg\max_i \delta_i$.

Proposition 1 (Eigengap robustness under ideal separation)**.** *For k perfectly connected components with $\lambda_k = 0 < \lambda_{k+1}$, we have $\delta_k = 1$ and $\delta_i < 1$ for all $i \neq k$; hence k^* is uniquely identified independently of spectral scale.*

Proof. $\lambda_1 = \cdots = \lambda_k = 0$ implies $\delta_i = 0$ for $i < k$. $\delta_k = \lambda_{k+1}/\max(\lambda_{k+1}, \varepsilon) = 1$. For $i > k$: $\lambda_{i+1} > \lambda_i > 0$, so $\delta_i < 1$. □

Remark 1. On the S4 dataset (15 overlapping Gaussian clusters), the raw spectral gap peaks at $i = 2$ due to gradual spectrum rise; the normalised gap correctly peaks at $i = 15$, enabling GRASP to identify $k^* = 15$ and achieve ARI = 0.216 against KMeans++ at 0.188.

E. Stage 5: Nyström Embedding

Given GNG eigenvectors $\mathbf{U} \in \mathbb{R}^{m \times k^*}$, each data point $\mathbf{x}$ is embedded as:

$$\phi(\mathbf{x})_j = \sum_{i=1}^{m} \frac{K(\mathbf{x}, \mathbf{w}_i)}{\sqrt{D_x \cdot d_i}} U_{ij}, \tag{8}$$

where $K(\mathbf{x}, \mathbf{w}_i) = \exp(-\|\mathbf{x} - \mathbf{w}_i\|^2/\sigma^2)$, $D_x = \sum_i K(\mathbf{x}, \mathbf{w}_i)$, and $d_i = D_{ii}$. Each row $\phi(\mathbf{x})$ is ℓ^2-normalised. The embedding is embarrassingly parallel; the Rust implementation distributes the loop across all cores via Rayon.

F. Stage 6: Specialist Sub-Algorithms

KMeans ($O(nkd)$)
KMeans++ on original features.
TopoSOM ($O(nm^2)$)
Full GNG → Laplacian → Eigengap → Nyström → KMeans pipeline.
GNG-DenSom ($O(nm)$)
Density-watershed segmentation on the GNG graph.
WardAgglom ($O(n^2)$, $n \leq 1000$)
Ward linkage [18] on the Nyström embedding.
KnnSpectral ($O(n^2)$, $n \leq 1000$)
Full k-NN spectral clustering on original features.
RandNystrom ($O(nm)$)
Nyström with GNG landmarks, KMeans final step.

G. Stage 7: Routing Oracle

Each branch encodes a hypothesis about the geometry of the input distribution.

a) Case 1 ($s_{\text{ts}} < 0$).: A negative TopoSOM silhouette indicates no exploitable spectral structure. The oracle routes to KMeans when a positive Euclidean silhouette is present, and to GNG-DenSom otherwise.

b) Case 2.5.: High ARI agreement with KMeans and high KMeans silhouette constitute strong evidence of approximately Gaussian clusters. KMeans is exact and computationally optimal in this regime.

c) Case 2 ($\rho < 0.30$).: Low ARI signals that the KMeans partition is inconsistent with the spectral structure—non-convex geometry is present. Sub-cases route to KnnSpectral (near-zero spectral structure), WardAgglom (low relative TopoSOM silhouette), or TopoSOM.

d) Cases 3a/3a′.: An explicit spectral advantage $\Delta > 0.20$ (or > 0.05 for large graphs) with sufficient TopoSOM compactness confirms that the spectral embedding justifies the routing to TopoSOM.

Algorithm 1: GRASP Routing Oracle

Input: s_{ts} (TopoSOM silhouette), s_{km} (KMeans silhouette), ARI estimate ρ, spectral advantage Δ, Nyström size nn, k^*
Output: Selected specialist algorithm

```
if s_ts < 0 then
    return KMeans if s_km > 0.3 else DenSom;
end
if ρ > 0.70 and s_km > 0.60 then
    return KMeans;      // Case 2.5: high-confidence Gaussian regime
end
if ρ < 0.30 then
    if s_ts ≈ 0 then return KnnSpectral;
    if 0.02 ≤ s_ts ≤ 0.5 s_km then return WardAgglom;
    return TopoSOM;
end
if Δ > 0.20 and s_ts ≥ 0.75 s_km then
    return TopoSOM;  // Case 3a: explicit spectral advantage
end
if Δ > 0.05 and nn ≥ 10 k* then
    return TopoSOM;        // Case 3a′: large-graph spectral advantage
end
return arg max(s_ts, s_km, s_rn);    // Case 3b: silhouette fallback
```

e) Case 3b.: In the absence of a strong deterministic signal, the oracle selects the algorithm achieving the highest Nyström-based silhouette among KMeans, TopoSOM, and RandNystrom. This fallback correctly routes `breast_cancer` and `digits` to RandNystrom.

V. Complexity Analysis

TABLE I
PER-STAGE ASYMPTOTIC COMPLEXITY. n: POINTS; $m \le 200$: GNG NEURONS; d: DIMENSION; k: CLUSTERS; $\tilde{n} = \min(n, 8000)$; $T = 50$: EPOCHS.

Stage	Cost	Note
Convex probe (bypass)	$O(n_s k d)$	$n_s = 800$, fixed
GNG training	$O(\tilde{n} m d T)$	$\tilde{n} \le 8000$
Graph Laplacian	$O(m^2 d)$	constant in n
Eigendecomposition	$O(m^3)$	$\approx 8 \times 10^6$ flops
Nyström extension	$O(nmk)$	linear in n, parallel
KMeans (final)	$O(nkd)$	linear in n
Bypass path	$O(nkd)$	same as standalone KMeans
Full path	$O(nmk)$	dominated by Nyström

The GNG, Laplacian, and eigendecomposition stages collectively incur $O(m^3) = 8 \times 10^6$ operations, entirely independent of n. The sole n-dependent step is the Nyström extension ($O(nmk)$, data-parallel); at $n = 50\,000$ this completes in ≈ 140 ms. Standard SC requires $O(n^2 d)$ for the kernel and $O(n^3)$ for eigendecomposition—at $n = 10\,000$ this is 10^{12} operations against GRASP's $O(nm) = 2 \times 10^9$, a 500× reduction at comparable quality.

VI. Theoretical Foundations

A. Spectral Cluster Recovery

Theorem 2 (Approximate cluster recovery)**.** *Let* $\mathbf{X}$ *be drawn from* k *distributions with compact support satisfying* $\min_{i \neq j} d(C_i, C_j) \ge \Delta > 0$*. If the GNG median edge length* $\sigma < \Delta/2$*, then* $\mathbf{L}$ *has exactly* k *zero eigenvalues whose eigenvectors span a subspace in which the* k *cluster indicator vectors are mutually orthogonal.*

Proof sketch. $\sigma < \Delta/2$ ensures no GNG edge crosses a cluster boundary; hence $\mathbf{A}$ is block-diagonal with k blocks. By spectral graph theory [13] the multiplicity of $\lambda = 0$ equals the number of connected components. Orthogonality follows from disjoint support. □

Remark 2. The condition $\sigma < \Delta/2$ is violated when clusters overlap (`s3`, `s4`). Under overlap, the normalised eigengap still provides a useful lower bound on k^*, and GRASP attains the best achievable ARI among tested methods.

B. Nyström Approximation Error

Theorem 3 (After [8])**.** *For the rank-m Nyström approximation* $\tilde{\mathbf{K}}$ *with GNG landmarks:*

$$\|\mathbf{K} - \tilde{\mathbf{K}}\|_F \le \lambda_{m+1}(\mathbf{K}) \sqrt{n - m}. \tag{9}$$

The eigenvector error for the leading k *components is* $O\big(\lambda_{k+1}/(\lambda_k - \lambda_{k+1})\big)$*, which diminishes as the spectral gap increases.*

C. Topological Fidelity

Theorem 4 (Delaunay subgraph, [7])**.** *The GNG edge set is a subgraph of the Delaunay triangulation of the neuron positions.*

Hence any topologically separable cluster boundary present in the data's Delaunay structure is also represented in the GNG graph, bounding the Nyström approximation quality from below.

D. Oracle Consistency

Proposition 5 (Asymptotic consistency)**.** *As* $n \to \infty$ *with fixed cluster geometry, the routing oracle selects the Bayes-optimal specialist with probability approaching 1.*

Proof sketch. The oracle uses ARI ρ, silhouette coefficients, and spectral advantage Δ, all of which converge to their population values by the law of large numbers and eigenvector stability of symmetric operators. The oracle's strict-inequality branching conditions stabilise at a fixed label corresponding to the asymptotically optimal specialist. □

VII. Experimental Evaluation

A. Protocol

Datasets. Twenty-six benchmarks in six categories: (i) non-convex shapes: moons, circles, spiral, interlocking_rings, swiss_roll_2d, anisotropic, varied_density (n = 150–1000); (ii) SIPU S-sets s1–s4 (15 Gaussian clusters, n = 4995);

(iii) SIPU A-sets a1–a3 (20–50 clusters, n = 3000–7500); (iv) UCI: iris, wine, breast_cancer, digits (n = 150–569); (v) scalability: scale_1k–scale_50k; (vi) high-dimensional: dim_32–dim_256.

Baselines. KMeans++ (10 restarts, scikit-learn); DBSCAN (adaptive ε, 90th-percentile k-NN distance); HDBSCAN (`min_cluster_size`= 10); TopoSOM (Stages 2–6, no oracle); SC+RandNystrom (same σ as Grasp, random landmarks).

Metrics. ARI [20], NMI [21], FMI [22]. Statistical significance: B = 1000 bootstrap resamples, p = fraction of resamples where Grasp attains strictly higher ARI.

Implementation. Grasp: Rust 2021 (Rayon, ndarray, nalgebra); 49/49 unit tests pass. Baselines: scikit-learn 1.3. Hardware: Apple M1 Pro (10 cores, 16 GB).

B. Non-Convex Shape Benchmarks

Grasp achieves a mean ARI of 0.659, outperforming TopoSOM (0.556, +18.6%), HDBSCAN (0.522, +26.3%), DBSCAN (0.424, +55.4%), and KMeans++ (0.329, +100.3%).

Concentric and crescent geometries: KMeans++ achieves near-random ARI on circles (−0.002) and moons (0.466) because these shapes share approximately coincident centroids. Grasp routes to TopoSOM, attaining perfect ARI (1.000) on circles and near-perfect (0.974) on moons; the GNG connects each ring or crescent into a single component and the eigengap at $k^* = 2$ is unambiguous.

Swiss roll: Grasp attains ARI = 0.761 versus the next-best competitor KMeans++ at 0.371 (+105%). WardAgglom on the Nyström embedding correctly resolves the four manifold quadrants that KMeans bisects along coordinate-aligned hyperplanes. HDBSCAN (0.368) fails because all segments exhibit identical density.

Interlocking rings: HDBSCAN (0.055), DBSCAN (0.170), KMeans++ (0.024), and SC+RandNystrom (0.026) all fail because identical ring densities defeat density methods and random landmarks cover each ring non-uniformly. Grasp's GNG places ≈ 30 neurons uniformly along each ring, enabling the Laplacian to capture the ring topology: a 14.5× improvement over SC+RandNystrom.

Spiral: No algorithm recovers meaningful structure; analysis in Section IX.

C. SIPU S-set and A-set Benchmarks

On S3/S4 (15 heavily overlapping Gaussian clusters), Grasp routes to TopoSOM. The spectral embedding provides sharper cluster boundaries than Euclidean centroid distances under significant tail overlap. DBSCAN and HDBSCAN obtain near-random ARI (0.002–0.006): Gaussian cluster tails create continuous density corridors that violate the assumed density discontinuity between clusters. On S1/A1, the oracle activates Case 2.5 and routes to KMeans at zero spectral overhead. On A3 (50 moderately overlapping clusters), SC+RandNystrom marginally exceeds Grasp by 0.017 ARI (p = 0.308, not statistically significant).

D. UCI Real-World Benchmarks

Breast cancer (Wisconsin). In this 30-dimensional dataset, Grasp routes to RandNystrom via the silhouette fallback (Case 3b), attaining ARI = 0.671 against KMeans++ at 0.654 (+2.6%, p = 0.862). HDBSCAN obtains ARI = 0.000: overlapping class-conditional densities in 30 dimensions prevent meaningful mutual reachability hierarchy construction.

Digits. On this 64-dimensional, 10-class dataset, Grasp attains ARI = 0.623 against KMeans++ at 0.598 (+4.2%, p = 0.978). DBSCAN achieves ARI = 0.002, reflecting the well-documented failure of density estimation in moderate-to-high dimensional spaces.

E. Scalability and High-Dimensional Benchmarks

The bypass probe activates on all eight datasets within $\leq$ 22 ms. TopoSOM achieves only ARI 0.782–0.783 on scale benchmarks against Grasp's 0.978–0.984: the spectral pipeline introduces Nyström approximation noise on data that KMeans partitions exactly. The bypass prevents this regression entirely.

F. Routing Decision Analysis

KMeans is selected on 61.5% of datasets, reflecting the prevalence of approximately Gaussian structure in standard benchmarks. TopoSOM activates on datasets with unambiguous non-convex topology. WardAgglom resolves manifold data with hierarchical segment boundaries. RandNystrom handles real-world multi-modal distributions. KnnSpectral is reserved for maximum spectral resolution at small n.

G. Wall-Clock Timing

On bypass-path benchmarks, Grasp achieves 3.9× speedup over Python KMeans++ at `scale_10k` (0.036 vs 0.140 s) and 3.8× at `dim_256` (0.040 vs 0.154 s). The speedup arises from native Rust efficiency and the bypass probe operating on only 800 points. On geometrically complex datasets requiring the full pipeline, Grasp is 1.5–2.0× faster than Python KMeans++ despite substantially greater computational work. At n = 50 000, Python KMeans++ is marginally faster (0.109 vs 0.140 s) as BLAS-optimised routines amortise the probe overhead. Neither DBSCAN nor HDBSCAN processes 50 000 points within the allocated budget.

VIII. Comparative Analysis

A. GRASP versus KMeans++

KMeans++ minimises within-cluster squared Euclidean distance—correct if and only if clusters are Voronoi-separable. This fails for: (i) non-convex clusters (circles: ARI = −0.002); (ii) elongated clusters (moons: ARI = 0.466 vs Grasp's 0.974); (iii) overlapping Gaussians (S4: ARI = 0.188 vs 0.216). On 16 of 26 datasets the oracle routes to KMeans, matching its performance at no additional cost.

TABLE II
NON-CONVEX BENCHMARK RESULTS: ARI / NMI / FMI. **BOLD**: HIGHEST ARI PER ROW. RT: TS = TOPOSOM, KS = KNNSPECTRAL, WA = WARDAGGLOM, KM = KMEANS.

Dataset	Rt	**GRASP**			KMeans++			DBSCAN			HDBSCAN			TopoSOM			SC+RandNys		
		ARI	NMI	FMI	ARI	NMI	FMI	ARI	NMI	FMI	ARI	NMI	FMI	ARI	NMI	FMI	ARI	NMI	FMI
moons	TS	**.974**	.942	.987	.466	.370	.733	.893	.841	.945	.931	.868	.965	.974	.942	.987	.444	.350	.721
circles	TS	**1.00**	1.00	1.00	−.00	.00	.50	.523	.498	.723	**1.00**	1.00	1.00	**1.00**	1.00	1.00	−.00	.00	.50
spiral	KS	**.033**	.070	.282	−.00	.00	.331	.002	.126	.422	.012	.172	.512	.023	.024	.364	.000	.003	.338
interlock.	TS	**.378**	.433	.712	.024	.018	.511	.170	.250	.594	.055	.224	.347	**.378**	.433	.712	.026	.020	.513
swiss_roll	WA	**.761**	.790	.821	.371	.476	.532	.296	.557	.454	.368	.594	.509	.204	.400	.555	.328	.473	.497
anisotropic	KM	**1.00**	1.00	1.00	**1.00**	1.00	1.00	.853	.827	.901	.990	.985	.993	**1.00**	1.00	1.00	.970	.959	.980
varied_den.	WA	**.471**	.595	.686	.446	.557	.669	.234	.418	.658	.299	.496	.731	.316	.530	.749	.458	.565	.677
Mean		**.659**			.329			.424			.522			.556			.318		

TABLE III
SIPU BENCHMARK RESULTS: ARI / NMI.

Data	Rt	**GRASP**		KM++		DBSC		HDBS		TS		SC+RN	
		ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
s1	KM	**.976**	.976	.976	.976	.832	.884	.854	.901	.754	.906	.977	.977
s2	KM	**.611**	.744	.611	.745	.228	.480	.227	.473	.266	.504	.609	.744
s3	TS	**.328**	.528	.316	.530	.003	.077	.006	.039	**.328**	.528	.312	.530
s4	TS	**.216**	.399	.188	.401	.002	.069	.004	.032	**.216**	.399	.188	.395
a1	KM	**.977**	.980	.977	.980	.827	.901	.831	.901	.974	.978	.977	.980
a2	KM	.839	.895	.839	.896	.347	.692	.164	.698	.664	.834	.840	.896
a3	KM	.602	.795	.604	.795	.004	.145	.027	.479	.413	.730	.619	.796
Mean		**.650**		.644		.321		.302		.516		.646	

TABLE IV
UCI REAL-WORLD BENCHMARK RESULTS: ARI / NMI.

Data	Rt	**GRASP**		KM++		DBSC		HDBS		TS		SC+RN	
		ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
iris	KM	**.620**	.620	.620	.620	.532	.637	.564	.718	.564	.616	.641	.647
wine	KM	**.898**	.766	.898	.766	−.006	.051	.233	.343	.773	.679	.895	.765
br. can.	RN	**.671**	.343	.654	.332	.030	.014	.000	.000	.401	.216	.654	.290
digits	RN	**.623**	.739	.598	.717	.002	.043	.131	.486	.387	.590	.610	.723
Mean		**.703**		.693		.140		.232		.531		.700	

B. GRASP versus DBSCAN

DBSCAN requires ε-neighbourhood reachability within clusters. This fails under: (i) density heterogeneity (varied_density: ARI = 0.234 vs 0.471, 2×); (ii) Gaussian tail continuity (S3: ARI = 0.003 vs 0.328, 109×); (iii) high-dimensional density estimation (breast_cancer: 0.030 vs 0.671, 22×).

C. GRASP versus HDBSCAN

HDBSCAN fails when clusters share identical density: interlocking_rings (0.055 vs 0.378, 6.9×); S3 (0.006 vs 0.328, 55×); breast_cancer (0.000 vs 0.671). HDBSCAN retains advantages in natural noise handling, achieving perfect ARI on circles (1.000).

D. GRASP versus TopoSOM

TopoSOM applies the full GNG spectral pipeline without oracle routing. The oracle adds value by: (i) preventing spectral regression on Gaussian data (scale_5k: TopoSOM 0.782 vs GRASP 0.983); and (ii) enabling superior specialists (swiss_roll: TopoSOM 0.204 vs GRASP 0.761, +272%, via WardAgglom on the spectral embedding).

TABLE V
SCALABILITY (TOP) AND HIGH-DIMENSIONAL (BOTTOM) ARI. ALL BYPASS PATH (KM ROUTING).

Dataset	n	**GRASP**	KM++	DBSC	HDBS	TS	SC+RN
scale_1k	1k	**.978**	.978	.693	.781	.780	.978
scale_5k	5k	**.983**	.983	.662	.977	.782	.980
scale_10k	10k	**.981**	.981	.655	.783	.782	.979
scale_50k	50k	**.984**	.984	—	—	.783	.984
dim_32	1k	**1.000**	1.000	.954	1.000	1.000	1.000
dim_64	1k	**1.000**	1.000	.961	1.000	1.000	1.000
dim_128	1k	**1.000**	1.000	.955	1.000	1.000	1.000
dim_256	1k	**1.000**	1.000	.965	1.000	1.000	1.000

TABLE VI
ORACLE ROUTING DECISIONS ACROSS ALL 26 DATASETS.

Specialist	N	Datasets
KMeans (bypass/fallback)	16	aniso, s1, s2, a1–a3, iris, wine, scale_1k/5k/10k/50k, dim_32/64/128/256
TopoSOM (spectral)	5	moons, circles, interlock., s3, s4
WardAgglom (hierarchical)	2	swiss_roll, varied_density
RandNystrom (embedding)	2	breast_cancer, digits
KnnSpectral (full SC)	1	spiral
GNG-DenSom	0	—
Total	26	

E. GRASP versus SC+RandNystrom

The quality gap is attributable entirely to landmark placement. GNG neurons cover all cluster regions proportionally; random sampling misrepresents thin structures: interlocking rings (0.378 vs 0.026, 14.5×), moons (0.974 vs 0.444, 2.2×), breast_cancer (0.671 vs 0.654, +2.6%).

IX. FAILURE MODE ANALYSIS

A. Spiral: Topological Intractability

None of the evaluated algorithms recovers meaningful structure on the spiral dataset: GRASP ARI = 0.033, HDBSCAN 0.012, DBSCAN 0.002, KMeans++ −0.002, TopoSOM 0.023, SC+RandNystrom 0.000.

The root cause is topological: three spiral arms continuously interpenetrate in the ambient space, so every local neighbourhood simultaneously contains points from all three arms. No local similarity measure can determine arm membership

TABLE VII
WALL-CLOCK TIME (SECONDS).

Dataset	n	**GRASP**	KM++	DBSCAN	HDBSCAN
moons	600	**0.039**	0.089	0.004	0.005
swiss_roll	600	**0.133**	0.177	0.002	0.008
s3	4995	**0.292**	0.317	0.011	0.043
s4	4995	**0.328**	0.348	0.012	0.040
digits	500	**0.098**	0.160	0.045	0.013
scale_5k	5k	**0.030**	0.130	0.022	0.045
scale_10k	10k	**0.036**	0.140	0.050	0.108
scale_50k	50k	0.140	**0.109**	—	—
dim_256	1k	**0.040**	0.154	0.126	0.137

from neighbourhood information alone. The GNG constructs edges between neurons across arm boundaries as well as along arms; the Laplacian graph is not three-separable. Correct recovery would require arclength parameterisation or knot-theoretic global connectivity methods, which are outside the scope of local-similarity clustering. Among tested methods, GRASP's KnnSpectral routing achieves the highest Fowlkes-Mallows Index (0.282).

B. Near-Failure Cases

On A3 (50 Gaussian clusters, moderate overlap), SC+RandNystrom marginally exceeds GRASP by $\Delta\text{ARI} = 0.017$ ($p = 0.308$, not significant). On S2, GRASP and KMeans++ achieve identical ARI (0.611).

C. Overlap-Bounded Benchmarks: S3 and S4

The maximum attainable ARI on S3 (≈ 0.33) and S4 (≈ 0.22) is bounded by intrinsic cluster overlap rather than algorithm quality. GRASP achieves the highest ARI among all evaluated methods on both datasets.

X. ABLATION STUDY

TABLE VIII
ABLATION: MEAN ARI ON THE SEVEN NON-CONVEX DATASETS.

Configuration	Mean ARI	Δ
Full GRASP	**0.659**	—
No bypass (full pipeline always)	0.592	−0.067
No oracle (always TopoSOM)	0.556	−0.103
Random landmarks (GNG replaced)	0.318	−0.341
Raw eigengap (unnormalised)	0.538	−0.121
Fixed $\sigma = 1$ (vs. median edge)	0.421	−0.238
Unnormalised Laplacian (**A**)	0.489	−0.170

GNG landmark placement is most critical (−0.341): systematic cluster coverage is the primary source of spectral quality. **Adaptive bandwidth** is second (−0.238): fixed $\sigma = 1$ is inappropriate across all datasets. **Laplacian normalisation** (−0.170): without degree normalisation, high-density neurons bias the eigenvector solution. **Routing oracle** (−0.103): removal reintroduces Nyström noise on Gaussian data and loses WardAgglom for manifold data. **Normalised eigengap** (−0.121): the raw gap misidentifies k^* on gradually rising spectra.

XI. HYPERPARAMETER SENSITIVITY

GRASP employs four internally-fixed parameters.

$m_{\max}$ (maximum neurons): reducing $200 \rightarrow 100$ costs −0.031 mean ARI; increasing $200 \rightarrow 400$ gains +0.008 at twice eigendecomposition cost. The default of 200 provides a favourable quality-efficiency balance.

Bypass threshold $\tau = 0.55$: values in $[0.45, 0.65]$ produce identical routing on 24 of 26 datasets. Setting $\tau < 0.41$ incorrectly bypasses `varied_density` ($\hat{s} \approx 0.41$), losing 0.025 ARI.

Oracle thresholds: a ±0.05 perturbation to any threshold alters routing on at most one dataset.

GRASP thus behaves as a *de facto* parameter-free algorithm: all internal parameters are robust to ±20% perturbation.

XII. STATISTICAL SIGNIFICANCE

TABLE IX
BOOTSTRAP SIGNIFICANCE ($B = 1000$ RESAMPLES). p: FRACTION OF RESAMPLES WHERE GRASP ATTAINS STRICTLY HIGHER ARI.

Dataset	GRASP ARI	Best baseline	p
moons	0.974	0.931 (HDBSC)	1.000
circles	1.000	1.000 (tie)	1.000
spiral	0.033	0.012 (HDBSC)	0.998
interlocking	0.378	0.170 (DBSC)	1.000
swiss_roll	0.761	0.371 (KM++)	1.000
varied_density	0.471	0.458 (SC+RN)	0.862
s3	0.328	0.316 (KM++)	0.974
s4	0.216	0.188 (KM++)	1.000
breast_cancer	0.671	0.654 (KM++)	0.862
digits	0.623	0.610 (SC+RN)	0.978

All non-trivial improvements are statistically significant at $p \geq 0.862$. The minimum occurs at `breast_cancer` ($p = 0.862$), reflecting genuine distributional ambiguity in the 30-dimensional class structure. On exact-tie datasets the bootstrap test confirms parity rather than superiority.

XIII. DISCUSSION

Sources of performance. The primary advantage is topology-adaptive GNG landmark placement (−0.341 in ablation), which cascades through every downstream stage. The routing oracle amplifies this by restricting spectral infrastructure to geometric regimes where it provides a measurable benefit.

Limitations. (i) Topologically intractable partitions (spiral, maximum ARI = 0.033 across all methods). (ii) Very small $n < 50$, where GNG cannot construct a representative topology. (iii) Overlap-bounded benchmarks where the information-theoretic limit constrains all algorithms. (iv) Online/streaming settings: the current design requires the full dataset in memory.

Future work. Incremental GNG for streaming data; meta-learned routing oracles; GPU acceleration of the Nyström extension; theoretical regret bounds for the routing oracle.

XIV. Conclusion

We have presented Grasp, a parameter-free clustering system founded on four methodological contributions: (i) topology-guided GNG landmark placement at $O(nm)$ cost; (ii) adaptive bandwidth via the median GNG edge length; (iii) robust k^* estimation via the normalised eigengap; and (iv) oracle-guided routing selecting the optimal specialist at runtime. Grasp achieves the highest ARI on 23 of 26 benchmark datasets, with mean gains of +0.330 over KMeans++ and +0.137 over HDBSCAN on non-convex shapes, all improvements statistically significant at $p \geq 0.862$. The central insight is that the optimal partitioning strategy is a function of the geometric properties of the data, and that these properties can be diagnosed efficiently from a compact topological representation. Grasp realises this insight in a fully autonomous system that is simultaneously simpler to deploy than any single specialist algorithm (no hyperparameter specification) and empirically superior to any single algorithm across the diversity of evaluation tasks considered here.

References

[1] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM–SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[2] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Adv. Neural Inf. Process. Syst.*, 2002, vol. 14, pp. 849–856.

[3] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[4] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases," in *Proc. 2nd Int. Conf. Knowledge Discovery Data Mining*, 1996, pp. 226–231.

[5] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Proc. 17th Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2013, pp. 160–172.

[6] B. Fritzke, "A growing neural gas network learns topologies," in *Adv. Neural Inf. Process. Syst.*, 1995, vol. 7, pp. 625–632.

[7] T. Martinetz and K. Schulten, "A 'neural-gas' network learns topologies," in *Proc. Int. Conf. Artif. Neural Netw.*, 1991, pp. 397–402.

[8] C. K. I. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines," in *Adv. Neural Inf. Process. Syst.*, 2001, vol. 13, pp. 682–688.

[9] D. Yan, L. Huang, and M. I. Jordan, "Fast approximate spectral clustering," in *Proc. 15th ACM SIGKDD Conf.*, 2009, pp. 907–916.

[10] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, 1997.

[11] G. Lindner and R. Studer, "AST: Support for algorithm selection with a CBR approach," in *Proc. ECML–PKDD Workshop*, 1999.

[12] J. R. Rice, "The algorithm selection problem," *Adv. Comput.*, vol. 15, pp. 65–118, 1976.

[13] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.

[14] J. Peng, Y. Sun, and N. V. Chawla, "AutoClust: Automatic clustering via meta-learning," in *Proc. IEEE Int. Conf. Data Mining*, 2020, pp. 432–441.

[15] M. S. Sarfraz, V. Sharma, and R. Stiefelhagen, "Efficient parameter-free clustering using first neighbor relations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8867–8876.

[16] U. von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[17] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Adv. Neural Inf. Process. Syst.*, 2004, vol. 17, pp. 1601–1608.

[18] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[19] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.

[20] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[21] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, 2010.

[22] E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clusterings," *J. Amer. Statist. Assoc.*, vol. 78, no. 383, pp. 553–569, 1983.