Resolution-Aware Structural Density Peak Clustering

Jie Yang
The University of Sydney
Sydney, Australia
jie.yang1@sydney.edu.au

Hsiang-Ting Chen
Adelaide University
Adelaide, Australia
tim.chen@adelaide.edu.au

Yan Ma
Shanghai Normal University
Shanghai, China
ma-yan@shnu.edu.cn

Xinyan Liang
Shanxi University
Taiyuan, China
liangxinyan48@163.com

Avinash Kumar Singh
University of Technology Sydney
Sydney, Australia
avinash.singh@uts.edu.au

Liang Du
Shanxi University
Taiyuan, China
duliang@sxu.edu.cn

Cheng-You Lu
University of Technology Sydney
Sydney, Australia
cheng-you.lu@student.uts.edu.au

Chenglong Zhang
Nanjing University
Nanjing, China
clzhang123@163.com

Bingbing Jiang*
Hangzhou Normal University
Hangzhou, China
jiangbb@hznu.edu.cn

Weiping Ding
Nantong University
Nantong, China
dwp9988@163.com

Wei Chen
The University of Sydney
Sydney, Australia
wei.chenbme@sydney.edu.au

Abstract

Density Peak Clustering (DPC) identifies centers and propagates labels at the sample level, which makes density estimation sensitive to the cutoff and leaves assignment prone to errors along long chains. Performance therefore suffers under heterogeneous densities and complex cluster structures. This paper argues that density peaks inference should instead be performed on *resolution-controlled structural units*. On such units density becomes a stable count and relative distance follows the separation of the enclosing hierarchy, while assignment runs over a compact graph of units instead of long sample-to-sample chains. This view is realized in **Resolution-Aware Structural Density Peak Clustering (RSDP)**, which builds a constrained hierarchy, selects an admissible level via a single resolution parameter, and applies a one-pass DPC rule on the selected structural units. Under a faithful-realization condition, we prove that RSDP's empirical γ-ranking recovers the dominant population units and that its one-pass assignment recovers the sampled ground-truth partition. On seven real-world benchmarks, RSDP shows a substantial advantage over ten state-of-the-art DPC variants in average ACC and NMI.

1 Introduction

Density Peak Clustering (DPC) identifies cluster centers as samples whose local density ρ is high and whose relative distance δ to any higher-density sample is large; remaining samples are assigned by following nearest-higher-density links [Rodriguez and Laio, 2014]. DPC is popular because it is

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

simple. Centers are read off a two-dimensional decision plot, labels are assigned without iteration, and outliers stand apart from the peaks. Every step, however, works on individual samples, from density estimation and peak detection to label propagation along chains of samples. As a result, DPC is sensitive to the cutoff when densities are heterogeneous, and one wrong link can propagate along a long chain [Wang et al., 2024].

Most DPC variants inherit this pointwise view and improve one component (density estimation, neighborhood construction, center detection, or propagation rules) while leaving the sample as the unit of inference [Liu et al., 2018, Sun et al., 2022, Zhang et al., 2023, Zhou et al., 2022, Ding et al., 2025a, Jiang et al., 2026, Wang et al., 2025, Fei et al., 2025b,a, Kim et al., 2024]. A separate line of research introduces sub-clusters or micro-clusters, but uses them as auxiliary structures for merging, compression, or efficiency rather than as the objects on which density and peak scores are *defined* [Xu et al., 2016, Li et al., 2023, Nie et al., 2026, Zhai et al., 2026, Xu and Pham, 2024]. Either way, the basic unit of inference remains the sample.

We argue that density peaks should be inferred at a *resolution-controlled structural level* rather than at the sample level. Once density, relative distance, and the peak score are defined on the units of a hierarchy, density becomes an integer count governed by Binomial concentration, relative distance becomes a gap between units that the hierarchy has already separated, and assignment runs over a graph of $M \ll n$ units, so chains are short. Resolution then decides the level at which the whole DPC procedure operates, rather than acting as one more tuning parameter on a pointwise method.

We implement this idea in **Resolution-Aware Structural Density Peak Clustering (RSDP)**. RSDP builds a constrained hierarchy on the data and selects an admissible level via a single resolution parameter λ. Every sub-cluster at that level is treated as a *structural point*, and the standard DPC operations (peak selection and one-pass assignment) run entirely at this structural level. Sample labels are recovered by propagating structural labels back to constituent samples, with no point-level refinement and no sample-level chains.

Contributions. **(C1)** A structure-based reformulation of DPC: ρ, δ, and γ are defined on hierarchical structural units rather than on points or point-based surrogates. Moving the unit of inference from samples to structural units gives stable recovery under heterogeneous densities and complex cluster structures, and is the basis of our theoretical analysis. **(C2)** Resolution as a single tunable parameter: one λ selects the structural level on which ρ, δ, and γ are jointly defined, replacing the multiple bandwidth and neighborhood parameters of prior DPC variants. **(C3)** A conditional inference guarantee: under a faithful-realization condition and a density-distance margin, RSDP's empirical γ-ranking recovers the dominant population units and its one-pass assignment recovers the sampled ground-truth partition. The guarantee concerns only the structural inference layer and does not depend on the stability of the underlying hierarchy. Across seven real-world benchmarks, RSDP shows a substantial accuracy advantage over ten DPC baselines.

2 Related Work

DPC and pointwise improvements. Prior DPC variants retain samples as the unit of inference and redesign one component of the pipeline. SNN-DPC augments density with shared-nearest-neighbor counts [Liu et al., 2018]; adaptive-neighbor DPC replaces the global cutoff with local neighborhoods [Sun et al., 2022]; BC-DPC balances density against connectivity across heterogeneous regions [Zhang et al., 2023]; Core-point DPC identifies core points through kNN density and adds a non-core integration step [Zhou et al., 2022]; WANN-DPC weights adaptive neighbors with correction-based sequential selection [Xie et al., 2026]; CD-DPC introduces a centrifugal degree along nearest-higher-density paths [Ma et al., 2026]. These contributions improve density estimation or propagation but keep ρ, δ, and γ at the sample level.

Hierarchical structures used externally. A second line introduces hierarchical or sub-cluster structures. DenPEHC embeds density peaks into a hierarchical procedure [Xu et al., 2016]; sub-cluster-merging DPC over-segments and recovers the partition by merging [Li et al., 2023, Nie et al., 2026]; representative- and micro-cluster-based variants documented in [Wang et al., 2024] use structural objects to support aggregation or post-hoc merging, yet density, δ, and γ continue to be computed at or through the sample level.

Hierarchy and resolution as modeling variables. A broader literature treats hierarchy and resolution as central modeling variables [Kim et al., 2024, Yang et al., 2025, 2026]. EHKM separates granularity from resolution and supports partition extraction at arbitrary resolutions [Zhao et al., 2026]. FINCH builds a hierarchy from first-neighbor relations with no preset thresholds [Sarfraz et al., 2019]. TORC constructs a hierarchy via constrained mergers and prunes it using mass–distance peaks [Yang and Lin, 2025b,a]. CoHiClust distills an embedding into a binary tree carrying cluster labels [Pach et al., 2026], and HMKC learns a sequence of increasingly coarse partitions across multiple kernels [Liu et al., 2021]. Granular methods [Li et al., 2025, Wang and Du, 2026] replace samples with information units linked through soft affiliation graphs, and GB-DP runs density peaks on spherical granular balls whose density and separation are derived from ball centers and radii [Cheng et al., 2024]. Recent multi-view and deep clustering methods also exploit structure beyond individual samples, e.g., cross-view topology [Dong et al., 2025], graph matching [Dong et al., 2026], label diffusion [Chen et al., 2026], and soft neighbors [Duan et al., 2025]. RSDP brings these ideas into density peaks inference itself, defining ρ, δ, and γ on the same structural units and letting resolution select the level at which inference takes place. Appendix B discusses the specific differences from FINCH and TORC.

3 Method

We consider clustering with a target number of clusters K given. Let $\mathcal{X} = \{x_i\}_{i=1}^n$ denote the dataset with metric d. RSDP builds a constrained hierarchy on $\mathcal{X}$, selects an admissible level, and performs density-peaks inference on the structural units at that level. For readability, this section uses unhatted notation; hats are introduced in §4, where empirical and population quantities are distinguished.

3.1 Design principles

Three principles distinguish RSDP from prior DPC variants: **(P1)** *Density on structural units.* Sample-level ρ inherits the instability of pointwise density estimation. Treating each sub-cluster of a hierarchy as a structural point turns ρ into a parameter-free integer count whose fluctuations are controlled by Binomial concentration. **(P2)** *Resolution as modeling variable.* A single $\lambda \in (0, 1]$ selects the hierarchy level, fixing the structural units on which ρ, δ, γ are jointly defined; coarser levels yield fewer, larger units, finer levels approach sample-level units as a degenerate limit. **(P3)** *One-pass assignment.* Sample-level nearest-higher-density chains have length up to $n-1$, so a single mis-link can mislabel a long tail. Structural chains have length at most $M-1 \ll n$; one-pass assignment suffices, with no correction. These principles map respectively onto §3.2, §3.3, and §3.4.

3.2 Constrained hierarchical clustering

Starting from the singleton partition, constrained hierarchical clustering (CHC) [Yang and Lin, 2023, 2024] merges clusters under a size-constrained single-linkage rule. For an intermediate partition $\mathcal{C} = \{C_1, \ldots, C_M\}$, the single-linkage distance between clusters is defined as

$$d(C_i, C_j) = \min_{x \in C_i,\, y \in C_j} d(x, y), \tag{1}$$

and each cluster C_i connects to its nearest neighbor only when the latter is no smaller:

$$A_{ij} = \mathbf{1}\left(j = \arg\min_{k \neq i} d(C_i, C_k) \ \wedge \ |C_i| \leq |C_j| \right). \tag{2}$$

The connected components of the graph induced by A form the next coarser partition; iterating yields a fine-to-coarse hierarchy $\mathcal{H} = \{\mathcal{C}^{(1)}, \ldots, \mathcal{C}^{(L)}\}$, where $\mathcal{C}^{(1)}$ is the singleton partition and $\mathcal{C}^{(L)}$ is the coarsest. Throughout, d denotes the single-linkage distance (1) applied at whatever scale is needed (between samples, or between sub-clusters). The size constraint in (2) matters because it biases the hierarchy toward absorbing smaller units into no-smaller neighbors, which yields stable structural units whose integer sizes serve as parameter-free density surrogates. Unlike DPC variants that use a hierarchy as an external aid (e.g., DenPEHC), RSDP performs density peaks inference on the hierarchy itself and never computes a point-level ρ.

3.3 Resolution-aware structural density peaks

Level selection. Given $\lambda \in (0, 1]$, define the target level $\widetilde{l} = \max(1, \lfloor \lambda L + 0.5 \rfloor)$. The selected level is the largest level not exceeding $\widetilde{l}$ at which the partition has more than K units:

$$l^* = \max\{l \in [1, \widetilde{l}] \ : \ |\mathcal{C}^{(l)}| > K\}. \tag{3}$$

This is well-defined since $|\mathcal{C}^{(1)}| = n > K$ for any non-trivial dataset. Smaller λ corresponds to a finer level. Write $\mathcal{C}^{(l^*)} = \{C_1, \ldots, C_M\}$; each C_i is a *structural point*.

Structural density. We define the structural density of unit C_i as the number of samples it contains:

$$\rho(C_i) = |C_i|. \tag{4}$$

This integer count needs no d_c or kernel bandwidth, and it estimates the population structural density $q_i = P(Q_i)$ in the sense that $\rho(C_i)/n \to q_i$ in probability under mild concentration.

Structural distance. The structural distance is the single-linkage distance in (1) applied at the sub-cluster level. For a predecessor set $\mathcal{P}(C_i)$ defined below, we take the minimum to any predecessor when one exists, and otherwise the maximum to any other unit (the standard DPC convention for the overall density-maximizer):

$$\delta(C_i) = \begin{cases} \min\limits_{C_j \in \mathcal{P}(C_i)} d(C_i, C_j), & \mathcal{P}(C_i) \neq \emptyset, \\ \max\limits_{C_j \neq C_i} d(C_i, C_j), & \mathcal{P}(C_i) = \emptyset, \end{cases} \tag{5}$$

where the predecessor set $\mathcal{P}(C_i) = \{C_j : C_j \succ C_i\}$ uses a strict priority order $\succ$: $C_j \succ C_i$ if either $\rho(C_j) > \rho(C_i)$, or $\rho(C_j) = \rho(C_i)$ and a fixed injective index ι satisfies $\iota(C_j) < \iota(C_i)$ (deterministic tie-breaking).

Structural peak score. The peak score combines structural density and structural distance:

$$\gamma(C_i) = \rho(C_i)\, \delta(C_i). \tag{6}$$

All three quantities in (4), (5), and (6) are defined on the same M units at the same resolution, with no point-level surrogate. The number of comparisons therefore drops from $\Theta(n^2)$ to $\Theta(M^2)$ with $M \ll n$, and γ inherits a combinatorial-geometric margin between dominant and non-dominant within-cluster units, which is made precise in §4.

3.4 Center selection and structural assignment

Center selection. The K structural units with the largest peak scores are taken as centers, each receiving a unique label [Yang et al., 2017, 2021]:

$$\mathcal{S} = \mathrm{TopK}_{C_i \in \mathcal{C}^{(l^*)}}\, \gamma(C_i). \tag{7}$$

Ties in γ are broken by the priority order $\succ$, so the highest-priority unit, which has the largest γ and no predecessor, is always a center.

Structural assignment. Each non-center unit $C_i \notin \mathcal{S}$ inherits the label of its nearest predecessor:

$$\pi(C_i) = \arg \min_{C_j \in \mathcal{P}(C_i)} d(C_i, C_j), \qquad y(C_i) = y(\pi(C_i)), \tag{8}$$

processed in descending priority order so that $\pi(C_i)$ has already been labeled when C_i is processed.

Sample-level recovery. Each sample inherits the label of its containing structural unit:

$$y(x) = y(C_i) \qquad \text{for } x \in C_i. \tag{9}$$

No point-level refinement is applied. The structural chain induced by (8) has length at most $M-1$, and each link compares a unit to a higher-priority unit by a single-linkage distance. Short chains limit how far a single wrong link can spread, and it is this long-tail mislabeling that motivates corrective passes in sample-level DPC.

Algorithm 1 Resolution-Aware Structural Density Peak Clustering (RSDP)

Require: Dataset $\mathcal{X}$, target K, resolution $\lambda \in (0, 1]$
Ensure: Sample labels $\{y(x_i)\}_{i=1}^n$
1: Build CHC hierarchy $\mathcal{H} = \{\mathcal{C}^{(1)}, \ldots, \mathcal{C}^{(L)}\}$ via (1)–(2)
2: Pick admissible level l^* by (3); set $\mathcal{C}^{(l^*)} = \{C_1, \ldots, C_M\}$
3: Compute $\rho(C_i)$, $\delta(C_i)$, $\gamma(C_i)$ by (4), (5), (6)
4: Centers: $\mathcal{S} \leftarrow \mathrm{TopK}_{C_i}\, \gamma(C_i)$; assign each $C \in \mathcal{S}$ a distinct label
5: **for** each non-center C_i in descending priority $\succ$ **do**
6: $\quad y(C_i) \leftarrow y(\pi(C_i))$ via (8)
7: **end for**
8: **return** $y(x) \leftarrow y(C_i)$ for each $x \in C_i$

3.5 Complexity

The full procedure is summarized in Algorithm 1. The hierarchy is built from iterated 1-NN graphs over the current partition. With cached pairwise distances or appropriate nearest-neighbor maintenance, hierarchy construction can be implemented within $O(n^2)$ time and memory in general metrics. Computing δ, γ, and the one-pass assignment at level l^* each cost $O(M^2)$. We therefore state the complexity conservatively as $O(n^2 + M^2)$. Nearest-neighbor indices can lower the empirical cost of hierarchy construction, e.g., kd-trees in low-dimensional Euclidean spaces and metric trees or approximate nearest-neighbor indices in high dimensions. Hierarchy construction remains the quadratic stage, whereas structural inference on $M \ll n$ units is substantially cheaper than sample-level DPC.

4 Theoretical Analysis

We show that the three principles in §3.1 lead to a *conditional inference guarantee*. Assuming the underlying hierarchy realizes a faithful structural partition at level l^*, the RSDP inference layer ($\widehat{\rho}, \widehat{\delta}, \widehat{\gamma}$ definition, top-K selection, one-pass assignment) recovers the sampled ground-truth partition. Our analysis therefore concerns the inference layer rather than the full pipeline of hierarchy construction and RSDP. Whether the hierarchy achieves a faithful realization is a question of cluster stability for the underlying construction [Hartigan, 1981, Chaudhuri and Dasgupta, 2010], which lies outside the inference layer analyzed here. The setup, intermediate lemmas, and proofs are deferred to Appendix A.

Setup. Let $(\mathcal{X}, d)$ be a separable metric space and P a Borel probability measure on $\mathcal{X}$. Data $X_1, \ldots, X_n$ are i.i.d. from P, and $G_1, \ldots, G_K$ form the ground-truth measurable partition of $\mathrm{supp}(P)$ with $\pi_k = P(G_k) > 0$ and $K \geq 2$. At resolution λ, fix a measurable *population structural partition* $\mathcal{Q} = \{Q_1, \ldots, Q_M\}$ refining $\{G_k\}$, with $Q_i \subseteq G_{k(i)}$, $q_i := P(Q_i) > 0$, and $M \geq K+1$. $\mathcal{Q}$ plays the role of $\mathcal{C}^{(l^*)}$ in the population: the target the empirical level-l^* partition attempts to realize.

To distinguish empirical from population objects, we use hats: let $\widehat{\mathcal{C}} = \{\widehat{C}_1, \ldots, \widehat{C}_{\widehat{M}}\}$ be the empirical structural partition at level l^*, with empirical single-linkage distance $\widehat{d}$ and empirical counterparts $\widehat{\delta}, \widehat{\gamma}$ of (4), (5), (6). Define the *faithful realization event*

$$\mathcal{E}_n = \big\{\widehat{M} = M \text{ and } \exists\, \sigma \in S_M :\ \widehat{C}_{\sigma(i)} = \{X_j : X_j \in Q_i\}\ \forall i\big\},$$

and the population structural distance $d^*(Q_i, Q_j) = \inf\{d(x, y) : x \in Q_i, y \in Q_j\}$.

Population statistics. For each ground-truth cluster G_k, let $\mathcal{Q}_k = \{Q \in \mathcal{Q} : Q \subseteq G_k\}$, $P_k^* = \arg\max_{Q \in \mathcal{Q}_k} q_Q$, $a_k = q_{P_k^*}$, and b_k the second-largest population structural density in $\mathcal{Q}_k$ (or 0 if $|\mathcal{Q}_k| = 1$). Let R denote the maximum within-cluster structural radius, and define the symmetric cluster separation

$$\Delta_k = \min_{Q \in \mathcal{Q}_k,\, Q' \notin \mathcal{Q}_k} d^*(Q, Q'), \tag{10}$$

with $a_{\min} = \min_k a_k$, $b_{\max} = \max_k b_k$, $\Delta = \min_k \Delta_k$. The *global density–distance margin* is

$$m := a_{\min}\Delta - b_{\max}R, \tag{11}$$

capturing the combinatorial-geometric gap between dominant and non-dominant within-cluster units. Remaining cluster-wise quantities are defined formally in Appendix A.1.

Assumptions. **(A1)** *Dominance:* $a_k > b_k$, with P_k^* unique $\forall k$. **(A2)** *Margin:* $m > 0$. **(A3)** *Ascending connectivity:* for each k and $Q \in \mathcal{Q}_k \setminus \{P_k^*\}$, there exists $Q' \in \mathcal{Q}_k$ with $q_{Q'} > q_Q$ and $d^*(Q, Q') < \Delta_k$. **(D)** *Distance stability:* there exists a sequence $\nu_n \to 0$ and an event $\mathcal{D}_n$ with $\Pr(\mathcal{D}_n^c \cap \mathcal{E}_n) \le o(1)$ such that on $\mathcal{E}_n \cap \mathcal{D}_n$, $\sup_{i \neq j} |\widehat{d}(\widehat{C}_{\sigma(i)}, \widehat{C}_{\sigma(j)}) - d^*(Q_i, Q_j)| \le \nu_n$.

(A1) and (A2) are geometric conditions on $(P, \mathcal{Q})$. (A3) replaces a stronger parent-consistency assumption with the weaker statement that each non-dominant unit has *some* higher-density within-cluster unit closer to it than any out-of-cluster unit. (D) isolates the gap between population structural geometry and its empirical realization, discussed in Appendix A.2.

Main results. The first theorem says the empirical γ-ranking recovers the dominant population units; the second says the resulting one-pass assignment recovers the sampled ground-truth partition.

Theorem 1 (Empirical separability under faithful realization)**.** *Assume* (A1), (A2), (D)*, and* $K \ge 2$*. If* $b_{\max}\nu_n < m/2$ *for all sufficiently large* n*, then*

$$\Pr\Big(\mathcal{E}_n \cap \big\{\textit{top-}K\textit{ empirical structural units ranked by } \widehat{\gamma} \textit{ are exactly } \{\widehat{C}_{\sigma(P_k^*)}\}_{k=1}^K\big\}\Big) \ge \Pr(\mathcal{E}_n) - o(1).$$

Theorem 2 (Assignment correctness)**.** *Assume* (A1), (A2), (A3), (D)*,* $K \ge 2$*, and* $\nu_n < \eta/2$ *for all sufficiently large* n *(with* η *as in Lemma 4). Then*

$$\Pr\Big(\mathcal{E}_n \cap \big\{\textit{RSDP labels are constant on } \{X_j : X_j \in G_k\} \textit{ for each } k, \textit{ in one-to-one correspondence with } G_1, \dots, G_K\big\}\Big) \ge \Pr(\mathcal{E}_n) - o(1).$$

Interpretation. The bound $\Pr(\text{success} \cap \mathcal{E}_n) \ge \Pr(\mathcal{E}_n) - o(1)$ is a *conditional inference guarantee*: on the faithful-realization event, RSDP succeeds except on auxiliary density- and distance-instability events of probability $o(1)$. Whether $\Pr(\mathcal{E}_n) \to 1$ depends on the stability of the underlying hierarchy, which is separate from the structural inference layer analyzed here.

Proof sketch. Theorem 1 combines three ingredients: (i) unconditional structural-density concentration for the population counts $N_Q = \sum_{j=1}^n \mathbf{1}\{X_j \in Q\} \sim \text{Binomial}(n, q_Q)$, which transfers to $|\widehat{C}_{\sigma(Q)}| = N_Q$ on $\mathcal{E}_n$ and controls $\widehat{\rho}$; (ii) the deterministic lower bound $\widehat{d} \ge d^*$ on $\mathcal{E}_n$, since empirical infima are taken over subsets of the population pairs; (iii) population separability $\min_k \gamma^*(P_k^*) \ge a_{\min}\Delta > b_{\max}R$, from (A1)–(A2) and the symmetric definition of Δ_k. Theorem 2 adds a parent-consistency lemma from (A3): every empirical non-center unit's nearest predecessor lies in its own ground-truth cluster. Each design principle of §3.1 plays a part in this argument. P1 (density on structural units) enables ingredient (i), since structural densities are Binomial counts of population units that need no cutoff or bandwidth. P2 (resolution as variable) enables ingredient (iii), since m is a property of the resolution-selected partition. P3 (one-pass) makes Theorem 2 a one-pass statement: parent consistency propagates through chains of length at most $M-1$ without correction.

5 Experiments

We evaluate RSDP on three synthetic datasets that isolate typical DPC failure modes, and on seven real-world datasets against ten DPC variants. The details of the real-world datasets are reported in Table 1. The compared baselines include DPC [Rodriguez and Laio, 2014], DPCDVND [Zang et al., 2024], KSFDPC [Li et al., 2023], DPCFSC [Li et al., 2022], SMMP [Guan et al., 2023b], DBSCAN-DPC (DBDPC) [Hou et al., 2024], VDPC [Wang et al., 2023], DEMOS [Guan et al., 2023a], DPCMFP [Zang et al., 2025], and FDPAN [Ding et al., 2025b]. The Accuracy (ACC) and Normalized Mutual Information (NMI) [Strehl and Ghosh, 2002] are reported, and the hyperparameters of all baselines are tuned according to their original publications.

Table 1: Real-world datasets used in the experiments.

Datasets	#Samples	#Dimensions	#Clusters
COIL100 [Nene et al., 1996]	7200	49152	100
NH [Cao et al., 2015]	4660	6750	5
UMIST [Graham and Allinson, 1998]	575	10304	20
CMUPIE [Sim et al., 2002]	2856	1024	68
MNIST [LeCun et al., 1998]	10000	784	10
Pendigits [Alimoglu and Alpaydin, 1997]	10992	16	10
Miceprotein [Higuera et al., 2015]	1077	77	8

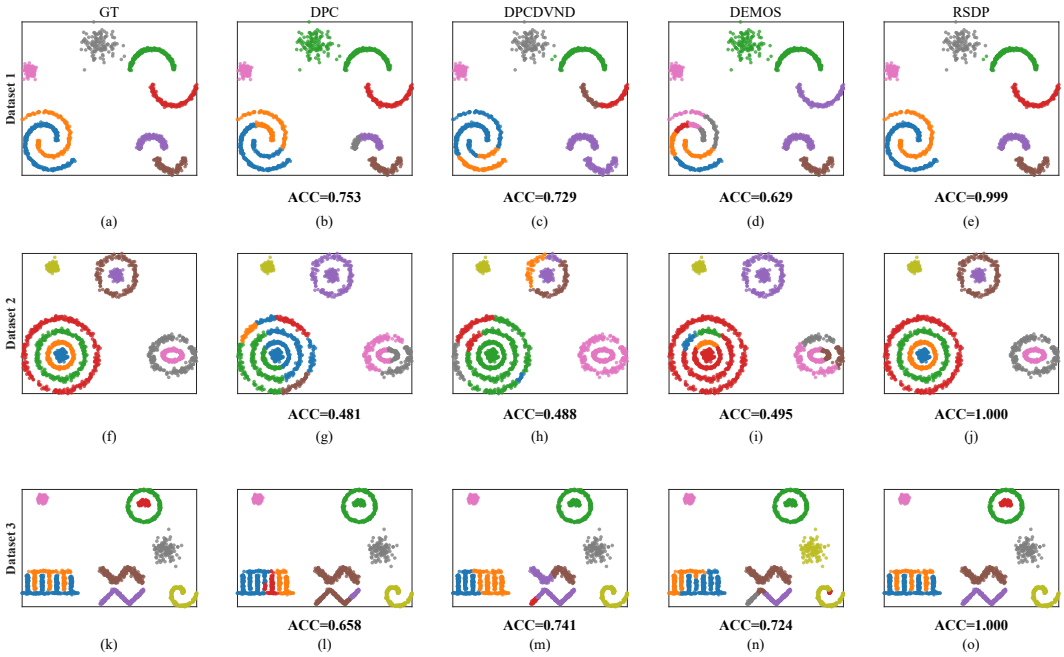


Figure 1: Comparison of clustering predictions on three synthetic datasets with heterogeneous, interlocking, and nested cluster geometries. Each column shows one method's prediction (or ground truth in the leftmost column). Predicted labels are aligned to ground-truth colors via Hungarian matching. RSDP attains near-perfect recovery on all three datasets.

5.1 Experiments on synthetic datasets

We construct three synthetic datasets whose clusters exhibit heterogeneous density and multi-modal sub-structures (Figure 1(a), (f), (k)). DPC (Figure 1(b), (g), (l)) chains each non-center to its nearest denser sample, so a low-density cluster adjacent to a denser one collapses through a single edge, and modal peaks pass as centers. DEMOS (Figure 1(d), (i), (n)) lifts inference to density mounts via a multi-valley-link term that assumes intra-cluster mounts are more connected than inter-cluster ones; once an intra-cluster valley matches a genuine gap, the term splits clusters into competing manual-pick candidates. Although DPCDVND (Figure 1(c), (h), (m)) patches the sample-level decisions with two diffusion rules gated by a global density threshold and a per-point kNN radius, neither adapts to the unequal cluster density: diffusion bleeds across genuine gaps in sparse regions and stops short inside multi-modal ones. RSDP (Figure 1(e), (j), (o)) avoids this problem by making the decision on hierarchical structural units instead of samples or mounts. Modal variation is absorbed at a lower resolution before any center is scored, so the gap statistic responds to topology rather than to local contrast. These results confirm that RSDP recovers ground truth on all three datasets, whereas the selected baselines have failed to do so to varying degrees.

Table 2: ACC and NMI of RSDP against ten state-of-the-art DPC variants on the real-world datasets. **Bold** marks the best, underline marks the second-best.

	COIL100		NH		UMIST		CMUPIE		MNIST		Pendigits		Miceprotein	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
DPC [2014]	0.5329	0.8641	0.7798	0.8207	0.5096	0.7701	0.7777	0.9525	0.4408	0.5293	0.7396	0.7849	0.4708	0.4849
DPCDVND [2024]	0.8029	0.9469	0.6028	0.5080	0.7252	0.8737	0.2174	0.4730	0.4153	0.4723	0.7621	0.7722	0.4002	0.4664
KSFDPC [2023]	0.5969	0.8416	0.5697	0.7367	0.7026	0.8521	0.9443	0.9725	0.7335	0.7959	0.7831	0.8081	0.3203	0.4599
DPCFSC [2022]	0.6551	0.8944	0.7940	0.8444	0.5791	0.7412	0.1642	0.3998	0.6487	0.6907	0.7746	0.7695	0.4568	0.5397
SMMP [2023b]	0.2400	0.7292	0.5464	0.6245	0.3826	0.7210	0.0147	0.0000	0.5309	0.5569	0.8417	0.8070	0.4847	0.5250
DBDPC [2024]	0.5976	0.8488	0.4976	0.6793	0.5670	0.7178	0.6838	0.8753	0.7547	0.7923	0.7416	0.7246	0.4123	0.5687
VDPC [2023]	0.5428	0.8662	0.7717	0.6315	0.5148	0.7720	0.7731	0.8166	0.6403	0.6928	0.7928	0.7657	0.4345	0.4209
DEMOS [2023a]	0.2300	0.7215	0.7219	0.6337	0.3322	0.6874	0.3582	0.6426	0.5175	0.5388	0.6655	0.7603	0.5404	0.5359
DPCMFP [2025]	0.5406	0.8279	0.6376	0.6713	0.4243	0.6897	0.2696	0.6363	0.5107	0.5899	0.7757	0.7958	0.4308	0.5352
FDPAN [2025b]	0.5233	0.8177	0.5170	0.5762	0.5774	0.8002	0.2581	0.5144	0.4991	0.4939	0.8634	0.8216	0.4559	0.5158
RSDP	**0.9144**	**0.9788**	**0.9629**	**0.9397**	**0.8191**	**0.9264**	**1.0000**	**1.0000**	**0.8958**	**0.8297**	**0.8753**	**0.8412**	**0.5766**	**0.5876**

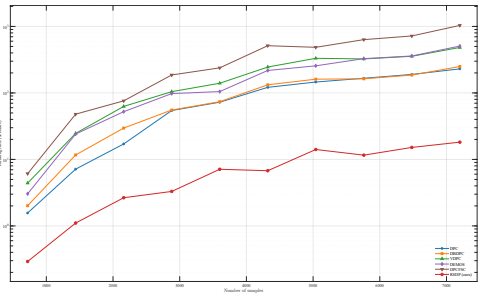
5.2 Experiments on real-world datasets

Table 2 reports the ACC and NMI of all the methods on real-world datasets, in which RSDP achieves the best score and exceeds the strongest competitor, KSFDPC, by 19.91 and 9.09 percentage points in average ACC and NMI. We attribute the improvement to estimating density and distance on hierarchical units, which absorb sample-level perturbations and adapt their granularity to the local density before any peak is selected. The original DPC relies on a single global cutoff d_c over raw samples and fails to simultaneously resolve dense low-dimensional Pendigits clusters and sparse high-dimensional UMIST manifolds, since no single scalar bandwidth adapts well to both regimes. Three kNN-based variants, KSFDPC, DPCMFP, and FDPAN, use a fixed neighborhood size k and are ineffective on COIL100, which has only 72 high-dimensional samples per class, so the kNN graph either fragments true objects under sampling noise or links distinct classes. The density-stratified variants (DBDPC and VDPC) assume globally separable density levels, and their ACC degrades on Miceprotein, whose imbalanced protein-expression clusters span overlapping density bands resisting hard stratification. DPCFSC and DEMOS rely on local similarity or mount-level aggregations that might be less compatible with the multi-pose, manifold-structured CMUPIE dataset, where intra-class variance is large. Although these datasets differ in dimensionality, class balance, and density heterogeneity, RSDP gives the best ACC and NMI on every one of them.

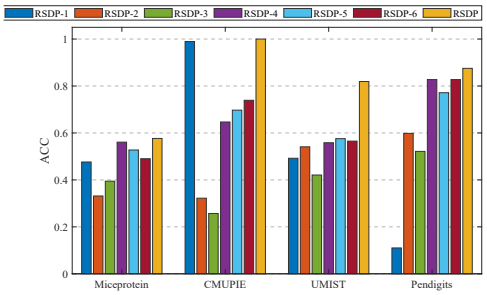
5.3 Running time and ablation

Running time. We measure runtime against five representative DPC variants (DPC, DBDPC, VDPC, DEMOS, and DPCFSC) on COIL100 nested subsets (Figure 2(a)). All baselines grow steeply with data size: DPCFSC reaches 1036s; DPC, DBDPC, VDPC, and DEMOS range from 230s to 507s when $n = 7200$. RSDP finishes the same task in 18.2s. Two factors contribute to this gap: (i) nearest-neighbor indices can reduce the empirical cost of the iterated 1-NN hierarchy construction, and (ii) all subsequent stages, i.e., peak finding and cluster assignment, operate on $M \ll n$ structural units instead of samples. Across all subsets, RSDP's runtime grows more slowly with n than that of the sample-level methods.

Ablation. As shown in Figure 2(b), RSDP achieves the highest ACC on all four datasets. The constrained hierarchy outperforms generic linkages because its size-asymmetric merge rule lets larger sub-clusters absorb smaller ones, propagating connectivity along manifolds and suppressing the early incorrect merges that fragment single-linkage and over-coarsen average/complete-linkage. Single-linkage $\widehat{d}$ is the empirical analogue of the population infimum d^* and converges to it under sufficient boundary coverage (Proposition 1); the centroid/average/complete alternatives estimate different population functionals (mean-pair, expected-pair, sup-pair distances) and induce biased empirical gaps that violate the margin condition (A2). Replacing the hierarchy (RSDP-1/2/3) costs more on average than changing only the distance functional (RSDP-4/5/6), yet even the latter stay below RSDP on all four datasets. Both the size-constrained hierarchy and a distance functional aligned with the population infimum used in the theory therefore contribute.

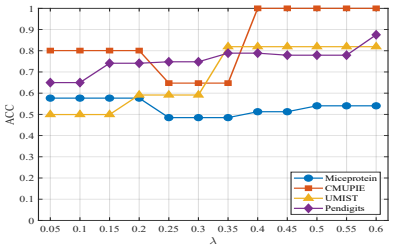


(a) Running time vs. data size on COIL100 (log-scale).

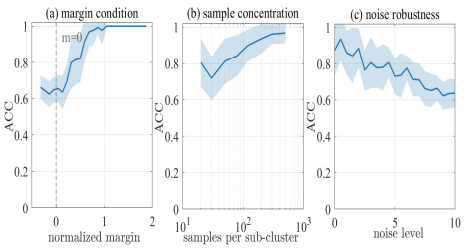


(b) Ablation of CHC hierarchy and single-linkage $\widehat{d}$.

Figure 2: (a) RSDP shows better empirical scaling than five sample-level DPC baselines on COIL100 nested subsets ($n = 720, \ldots, 7200$). (b) RSDP-1/2/3 replace the CHC hierarchy with single-/average-/complete-linkage at matched granularity; RSDP-4/5/6 replace the single-linkage $\widehat{d}$ with centroid/average/complete distances on the same hierarchy. Each variant uses the same λ as baseline RSDP on the corresponding dataset.



(a) ACC vs. λ on four datasets.



(b) Synthetic study of Theorem 1.

Figure 3: (a) ACC of RSDP across $\lambda \in [0.05, 0.60]$ on four real-world datasets; performance is strongest in the mid-range $\lambda \in [0.4, 0.6]$ on CMUPIE, UMIST, and Pendigits. (b) Results on synthetic $K{=}4$ Gaussian mixtures with controllable population statistics: mean ACC over 20 random seeds; shaded bands denote ± 1 standard deviation; $\lambda{=}0.10$ throughout. Sub-panels test the margin condition (A2), structural-density concentration, and the relaxation of distance-stability (D), respectively.

5.4 Hyperparameter sensitivity and synthetic study

Sensitivity to λ. Figure 3(a) reports the ACC of RSDP across $\lambda \in [0.05, 0.60]$ on four datasets. Performance is strongest in the mid-range $\lambda \in [0.4, 0.6]$ on CMUPIE, UMIST, and Pendigits, which we therefore recommend, whereas Miceprotein peaks at $\lambda \leq 0.2$. The two regimes outside this range can fail for distinct reasons. A smaller λ selects a fine hierarchy level whose units are dominated by sample-level noise and outliers; the empirical $\widehat{\rho}$ is then poorly concentrated, eroding the γ-margin required by Theorem 1. Large λ selects a coarse level whose units can span multiple ground-truth clusters, violating the refinement assumption $Q_i \subseteq G_{k(i)}$. The plateaus in $[0.4, 0.6]$ on CMUPIE and UMIST indicate that performance is not overly sensitive to λ within this range.

Synthetic study. Figure 3(b) examines the three ingredients of Theorem 1 on synthetic data. RSDP is run with $\lambda{=}0.10$ (a fine resolution chosen to expose the controlled sub-cluster geometry of the synthetic mixture). As the normalized margin $m/(a_{\min}R)$, with $m = a_{\min}\Delta - b_{\max}R$, grows beyond zero (dashed line), ACC rises from about 0.65 to near-perfect recovery, consistent with the role of the margin condition (A2). ACC increases overall with the sub-cluster sample size n on a log scale, consistent with the Binomial concentration of $\widehat{\rho}/n$ around its population structural density at rate $O(n^{-1/2})$. ACC degrades gradually as the noise-to-blob ratio $\sigma/\sigma_{\text{blob}}$ grows, reflecting the relaxation of the distance-stability condition (D) as ν_n grows. Together, the three panels illustrate how these conditions affect recovery in practice.

Table 3: ACC and NMI of RSDP against agglomerative clustering with single, average, and Ward linkage, and against the density-connected methods DBSCAN and HDBSCAN on the real-world datasets. **Bold** marks the best, underline marks the second-best.

	COIL100		NH		UMIST		CMUPIE		MNIST		Pendigits		Miceprotein	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Single [1957]	0.3504	0.6990	0.3230	0.1185	0.4261	0.6685	0.9496	0.9897	0.1142	0.0144	0.1103	0.0475	0.1551	0.0853
Average [1958]	0.2794	0.7256	0.4436	0.4923	0.3843	0.6298	0.1961	0.5060	0.1157	0.0217	0.4639	0.6058	0.3779	0.3584
Ward [1963]	0.6797	0.8660	0.8998	0.8456	0.4261	0.6456	0.2728	0.6401	0.6946	0.7300	0.6723	0.7179	0.4847	0.5135
DBSCAN [1996]	0.3728	0.7240	0.4682	0.6895	0.3200	0.5503	0.4093	0.7020	0.1135	0.0000	0.2999	0.5453	0.3315	0.4702
HDBSCAN [2013]	0.7556	0.9180	0.0987	0.5630	0.6452	0.8390	0.9839	0.9865	0.6141	0.6689	0.7807	0.8325	0.1848	**0.6457**
RSDP	**0.9144**	**0.9788**	**0.9629**	**0.9397**	**0.8191**	**0.9264**	**1.0000**	**1.0000**	**0.8958**	**0.8297**	**0.8753**	**0.8412**	**0.5766**	0.5876

5.5 Additional comparisons

Since RSDP builds a hierarchy and infers density, we further compare it with standard agglomerative clustering (single-, average-, and Ward-linkage [Sneath, 1957, Sokal and Michener, 1958, Ward, 1963]) and density-connected clustering (DBSCAN [Ester et al., 1996] and HDBSCAN [Campello et al., 2013]) in Table 3. RSDP is the best in 13 of the 14 dataset–metric pairs, the exception being NMI on Miceprotein, where HDBSCAN reaches 0.6457 against 0.5876. Averaged over the seven datasets, RSDP attains 0.8634 ACC and 0.8719 NMI, compared with 0.5900 and 0.7084 for Ward-linkage and 0.5804 and 0.7791 for HDBSCAN, the two strongest of these baselines. HDBSCAN tends to over-fragment high-dimensional or density-heterogeneous data, most visibly on NH and Miceprotein, whereas RSDP compares structural units at one common resolution.

6 Conclusion

We have argued that density peaks inference should be performed on resolution-controlled structural units rather than on samples. RSDP puts this into practice by building a constrained hierarchy, using a single resolution parameter to choose the level on which ρ, δ, and γ are defined, and assigning labels in one pass over the units at that level. Conditional on a faithful structural realization, RSDP's inference layer recovers the dominant population units and the sampled ground-truth partition. Empirically, RSDP shows a substantial accuracy advantage over DPC variants across the seven real-world datasets and better runtime scaling than sample-level baselines. As in standard DPC, RSDP takes the number of clusters K as input, together with a single resolution λ. Natural directions for future work include estimating K from the gap in the descending γ ordering, selecting λ without labels, and scale-aware structural densities for units of very different spatial extent.

Acknowledgment

The authors acknowledge the support of the School of Biomedical Engineering, The University of Sydney. This work is also supported in part by the Zhejiang Provincial Natural Science Foundation of China (LMS26F020035), the National Natural Science Foundation of China (62576178 and U2433216), and the Chongqing Postdoctoral Research Special Funding Project (2025CQBSHTB3038).

The authors declare no competing interests.

References

Fevzi Alimoglu and Ethem Alpaydin. Combining multiple representations and classifiers for pen-based handwritten digit recognition. In *Proceedings of the International Conference on Document Analysis and Recognition*, volume 2, pages 637–640, 1997.

Ricardo J. G. B. Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining (PAKDD 2013)*, volume 7819 of *Lecture Notes in Computer Science*, pages 160–172. Springer, 2013.

Xiaochun Cao, Changqing Zhang, Chengju Zhou, Huazhu Fu, and Hassan Foroosh. Constrained multi-view video face clustering. *IEEE Transactions on Image Processing*, 24(11):4381–4393, 2015.

Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for the cluster tree. In *Advances in Neural Information Processing Systems*, volume 23, pages 343–351, 2010.

Zhikui Chen, Meng Liu, Yuzhe Li, Hao Zheng, Zhenjiao Liu, Hongwei Zhang, and Liang Zhao. Double-incomplete multi-view clustering with self-induced semantic label diffusion. *Information Sciences*, 753:123626, 2026.

Dongdong Cheng, Ya Li, Shuyin Xia, Guoyin Wang, Jinlong Huang, and Sulan Zhang. A fast granular-ball-based density peaks clustering algorithm for large-scale data. *IEEE Transactions on Neural Networks and Learning Systems*, 35(12):17202–17215, 2024.

Ling Ding, Chao Li, Shifei Ding, Xiao Xu, Lili Guo, and Xindong Wu. Robust density peaks clustering for manifold data with multiple peaks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11):10696–10708, 2025a.

Shifei Ding, Chao Li, Xiao Xu, Lili Guo, Ling Ding, and Xindong Wu. Fast density peaks clustering algorithm based on approximate k-nearest neighbors. *IEEE Transactions on Knowledge and Data Engineering*, 37(10):5878–5889, 2025b.

Zhibin Dong, Dayu Hu, Jiaqi Jin, Siwei Wang, Xinwang Liu, and En Zhu. Selective cross-view topology for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34: 4792–4805, 2025.

Zhibin Dong, Shengju Yu, Siwei Wang, Xinwang Liu, and En Zhu. Cross-view graph matching for unsupervised learning with unaligned multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 38(4):2221–2233, 2026.

Yu Duan, Huimin Chen, Runxin Zhang, Rong Wang, Feiping Nie, and Xuelong Li. Soft neighbors supported contrastive clustering. *IEEE Transactions on Image Processing*, 34:4315–4327, 2025.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231. AAAI Press, 1996.

Zexuan Fei, Yan Ma, Jinfeng Zhao, Bin Wang, and Jie Yang. KNEG-CL: Unveiling data patterns using a k-nearest neighbor evolutionary graph for efficient clustering. *Information Sciences*, 690: 121602, 2025a.

Zexuan Fei, Haoyu Zhai, Jie Yang, Bin Wang, and Yan Ma. Discovering generalized clusters with adaptive mixture density-based clustering. *Knowledge-Based Systems*, 314:113250, 2025b.

Daniel B. Graham and Nigel M. Allinson. Characterising virtual eigensignatures for general purpose face recognition. In H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman Soulié, and T. S. Huang, editors, *Face Recognition: From Theory to Applications*, volume 163 of *NATO ASI Series F*, pages 446–456. Springer, Berlin, Germany, 1998.

Junyi Guan, Sheng Li, Xiaojun Chen, Xiongxiong He, and Jiajia Chen. DEMOS: Clustering by pruning a density-boosting cluster tree of density mounts. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10814–10830, 2023a.

Junyi Guan, Sheng Li, Xiongxiong He, Jinhui Zhu, Jiajia Chen, and Peng Si. SMMP: A stable-membership-based auto-tuning multi-peak clustering algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6307–6319, 2023b.

John A. Hartigan. Consistency of single linkage for high-density clusters. *Journal of the American Statistical Association*, 76(374):388–394, 1981.

Clara Higuera, Katheleen J. Gardiner, and Krzysztof J. Cios. Self-organizing feature maps identify proteins critical to learning in a mouse model of Down syndrome. *PLoS One*, 10(6):e0129126, June 2015.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Jian Hou, Houshen Lin, Huaqiang Yuan, and Marcello Pelillo. Flexible density peak clustering for real-world data. *Pattern Recognition*, 156:110772, 2024.

Bingbing Jiang, Zhongli Wang, Jie Yang, Guang-Kui Xu, Wei Chen, Chenglong Zhang, Xinyan Liang, Peng Zhou, Weiguo Sheng, and Weiping Ding. Self-enhanced density clustering for high dimension and low sample size data. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 508–519, 2026.

Kwangho Kim, Jisu Kim, Larry A. Wasserman, and Edward H. Kennedy. Hierarchical and density-based causal clustering. In *Advances in Neural Information Processing Systems*, pages 30363–30393, 2024.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Chao Li, Shifei Ding, Xiao Xu, Haiwei Hou, and Ling Ding. Fast density peaks clustering algorithm based on improved mutual K-nearest-neighbor and sub-cluster merging. *Information Sciences*, 647:119470, 2023.

Wentao Li, Lingwei Wei, Witold Pedrycz, Weiping Ding, Chao Zhang, Tao Zhan, and Shuyin Xia. Granular-ball regeneration clustering with principle of justifiable granularity. *IEEE Transactions on Neural Networks and Learning Systems*, 36(10):18173–18187, 2025.

Yan Li, Lingyun Sun, and Yongchuan Tang. DPC-FSC: An approach of fuzzy semantic cells to density peaks clustering. *Information Sciences*, 616:88–107, 2022.

Jiyuan Liu, Xinwang Liu, Siwei Wang, Sihang Zhou, and Yuexiang Yang. Hierarchical multiple kernel clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8671–8679, 2021.

Rui Liu, Hong Wang, and Xiaomei Yu. Shared-nearest-neighbor-based clustering by fast search and find of density peaks. *Information Sciences*, 450:200–226, 2018.

Linlin Ma, Hui Li, Xincheng Liu, Huihui Chu, Yue Guan, Yuzhen Zhao, Yawen Chen, Da Wang, and Wenke Zang. CD-DPC: Centrifugal degree based density peaks clustering algorithm. *Pattern Recognition*, 176:113223, 2026.

Sameer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (COIL-100). Technical Report CUCS-006-96, Department of Computer Science, Columbia University, New York, NY, USA, 1996.

Feiping Nie, Yitao Song, Qilong Qiu, Jingjing Xue, Rong Wang, and Xuelong Li. Clustering based on density propagation and subcluster merging. *IEEE Transactions on Knowledge and Data Engineering*, 38(3):1541–1559, 2026.

Mateusz Pach, Przemysław Rola, Michał Znalezniak, Patryk Kaszuba, Marcin Przewiezlikowski, Jacek Tabor, and Marek Śmieja. Adaptive contrastive hierarchical clustering with nodes visualization. *Artificial Intelligence Review*, 59(3):99, 2026.

Alex Rodriguez and Alessandro Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

Saquib Sarfraz, Vivek Sharma, and Rainer Stiefelhagen. Efficient parameter-free clustering using first neighbor relations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8926–8935, 2019.

Terence Sim, Simon Baker, and Maan Bsat. The CMU pose, illumination, and expression (PIE) database. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 53–58, 2002.

Peter H. A. Sneath. The application of computers to taxonomy. *Journal of General Microbiology*, 17 (1):201–226, 1957.

Robert R. Sokal and Charles D. Michener. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38:1409–1438, 1958.

Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.

Lin Sun, Xiaoying Qin, Weiping Ding, and Jiucheng Xu. Nearest neighbors-based adaptive density peaks clustering with optimized allocation strategy. *Neurocomputing*, 473:159–181, 2022.

Qijia Wang and Mingjing Du. Structure-aware granular ball clustering. *Information Sciences*, 740: 123224, 2026.

Yizhang Wang, Di Wang, You Zhou, Xiaofeng Zhang, and Chai Quek. VDPC: Variational density peak clustering algorithm. *Information Sciences*, 621:627–651, 2023.

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.

Zhongli Wang, Jie Yang, Junyi Guan, Chenglong Zhang, Xinyan Liang, Bingbing Jiang, and Weiguo Sheng. Enhanced density peak clustering for high-dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 21411–21419, 2025.

Joe H. Ward, Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

Juanying Xie, Huan Yan, Mingzhao Wang, Philip W. Grant, and Witold Pedrycz. WANN-DPC: Density peaks finding clustering based on weighted adaptive nearest neighbors. *Pattern Recognition*, 170:111953, 2026.

HaoChuan Xu and Ninh Pham. Scalable DBSCAN with random projections. In *Advances in Neural Information Processing Systems*, pages 27978–28008, 2024.

Ji Xu, Guoyin Wang, and Weihui Deng. DenPEHC: Density peak based efficient hierarchical clustering. *Information Sciences*, 373:200–218, 2016.

Jie Yang and Chin-Teng Lin. Multi-view adjacency-constrained hierarchical clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(4):1126–1138, 2023.

Jie Yang and Chin-Teng Lin. Enhanced adjacency-constrained hierarchical clustering using fine-grained pseudo labels. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8 (3):2481–2492, 2024.

Jie Yang and Chin-Teng Lin. Toward autonomous distributed clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(2):2065–2072, 2025a.

Jie Yang and Chin-Teng Lin. Autonomous clustering by fast find of mass and distance peaks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):5336–5349, 2025b.

Jie Yang, Yan Ma, Xiangfen Zhang, Shunbao Li, and Yuping Zhang. An initialization method based on hybrid distance for k-means algorithm. *Neural Computation*, 29(11):3094–3117, 2017.

Jie Yang, Yu-Kai Wang, Xin Yao, and Chin-Teng Lin. Adaptive initialization method for K-means algorithm. *Frontiers in Artificial Intelligence*, 4:740817, 2021.

Jie Yang, Wei Chen, Feng Liu, Peng Zhou, Zhongli Wang, Xinyan Liang, and Bingbing Jiang. Multi-view clustering via multi-granularity ensemble. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 6794–6802, 2025.

Jie Yang, Cheng-You Lu, Zhongli Wang, Hsiang-Ting Chen, Guang-Kui Xu, Chenglong Zhang, Shuting Dong, Xinyan Liang, and Bingbing Jiang. Multi-view clustering with granularity-aware pseudo supervision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 27538–27546, 2026.

Wenke Zang, Jing Che, Linlin Ma, Xincheng Liu, Aoyu Song, Jingwen Xiong, Yuzhen Zhao, Xiyu Liu, Yawen Chen, and Hui Li. Density peaks clustering based on density voting and neighborhood diffusion. *Information Sciences*, 681:121209, 2024.

Wenke Zang, Xincheng Liu, Linlin Ma, Minghe Sun, Jing Che, Yuzhen Zhao, Yuanhua Wang, Da Wang, and Xiyu Liu. DPC-MFP: An adaptive density peaks clustering algorithm with multiple feature points. *Neurocomputing*, 618:129060, 2025.

Haoyu Zhai, Jie Yang, Hantao Guo, Bin Wang, and Yan Ma. Density-increment and cut-edge optimized clustering via minimum spanning forest. *Neurocomputing*, 674:132957, 2026.

Qinghua Zhang, Yongyang Dai, and Guoyin Wang. Density peaks clustering based on balance density and connectivity. *Pattern Recognition*, 134:109052, 2023.

Lin Zhao, Yuan Yuan, and Feiping Nie. Efficient hierarchical multi-resolution k-means clustering. *Pattern Recognition*, 174:112930, 2026.

Zhou Zhou, Gangquan Si, Haodong Sun, Kai Qu, and Weicheng Hou. A robust clustering algorithm based on the identification of core points and KNN kernel density estimation. *Expert Systems with Applications*, 195:116573, 2022.

APPENDIX

A Theoretical Setup, Lemmas, and Proofs

This appendix expands the formal setup of §4, discusses sufficient conditions under which the distance-stability assumption (D) holds, states four intermediate lemmas, and gives complete proofs of Theorems 1 and 2.

A.1 Setup details

We restate notation from §4 and add the cluster-wise quantities deferred to this appendix. $(\mathcal{X}, d)$ is a separable metric space, P a Borel probability measure on $\mathcal{X}$, and $X_1, \dots, X_n$ are i.i.d. from P. The ground-truth partition $G_1, \dots, G_K$ has $\pi_k = P(G_k) > 0$. At resolution λ, we fix a population structural partition $\mathcal{Q} = \{Q_1, \dots, Q_M\}$ with $Q_i \subseteq G_{k(i)}$ and $q_i = P(Q_i) > 0$ for all i, refining the ground-truth partition with $M \geq K+1$.

For each G_k, let $\mathcal{Q}_k = \{Q \in \mathcal{Q} : Q \subseteq G_k\}$. Order the units in $\mathcal{Q}_k$ by population structural density and let $P_k^* = \arg\max_{Q \in \mathcal{Q}_k} q_Q$ be the *dominant* unit, $a_k = q_{P_k^*}$, and b_k the second-largest population structural density in $\mathcal{Q}_k$ (or 0 if $|\mathcal{Q}_k| = 1$). The within-cluster *structural radius* and inter-cluster separation are

$$R_k = \max_{Q,Q' \in \mathcal{Q}_k} d^*(Q, Q'), \;\; R = \max_k R_k, \;\; \Delta_k = \min_{Q \in \mathcal{Q}_k, Q' \notin \mathcal{Q}_k} d^*(Q, Q'), \;\; \Delta = \min_k \Delta_k. \tag{12}$$

We write $a_{\min} = \min_k a_k$, $b_{\max} = \max_k b_k$, $q_{\min} = \min_i q_i$. The global structural margin is $m = a_{\min}\Delta - b_{\max}R$.

For each population unit Q_i, define the (unconditional) count

$$N_i = \sum_{j=1}^{n} \mathbf{1}\{X_j \in Q_i\} \sim \text{Binomial}(n, q_i), \qquad \textstyle\sum_i N_i = n. \tag{13}$$

The empirical structural partition at level l^* is $\widehat{\mathcal{C}} = \{\widehat{C}_1, \dots, \widehat{C}_{\widehat{M}}\}$, with single-linkage distance $\widehat{d}$ and empirical $(\widehat{\rho}, \widehat{\delta}, \widehat{\gamma})$ defined by (4), (5), (6). The faithful realization event $\mathcal{E}_n$ requires $\widehat{M} = M$ and a permutation $\sigma \in S_M$ such that $\widehat{C}_{\sigma(i)} = \{X_j : X_j \in Q_i\}$ for all i. On $\mathcal{E}_n$, $|\widehat{C}_{\sigma(i)}| = N_i$.

We assume throughout that R and Δ are finite.

For each empirical unit, the *population peak score* is

$$\gamma^*(Q_i) = q_i \cdot \delta^*(Q_i), \qquad \delta^*(Q_i) = \begin{cases} \min_{Q_j \succ_P Q_i} d^*(Q_i, Q_j), & \text{if predecessor exists,} \\ \max_{Q_j \neq Q_i} d^*(Q_i, Q_j), & \text{otherwise,} \end{cases} \tag{14}$$

where $\succ_P$ is the population priority order ranked by population structural density with the same fixed tie-breaking ι.

A.2 Sufficient conditions for (D)

Assumption (D) requires the existence of an event $\mathcal{D}_n$ with $\Pr(\mathcal{D}_n^c \cap \mathcal{E}_n) \le o(1)$ on which $\sup_{i\neq j} |\widehat{d}(\widehat{C}_{\sigma(i)}, \widehat{C}_{\sigma(j)}) - d^*(Q_i, Q_j)| \le \nu_n$. Since $\widehat{d}$ is the minimum pairwise sample distance and d^* is the population infimum, we always have $\widehat{d} \ge d^*$ on $\mathcal{E}_n$ (Lemma 2 below). So (D) amounts to controlling the upper deviation $\widehat{d} - d^*$ on a high-probability event jointly with $\mathcal{E}_n$.

Proposition 1 (Sufficient condition for (D))**.** *Suppose for every pair (Q_i, Q_j) with $i \neq j$ and every $t > 0$, there exist measurable boundary neighborhoods $A_{ij}(t) \subseteq Q_i$ and $B_{ij}(t) \subseteq Q_j$ such that*

$$\sup_{x\in A_{ij}(t),\, y\in B_{ij}(t)} d(x,y) \le d^*(Q_i,Q_j) + t, \quad \min\left\{\frac{P(A_{ij}(t))}{q_i},\, \frac{P(B_{ij}(t))}{q_j}\right\} \ge g_{ij}(t), \tag{15}$$

where g_{ij} is non-decreasing and right-continuous, with $g_{ij}(t) > 0$ for all $t > 0$ and $\lim_{t\to 0^+} g_{ij}(t) = 0$. Define the Chernoff event $\mathcal{C}_n = \bigcap_{i=1}^M \{N_i \ge nq_i/2\}$ and let

$$\nu_n = \inf\big\{t > 0 :\, n\, q_{\min}\, g_{\min}(t) \ge 4\log(M^2 n)\big\}, \quad g_{\min} = \min_{i\neq j} g_{ij}. \tag{16}$$

Then there exists a boundary coverage event $\mathcal{B}_n$ with $\Pr(\mathcal{B}_n^c) \le O(1/n)$ such that on $\mathcal{E}_n \cap \mathcal{C}_n \cap \mathcal{B}_n$,

$$\sup_{i\neq j} \big|\widehat{d}(\widehat{C}_{\sigma(i)}, \widehat{C}_{\sigma(j)}) - d^*(Q_i,Q_j)\big| \le \nu_n. \tag{17}$$

Sketch. The lower bound is Lemma 2. For the upper bound, define the boundary coverage event using the raw sample sets in each population unit (independent of the hierarchy output):

$$\mathcal{B}_n = \bigcap_{i\neq j} \Big\{ \big\{X_\ell : X_\ell \in Q_i\big\} \cap A_{ij}(\nu_n) \neq \emptyset \ \wedge\ \big\{X_\ell : X_\ell \in Q_j\big\} \cap B_{ij}(\nu_n) \neq \emptyset \Big\}. \tag{18}$$

The events $\{X_\ell \in Q_i\}$ and $\{X_\ell \in A_{ij}(\nu_n)\}$ are unconditionally independent across ℓ, and a sample drawn from $P|_{Q_i}$ lies in $A_{ij}(\nu_n)$ with probability $P(A_{ij}(\nu_n))/q_i \ge g_{ij}(\nu_n)$. Hence

$$\Pr\big(\{X_\ell : X_\ell \in Q_i\} \cap A_{ij}(\nu_n) = \emptyset \,\big|\, N_i\big) \le (1 - g_{ij}(\nu_n))^{N_i} \le \exp(-N_i\, g_{ij}(\nu_n)).$$

On $\mathcal{C}_n$, $N_i \ge nq_{\min}/2$, so this is at most $\exp(-nq_{\min}g_{ij}(\nu_n)/2) \le 1/(M^2 n)$ by the definition of ν_n. Analogously for $B_{ij}(\nu_n)$. Union bound over the two sub-events and the $\binom{M}{2}$ pairs gives $\Pr(\mathcal{B}_n^c \cap \mathcal{C}_n) \le O(1/n)$, so $\Pr(\mathcal{B}_n^c) \le \Pr(\mathcal{C}_n^c) + O(1/n) = o(1)$.

On $\mathcal{E}_n$, the empirical cluster $\widehat{C}_{\sigma(i)}$ coincides with the raw sample set $\{X_\ell : X_\ell \in Q_i\}$ by definition of $\mathcal{E}_n$. Therefore on $\mathcal{E}_n \cap \mathcal{C}_n \cap \mathcal{B}_n$, every pair (i,j) has at least one cross-pair $(x,y) \in A_{ij}(\nu_n) \times B_{ij}(\nu_n)$ with both $x \in \widehat{C}_{\sigma(i)}$ and $y \in \widehat{C}_{\sigma(j)}$, yielding $\widehat{d} \le d^* + \nu_n$. □

The proposition shows (D) holds whenever any positive-measure neighborhoods of the boundary infimum exist on both sides of every pair: a mild regularity condition met by all distributions with continuous, strictly positive density on a neighborhood of population unit boundaries.

A.3 Lemmas

Lemma 1 (Structural-density concentration)**.** *For any $\varepsilon > 0$, define the density concentration event*

$$\mathcal{H}_n(\varepsilon) = \bigcap_{i=1}^M \big\{|N_i/n - q_i| \le \varepsilon\big\}. \tag{19}$$

Then $\Pr(\mathcal{H}_n(\varepsilon)) \ge 1 - 2M\exp(-2n\varepsilon^2)$. On $\mathcal{E}_n \cap \mathcal{H}_n(\varepsilon)$, the empirical structural densities satisfy $|\widehat{\rho}(\widehat{C}_{\sigma(i)})/n - q_i| \le \varepsilon$ for all $i \in [M]$.

Proof. $N_i = \sum_{j=1}^n \mathbf{1}\{X_j \in Q_i\}$ is unconditionally a sum of n i.i.d. Bernoulli(q_i) random variables, so Hoeffding's inequality [Hoeffding, 1963] gives $\Pr(|N_i/n - q_i| > \varepsilon) \le 2\exp(-2n\varepsilon^2)$; the union bound over $i \in [M]$ yields the claim. On $\mathcal{E}_n$, $\widehat{\rho}(\widehat{C}_{\sigma(i)}) = N_i$. □

Lemma 2 (Deterministic distance lower bound)**.** *On $\mathcal{E}_n$, $\widehat{d}(\widehat{C}_{\sigma(i)}, \widehat{C}_{\sigma(j)}) \geq d^*(Q_i, Q_j)$ for all $i \neq j$.*

Proof. On $\mathcal{E}_n$, every $x \in \widehat{C}_{\sigma(i)}$ satisfies $x \in Q_i$ and similarly $y \in Q_j$. Therefore $\widehat{d} = \min_{x \in \widehat{C}_{\sigma(i)}, y \in \widehat{C}_{\sigma(j)}} d(x, y) \geq \inf_{x \in Q_i, y \in Q_j} d(x, y) = d^*(Q_i, Q_j)$, since the empirical infimum is taken over a (random) subset of the population pairs. □

Lemma 3 (Population separability)**.** *Under (A1), (A2), for every $k \in [K]$,*

$$\gamma^*(P_k^*) \geq a_{\min}\Delta > b_{\max}R \geq \gamma^*(Q) \quad \text{for all } Q \in \mathcal{Q} \setminus \{P_1^*, \ldots, P_K^*\},$$

so the top-K population units ranked by γ^* are exactly $\{P_k^*\}_{k=1}^K$.

Proof. Lower bound on $\gamma^*(P_k^*)$: Under (A1), P_k^* is the unique density-maximizer in $\mathcal{Q}_k$, so any predecessor of P_k^* under $\succ_P$ has no lower population structural density and hence lies outside $\mathcal{Q}_k$. The symmetric definition of Δ_k in (10) gives $\delta^*(P_k^*) \geq \Delta_k \geq \Delta$. If P_k^* has no predecessor under $\succ_P$, the $\max_{Q \neq P_k^*} d^*(P_k^*, Q) \geq \Delta$ branch applies trivially. Combining with $q_{P_k^*} = a_k \geq a_{\min}$ yields $\gamma^*(P_k^*) \geq a_{\min}\Delta$.

Upper bound on $\gamma^*(Q)$ for non-dominant Q: Any $Q \in \mathcal{Q} \setminus \{P_1^*, \ldots, P_K^*\}$ belongs to some $\mathcal{Q}_k$ with $Q \neq P_k^*$, so $q_Q \leq b_k \leq b_{\max}$. Since P_k^* has strictly higher density, P_k^* is a predecessor of Q under $\succ_P$, giving $\delta^*(Q) \leq d^*(Q, P_k^*) \leq R_k \leq R$. Thus $\gamma^*(Q) \leq b_{\max}R$.

(A2) gives $a_{\min}\Delta > b_{\max}R$, completing the strict separation. □

Lemma 4 (Parent consistency)**.** *Define the* local parent gap

$$\eta = \min_{k \in [K]} \min_{Q \in \mathcal{Q}_k \setminus \{P_k^*\}} \Big[\Delta_k - \min_{Q' \in \mathcal{Q}_k :\, q_{Q'} > q_Q} d^*(Q, Q')\Big]. \tag{20}$$

Under (A3), for each non-dominant $Q \in \mathcal{Q}_k \setminus \{P_k^*\}$, the inner minimum is well-defined and $d^*(Q, Q') < \Delta_k$ for the witness Q', so each bracket is strictly positive. Since $\mathcal{Q}$ is finite, the minimum of finitely many strictly positive gaps is strictly positive, hence $\eta > 0$. Assume (A1), (A2), (A3), (D), $K \geq 2$, and suppose $\nu_n < \eta/2$ for all sufficiently large n. Then there exists an event $\mathcal{P}_n$ with $\Pr(\mathcal{P}_n^c \cap \mathcal{E}_n) \leq o(1)$ such that on $\mathcal{E}_n \cap \mathcal{P}_n \cap \mathcal{D}_n$, for every non-dominant unit $Q \in \mathcal{Q}_k \setminus \{P_k^*\}$ the empirical nearest predecessor of $\widehat{C}_{\sigma(Q)}$ lies in $\widehat{C}_{\sigma(\mathcal{Q}_k \setminus \{Q\})}$, i.e., within the same ground-truth cluster G_k.

Proof. Fix k and $Q \in \mathcal{Q}_k \setminus \{P_k^*\}$. By (A3), choose any

$$Q'_\star \in \arg\min_{Q' \in \mathcal{Q}_k :\, q_{Q'} > q_Q} d^*(Q, Q')$$

(ties resolved arbitrarily); $d^*(Q, Q'_\star) \leq \Delta_k - \eta$ by definition of η.

Step 1 (empirical priority). By Lemma 1, the ordering $\widehat{\rho}(\widehat{C}_{\sigma(Q'_\star)}) > \widehat{\rho}(\widehat{C}_{\sigma(Q)})$ holds with probability $1 - o(1)$, since the population gap $q_{Q'_\star} - q_Q > 0$ is fixed and the Hoeffding deviations are $O_p(n^{-1/2})$. Hence $\widehat{C}_{\sigma(Q'_\star)}$ lies in the empirical predecessor set of $\widehat{C}_{\sigma(Q)}$ on a $1 - o(1)$ event.

Step 2 (within-cluster distance bound). By (D), $\widehat{d}(\widehat{C}_{\sigma(Q)}, \widehat{C}_{\sigma(Q'_\star)}) \leq d^*(Q, Q'_\star) + \nu_n \leq \Delta_k - \eta + \nu_n$.

Step 3 (cross-cluster distance bound). For any $Q'' \notin \mathcal{Q}_k$, by Lemma 2, $\widehat{d}(\widehat{C}_{\sigma(Q)}, \widehat{C}_{\sigma(Q'')}) \geq d^*(Q, Q'') \geq \Delta_k$.

Step 4 (separation). Combining steps 2 and 3,

$$\widehat{d}(\widehat{C}_{\sigma(Q)}, \widehat{C}_{\sigma(Q'')}) - \widehat{d}(\widehat{C}_{\sigma(Q)}, \widehat{C}_{\sigma(Q'_\star)}) \geq \eta - \nu_n > \eta/2 > 0$$

under the hypothesis $\nu_n < \eta/2$. Therefore no cross-cluster $\widehat{C}_{\sigma(Q'')}$ can be the empirical nearest predecessor of $\widehat{C}_{\sigma(Q)}$; the nearest predecessor must lie in $\widehat{C}_{\sigma(\mathcal{Q}_k \setminus \{Q\})}$. Union bound over the $\sum_k (|\mathcal{Q}_k| - 1) \leq M - K$ non-dominant units yields $1 - o(1)$. □

A.4 Proofs of main theorems

Proof of Theorem 1. We work on the event $\mathcal{E}_n \cap \mathcal{H}_n(\varepsilon_n)$ for an $\varepsilon_n \to 0$ to be fixed below. By Lemma 1, $\Pr(\mathcal{H}_n(\varepsilon_n)) \to 1$ whenever $n\varepsilon_n^2 \to \infty$, so it suffices to choose any such ε_n. On $\mathcal{E}_n \cap \mathcal{H}_n(\varepsilon_n)$, $\widehat{\rho}(\widehat{C}_{\sigma(i)})/n = N_i/n \in [q_i - \varepsilon_n, q_i + \varepsilon_n]$ for all $i \in [M]$.

Lower bound on $\widehat{\gamma}$ at dominant units. For P_k^*, $\widehat{\rho}(\widehat{C}_{\sigma(P_k^*)}) = N_{i^*} \geq n(a_k - \varepsilon_n)$ on $\mathcal{H}_n$. We bound $\widehat{\delta}$ by case analysis on whether $\widehat{C}_{\sigma(P_k^*)}$ has an empirical predecessor:

1. If an empirical predecessor exists, then by Lemma 1, no within-cluster non-dominant unit $Q \in \mathcal{Q}_k \setminus \{P_k^*\}$ can be the predecessor for sufficiently large n, since the strict density ordering $\widehat{\rho}(\widehat{C}_{\sigma(P_k^*)}) > \widehat{\rho}(\widehat{C}_{\sigma(Q)})$ is preserved on $\mathcal{H}_n$ whenever $\varepsilon_n < (a_k - q_Q)/2$. Hence the predecessor lies outside $\mathcal{Q}_k$ with probability $1 - o(1)$, and $\widehat{\delta}(\widehat{C}_{\sigma(P_k^*)}) \geq \Delta_k \geq \Delta$ by Lemma 2.

2. If no empirical predecessor exists (i.e., P_k^* is the empirical density-maximizer overall), $\widehat{\delta}$ takes the maximum distance to any other empirical unit. Since $K \geq 2$, there exists $k' \neq k$, so at least one population unit lies outside $\mathcal{Q}_k$. By Lemma 2, this maximum is at least $\Delta_k \geq \Delta$.

In both cases, after rescaling by n,

$$\frac{1}{n}\widehat{\gamma}(\widehat{C}_{\sigma(P_k^*)}) \geq (a_k - \varepsilon_n)\Delta \geq a_{\min}\Delta - \varepsilon_n\Delta.$$

Upper bound on $\widehat{\gamma}$ at non-dominant units. For $Q \in \mathcal{Q} \setminus \{P_k^*\}_{k=1}^K$, Q belongs to some $\mathcal{Q}_k$ with $q_Q \leq b_k \leq b_{\max}$, so $\widehat{\rho}(\widehat{C}_{\sigma(Q)})/n \leq q_Q + \varepsilon_n \leq b_{\max} + \varepsilon_n$. Since P_k^* has a strictly larger population structural density than Q, by Lemma 1 the empirical realization $\widehat{C}_{\sigma(P_k^*)}$ is in the empirical predecessor set of $\widehat{C}_{\sigma(Q)}$ with probability $1 - o(1)$. Therefore by (D) and the within-cluster radius bound,

$$\widehat{\delta}(\widehat{C}_{\sigma(Q)}) \leq \widehat{d}(\widehat{C}_{\sigma(Q)}, \widehat{C}_{\sigma(P_k^*)}) \leq d^*(Q, P_k^*) + \nu_n \leq R + \nu_n.$$

Combining,

$$\frac{1}{n}\widehat{\gamma}(\widehat{C}_{\sigma(Q)}) \leq (b_{\max} + \varepsilon_n)(R + \nu_n) = b_{\max}R + b_{\max}\nu_n + \varepsilon_n R + \varepsilon_n\nu_n.$$

Separation. The $\widehat{\gamma}$-gap between dominant and non-dominant units is at least

$$a_{\min}\Delta - b_{\max}R - \big[\varepsilon_n\Delta + b_{\max}\nu_n + \varepsilon_n R + \varepsilon_n\nu_n\big] = m - \big[\varepsilon_n(\Delta + R + \nu_n) + b_{\max}\nu_n\big].$$

By (A2) and Lemma 3, $m > 0$. The bracket tends to 0 as $n \to \infty$ since $\varepsilon_n, \nu_n \to 0$. The hypothesis $b_{\max}\nu_n < m/2$ guarantees $b_{\max}\nu_n$ alone is $< m/2$; choosing ε_n so that $\varepsilon_n(\Delta + R + \nu_n) < m/2 - b_{\max}\nu_n$ makes the gap strictly positive for all sufficiently large n. Such a choice satisfies $\Pr(\mathcal{H}_n(\varepsilon_n)) \to 1$ by Lemma 1. Therefore on $\mathcal{E}_n \cap \mathcal{H}_n(\varepsilon_n)$ intersected with the (D)-event, the top-K empirical $\widehat{\gamma}$-units coincide with $\{\widehat{C}_{\sigma(P_k^*)}\}_{k=1}^K$, and this intersection has probability $\Pr(\mathcal{E}_n) - o(1)$. □

Proof of Theorem 2. By Theorem 1, on $\mathcal{E}_n$ except on an auxiliary event of probability $o(1)$, the centers selected by RSDP are exactly the empirical realizations of the population dominant units, $\{\widehat{C}_{\sigma(P_k^*)}\}_{k=1}^K$, each receiving a distinct label corresponding bijectively to $G_1, \ldots, G_K$.

By Lemma 4, on $\mathcal{E}_n$ except on a further auxiliary event of probability $o(1)$, every empirical non-center unit $\widehat{C}_{\sigma(Q)}$ with $Q \in \mathcal{Q}_k$ has its nearest predecessor in $\widehat{C}_{\sigma(\mathcal{Q}_k \setminus \{Q\})}$. Processing non-centers in descending priority order, the assignment rule (8) thus propagates the label of G_k from $\widehat{C}_{\sigma(P_k^*)}$ down through the within-cluster predecessor chain to every $\widehat{C}_{\sigma(Q)}$ with $Q \in \mathcal{Q}_k$. This chain has length at most $|\mathcal{Q}_k| - 1 \leq M - 1$.

Finally, sample-level recovery (9) pushes the structural label down to constituent samples; on $\mathcal{E}_n$ each empirical unit $\widehat{C}_{\sigma(i)}$ contains exactly the samples in Q_i, so each $X_j \in Q_i \subseteq G_{k(i)}$ receives the label of $G_{k(i)}$. The resulting labels are constant on $\{X_j : X_j \in G_k\}$ for each k, with the predicted labels in one-to-one correspondence with $G_1, \ldots, G_K$. □

Remark on the margin regime. The margin condition $b_{\max}\nu_n < m/2$ in Theorem 1 is essentially sufficient: it captures the regime in which the empirical $\widehat{\gamma}$-gap remains positive. When the empirical distance error ν_n is of the same order as $m/b_{\max}$, the empirical $\widehat{\gamma}$-gap can shrink to zero, in which case, the top-K ranking may become unstable. The condition therefore identifies the asymptotic regime in which RSDP's structural inference layer is consistent: $\nu_n = o(m/b_{\max})$, with $m, b_{\max}$ treated as fixed properties of the population $(P, \mathcal{Q})$.

B Distinction from FINCH and TORC

The two methods closest to RSDP differ in what they infer about the hierarchy and how the partition is chosen. FINCH builds a first-neighbor hierarchy and returns the partition at one of its discrete levels; it does not perform density peaks inference on the selected level, so ρ, δ, and γ never enter its formulation. TORC builds a constrained hierarchy and prunes it autonomously using mass–distance peaks, returning a parameter-free partition without using a target K. RSDP differs from both in three specific ways: (i) it applies the full DPC inference template (ρ, δ, top-K selection, one-pass assignment) on hierarchical structural units rather than on samples; (ii) it is a known-K method whose resolution parameter λ controls which level the inference is performed on, decoupling resolution from hierarchy construction; and (iii) its theoretical guarantee is a conditional inference statement about the structural DPC layer (§4), not an end-to-end consistency claim for the hierarchy.

C Experimental Details

Table 4 lists the main hyperparameters of RSDP and the ten DPC baselines together with their search ranges, which follow the settings suggested in the original publications and public implementations.

Table 4: Main hyperparameters and their search ranges (n: number of samples).

Method	Hyperparameter	Search range
DPC [2014]	cutoff percentage p for d_c	$p \in \{0.010, 0.011, \ldots, 0.020\}$
DPCDVND [2024]	number of neighbors k	$k \in \{2, 3, \ldots, 50\}$
KSFDPC [2023]	number of neighbors K	$K \in \{5, 10, \ldots, 200\}$
DPCFSC [2022]	density shape parameter a (with $0.1n$ neighbors)	$a \in \{0.01, 0.02, \ldots, 0.99\}$
SMMP [2023b]	none	–
DBDPC [2024]	neighborhood ratio ϵ; neighbor rank κ for Eps	$\epsilon \in \{0.010, 0.011, \ldots, 0.040\}$; $\kappa \in \{5, 6, \ldots, 20\}$
VDPC [2023]	cutoff percentage pct; representative threshold d_t	$\mathrm{pct} \in \{0.01, 0.1, 0.5, 1, 2, 5, 10, 50, 100\}$ %; d_t over the range of δ
DEMOS [2023a]	number of neighbors k	$k = \lceil \alpha\sqrt{n} \rceil$, $\alpha \in \{0.25, 0.5, \ldots, 2\}$
DPCMFP [2025]	number of neighbors k; feature-point parameter l	$k \in \{1, 2, \ldots, 50\}$; $l \in \{1, 2, \ldots, 30\}$
FDPAN [2025b]	tree depth l; number of neighbors k_v	l from $\lfloor \log_2(3n/100) \rfloor$ upward; $k_v \in \{1, 2, \ldots, k\}$ with k the approximate-kNN size
RSDP (ours)	resolution λ	$\lambda \in \{0.1, 0.2, \ldots, 1\}$

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The contributions (C1)–(C3) listed in Section 1 are matched one-to-one by Section 3 (algorithm), Section 4 (theory), and Section 5 (empirical results).

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The conditional scope of the theoretical guarantee, the sensitivity to λ, and the complexity caveats are explicitly discussed in Sections 4, 5.4, and 3.5, respectively; Section 6 also notes the required inputs K and λ and directions for future work.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All assumptions (A1)–(A3) and (D) are stated in Section 4; full lemmas, proposition, and proofs of Theorems 1 and 2 are provided in Appendix A.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Algorithm 1 fully specifies RSDP with a single hyperparameter λ; all datasets are public benchmarks cited in Table 1, and baseline configurations follow each method's original publication.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: All datasets are publicly available; Algorithm 1 fully specifies the method; and the experimental settings are provided in Appendix C.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The single hyperparameter λ, its recommended range, evaluation metrics (ACC, NMI), and Hungarian label alignment are reported in Sections 5 and 5.4; the tuning ranges of RSDP and the DPC baselines are given in Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Figure 3(b) reports mean ACC over 20 random seeds (dataset-generation seeds for the synthetic Gaussian mixtures) with shaded bands denoting ± 1 standard deviation. RSDP and all the cited DPC variants are deterministic algorithms given fixed hyperparameters, so Table 2 reports single-run scores.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Section 5.3 reports per-method wall-clock runtimes on COIL100 nested subsets ($n = 720, \ldots, 7200$), with RSDP at 18.2s versus baselines at 230–1036s.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This work develops an unsupervised clustering algorithm evaluated on standard public benchmarks; it involves no human subjects, scraped data, or high-risk deployment context.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This is foundational research on a clustering algorithm with no direct path to specific applications or deployments carrying societal impact.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper releases neither pretrained models nor scraped datasets and poses no identifiable misuse risk.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All datasets and baseline methods are cited to their original publications in Tables 1 and 2.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new datasets, models, or pretrained checkpoints.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve human subjects research and therefore does not require IRB approval.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not a component of the proposed method or its evaluation; any LLM use was limited to editing and proofreading assistance.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.