MULTIPLICITY IS NOT DIVERSITY: REPLICATION-INVARIANT ENSEMBLE CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Ensemble clustering combines multiple base partitions into a consensus, yet most formulations operate on the ensemble as a multiset. Presenting the same decision repeatedly can change the result even though no new decision information is introduced. We study this multiplicity failure in the decision-only setting and formulate *replication invariance*. It requires the optimized consensus to remain unaffected by arbitrary copies of an observed partition. The challenge is to achieve replication invariance while retaining contributions from multiple distinct partitions. We propose LINE, a rep**L**ication-**I**nvariant e**N**semble clust**E**ring method that addresses this challenge by coupling consensus fidelity with a geometry of decision redundancy. Each partition is represented by a normalized centered co-assignment kernel, while its fidelity to the consensus is measured by normalized conditional code length. We show that the resulting objective admits a quotient representation over decision-equivalence classes, establishing exact replication invariance. We further derive a redundancy-adjusted effective ensemble size and establish stability under near replication, with a perturbation bound independent of replication multiplicity. In addition, LINE has linear time complexity with respect to the number of samples, supporting efficient large-scale ensemble clustering. Experiments on twelve datasets demonstrate strong clustering performance and substantially improved robustness to both exact and near replication. The source code is provided in the supplementary material.

1 INTRODUCTION

Ensemble clustering seeks a consensus partition from multiple base clustering results (Li et al., 2026; Wei et al., 2025; Zhou et al., 2025). It is particularly useful when the original features are heterogeneous, private, or no longer available, because the consensus can be learned directly from decision information (Zheng et al., 2025; Zhang et al., 2025b; Chen et al., 2025). A common intuition is that a larger ensemble provides stronger evidence: agreement across many partitions should make the recovered structure more reliable. This intuition is meaningful only when ensemble cardinality reflects the amount of distinct decision evidence.

In practice, however, a larger ensemble does not necessarily contain more distinct decision evidence (Xie et al., 2024). Identical or highly redundant decisions may occur through repeated generation or a replication attack, in which an adversary copies an existing partition to amplify its influence. Such an attack increases ensemble cardinality without adding decision evidence. Nevertheless, additive consensus losses, co-association averages, and voting rules count every occurrence separately, allowing the consensus to change solely with the multiplicity of an existing decision.

This distinction motivates *replication invariance* (RI). Two ensembles that contain the same distinct decisions but differ only in how many times some of them are repeated should represent the same decision evidence. We therefore require ensembles with the same distinct partitions to induce the same set of optimal consensus solutions, regardless of partition multiplicities. However, replication invariance alone does not guarantee meaningful ensemble aggregation. A trivial way to satisfy RI is to rely on only one base partition, which avoids multiplicity effects but also discards the complementary information carried by other decisions (Xu et al., 2024; Ji et al., 2025). Our goal is therefore more demanding: genuinely distinct decisions should be allowed to contribute jointly, while repeated versions of the same decision should not gain additional influence. We further require

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

the consensus to remain stable when such redundant decisions are only approximately, rather than exactly, equivalent.

To meet these requirements, we develop LINE (rep**L**ication-**I**nvariant e**N**semble clust**E**ring). For each base partition, we measure its fidelity to the latent consensus by a normalized conditional coding distortion, making partitions with different granularities comparable. We then characterize redundancy among partitions through a label-permutation-invariant geometry of normalized centered co-assignment kernels. The source weights are optimized jointly with the consensus under this geometry, allowing genuinely distinct decisions to contribute together without giving extra influence to exact replicas simply because they are repeated. The resulting dependency-aware rate–distortion formulation admits a quotient representation over decision-equivalence classes. Consequently, replicating an existing decision changes only how its weight is distributed within the corresponding equivalence class, while leaving the reduced objective and its globally optimal consensus set unchanged. This establishes exact RI. The same geometry further induces a redundancy-adjusted effective ensemble size and yields a near-replication stability bound that is independent of the number of replicas. In addition, LINE scales linearly with the number of samples for fixed ensemble and consensus sizes, making it suitable for large-scale ensemble clustering. The main contributions of this paper are:

- We expose a previously overlooked multiplicity failure in decision-only ensemble clustering: repeating a clustering decision can alter the consensus without introducing new decision information. We formalize invariance to such repetitions as *replication invariance*.
- We propose LINE, which separates consensus fidelity from decision redundancy. Normalized conditional coding makes heterogeneous partitions comparable, while a permutation-invariant co-assignment geometry allows distinct decisions to contribute jointly without rewarding repeated ones.
- We establish theoretical properties of replication-invariant aggregation. We derive a quotient formulation over decision-equivalence classes and prove exact replication leaves the globally optimal consensus set unchanged. We characterize a redundancy-adjusted effective ensemble size and establish a near-replication stability bound independent of replication multiplicity.
- We develop an efficient optimization scheme whose computation is linear in the number of samples for fixed ensemble and consensus sizes. Experiments on twelve datasets demonstrate competitive clustering performance and robustness to exact and near replication.

2 Related Work

Ensemble clustering. Ensemble clustering constructs a consensus through label alignment, co-association, graph partitioning, or spectral methods (Zhou et al., 2024b; Treder-Tschechlov et al., 2024; Hao et al., 2024). Large-scale approaches improve efficiency through sparse bipartite, anchor, and hypergraph representations (Zhang et al., 2025a; Xu et al., 2026a; Chen et al., 2023; Xu et al., 2026b; 2025), while weighting and ensemble selection methods exploit partition reliability and diversity (Bian et al., 2024; Zhang, 2022). Recent theory further relates finite-ensemble performance to bias and diversity (Zhang et al., 2025d). These approaches improve consensus quality and scalability, but generally treat the observed multiplicities of base partitions as part of the given ensemble (Qin et al., 2025; Zhang et al., 2025c; Zhou et al., 2024a). Although ensemble selection (Ohl & Lindsten, 2025; Golalipour et al., 2024) can reduce redundancy by retaining a subset of base partitions, it does not require the consensus to remain unchanged when a retained decision is replicated. Replication invariance addresses this different question by requiring arbitrary copies of an observed decision to leave the globally optimal consensus unchanged.

Information-theoretic clustering. Information-theoretic clustering uses entropy, mutual information, and information-bottleneck objectives to learn compact representations that preserve cluster-relevant structure (Hu et al., 2024; 2026; Lou et al., 2026; Guo et al., 2026; Hu et al., 2025). Recent multi-view studies use variational mutual-information bounds to characterize shared, complementary, and redundant information across views (Huang et al., 2023b; Cui et al., 2023; Gao & Pu, 2025; Geng et al., 2025; Dufumier et al., 2025). Later work improves the tractability and reliability of these objectives through deterministic estimation and information-guided modality weighting (Yan et al., 2024; Lou et al., 2025a), while newer formulations explore contrastive bottlenecks, calibration, sufficiency, and balance (Lou et al., 2025b; Hu et al., 2026; Ren et al., 2026; Guo et al., 2026).

These methods primarily operate on feature views and study what information should be retained or combined during representation learning. Our setting instead observes only clustering decisions and asks whether their arbitrary multiplicities can change the resulting consensus.

3 Replication-Invariant Consensus Learning

3.1 Problem Formulation

Let $\Pi = \{\pi^{(1)}, \ldots, \pi^{(M)}\}$ be an observed ensemble of M partitions over n samples. Partition v contains c_v clusters and assigns sample i a label $y_i^{(v)} \in [c_v]$. The decision-only input is $\mathbf{Y} = [\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(M)}]$, where $\boldsymbol{y}^{(v)} \in [c_v]^n$. LINE aims to infer a consensus partition with K clusters from $\mathbf{Y}$ that is invariant to the multiplicity of equivalent decisions. Two partitions $\pi^{(u)}$ and $\pi^{(v)}$ are *decision-equivalent*, denoted by $\pi^{(u)} \simeq \pi^{(v)}$, if they induce the same co-assignment relation:

$$\mathbb{I}[y_i^{(u)} = y_j^{(u)}] = \mathbb{I}[y_i^{(v)} = y_j^{(v)}], \qquad \forall i, j. \tag{1}$$

Thus partitions that differ only by a permutation of their cluster labels represent the same decision. Let Π_j^{+L} be obtained by replacing $\pi^{(j)}$ with L decision-equivalent copies. The two ensembles differ only in the multiplicity of partitions equivalent to $\pi^{(j)}$.

Definition 1 (Replication invariance)**.** *An ensemble clustering objective is replication invariant if, for every Π, every $\pi^{(j)} \in \Pi$, and every $L \geq 1$, the optimal consensus set is identical for Π and Π_j^{+L}.*

Definition 1 concerns the optimized consensus objective rather than a particular split of source weights. It first requires a representation in which decision-equivalent partitions induce the same signal for every candidate consensus. A useful ensemble rule must satisfy an additional requirement not contained in RI itself: it should combine support from multiple distinct decisions rather than satisfy the axiom by selecting a single source. We therefore proceed in three steps. We first construct a decision-invariant and granularity-comparable fidelity; we then show why exact RI and non-degenerate source weighting are in tension under index-wise regularization. Finally, we regularize the weighted mean of partition embeddings, so that exact replicas contribute only through their aggregate weight.

3.2 Decision-Invariant Consensus Fidelity

Replication invariance first requires equivalent decisions to exert the same fidelity signal on any candidate consensus. A direct comparison of cluster labels is unsuitable because the indices are arbitrary, while unnormalized fitting losses are not comparable when partitions have different cluster numbers or label uncertainties. We therefore characterize consensus fidelity through conditional uncertainty: how much of each partition's uncertainty remains unexplained by the shared consensus?

For sample i, let $Z_i \in [K]$ be a latent consensus state with posterior $\boldsymbol{q}_i \in \Delta_K$. The posterior assignment of sample i is $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{iK})$, and all posterior assignments form $\mathbf{Q} = [q_{ic}] \in \mathbb{R}^{n \times K}$. Because the consensus states and the clusters in each observed partition have unrelated indices and may have different cardinalities, imposing a one-to-one correspondence would be inappropriate. Instead, partition v is interpreted as an observation of Z through its own conditional channel:

$$\theta_{r|c}^{(v)} = p(Y^{(v)} = r \mid Z = c), \qquad \boldsymbol{\theta}_{\cdot|c}^{(v)} \in \Delta_{c_v}. \tag{2}$$

This construction accommodates $c_v \neq K$ without aligning numeric cluster labels across partitions. Its expected conditional code length is:

$$\mathcal{C}_v(\mathbf{Q}, \boldsymbol{\Theta}^{(v)}) = -\frac{1}{n} \sum_{i=1}^{n} \sum_{c=1}^{K} q_{ic} \log \theta_{y_i^{(v)}|c}^{(v)}. \tag{3}$$

The raw code length is still not comparable across partitions: a partition with larger marginal label entropy has a larger coding scale even when the same fraction of its structure is explained. To measure explanatory quality rather than absolute label uncertainty, we normalize by the marginal entropy:

$$D_v(\mathbf{Q}, \boldsymbol{\Theta}^{(v)}) = \frac{\mathcal{C}_v(\mathbf{Q}, \boldsymbol{\Theta}^{(v)})}{H(Y^{(v)})}, \qquad H(Y^{(v)}) = -\sum_{r=1}^{c_v} p_r^{(v)} \log p_r^{(v)}, \tag{4}$$

where $p_r^{(v)} = n^{-1} \sum_i \mathbb{I}[y_i^{(v)} = r]$ is the empirical cluster proportion. Partitions with $H(Y^{(v)}) = 0$ contain no clustering decision and are removed. The fidelity of partition v is then the shortest normalized description attainable through its channel:

$$d_v(\mathbf{Q}) := \min_{\boldsymbol{\Theta}^{(v)}} D_v(\mathbf{Q}, \boldsymbol{\Theta}^{(v)}) = \frac{H_{\mathbf{Q}}(Y^{(v)} \mid Z)}{H(Y^{(v)})} \in [0, 1]. \tag{5}$$

The conditional entropy is evaluated under the soft joint distribution induced by $\mathbf{Q}$ and the observed labels. Thus $d_v(\mathbf{Q})$ is the fraction of partition uncertainty left unexplained by the consensus: zero means that Z determines partition v, whereas one means that Z explains none of its uncertainty. Crucially, conditional and marginal entropies are unchanged by a permutation of the observed labels. Hence decision-equivalent partitions induce the same function $d_v(\mathbf{Q})$, satisfying the first requirement imposed by replication invariance.

Fidelity, however, only measures how well the observed decisions are predicted from Z; it does not express how much object-specific assignment information the consensus retains for that fit. Two consensus posteriors can therefore achieve comparable partition fidelity using different amounts of information. To prefer a compact shared explanation, let I be a uniformly sampled object index, define $p_{\mathbf{Q}}(I = i, Z = c) = q_{ic}/n$, and control the retained assignment information by the rate:

$$\mathcal{R}(\mathbf{Q}) = I_{\mathbf{Q}}(I; Z). \tag{6}$$

With all partition distortions on a common scale, the most direct criterion distributes a unit weight vector $\boldsymbol{w} \in \Delta_M$ over them:

$$\mathcal{F}_{\mathrm{RD}}(\mathbf{Q}, \boldsymbol{w}) = \tau I_{\mathbf{Q}}(I; Z) + \boldsymbol{d}_\Pi(\mathbf{Q})^\top \boldsymbol{w}, \tag{7}$$

where $\boldsymbol{d}_\Pi(\mathbf{Q}) = (d_1(\mathbf{Q}), \ldots, d_M(\mathbf{Q}))^\top$. This formulation already reveals why exact RI alone is insufficient for useful ensemble aggregation.

Proposition 1 (Why exact RI alone is insufficient)**.** *Fix any feasible* $\mathbf{Q}$*.*

(i) The fidelity-only weight problem satisfies

$$\min_{\boldsymbol{w} \in \Delta_M} \boldsymbol{d}_\Pi(\mathbf{Q})^\top \boldsymbol{w} = \min_{v \in [M]} d_v(\mathbf{Q}). \tag{8}$$

Hence exact replication leaves this value unchanged for every $\mathbf{Q}$*, so the fidelity-only reduced objective is exactly RI. If the minimum distortion is unique, however, the unique optimizer is the corresponding simplex vertex, so the rule collapses to a single source.*

(ii) Consider instead the isotropically regularized problem $\boldsymbol{d}_\Pi(\mathbf{Q})^\top \boldsymbol{w} + \lambda \|\boldsymbol{w}\|_2^2$ *with* $\lambda > 0$*. If one decision carrying aggregate mass* α *is represented by* L *equivalent copies, the minimum within-group quadratic contribution is* $\lambda\alpha^2/L$*, rather than* $\lambda\alpha^2$ *for one presentation. Thus the regularized objective is multiplicity-sensitive in general.*

Proposition 1 shows that exact RI alone is insufficient for useful aggregation. Fidelity-only simplex weighting satisfies exact RI through minimum-distortion source selection, whereas generic anti-concentration regularization can reward replication by operating on source indices. We therefore need a regularizer that spreads mass across genuinely distinct decisions, with exact replicas entering only through their aggregate mass.

3.3 Decision Redundancy Geometry

To determine whether a well-explained partition contributes new or repeated support, redundancy must be measured on clustering decisions rather than on their numeric labels. A co-assignment relation provides such a representation because it is unchanged by label permutations and remains well defined when cluster numbers differ. Let $\mathbf{H}^{(v)} \in \{0, 1\}^{n \times c_v}$ be the one-hot assignment matrix of partition v, and define:

$$\mathbf{K}^{(v)} = \mathbf{H}^{(v)} \mathbf{H}^{(v)\top}, \qquad \mathbf{P} = \mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top, \qquad \widetilde{\mathbf{K}}^{(v)} = \mathbf{P}\mathbf{K}^{(v)}\mathbf{P}. \tag{9}$$

The raw co-assignment kernel contains partition-dependent baseline and magnitude effects induced by cluster sizes and label granularity. Centering removes the constant component, while unit

normalization prevents these scale effects from being mistaken for structural agreement. For every non-degenerate partition, define

$$\boldsymbol{\phi}_v = \frac{\mathrm{vec}(\widetilde{\mathbf{K}}^{(v)})}{\|\widetilde{\mathbf{K}}^{(v)}\|_F}, \qquad R_{uv} = \langle \boldsymbol{\phi}_u, \boldsymbol{\phi}_v \rangle = \frac{\langle \widetilde{\mathbf{K}}^{(u)}, \widetilde{\mathbf{K}}^{(v)} \rangle_F}{\|\widetilde{\mathbf{K}}^{(u)}\|_F \|\widetilde{\mathbf{K}}^{(v)}\|_F}. \tag{10}$$

The entry R_{uv} therefore measures alignment between two decision structures without requiring equal cluster numbers or aligned labels. Exact replicas occupy the same point in this geometry, while near replicas remain close.

Proposition 2 (Redundancy geometry)**.** *For all non-degenerate active partitions,*

$$\mathbf{R} \succeq 0, \qquad 0 \le R_{uv} \le 1, \qquad R_{vv} = 1. \tag{11}$$

Moreover, $\pi^{(u)} \simeq \pi^{(v)}$ *implies* $\boldsymbol{\phi}_u = \boldsymbol{\phi}_v$ *and* $R_{uv} = 1$.

Proposition 2 supplies the structure needed to resolve Proposition 1. Equivalent decisions occupy the same point, while non-equivalent decisions can occupy different directions; positive semidefiniteness further allows their concentration to be controlled by a convex quadratic form. The key is to regularize the *mean decision direction* rather than the Euclidean norm of the coordinate vector $\boldsymbol{w}$. Exact replicas can then share mass without reducing the penalty simply because more coordinates are available.

3.4 Replication-Invariant Decision Aggregation

We interpret $\boldsymbol{w} \in \Delta_M$ as one fixed unit of *decision mass*. Its image in the redundancy geometry is the weighted decision mean

$$\boldsymbol{\mu}_{\boldsymbol{w}} = \sum_{v=1}^{M} w_v \boldsymbol{\phi}_v, \qquad \Omega(\boldsymbol{w}; \mathbf{R}) = \|\boldsymbol{\mu}_{\boldsymbol{w}}\|_2^2 = \boldsymbol{w}^\top \mathbf{R} \boldsymbol{w}. \tag{12}$$

The penalty Ω discourages concentration on mutually aligned decision directions while allowing mass to spread across less redundant support. Unlike $\|\boldsymbol{w}\|_2^2$, it is insensitive to how a fixed mass is split among exact replicas. Indeed, for any equivalence class $\mathcal{C}_g$ with aggregate mass $\alpha_g = \sum_{v \in \mathcal{C}_g} w_v$,

$$\sum_{u,v \in \mathcal{C}_g} w_u w_v R_{uv} = \left(\sum_{v \in \mathcal{C}_g} w_v \right)^2 = \alpha_g^2, \tag{13}$$

because $R_{uv} = 1$ within the class. The fidelity term has the same class-mass property because all equivalent members share the same $d_v(\mathbf{Q})$. This geometry-aware regularization therefore discourages single-source collapse without rewarding additional presentations of an existing decision. The resulting LINE objective is:

$$\begin{aligned} \min_{\mathbf{Q}, \boldsymbol{w}} \quad & \overline{\mathcal{F}}_\Pi(\mathbf{Q}, \boldsymbol{w}) = \tau I_{\mathbf{Q}}(I; Z) + \boldsymbol{d}_\Pi(\mathbf{Q})^\top \boldsymbol{w} + \lambda \boldsymbol{w}^\top \mathbf{R} \boldsymbol{w} \\ & \text{s.t.} \quad \boldsymbol{q}_i \in \Delta_K, \qquad \boldsymbol{w} \in \Delta_M. \end{aligned} \tag{14}$$

The rate term controls the amount of object-specific information retained by the consensus, the distortion term measures how well the consensus explains each observed decision, and the quadratic term regulates how the single unit of decision mass is distributed over decision space. The same $\boldsymbol{w}$ governs fidelity and redundancy, so a source cannot receive one weight for fit and another for dependence. To determine whether replication can alter the optimized consensus, we minimize over the partition weights:

$$\mathcal{J}_\Pi(\mathbf{Q}) = \min_{\boldsymbol{w} \in \Delta_M} \overline{\mathcal{F}}_\Pi(\mathbf{Q}, \boldsymbol{w}). \tag{15}$$

Let $\mathcal{C}_1, \ldots, \mathcal{C}_G$ be the equivalence classes of Π under $\simeq$. All members of $\mathcal{C}_g$ share the same distortion function $d_g(\mathbf{Q})$ and embedding $\boldsymbol{\phi}_g$. Their aggregate mass α_g and class-level Gram matrix S_{gh} are:

$$\alpha_g = \sum_{v \in \mathcal{C}_g} w_v, \qquad \boldsymbol{\alpha} \in \Delta_G, \qquad S_{gh} = \langle \boldsymbol{\phi}_g, \boldsymbol{\phi}_h \rangle. \tag{16}$$

The fidelity and redundancy terms therefore depend on each equivalence class only through its aggregate mass α_g. Changing its multiplicity can alter how this mass is distributed among equivalent members, but not its class-level contribution.

Theorem 1 (Quotient representation). *For every feasible consensus posterior* $\mathbf{Q}$*,*

$$\mathcal{J}_{\Pi}(\mathbf{Q}) = \tau I_{\mathbf{Q}}(I;Z) + \min_{\boldsymbol{\alpha}\in\Delta_G} \left\{ \sum_{g=1}^{G} \alpha_g d_g(\mathbf{Q}) + \lambda \boldsymbol{\alpha}^{\top} \mathbf{S} \boldsymbol{\alpha} \right\}. \tag{17}$$

Hence $\mathcal{J}_{\Pi}$ *depends on the distinct decision classes and their geometry, but not on their observed multiplicities.*

Every feasible class mass can be distributed arbitrarily among the members of its class without changing either source-dependent term. Replication therefore changes only the within-class parameterization of $\boldsymbol{w}$, which yields the desired invariance.

Corollary 1 (Exact replication invariance). *For any* Π*,* $\pi^{(j)} \in \Pi$*, and* $L \geq 1$*,*

$$\mathcal{J}_{\Pi_j^{+L}}(\mathbf{Q}) = \mathcal{J}_{\Pi}(\mathbf{Q}) \qquad \textit{for every } \mathbf{Q}. \tag{18}$$

Consequently, the two ensembles have identical globally optimal consensus sets and objective values.

Replication invariance is an observational property of the supplied decisions. In the decision-only setting, exact replicas introduce no new decision structure beyond the partition already observed and therefore provide no additional decision evidence. Accordingly, LINE prevents repeated presentations of the same decision from gaining additional influence through multiplicity alone. Corollary 1 is also a statement about the reduced objective and its global minimizers.

4 Experiments

4.1 Experimental Setup

Datasets and Ensemble Construction. We evaluate LINE on twelve benchmark datasets spanning different data domains and scales. Their statistics are summarized in Table 1, and detailed descriptions are provided in the Appendix. Following the widely adopted setting in ensemble clustering (Huang et al., 2021; Xu et al., 2024), we independently run k-means 100 times on each dataset to construct a pool of candidate base partitions. An ensemble contains $M = 20$ base partitions sampled from this pool. We independently construct ten such ensembles and report the mean and standard deviation over the ten trials. Within each trial, all competing methods receive exactly the same set of observed partitions, thereby removing differences caused by the random generation of the ensemble. **Comparison Methods.** We compare LINE with eleven representative ensemble clustering methods: SEC (Liu et al., 2015), PTGP (Huang et al., 2016), LWGP (Huang et al., 2018), U-SENC (Huang et al., 2020), ECPCS-MC (Huang et al., 2021), FastMICE (Huang et al., 2023a), FSEC (Zhang et al., 2025a), CEHM (Li et al., 2025), YACHT (Xu et al., 2025), RANGE (Xu et al., 2026a), and DREAM (Xu et al., 2026b). For all competing methods, the hyperparameters are set according to the recommendations in the original papers or released implementations. **Evaluation Metrics.** Clustering performance is evaluated using three widely adopted metrics: clustering accuracy (ACC), normalized mutual information (NMI), and adjusted Rand index (ARI). Higher values indicate better clustering performance. The reported results are presented as the mean and standard deviation over ten independent trials. **Implementation Details.** For LINE, τ is varied over $\{0.01, 0.02, 0.04, 0.08, 0.1, 0.2\}$ and λ over $\{0.1, 0.5, 1, 2, 5, 10\}$. Their sensitivity is further investigated in Section 4.4. The optimization is terminated when the relative change of the objective value falls below a predefined tolerance or the maximum number of iterations is reached. All methods are evaluated under the same computing environment.

Table 1: Characteristics of all datasets.

ID	Dataset	#Samples	#Features	#Classes
D1	Optdigits	5620	64	10
D2	GesturePhase	9873	32	5
D3	ALOI100	10000	128	100
D4	Pendigits	10992	16	10
D5	Nursery	12960	8	5
D6	DryBean	13611	16	7
D7	JungleChess2pcs	44819	6	3
D8	SensIT	98528	50	3
D9	EMNIST	280000	784	10
D10	MHEALTH	343195	23	12
D11	NF-ToN-IoT	1157994	10	10
D12	KDDCup99	4898431	41	23

4.2 Overall Clustering Performance

Table 2 reports the clustering performance on the original ensembles. Overall, LINE achieves the best mean result in 29 of the 36 dataset–metric comparisons, indicating that its replication-invariant

Table 2: Clustering performance on D1–D12, measured by ACC, ARI, and NMI (%). The best results are highlighted in bold.

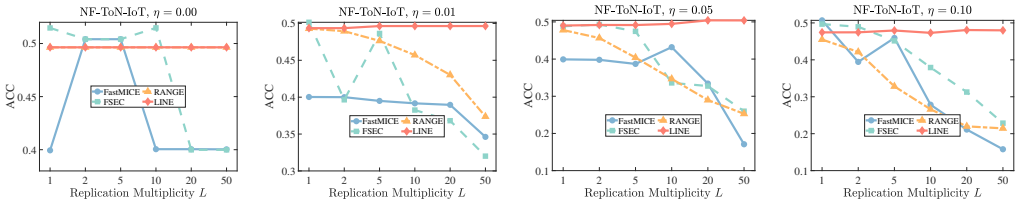
Method	Optdigits			GesturePhase			ALOI100		
	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
k-means (base clusterings)	75.55 ± 6.15	63.47 ± 5.92	73.36 ± 2.67	35.51 ± 1.19	3.19 ± 0.62	9.82 ± 0.38	64.80 ± 1.76	56.66 ± 1.74	83.76 ± 0.52
SEC [*KDD*, 2015]	60.20 ± 6.46	46.08 ± 8.65	65.90 ± 3.27	33.70 ± 3.02	1.55 ± 1.22	5.83 ± 3.68	54.49 ± 2.82	25.61 ± 7.68	78.14 ± 1.87
PTGP [*TKDE*, 2016]	76.61 ± 3.43	64.46 ± 3.03	73.23 ± 1.40	34.88 ± 0.07	2.92 ± 0.06	9.66 ± 0.09	57.41 ± 1.29	52.23 ± 0.96	82.10 ± 0.33
LWGP [*TCYB*, 2018]	78.94 ± 3.00	66.81 ± 2.85	74.08 ± 1.33	34.94 ± 0.05	3.05 ± 0.07	9.78 ± 0.10	68.65 ± 0.98	59.73 ± 0.70	84.84 ± 0.21
USENC [*TKDE*, 2020]	78.18 ± 3.03	66.01 ± 2.50	74.07 ± 0.83	35.04 ± 0.30	3.20 ± 0.27	9.88 ± 0.14	69.95 ± 1.92	61.09 ± 1.68	84.98 ± 0.42
ECPCS-MC [*TSMC*, 2021]	79.09 ± 2.85	66.65 ± 3.23	74.37 ± 1.16	34.89 ± 0.04	2.92 ± 0.06	9.65 ± 0.10	67.53 ± 1.02	59.17 ± 1.09	84.64 ± 0.31
FastMICE [*TKDE*, 2023]	78.32 ± 3.17	65.98 ± 3.20	74.03 ± 1.02	35.04 ± 0.30	3.20 ± 0.27	9.88 ± 0.14	64.54 ± 1.49	56.85 ± 0.77	83.72 ± 0.21
FSEC [*Inf. Fusion*, 2025]	79.31 ± 1.58	66.44 ± 2.40	74.30 ± 0.54	35.04 ± 0.30	3.20 ± 0.27	9.88 ± 0.14	70.80 ± 1.11	61.85 ± 0.84	85.18 ± 0.19
CEHM [*AAAI*, 2025]	74.96 ± 2.73	61.95 ± 3.05	72.74 ± 1.03	34.85 ± 0.04	2.87 ± 0.08	9.62 ± 0.10	68.02 ± 0.73	59.08 ± 0.65	84.77 ± 0.25
YACHT [*IJCAI*, 2025]	79.50 ± 3.13	66.73 ± 3.80	74.26 ± 1.44	36.59 ± 0.70	4.57 ± 0.97	10.07 ± 0.59	71.94 ± 1.18	62.55 ± 1.17	85.06 ± 0.33
RANGE [*CVPR*, 2026]	79.01 ± 2.40	66.24 ± 2.13	73.91 ± 1.27	36.53 ± 0.82	$\mathbf{5.41 \pm 0.97}$	10.47 ± 0.63	73.21 ± 0.83	63.35 ± 1.28	85.36 ± 0.29
DREAM [*TKDE*, 2026]	79.62 ± 2.05	66.94 ± 2.71	74.24 ± 0.98	36.03 ± 0.55	3.24 ± 0.55	10.10 ± 0.23	73.78 ± 1.19	63.63 ± 1.21	85.38 ± 0.36
LINE (Ours)	$\mathbf{81.38 \pm 3.37}$	$\mathbf{68.79 \pm 3.06}$	$\mathbf{75.36 \pm 1.69}$	$\mathbf{37.29 \pm 0.88}$	4.95 ± 1.06	$\mathbf{10.83 \pm 0.71}$	$\mathbf{74.14 \pm 0.85}$	$\mathbf{63.99 \pm 0.82}$	$\mathbf{85.57 \pm 0.24}$

Method	Pendigits			Nursery			DryBean		
	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
k-means (base clusterings)	68.85 ± 5.14	54.84 ± 3.82	67.92 ± 1.38	10.03 ± 7.20	3.96 ± 3.18	19.31 ± 5.46	56.32 ± 1.70	38.66 ± 1.90	51.62 ± 0.42
SEC [*KDD*, 2015]	57.05 ± 7.36	36.40 ± 11.10	58.39 ± 5.38	32.00 ± 4.54	3.04 ± 5.30	5.77 ± 7.20	51.35 ± 3.87	31.64 ± 6.49	43.98 ± 5.95
PTGP [*TKDE*, 2016]	72.19 ± 3.99	56.48 ± 2.52	68.44 ± 0.94	33.93 ± 5.54	4.51 ± 5.01	7.55 ± 7.37	55.64 ± 0.48	36.32 ± 1.55	51.09 ± 0.32
LWGP [*TCYB*, 2018]	70.22 ± 4.76	54.96 ± 3.37	67.99 ± 1.15	34.92 ± 5.11	5.76 ± 5.73	10.69 ± 8.39	55.12 ± 0.03	37.70 ± 0.01	51.29 ± 0.02
USENC [*TKDE*, 2020]	69.01 ± 3.30	54.26 ± 1.83	68.02 ± 0.26	30.06 ± 4.82	3.59 ± 4.55	5.66 ± 5.70	55.77 ± 1.13	37.77 ± 1.18	51.22 ± 0.39
ECPCS-MC [*TSMC*, 2021]	68.38 ± 3.88	54.03 ± 2.54	67.97 ± 0.61	35.19 ± 8.00	$\mathbf{12.02 \pm 10.39}$	16.12 ± 12.65	55.12 ± 0.03	37.70 ± 0.01	51.30 ± 0.00
FastMICE [*TKDE*, 2023]	70.18 ± 4.25	54.94 ± 3.09	68.00 ± 1.00	28.95 ± 4.52	3.47 ± 4.43	5.16 ± 5.57	55.12 ± 0.03	37.70 ± 0.01	51.29 ± 0.02
FSEC [*Inf. Fusion*, 2025]	69.42 ± 4.07	54.67 ± 2.56	68.14 ± 0.48	29.00 ± 4.54	3.62 ± 4.45	5.59 ± 5.60	56.82 ± 1.81	38.35 ± 0.75	50.98 ± 0.38
CEHM [*AAAI*, 2025]	67.33 ± 2.52	52.82 ± 2.35	67.17 ± 1.32	32.66 ± 4.86	3.12 ± 5.27	6.63 ± 7.94	55.12 ± 0.03	37.70 ± 0.01	51.30 ± 0.00
YACHT [*IJCAI*, 2025]	71.30 ± 3.65	56.11 ± 3.26	67.26 ± 1.28	33.43 ± 9.92	7.96 ± 12.25	11.30 ± 13.35	55.42 ± 3.81	37.12 ± 1.49	50.67 ± 0.22
RANGE [*CVPR*, 2026]	72.96 ± 3.42	55.65 ± 2.92	67.94 ± 1.08	30.91 ± 5.59	5.27 ± 5.91	7.72 ± 8.49	55.87 ± 2.42	37.54 ± 0.89	50.95 ± 0.30
DREAM [*TKDE*, 2026]	72.78 ± 4.47	56.58 ± 3.68	67.94 ± 1.68	32.97 ± 5.65	4.94 ± 4.46	8.00 ± 7.77	59.33 ± 2.15	38.91 ± 2.71	50.67 ± 0.62
LINE (Ours)	$\mathbf{74.34 \pm 0.24}$	$\mathbf{57.80 \pm 0.48}$	$\mathbf{68.63 \pm 0.54}$	$\mathbf{37.32 \pm 9.75}$	11.36 ± 11.63	$\mathbf{17.25 \pm 17.41}$	$\mathbf{60.14 \pm 0.32}$	$\mathbf{39.79 \pm 0.26}$	$\mathbf{51.87 \pm 0.13}$

Method	JungleChess2pcs			SensIT			EMNIST		
	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
k-means (base clusterings)	9.64 ± 9.06	1.85 ± 1.82	8.83 ± 1.42	65.70 ± 1.86	33.89 ± 1.76	29.07 ± 1.50	54.27 ± 3.89	37.27 ± 2.94	48.94 ± 1.82
SEC [*KDD*, 2015]	46.85 ± 3.72	2.62 ± 2.99	3.01 ± 1.30	60.36 ± 7.55	23.84 ± 15.64	21.56 ± 13.16	51.13 ± 3.94	34.27 ± 3.42	47.92 ± 1.60
PTGP [*TKDE*, 2016]	41.28 ± 1.67	1.84 ± 0.23	1.25 ± 0.03	65.71 ± 1.29	34.41 ± 1.10	29.32 ± 0.65	53.44 ± 2.95	36.99 ± 2.11	48.99 ± 1.02
LWGP [*TCYB*, 2018]	41.28 ± 1.67	1.84 ± 0.23	1.25 ± 0.03	64.74 ± 1.18	32.52 ± 1.21	30.12 ± 0.66	52.01 ± 3.22	35.27 ± 3.20	47.93 ± 1.69
USENC [*TKDE*, 2020]	40.95 ± 1.56	1.89 ± 0.21	1.26 ± 0.03	64.37 ± 0.00	32.14 ± 0.00	30.31 ± 0.00	53.92 ± 4.53	37.63 ± 3.02	49.49 ± 2.05
ECPCS-MC [*TSMC*, 2021]	40.95 ± 1.56	1.89 ± 0.21	1.26 ± 0.03	65.46 ± 1.43	33.41 ± 1.73	30.14 ± 0.25	55.35 ± 3.56	38.73 ± 2.97	50.11 ± 1.73
FastMICE [*TKDE*, 2023]	41.28 ± 1.67	1.84 ± 0.23	1.25 ± 0.03	64.58 ± 0.66	32.31 ± 0.52	30.30 ± 0.02	55.82 ± 3.01	38.89 ± 2.53	49.88 ± 1.48
FSEC [*Inf. Fusion*, 2025]	41.28 ± 1.67	1.84 ± 0.23	1.25 ± 0.03	64.58 ± 0.66	32.31 ± 0.52	30.30 ± 0.02	55.34 ± 3.50	38.81 ± 3.00	49.88 ± 1.77
CEHM [*AAAI*, 2025]	41.80 ± 1.95	2.11 ± 0.58	1.38 ± 0.28	64.07 ± 1.14	32.20 ± 1.13	30.32 ± 0.31	54.79 ± 1.87	37.93 ± 1.32	49.09 ± 0.77
YACHT [*IJCAI*, 2025]	45.55 ± 5.70	2.92 ± 2.78	4.37 ± 2.71	64.67 ± 1.57	33.76 ± 1.94	29.16 ± 1.18	55.81 ± 3.90	38.72 ± 2.91	50.13 ± 1.36
RANGE [*CVPR*, 2026]	44.17 ± 2.44	2.95 ± 1.31	4.27 ± 2.20	67.30 ± 1.10	34.80 ± 1.79	30.04 ± 0.28	57.64 ± 1.93	$\mathbf{40.46 \pm 2.20}$	50.75 ± 1.43
DREAM [*TKDE*, 2026]	$\mathbf{50.47 \pm 3.76}$	3.95 ± 3.79	5.41 ± 2.33	66.80 ± 2.23	35.68 ± 1.07	$\mathbf{30.81 \pm 0.02}$	56.95 ± 3.05	39.89 ± 2.65	50.61 ± 1.45
LINE (Ours)	49.97 ± 7.35	$\mathbf{7.36 \pm 6.82}$	$\mathbf{6.81 \pm 5.65}$	$\mathbf{67.60 \pm 0.14}$	$\mathbf{36.51 \pm 0.45}$	29.50 ± 0.20	$\mathbf{57.84 \pm 3.61}$	40.22 ± 2.50	$\mathbf{50.92 \pm 1.47}$

Method	MHEALTH			NF-ToN-IoT			KDDCup99		
	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
k-means (base clusterings)	24.45 ± 4.58	18.49 ± 6.32	37.92 ± 5.04	43.35 ± 4.96	3.53 ± 4.71	8.33 ± 10.73	58.99 ± 3.89	4.18 ± 13.26	5.48 ± 11.95
SEC [*KDD*, 2015]	24.90 ± 5.02	11.68 ± 4.72	30.43 ± 5.68	49.46 ± 3.40	8.84 ± 3.11	20.56 ± 7.23	69.93 ± 10.80	42.23 ± 37.40	39.37 ± 33.88
PTGP [*TKDE*, 2016]	21.22 ± 1.16	9.53 ± 1.48	30.45 ± 1.32	$\mathbf{50.26 \pm 0.32}$	10.25 ± 0.64	23.64 ± 1.08	62.64 ± 9.34	17.42 ± 30.90	17.99 ± 26.48
LWGP [*TCYB*, 2018]	23.05 ± 2.18	12.40 ± 2.27	32.32 ± 2.09	39.94 ± 0.03	0.10 ± 0.00	0.55 ± 0.00	57.64 ± 0.45	0.75 ± 1.11	1.64 ± 2.08
USENC [*TKDE*, 2020]	25.09 ± 2.49	14.27 ± 2.63	34.11 ± 2.08	47.28 ± 5.12	7.23 ± 4.96	16.69 ± 11.17	68.17 ± 10.91	36.32 ± 36.50	34.38 ± 30.70
ECPCS-MC [*TSMC*, 2021]	20.86 ± 0.92	9.00 ± 0.84	28.90 ± 0.75	48.31 ± 4.42	7.93 ± 4.13	18.49 ± 9.47	58.23 ± 0.08	1.82 ± 0.26	3.92 ± 0.47
FastMICE [*TKDE*, 2023]	25.75 ± 1.88	15.31 ± 2.13	34.66 ± 1.59	45.46 ± 5.48	5.26 ± 5.30	12.38 ± 11.79	61.98 ± 7.43	14.12 ± 24.49	15.62 ± 20.76
FSEC [*Inf. Fusion*, 2025]	24.42 ± 2.09	13.96 ± 2.45	33.79 ± 2.04	46.23 ± 5.48	6.25 ± 5.32	14.45 ± 11.99	70.36 ± 10.95	43.67 ± 36.58	40.58 ± 30.79
CEHM [*AAAI*, 2025]	16.69 ± 2.47	4.70 ± 1.91	23.49 ± 4.36	39.93 ± 0.06	0.08 ± 0.03	0.45 ± 0.16	57.72 ± 0.40	0.64 ± 0.63	1.28 ± 1.25
YACHT [*IJCAI*, 2025]	33.31 ± 4.89	22.22 ± 4.68	37.32 ± 3.58	48.99 ± 0.61	9.86 ± 1.00	22.31 ± 1.42	–	–	–
RANGE [*CVPR*, 2026]	22.70 ± 3.49	11.47 ± 3.31	31.29 ± 2.30	49.02 ± 0.38	9.41 ± 0.53	22.94 ± 1.03	70.35 ± 10.97	$\mathbf{44.03 \pm 36.37}$	40.74 ± 30.19
DREAM [*TKDE*, 2026]	31.45 ± 5.89	21.03 ± 5.50	37.54 ± 4.31	40.94 ± 3.33	1.07 ± 3.10	2.76 ± 7.10	58.17 ± 0.10	1.46 ± 0.19	2.99 ± 0.49
LINE (Ours)	$\mathbf{34.38 \pm 1.64}$	$\mathbf{23.67 \pm 2.03}$	$\mathbf{38.36 \pm 2.60}$	50.11 ± 0.39	$\mathbf{10.30 \pm 0.53}$	$\mathbf{23.75 \pm 1.01}$	$\mathbf{70.68 \pm 11.05}$	43.97 ± 36.35	$\mathbf{41.17 \pm 30.62}$

formulation does not compromise consensus quality in the standard setting. The advantage can be attributed to two aspects of the model. First, the normalized conditional-entropy fidelity places partitions with different cluster granularities on a comparable scale, avoiding a systematic preference induced by their intrinsic label uncertainty. Second, the redundancy geometry prevents highly overlapping decisions from receiving excessive aggregate support merely because they appear as separate ensemble members. In contrast, conventional graph- or weight-based consensus methods mainly operate on the observed ensemble members and do not explicitly distinguish distinct decision support from redundant evidence, which may bias the consensus toward repeatedly represented structures. This distinction is relevant even without artificial replication, since independently generated base partitions may still encode strongly overlapping decisions. Recent methods such as YACHT, RANGE, and DREAM remain competitive on several datasets, but their advantages are less consistent across the three metrics. For instance, on JungleChess2pcs and NF-ToN-IoT, LINE does not attain the highest ACC but achieves the best ARI and NMI, suggesting a better recovery of the overall partition structure. Overall, these results support the effectiveness of the proposed aggregation principle under ordinary ensembles, while its behavior under controlled replication is examined next.



(a) $\eta = 0.00$ (b) $\eta = 0.01$ (c) $\eta = 0.05$ (d) $\eta = 0.10$

Figure 1: ACC on NF-ToN-IoT across replication multiplicities L under different η values.

Table 3: Ablation study of the main components in LINE on all twelve datasets. Results are measured by ACC, ARI, and NMI (%). The best results are highlighted in bold.

Variant	Components				D1			D2			D3			D4			D5			D6		
	Norm.	Rate	Geom.	Quad.	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
w/o Normalization		✓	✓	✓	80.41	66.79	74.25	36.35	4.92	8.81	73.05	62.95	84.56	74.12	56.62	67.47	36.09	9.57	15.06	58.27	39.10	51.64
w/o Rate	✓		✓	✓	80.40	67.80	75.34	36.25	4.95	9.23	72.12	62.55	84.35	73.78	57.57	66.40	32.29	6.15	9.49	58.14	39.12	51.64
w/o Decision Geometry	✓	✓		✓	80.17	67.68	74.48	34.95	3.09	9.84	**74.25**	63.00	84.60	74.01	56.35	67.92	34.19	7.11	11.26	56.71	38.02	51.68
w/o Redundancy Penalty	✓	✓			80.23	67.85	74.74	34.88	2.93	9.66	66.85	57.55	83.98	73.20	56.10	67.11	32.89	6.19	9.73	56.72	38.02	51.68
LINE	✓	✓	✓	✓	**81.38**	**68.79**	**75.36**	**37.29**	**4.95**	**10.83**	74.14	**63.99**	**85.57**	**74.34**	**57.80**	**68.63**	**37.32**	**11.36**	**17.25**	**60.14**	**39.79**	**51.87**

Variant	Components				D7			D8			D9			D10			D11			D12		
	Norm.	Rate	Geom.	Quad.	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
w/o Normalization		✓	✓	✓	49.58	7.14	6.69	67.60	36.51	29.50	56.82	39.22	49.92	33.39	22.50	37.18	49.11	9.90	22.98	69.80	43.85	40.90
w/o Rate	✓		✓	✓	49.88	7.23	6.70	67.60	36.51	29.50	56.80	39.07	49.70	33.16	22.27	37.11	49.03	9.41	22.94	69.68	42.00	39.15
w/o Decision Geometry	✓	✓		✓	47.40	3.76	4.52	64.98	32.89	**30.20**	55.55	39.03	50.12	33.66	22.80	**41.38**	49.41	9.91	22.99	69.48	41.95	40.26
w/o Redundancy Penalty	✓	✓			47.74	3.75	4.39	64.98	32.89	30.20	55.62	38.64	49.80	26.23	12.86	32.64	49.02	9.91	23.02	69.37	42.84	40.37
LINE	✓	✓	✓	✓	**49.97**	**7.36**	**6.81**	**67.60**	**36.51**	29.50	**57.84**	**40.22**	**50.92**	**34.38**	**23.67**	38.36	**50.11**	**10.30**	**23.75**	**70.68**	**43.97**	**41.17**

4.3 Robustness to Exact and Near Replication

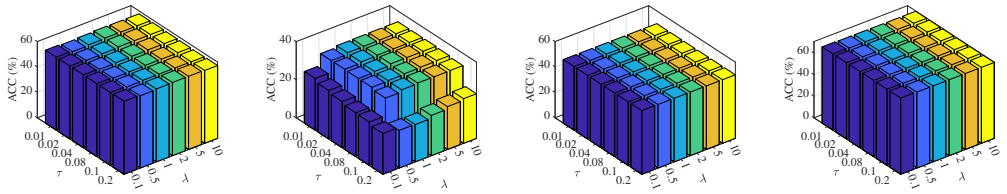
We evaluate robustness to decision multiplicity by comparing LINE with FastMICE, FSEC, and RANGE under exact and near replication. For each dataset, we fix an ensemble of $M = 20$ base partitions and randomly select one partition as the replication target. It is replaced with $L \in \{1, 2, 5, 10, 20, 50\}$ replicas, where $L = 1$ corresponds to the original ensemble without replication. We consider $\eta \in \{0, 0.01, 0.05, 0.10\}$, where $\eta = 0$ denotes exact replication and $\eta > 0$ perturbs the corresponding fraction of object assignments to generate near replicas.

As shown in Fig. 1, the performance of LINE remains largely stable as L increases under both exact and near replication, whereas the competing methods exhibit substantially larger variations. Under exact replication, repeated copies introduce no new clustering information but can still affect conventional aggregation methods. The consistent behavior under nonzero η further shows that LINE remains robust when replicated decisions are only approximately equivalent. This observation is consistent with Theorem 3, which bounds the effect of near replication through coding and embedding perturbations independently of the replication multiplicity L.

4.4 Ablation and Parameter Analysis

Ablation Study. We conduct ablation experiments on all datasets to examine the contribution of the main designs in LINE. Four variants are considered. **w/o Normalization** removes the entropy normalization in Eq. (5) and directly uses $H_{\mathbf{Q}}(Y^{(v)} \mid Z)$ as the partition distortion. **w/o Rate** removes the consensus rate term by setting $\tau = 0$. **w/o Decision Geometry** replaces $\mathbf{R}$ with the identity matrix, reducing the quadratic term to $\lambda\|\boldsymbol{w}\|_2^2$ while preserving generic weight regularization. **w/o Redundancy Penalty** removes the quadratic redundancy term by setting $\lambda = 0$.

As shown in Table 3, the complete model generally yields the most consistent performance across the twelve datasets. Removing entropy normalization or the rate term leads to degradation on most datasets, supporting their complementary roles in making heterogeneous partition fidelities comparable and controlling the complexity of the shared consensus. Removing the decision geometry generally leads to further degradation, indicating the benefit of explicitly modeling dependence among base partitions. Performance also decreases on several datasets when the redundancy penalty is removed, showing the importance of redundancy-aware aggregation when multiple partitions



(a) EMNIST (b) MHEALTH (c) NF-ToN-IoT (d) KDDCup99

Figure 2: ACC w.r.t. τ and λ.

convey overlapping decision information. Although a few individual metrics remain comparable or slightly improve under ablation, the full formulation provides the most stable overall performance. These results support the complementary contributions of normalized fidelity, consensus-rate control, and redundancy-aware aggregation in LINE.

Parameter Analysis. We further examine the sensitivity of LINE to its two trade-off parameters, with τ varied over $\{0.01, 0.02, 0.04, 0.08, 0.1, 0.2\}$ and λ over $\{0.1, 0.5, 1, 2, 5, 10\}$. As shown in Fig. 2 and Fig. 9, the performance remains relatively stable over a broad range of parameter combinations on most datasets, indicating that LINE is not particularly sensitive to parameter selection. More noticeable variations are observed on Nursery, ALOI100, and MHEALTH, suggesting that the balance between consensus complexity and redundancy-aware aggregation can be more dataset dependent in these cases. Overall, LINE exhibits favorable parameter robustness across the evaluated datasets.

4.5 Computational Efficiency

To evaluate computational efficiency, we compare the running times of different methods across datasets of varying scales. As shown in Fig. 3, LINE maintains competitive efficiency on small- and medium-scale datasets, while its advantage becomes increasingly pronounced on larger datasets. In contrast, the computational costs of several competing methods grow rapidly with the data scale, whereas LINE exhibits a substantially more moderate increase in running time. This favorable scaling behavior is consistent with the design of LINE, which avoids computationally expensive operations whose costs increase sharply with the number of samples. These results demonstrate that LINE achieves not only strong clustering performance but also favorable scalability, making it particularly suitable for large-scale ensemble clustering.

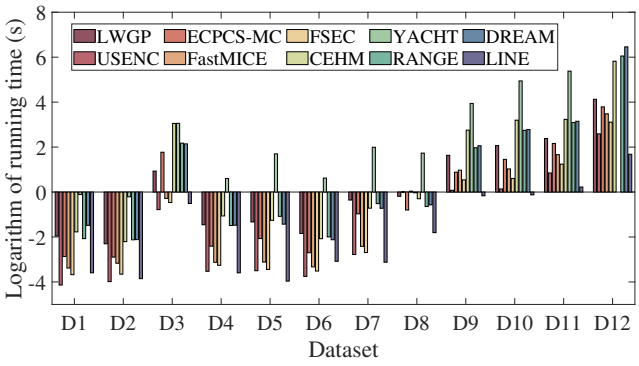


Figure 3: Running-time comparison of different methods.

5 Conclusion

We studied the multiplicity failure in decision-only ensemble clustering and formalized the desired robustness as replication invariance. To address this problem, we proposed LINE, which combines granularity-comparable consensus fidelity with a geometry of decision redundancy. We proved exact RI through a quotient formulation and further established a redundancy-adjusted effective ensemble size and near-replication stability independent of replication multiplicity. Experiments on twelve datasets demonstrate strong clustering performance and consistent robustness to exact and near replication. LINE also scales linearly with the number of samples for fixed ensemble and consensus sizes, supporting large-scale ensemble clustering.

AI USE STATEMENT

Generative AI tools were employed solely to improve the manuscript's language and formatting. All resulting edits were subsequently reviewed by the authors.

REFERENCES

Zekang Bian, Jia Qu, Jie Zhou, Zhibin Jiang, and Shitong Wang. Weighted adaptively ensemble clustering method based on fuzzy co-association matrix. *Information Fusion*, 103:102099, 2024.

Man-Sheng Chen, Jia-Qi Lin, Chang-Dong Wang, Wu-Dong Xi, and Dong Huang. On regularizing multiple clusterings for ensemble clustering by graph tensor learning. In *Proceedings of the ACM International Conference on Multimedia*, pp. 3069–3077, 2023.

Man-Sheng Chen, Jia-Qi Lin, Chang-Dong Wang, Dong Huang, and Jian-Huang Lai. Contrastive ensemble clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(8):14678–14690, 2025.

Chenhang Cui, Yazhou Ren, Jingyu Pu, Jiawei Li, Xiaorong Pu, Tianyi Wu, Yutao Shi, and Lifang He. A novel approach for effective multi-view clustering with information-theoretic perspective. In *Advances in Neural Information Processing Systems*, volume 36, pp. 44847–44859, 2023.

Benoit Dufumier, Javiera Castillo Navarro, Devis Tuia, and Jean-Philippe Thiran. What to align in multimodal contrastive learning? In *International Conference on Learning Representations*, volume 2025, pp. 5408–5432, 2025.

Xin Gao and Jian Pu. Deep incomplete multi-view learning via cyclic permutation of vaes. In *International Conference on Learning Representations*, volume 2025, pp. 99609–99643, 2025.

Chuanxing Geng, Aiyang Han, and Songcan Chen. Explicit view-labels matter: A multifacet complementarity study of multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):2520–2532, 2025.

Keyvan Golalipour, Ebrahim Akbari, and Homayun Motameni. Cluster ensemble selection based on maximum quality-maximum diversity. *Engineering Applications of Artificial Intelligence*, 131: 107873, 2024.

Shengzhao Guo, Jingyu Wang, Letian Zhang, Feiping Nie, and Xuelong Li. Ib2mc: Information bottleneck inspired balanced multiview clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–14, 2026.

Zhezheng Hao, Zhoumin Lu, Guoxu Li, Feiping Nie, Rong Wang, and Xuelong Li. Ensemble clustering with attentional representation. *IEEE Transactions on Knowledge and Data Engineering*, 36(2):581–593, 2024.

Shizhe Hu, Zhengzheng Lou, Xiaoqiang Yan, and Yangdong Ye. A survey on information bottleneck. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5325–5344, 2024.

Shizhe Hu, Jiahao Fan, Guoliang Zou, and Yangdong Ye. Multi-aspect self-guided deep information bottleneck for multi-modal clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17314–17322, 2025.

Shizhe Hu, Zhangwen Gou, Shuaiju Li, Jin Qin, Xiaoheng Jiang, Pei Lv, and Mingliang Xu. Calibrated information bottleneck for trusted multi-modal clustering. In *International Conference on Learning Representations*, 2026.

Dong Huang, Jian-Huang Lai, and Chang-Dong Wang. Robust ensemble clustering using probability trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 28(5):1312–1326, 2016.

Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. Locally weighted ensemble clustering. *IEEE Transactions on Cybernetics*, 48(5):1460–1473, 2018.

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

Dong Huang, Chang-Dong Wang, Hongxing Peng, Jianhuang Lai, and Chee-Keong Kwoh. Enhanced ensemble clustering via fast propagation of cluster-wise similarities. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1):508–520, 2021.

Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11388–11402, 2023a.

Weitian Huang, Sirui Yang, and Hongmin Cai. Generalized information-theoretic multi-view clustering. In *Advances in Neural Information Processing Systems*, volume 36, pp. 58752–58764, 2023b.

Xia Ji, Jiawei Sun, Jianhua Peng, Yue Pang, and Peng Zhou. Clustering ensemble based on fuzzy matrix self-enhancement. *IEEE Transactions on Knowledge and Data Engineering*, 37(1):148–161, 2025.

Feijiang Li, Jieting Wang, Liuya Zhang, Yuhua Qian, Shuai Jin, Tao Yan, and Liang Du. k-hyperedge medoids for clustering ensemble. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18271–18278, 2025.

Feijiang Li, Jieting Wang, Yuhua Qian, Hongren Yan, and Tingjin Luo. How relation enrichment improves clustering ensemble performance: A second order induced relation view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–17, 2026.

Hongfu Liu, Tongliang Liu, Junjie Wu, Dacheng Tao, and Yun Fu. Spectral ensemble clustering. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 715–724, 2015.

Zhengzheng Lou, Hang Xue, Chaoyang Zhang, and Shizhe Hu. A peer-review look on multi-modal clustering: An information bottleneck realization method. In *Proceedings of the International Conference on Machine Learning*, volume 267, pp. 40384–40399, 2025a.

Zhengzheng Lou, Ke Zhang, Yucong Wu, and Shizhe Hu. Super deep contrastive information bottleneck for multi-modal clustering. In *Proceedings of the International Conference on Machine Learning*, volume 267, pp. 40400–40410, 2025b.

Zhengzheng Lou, Yuhan Zhan, Mingyang Lv, Yingxuan Li, Yuyang Du, and Shizhe Hu. Structure-aware granular-ball based information bottleneck for multi-modal clustering. In *Proceedings of the International Conference on Machine Learning*, 2026.

Louis Ohl and Fredrik Lindsten. Discriminative ordering through ensemble consensus. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pp. 3252–3271, 2025.

Yalan Qin, Nan Pu, Nicu Sebe, and Guorui Feng. Latent space learning-based ensemble clustering. *IEEE Transactions on Image Processing*, 34:1259–1270, 2025.

Yazhou Ren, Zichen Wen, Junlong Ke, Chenhang Cui, Yonghao Huang, Xinyue Chen, Philip S. Yu, and Lifang He. Enhancing multi-view clustering: A sufficient information-theoretic approach for consistency acquisition and redundancy elimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(7):7420–7435, 2026.

Dennis Treder-Tschechlov, Manuel Fritz, Holger Schwarz, and Bernhard Mitschang. Ensemble clustering based on meta-learning and hyperparameter optimization. *Proceedings of the VLDB Endowment*, 17(11):2880–2892, 2024.

Paul Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001.

Wei Wei, Jianguo Wu, Xinyao Guo, Jing Yan, and Jiye Liang. Self-constrained clustering ensemble. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(12):11946–11960, 2025.

Fangyuan Xie, Feiping Nie, Weizhong Yu, and Xuelong Li. Parameter-free ensemble clustering with dynamic weighting mechanism. *Pattern Recognition*, 151:110389, 2024.

Jiaxuan Xu, Taiyong Li, and Lei Duan. Enhancing ensemble clustering with adaptive high-order topological weights. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16184–16192, 2024.

Jiaxuan Xu, Lei Duan, Xinye Wang, and Liang Du. Dynamic anchor-based ensemble clustering via hypergraph reconstruction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 6740–6748, 2025.

Jiaxuan Xu, Lei Duan, Xinye Wang, and Liang Du. Large-scale robust enhanced ensemble clustering via outlier decoupling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 39691–39700, 2026a.

Jiaxuan Xu, Lei Duan, Xinye Wang, Liang Du, Yidan Zhang, and Zhen Guo. Dynamic anchor-based one-step hypergraph ensemble clustering. *IEEE Transactions on Knowledge and Data Engineering*, 38(9):5798–5810, 2026b.

Xiaoqiang Yan, Zhixiang Jin, Fengshou Han, and Yangdong Ye. Differentiable information bottleneck for deterministic multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27435–27444, 2024.

Mimi Zhang. Weighted clustering ensemble: A review. *Pattern Recognition*, 124:108428, 2022.

Runxin Zhang, Shuaijun Hang, Zhensheng Sun, Feiping Nie, Rong Wang, and Xuelong Li. Anchor-based fast spectral ensemble clustering. *Information Fusion*, 113:102587, 2025a.

Runxin Zhang, Xia Wu, Huimin Chen, Guanxiong He, Zheng Wang, Rong Wang, and Feiping Nie. Toward balance adaptive weighted ensemble clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(6):5279–5292, 2025b.

Xu Zhang, Yuheng Jia, Mofei Song, and Ran Wang. Similarity and dissimilarity guided co-association matrix construction for ensemble clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(11):6694–6707, 2025c.

Xu Zhang, Haoye Qiu, Weixuan Liang, Hui Liu, Junhui Hou, and Yuheng Jia. Generalization performance of ensemble clustering: From theory to algorithm. In *Proceedings of the International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 75901–75927, 2025d.

Xuan Zheng, Yihang Lu, Rong Wang, Feiping Nie, and Xuelong Li. Structured graph-based ensemble clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(6):3728–3738, 2025.

Peng Zhou, Liang Du, Xinwang Liu, Zhaolong Ling, Xia Ji, Xuejun Li, and Yi-Dong Shen. Partial clustering ensemble. *IEEE Transactions on Knowledge and Data Engineering*, 36(5):2096–2109, 2024a.

Peng Zhou, Boao Hu, Dengcheng Yan, and Liang Du. Clustering ensemble via diffusion on adaptive multiplex. *IEEE Transactions on Knowledge and Data Engineering*, 36(4):1463–1474, 2024b.

Peng Zhou, Rongwen Li, Zhaolong Ling, Liang Du, and Xinwang Liu. Fair clustering ensemble with equal cluster capacity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(3): 1729–1746, 2025.

Summary: This appendix provides the experimental details and supplementary analyses used in the paper. Appendix B gives the optimization procedure and complexity analysis, Appendix C develops additional theoretical results, Appendices D–F provide derivations and proofs, and Appendix G presents additional experimental results.

A Experimental Details

A.1 Datasets

We conduct experiments on twelve datasets covering handwritten recognition, gesture analysis, object recognition, categorical decision making, sensor classification, human activity recognition, and network intrusion detection. Detailed information about the datasets is provided in Table 1.

Optdigits contains handwritten digit images represented by 64 numerical features and 10 classes. **GesturePhase** is the processed Gesture Phase Segmentation dataset, with five gesture-phase classes. **ALOI100** is the 100-class ALOI subset used in our experiments, containing 10,000 samples represented by 128-dimensional features. **Pendigits** contains handwritten digit trajectories represented by 16 numerical features and 10 digit classes. **Nursery** is a categorical decision-making dataset with 5 output classes. **DryBean** contains geometric measurements of dry-bean samples from 7 classes. **JungleChess2pcs** contains two-piece Jungle Chess states grouped into three outcome classes. **SensIT** is the 3-class sensor dataset used in our experiments, with 98,528 samples and 50 features. **EMNIST** denotes the processed 10-class digit data used in our experiments, containing 280,000 samples with 784 input dimensions. **MHEALTH** contains wearable-sensor measurements from 12 activity classes after the preprocessing used in our experiments. **NF-ToN-IoT** is a NetFlow-based intrusion dataset containing benign traffic and nine attack categories. **KDDCup99** is a large-scale network-intrusion dataset; our processed version contains 4,898,431 samples, 41 features, and 23 classes.

A.2 Optimization Procedure

The optimization procedure of LINE is summarized in Algorithm 1. It first constructs the decision representation and redundancy matrix, then alternately updates the model variables until convergence, after which the final consensus labels are obtained from the optimized posterior assignments.

Algorithm 1 LINE

1: **Input**: Base partitions $\Pi = \{\pi^{(v)}\}_{v=1}^{M}$, number of consensus clusters K, parameters τ and λ.
2: **Output**: Consensus clustering labels $\widehat{\boldsymbol{z}}$.
3: Construct the decision matrix $\mathbf{Y}$ and compute the marginal entropy $H(Y^{(v)})$ for each base partition.
4: Construct the redundancy matrix $\mathbf{R}$ from the normalized centered co-assignment representations by Eq. (10).
5: Initialize the soft consensus $\mathbf{Q}$, $\boldsymbol{\pi}$, $\{\boldsymbol{\Theta}^{(v)}\}_{v=1}^{M}$, and $\boldsymbol{w}$.
6: **while** not converged **do**
7: Update the consensus posterior $\mathbf{Q}$ by Eq. (24).
8: Update the consensus prior $\boldsymbol{\pi}$ by Eq. (21).
9: Update the partition channels $\{\boldsymbol{\Theta}^{(v)}\}_{v=1}^{M}$ by Eq. (22).
10: Compute the normalized distortions $\boldsymbol{d}$.
11: Update the partition weights $\boldsymbol{w}$ by solving Eq. (25).
12: **end while**
13: Obtain $\widehat{z}_i = \arg\max_{c\in[K]} q_{ic}$ for each sample i.

A.3 Computing Environment and Implementation Details

All experiments were conducted on the same workstation equipped with an AMD EPYC 7763 64-Core Processor at 2.45 GHz, two NVIDIA GeForce RTX 2080 Ti GPUs, and 128 GB of memory, running Windows 10 Professional. The algorithms were implemented in MATLAB R2024b, and all runtime comparisons were measured in this environment. The implementation used in the experiments is included in the supplementary material.

A.4 Complexity and Scalability Analysis

Table 4 compares the computational and memory requirements of recent representative ensemble clustering methods. Several methods, including ECAR, GPEC, LEADEC, FCE, and CE-SoCo, exhibit quadratic or higher-order dependence on the number of samples n, which limits their applicability to large-scale datasets. Recent scalable methods alleviate this sample-wise computational burden, but their complexities remain coupled with the total number of base clusters c. Specifically, CEHM requires $\mathcal{O}(nc + c^2)$ time and $\mathcal{O}(nc)$ space, while YACHT, DREAM, and RANGE exhibit stronger dependence on c through higher-order terms. In contrast, LINE has a time complexity of $\mathcal{O}(n + c^2)$ and a space complexity of $\mathcal{O}(n)$, yielding linear sample-dependent computation without constructing sample-wise pairwise matrices. Moreover, LINE is replication-invariant, ensuring that repeated equivalent partitions do not alter the consensus solely through their multiplicities. Together with its demonstrated scalability to millions of samples, these properties make LINE well suited for large-scale ensemble clustering in the presence of redundant decisions.

Table 4: Complexity comparison of recent representative methods.

Method	Time Complexity	Space Complexity	Replication Invariance	Scalability / Max. Reported (n)
ECAR [*TKDE*, 24]	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$	✗	✗/ 70,000
GPEC [*ICML*, 25]	$\mathcal{O}(n^3)$	$\mathcal{O}(n^2)$	✗	✗/ 5,456
LEADEC [*TIP*, 25]	$\mathcal{O}(n^3)$	$\mathcal{O}(n^2)$	✗	✗/ 7,797
FCE [*TPAMI*, 25]	$\mathcal{O}(n^2)$	$\mathcal{O}(n)$	✗	✗/ 10,299
CEHM [*AAAI*, 25]	$\mathcal{O}(nc + c^2)$	$\mathcal{O}(nc)$	✗	✓/ 43,500
YACHT [*IJCAI*, 25]	$\mathcal{O}(nc^2 + c^3)$	$\mathcal{O}(nc + c^2)$	✗	✓/ 280,000
CE-SoCo [*TPAMI*, 26]	$\mathcal{O}(c^2 + n^2 \log n)$	$\mathcal{O}(\max\{c^2, n^2\})$	✗	✓/ 10,000
DREAM [*TKDE*, 26]	$\mathcal{O}(nc^2 + c^3)$	$\mathcal{O}(nc + c^2)$	✗	✓/ 1,000,000
RANGE [*CVPR*, 26]	$\mathcal{O}(nc^2 + c^3)$	$\mathcal{O}(nc + c^2)$	✗	✓/ 2,000,000
LINE (Ours)	$\mathcal{O}(n + c^2)$	$\mathcal{O}(n)$	✓	✓/ 4,898,431

B Optimization and Complexity

B.1 Optimization

The information-theoretic objective in Eq. (14) admits an exact auxiliary-variable form. Introduce a consensus prior $\boldsymbol{\pi} \in \Delta_K$ and the partition channels $\{\boldsymbol{\Theta}^{(v)}\}_{v=1}^{M}$. The rate and fidelity terms satisfy

$$I_{\mathbf{Q}}(I; Z) = \min_{\boldsymbol{\pi} \in \Delta_K} \frac{1}{n} \sum_{i=1}^{n} D_{\mathrm{KL}}(\boldsymbol{q}_i \| \boldsymbol{\pi}), \tag{19}$$

$$d_v(\mathbf{Q}) = \min_{\boldsymbol{\Theta}^{(v)}} D_v(\mathbf{Q}, \boldsymbol{\Theta}^{(v)}). \tag{20}$$

Their minimizers are the empirical consensus prior

$$\pi_c^{\star} = \frac{1}{n} \sum_{i=1}^{n} q_{ic} \tag{21}$$

and the conditional assignment model

$$\theta_{r|c}^{(v)\star} = \frac{\sum_i q_{ic} \mathbb{I}[y_i^{(v)} = r]}{\sum_i q_{ic}}. \tag{22}$$

Substituting these representations into Eq. (14) yields the equivalent optimization problem

$$\begin{aligned} \min_{\mathbf{Q}, \boldsymbol{\pi}, \{\boldsymbol{\Theta}^{(v)}\}, \boldsymbol{w}} \quad & \mathcal{F} = \frac{\tau}{n} \sum_{i=1}^{n} D_{\mathrm{KL}}(\boldsymbol{q}_i \| \boldsymbol{\pi}) + \sum_{v=1}^{M} w_v D_v(\mathbf{Q}, \boldsymbol{\Theta}^{(v)}) + \lambda \boldsymbol{w}^{\top} \mathbf{R} \boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{q}_i \in \Delta_K, \quad \boldsymbol{\pi} \in \Delta_K, \quad \boldsymbol{\theta}_{\cdot|c}^{(v)} \in \Delta_{c_v}, \quad \boldsymbol{w} \in \Delta_M. \end{aligned} \tag{23}$$

The objective is not jointly convex, but each model block is tractable and the weight block is convex. With $\boldsymbol{\pi}$, $\{\boldsymbol{\Theta}^{(v)}\}$, and $\boldsymbol{w}$ held fixed, the consensus posterior is updated by

$$q_{ic} = \frac{\pi_c \exp\left[\frac{1}{\tau} \sum_{v=1}^{M} \frac{w_v}{H(Y^{(v)})} \log \theta_{y_i^{(v)}|c}^{(v)}\right]}{\sum_{k=1}^{K} \pi_k \exp\left[\frac{1}{\tau} \sum_{v=1}^{M} \frac{w_v}{H(Y^{(v)})} \log \theta_{y_i^{(v)}|k}^{(v)}\right]}. \tag{24}$$

The prior and channel parameters are then updated by Eqs. (21) and (22). Using the resulting normalized distortions $\boldsymbol{d} = (d_1, \ldots, d_M)^\top$, the partition weights solve

$$\min_{\boldsymbol{w} \in \Delta_M} \quad \boldsymbol{d}^\top \boldsymbol{w} + \lambda \boldsymbol{w}^\top \mathbf{R} \boldsymbol{w}. \tag{25}$$

This subproblem is convex because $\mathbf{R} \succeq 0$. We solve it by projected gradient descent with Euclidean simplex projection; its gradient Lipschitz constant is $2\lambda\lambda_{\max}(\mathbf{R})$. The implementation uses a tight inner tolerance and accepts a numerical weight update only when it does not increase the weight-block objective.

The procedure starts from a soft consensus initialized by weighted K-modes over unique decision patterns. It then alternates the posterior, model, and weight updates until the relative objective change falls below a tolerance. The final partition is obtained by

$$\widehat{z}_i = \arg\max_{c \in [K]} q_{ic}. \tag{26}$$

The initialization and all subsequent updates allow $c_v \neq K$ for every base partition.

Proposition 3 (Monotone exact-block optimization)**.** *Assume $\tau > 0$, non-degenerate active partitions, positive initialization for the probability variables, and exact minimization of every block. Then the objective sequence is non-increasing and converges to a finite value, and every accumulation point is blockwise stationary.*

Proposition 3 characterizes the ideal exact-block scheme. In the implementation, the convex weight block is solved to numerical tolerance by projected gradient descent with a descent safeguard; we therefore use the proposition as the optimization target rather than claiming exact finite-precision block minimization. The detailed update derivations and proof are given in Appendices D and E.

B.2 Complexity Analysis

Let $\mathbf{N}^{uv} = \mathbf{H}^{(u)\top}\mathbf{H}^{(v)}$ and $\boldsymbol{n}^{(v)} = \mathbf{H}^{(v)\top}\mathbf{1}$. Then

$$\mathbf{H}^{(u)\top}\mathbf{P}\mathbf{H}^{(v)} = \mathbf{N}^{uv} - \frac{\boldsymbol{n}^{(u)}\boldsymbol{n}^{(v)\top}}{n}, \tag{27}$$

$$\langle \widetilde{\mathbf{K}}^{(u)}, \widetilde{\mathbf{K}}^{(v)} \rangle_F = \left\| \mathbf{N}^{uv} - \frac{\boldsymbol{n}^{(u)}\boldsymbol{n}^{(v)\top}}{n} \right\|_F^2. \tag{28}$$

Hence every entry of $\mathbf{R}$ is computed exactly from a partition contingency table.

Objects with identical rows in $\mathbf{Y}$ have identical posterior updates. We therefore aggregate the n objects into $p \leq n$ unique decision patterns and retain their multiplicities; this compression changes neither the objective nor any sufficient statistic. Let $C = \sum_{v=1}^{M} c_v$. Constructing all pairwise contingency tables and their centered alignment statistics costs

$$O\big(pM^2 + C^2\big), \tag{29}$$

where the C^2 term accounts for operations on the contingency matrices and is independent of n. Weighted K-modes initialization costs $O(I_0 pMK)$, where I_0 is its iteration count, and each main iteration costs

$$O(pMK + T_w M^2), \tag{30}$$

where T_w is the number of projected-gradient steps in Eq. (25). Thus, for fixed ensemble size, partition granularities, and consensus size, the sample-dependent computation is linear in p and hence in n.

Including the compressed decision matrix, the working memory cost is

$$O\big(pM + pK + M^2 + KC\big). \tag{31}$$

Neither a sample co-association matrix nor a dense base-cluster graph is constructed.

C Additional Theoretical Analysis

Corollary 1 establishes exact invariance. We next study two consequences of the same geometry: how much distinct decision support remains after accounting for dependence, and how the objective changes when copies are only approximately equivalent.

C.1 Effective Decision Support

Equation (12) induces a redundancy-adjusted effective ensemble size.

Definition 2 (Effective ensemble size)**.** *For $\boldsymbol{w} \in \Delta_M$, define*

$$M_{\text{eff}}(\boldsymbol{w}, \mathbf{R}) = \frac{1}{\boldsymbol{w}^\top \mathbf{R} \boldsymbol{w}} = \frac{1}{\|\boldsymbol{\mu}_{\boldsymbol{w}}\|_2^2}. \tag{32}$$

By Theorem 1, $\boldsymbol{w}^\top \mathbf{R} \boldsymbol{w} = \boldsymbol{\alpha}^\top \mathbf{S} \boldsymbol{\alpha}$, so the quantity is controlled by class masses and inter-class alignment rather than replica counts.

Theorem 2 (Effective diversity over distinct decisions)**.** *Let G be the number of distinct decision classes. Then:*

(i) $1 \le M_{\text{eff}} \le G \le M$*;*

(ii) if the G class embeddings are orthogonal, then

$$M_{\text{eff}} = \frac{1}{\sum_{g=1}^{G} \alpha_g^2} = \exp(H_2(\boldsymbol{\alpha}))\,, \tag{33}$$

where $H_2(\boldsymbol{\alpha}) = -\log \sum_g \alpha_g^2$ is order-two Rényi entropy;

(iii) if all distinct classes have common alignment $\rho \in [0, 1]$, then

$$M_{\text{eff}} = \frac{1}{\rho + (1-\rho) \sum_g \alpha_g^2}. \tag{34}$$

For uniform class mass, this becomes

$$M_{\text{eff}} = \frac{G}{1 + (G-1)\rho}. \tag{35}$$

For orthogonal decision classes, Eq. (33) reduces to the inverse Simpson concentration, equivalently the order-two Hill number. The common-alignment case makes the role of dependence explicit: even uniformly weighted classes provide less effective support as their alignment increases. For corresponding solutions linked by Theorem 1, arbitrary replication leaves M_{eff} unchanged.

C.2 Stability under Near Replication

Exact replication is the zero-perturbation case. Let Π' be obtained by replacing partition j with L near replicas indexed by $j_1, \dots, j_L$. For each feasible $\mathbf{Q}$, define

$$\epsilon_D(\mathbf{Q}) = \max_{\ell \in [L]} \left| d'_{j_\ell}(\mathbf{Q}) - d_j(\mathbf{Q}) \right|, \tag{36}$$

and suppose

$$\max_{\ell \in [L]} \|\boldsymbol{\phi}'_{j_\ell} - \boldsymbol{\phi}_j\|_2 \le \epsilon_\phi. \tag{37}$$

To compare consensus posteriors without fixing arbitrary cluster indices, let $\mathfrak{P}_K$ be the set of $K \times K$ permutation matrices and define

$$\operatorname{dist}(\mathbf{Q}, \mathcal{A}) = \inf_{\overline{\mathbf{Q}} \in \mathcal{A}} \min_{\mathbf{P} \in \mathfrak{P}_K} \frac{1}{\sqrt{n}} \|\mathbf{Q} - \overline{\mathbf{Q}}\mathbf{P}\|_F. \tag{38}$$

Theorem 3 (Near-replication stability)**.** *For every feasible $\mathbf{Q}$,*

$$|\mathcal{J}_{\Pi'}(\mathbf{Q}) - \mathcal{J}_{\Pi}(\mathbf{Q})| \le \epsilon_D(\mathbf{Q}) + 2\lambda\epsilon_\phi. \tag{39}$$

If $\sup_{\mathbf{Q}} \epsilon_D(\mathbf{Q}) \le \bar{\epsilon}_D$ and $\delta = \bar{\epsilon}_D + 2\lambda\epsilon_\phi$, then

$$\left| \min_{\mathbf{Q}} \mathcal{J}_{\Pi'}(\mathbf{Q}) - \min_{\mathbf{Q}} \mathcal{J}_{\Pi}(\mathbf{Q}) \right| \le \delta. \tag{40}$$

Moreover, suppose $\mathcal{J}_\Pi$ satisfies

$$\mathcal{J}_\Pi(\mathbf{Q}) \ge \mathcal{J}_\Pi^\star + \kappa \operatorname{dist}^2(\mathbf{Q}, \mathcal{Q}_\Pi^\star) \tag{41}$$

for some $\kappa > 0$. Then every $\mathbf{Q}_{\Pi'}^\star \in \mathcal{Q}_{\Pi'}^\star$ satisfies

$$\operatorname{dist}(\mathbf{Q}_{\Pi'}^\star, \mathcal{Q}_\Pi^\star) \le \sqrt{\frac{2\delta}{\kappa}}. \tag{42}$$

The bound separates the effect of near replication into the change in normalized coding distortion and the displacement in decision geometry, and it is independent of the replication multiplicity L. When both perturbations vanish, Eq. (39) reduces to exact RI. The objective-value bound follows from uniform perturbation control, while the consensus-set bound additionally requires the quadratic-growth condition in Eq. (41). Appendix E.5 gives the proof, and Appendix F relates assignment edits to the two perturbation quantities.

D Derivations and Optimization Details

D.1 Consensus Prior

The auxiliary representation of the consensus rate depends on $\boldsymbol{\pi}$ through

$$\frac{\tau}{n}\sum_{i=1}^{n}\sum_{c=1}^{K} q_{ic}\log\frac{q_{ic}}{\pi_c}. \tag{43}$$

Discarding terms independent of $\boldsymbol{\pi}$ gives

$$\min_{\boldsymbol{\pi}\in\Delta_K} \quad -\sum_{c=1}^{K}\left(\sum_{i=1}^{n} q_{ic}\right)\log\pi_c. \tag{44}$$

The simplex-constrained maximum-likelihood solution is

$$\pi_c^\star = \frac{\sum_i q_{ic}}{\sum_{i,k} q_{ik}} = \frac{1}{n}\sum_i q_{ic}, \tag{45}$$

which is Eq. (21).

D.2 Conditional Assignment Models

For fixed $\mathbf{Q}$, partition v, and consensus state c, define the soft contingency counts

$$N_{vrc} = \sum_{i=1}^{n} q_{ic}\mathbb{I}[y_i^{(v)} = r]. \tag{46}$$

The conditional assignment block solves

$$\min_{\boldsymbol{\theta}_{\cdot|c}^{(v)}\in\Delta_{c_v}} \quad -\sum_{r=1}^{c_v} N_{vrc}\log\theta_{r|c}^{(v)}. \tag{47}$$

Its maximum-likelihood solution is

$$\theta_{r|c}^{(v)\star} = \frac{N_{vrc}}{\sum_{r'} N_{vr'c}} = \frac{\sum_i q_{ic}\mathbb{I}[y_i^{(v)} = r]}{\sum_i q_{ic}}, \tag{48}$$

which gives Eq. (22).

D.3 Reduced Information-Theoretic Objective

Under $p_{\mathbf{Q}}(I = i, Z = c) = q_{ic}/n$, the marginal distribution of Z is

$$p_{\mathbf{Q}}(Z = c) = \frac{1}{n}\sum_i q_{ic} = \pi_c^\star. \tag{49}$$

Therefore

$$\begin{aligned}\frac{1}{n}\sum_i D_{\mathrm{KL}}(\boldsymbol{q}_i\|\boldsymbol{\pi}^\star) &= \sum_{i,c}\frac{q_{ic}}{n}\log\frac{q_{ic}/n}{(1/n)\pi_c^\star}\\ &= I_{\mathbf{Q}}(I;Z).\end{aligned} \tag{50}$$

At the optimum of the conditional assignment model,

$$-\frac{1}{n}\sum_{i,c} q_{ic}\log \theta^{(v)\star}_{y_i^{(v)}|c} = -\sum_{c,r} p_{\mathbf{Q}}(Z=c, Y^{(v)}=r)\log p_{\mathbf{Q}}(Y^{(v)}=r \mid Z=c)$$
$$= H_{\mathbf{Q}}(Y^{(v)} \mid Z). \tag{51}$$

These identities verify the two auxiliary representations used in Appendix B.1. Substituting their minimizers into Eq. (23) recovers Eq. (14); minimizing over $\boldsymbol{w}$ then gives Eq. (15).

D.4 Consensus Posterior Update

For one object i, terms depending on $\boldsymbol{q}_i$ are

$$\min_{\boldsymbol{q}_i \in \Delta_K} \quad \tau \sum_{c=1}^{K} q_{ic}\log\frac{q_{ic}}{\pi_c} - \sum_{c=1}^{K} q_{ic}\sum_{v=1}^{M}\frac{w_v}{H(Y^{(v)})}\log \theta^{(v)}_{y_i^{(v)}|c}. \tag{52}$$

Adding a Lagrange multiplier η_i for $\sum_c q_{ic} = 1$ and differentiating yields

$$\tau\left(\log q_{ic} - \log \pi_c + 1\right) - \sum_{v=1}^{M}\frac{w_v}{H(Y^{(v)})}\log \theta^{(v)}_{y_i^{(v)}|c} + \eta_i = 0. \tag{53}$$

Exponentiating and normalizing over c gives Eq. (24).

D.5 Source-Weight Update

For fixed consensus and partition-model blocks, the source-dependent part of the objective is

$$f(\boldsymbol{w}) = \boldsymbol{d}^\top \boldsymbol{w} + \lambda \boldsymbol{w}^\top \mathbf{R}\boldsymbol{w}, \qquad \boldsymbol{w} \in \Delta_M. \tag{54}$$

Because $\mathbf{R} \succeq 0$, this is convex and

$$\nabla f(\boldsymbol{w}) = \boldsymbol{d} + 2\lambda \mathbf{R}\boldsymbol{w}. \tag{55}$$

The gradient is Lipschitz with constant

$$L_w = 2\lambda\lambda_{\max}(\mathbf{R}). \tag{56}$$

Projected gradient descent therefore takes the form

$$\boldsymbol{w}^{(s+1)} = \mathrm{Proj}_{\Delta_M}\left[\boldsymbol{w}^{(s)} - \gamma\left(\boldsymbol{d} + 2\lambda\mathbf{R}\boldsymbol{w}^{(s)}\right)\right]. \tag{57}$$

For $\lambda > 0$, any fixed step size $0 < \gamma < 2/L_w$ is valid, or a backtracking step can be used to ensure sufficient decrease. When $\lambda = 0$, the subproblem is linear over the simplex and is solved by a minimum-distortion vertex.

E Proofs

E.1 Proof of Proposition 1

For fixed $\mathbf{Q}$, write $d_v = d_v(\mathbf{Q})$. A linear function over the simplex attains its minimum at a simplex vertex, hence

$$\min_{\boldsymbol{w}\in\Delta_M} \boldsymbol{d}^\top \boldsymbol{w} = \min_v d_v. \tag{58}$$

Replicating a source only repeats an existing coordinate value and therefore cannot change this minimum. If the minimum is unique at $v^\star$, the only optimal simplex point is $\boldsymbol{e}_{v^\star}$, proving part (i).

For part (ii), suppose one decision is represented by L equivalent coordinates with nonnegative weights $w_1, \ldots, w_L$ summing to a fixed class mass α. Their linear fidelity contribution is αd and does not depend on the split. By Cauchy–Schwarz,

$$\sum_{\ell=1}^{L} w_\ell^2 \geq \frac{1}{L}\left(\sum_{\ell=1}^{L} w_\ell\right)^2 = \frac{\alpha^2}{L}, \tag{59}$$

with equality at $w_\ell = \alpha/L$. Thus the smallest isotropic quadratic cost attributable to that class is $\lambda\alpha^2/L$, which depends explicitly on its presentation multiplicity L. The corresponding reduced objective is therefore not replication invariant in general.

E.2 Proof of Proposition 2

By Eq. (10), $R_{uv} = \boldsymbol{\phi}_u^\top \boldsymbol{\phi}_v$. Let $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_M]$. Then

$$\mathbf{R} = \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \succeq 0. \tag{60}$$

Moreover,

$$\widetilde{\mathbf{K}}^{(v)} = \mathbf{P}\mathbf{H}^{(v)}\mathbf{H}^{(v)\top}\mathbf{P} \succeq 0. \tag{61}$$

For positive-semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathrm{tr}(\mathbf{AB}) \geq 0$. Hence

$$\langle \widetilde{\mathbf{K}}^{(u)}, \widetilde{\mathbf{K}}^{(v)} \rangle_F = \mathrm{tr}(\widetilde{\mathbf{K}}^{(u)}\widetilde{\mathbf{K}}^{(v)}) \geq 0. \tag{62}$$

It follows that $R_{uv} \geq 0$. Cauchy–Schwarz yields $R_{uv} \leq 1$, and normalization gives $R_{vv} = 1$.

For completeness, cyclicity of trace gives

$$\begin{aligned} \left\langle \widetilde{\mathbf{K}}^{(u)}, \widetilde{\mathbf{K}}^{(v)} \right\rangle_F &= \mathrm{tr}\Big(\mathbf{P}\mathbf{H}^{(u)}\mathbf{H}^{(u)\top}\mathbf{P}\mathbf{H}^{(v)}\mathbf{H}^{(v)\top}\mathbf{P}\Big) \\ &= \left\| \mathbf{H}^{(u)\top}\mathbf{P}\mathbf{H}^{(v)} \right\|_F^2, \end{aligned} \tag{63}$$

which, together with Eq. (27), proves Eq. (28).

E.3 Proof of Theorem 1

Fix $\mathbf{Q}$ and any $\boldsymbol{w} \in \Delta_M$. For class masses defined by Eq. (16),

$$\sum_{v=1}^{M} w_v d_v(\mathbf{Q}) = \sum_{g=1}^{G} \sum_{v \in \mathcal{C}_g} w_v d_g(\mathbf{Q}) = \sum_{g=1}^{G} \alpha_g d_g(\mathbf{Q}). \tag{64}$$

Similarly,

$$\sum_{v=1}^{M} w_v \boldsymbol{\phi}_v = \sum_{g=1}^{G} \alpha_g \boldsymbol{\phi}_g. \tag{65}$$

Using Eq. (12),

$$\boldsymbol{w}^\top \mathbf{R} \boldsymbol{w} = \left\| \sum_g \alpha_g \boldsymbol{\phi}_g \right\|_2^2 = \boldsymbol{\alpha}^\top \mathbf{S} \boldsymbol{\alpha}. \tag{66}$$

Thus every feasible $\boldsymbol{w}$ induces a feasible $\boldsymbol{\alpha}$ with the quotient objective value.

Conversely, let $\boldsymbol{\alpha} \in \Delta_G$. For each class, choose any nonnegative weights satisfying

$$\sum_{v \in \mathcal{C}_g} w_v = \alpha_g. \tag{67}$$

The resulting $\boldsymbol{w}$ lies in Δ_M and gives the same linear and quadratic terms. The original and quotient weight problems therefore have identical attainable objective sets and identical minima. Adding the consensus-rate term proves Eq. (17).

Replication changes the cardinality of one equivalence class but not the quotient variables $d_g(\mathbf{Q})$, $\boldsymbol{\phi}_g$, or $\mathbf{S}$. Corollary 1 follows immediately.

E.4 Proof of Theorem 2

Using the quotient representation,

$$\boldsymbol{w}^\top \mathbf{R} \boldsymbol{w} = \boldsymbol{\alpha}^\top \mathbf{S} \boldsymbol{\alpha} = \sum_g \alpha_g^2 + \sum_{g \neq h} \alpha_g \alpha_h S_{gh}. \tag{68}$$

Proposition 2 gives $0 \leq S_{gh} \leq 1$. Therefore

$$\boldsymbol{\alpha}^\top \mathbf{S} \boldsymbol{\alpha} \geq \sum_g \alpha_g^2 \geq \frac{1}{G}. \tag{69}$$

It also gives

$$\boldsymbol{\alpha}^\top \mathbf{S}\boldsymbol{\alpha} \le \sum_{g,h} \alpha_g \alpha_h = 1. \tag{70}$$

Taking reciprocals proves $1 \le M_{\text{eff}} \le G \le M$.

If the class embeddings are orthogonal, then $\mathbf{S} = \mathbf{I}$ and

$$M_{\text{eff}} = \left(\sum_g \alpha_g^2 \right)^{-1} = \exp(H_2(\boldsymbol{\alpha})). \tag{71}$$

If $S_{gg} = 1$ and $S_{gh} = \rho$ for $g \neq h$, then

$$\begin{aligned} \boldsymbol{\alpha}^\top \mathbf{S}\boldsymbol{\alpha} &= \sum_g \alpha_g^2 + \rho \sum_{g \neq h} \alpha_g \alpha_h \\ &= \rho + (1-\rho) \sum_g \alpha_g^2. \end{aligned} \tag{72}$$

Taking the reciprocal proves Eq. (34). For $\alpha_g = 1/G$, the stated uniform formula follows.

E.5 Proof of Theorem 3

Fix $\mathbf{Q}$. Consider any weight vector $\boldsymbol{w}'$ in the expanded ensemble Π' and aggregate the near-replica weights into

$$\alpha_j = \sum_{\ell=1}^{L} w'_{j\ell}. \tag{73}$$

All unmodified weights are copied directly to a feasible vector in the original ensemble. The change in the linear coding term is bounded by

$$\begin{aligned} \left| \sum_{\ell=1}^{L} w'_{j\ell} d'_{j\ell}(\mathbf{Q}) - \alpha_j d_j(\mathbf{Q}) \right| &\le \sum_{\ell=1}^{L} w'_{j\ell} \left| d'_{j\ell}(\mathbf{Q}) - d_j(\mathbf{Q}) \right| \\ &\le \alpha_j \epsilon_D(\mathbf{Q}) \le \epsilon_D(\mathbf{Q}). \end{aligned} \tag{74}$$

Let $\boldsymbol{\mu}'$ and $\boldsymbol{\mu}$ be the corresponding expanded and aggregated kernel means. Then

$$\begin{aligned} \|\boldsymbol{\mu}' - \boldsymbol{\mu}\|_2 &= \left\| \sum_{\ell=1}^{L} w'_{j\ell} (\boldsymbol{\phi}'_{j\ell} - \boldsymbol{\phi}_j) \right\|_2 \\ &\le \sum_{\ell=1}^{L} w'_{j\ell} \|\boldsymbol{\phi}'_{j\ell} - \boldsymbol{\phi}_j\|_2 \\ &\le \alpha_j \epsilon_\phi \le \epsilon_\phi. \end{aligned} \tag{75}$$

Both kernel means are convex combinations of unit vectors, so $\|\boldsymbol{\mu}'\|_2 \le 1$ and $\|\boldsymbol{\mu}\|_2 \le 1$. Consequently,

$$\begin{aligned} \left| \|\boldsymbol{\mu}'\|_2^2 - \|\boldsymbol{\mu}\|_2^2 \right| &\le (\|\boldsymbol{\mu}'\|_2 + \|\boldsymbol{\mu}\|_2) \|\boldsymbol{\mu}' - \boldsymbol{\mu}\|_2 \\ &\le 2\epsilon_\phi. \end{aligned} \tag{76}$$

Thus every feasible expanded weight vector has an aggregated original vector whose objective differs by at most

$$\delta(\mathbf{Q}) = \epsilon_D(\mathbf{Q}) + 2\lambda\epsilon_\phi. \tag{77}$$

Conversely, every original weight vector can be lifted to the expanded ensemble by assigning its j-th mass to any one near replica, or by splitting that mass arbitrarily. The same bound applies. Taking minima in both directions proves Eq. (39).

If $\epsilon_D(\mathbf{Q}) \le \bar{\epsilon}_D$ uniformly, then

$$\sup_{\mathbf{Q}} |\mathcal{J}_{\Pi'}(\mathbf{Q}) - \mathcal{J}_{\Pi}(\mathbf{Q})| \le \delta. \tag{78}$$

The elementary inequality

$$\left|\min_x f(x) - \min_x g(x)\right| \le \sup_x |f(x) - g(x)| \tag{79}$$

gives Eq. (40).

Finally, let $\mathbf{Q}^\star_{\Pi'}$ minimize $\mathcal{J}_{\Pi'}$. For any $\mathbf{Q}^\star_{\Pi} \in \mathcal{Q}^\star_{\Pi}$,

$$\begin{aligned}\mathcal{J}_{\Pi}(\mathbf{Q}^\star_{\Pi'}) &\le \mathcal{J}_{\Pi'}(\mathbf{Q}^\star_{\Pi'}) + \delta \\ &\le \mathcal{J}_{\Pi'}(\mathbf{Q}^\star_{\Pi}) + \delta \\ &\le \mathcal{J}_{\Pi}(\mathbf{Q}^\star_{\Pi}) + 2\delta.\end{aligned} \tag{80}$$

Combining this inequality with Eq. (41) yields

$$\kappa \operatorname{dist}^2(\mathbf{Q}^\star_{\Pi'}, \mathcal{Q}^\star_{\Pi}) \le 2\delta, \tag{81}$$

which proves Eq. (42).

E.6 Proof of Proposition 3

For fixed remaining blocks, the $\mathbf{Q}$-subproblem is a positive multiple of KL divergence plus a linear term and is minimized by Eq. (24). The consensus-prior and conditional-assignment updates are exact maximum-likelihood solutions. The weight subproblem is convex because $\mathbf{R} \succeq 0$. Exact cyclic block minimization therefore gives

$$\mathcal{F}^{(t+1)} \le \mathcal{F}^{(t)}. \tag{82}$$

Every term in Eq. (23) is nonnegative, so the objective sequence is bounded below and hence convergent. The simplex domains are compact, and the positivity assumptions keep the iterates in the region where all logarithmic block objectives are finite and continuous. Standard cyclic block-coordinate arguments then imply that every accumulation point of the exact-block sequence is blockwise stationary (Tseng, 2001).

F Observable Perturbation Bounds

Theorem 3 is stated directly in terms of coding and embedding changes because these are the two quantities used by the objective. They can also be controlled by changes to the cluster assignments.

F.1 From Decision Edits to Kernel-Embedding Distance

Proposition 4 (Embedding perturbation under label edits)**.** *Let partitions π and π' differ on at most m of the n objects. Let $\widetilde{\mathbf{K}} = \mathbf{PKP}$ and $\widetilde{\mathbf{K}}' = \mathbf{PK}'\mathbf{P}$, and assume*

$$\min\{\|\widetilde{\mathbf{K}}\|_F, \|\widetilde{\mathbf{K}}'\|_F\} \ge s > 0. \tag{83}$$

Then their normalized embeddings satisfy

$$\|\boldsymbol{\phi}' - \boldsymbol{\phi}\|_2 \le \frac{2\sqrt{2mn}}{s}. \tag{84}$$

Proof. An entry of the co-assignment matrix can change only if at least one of its two endpoint objects has been edited. There are at most $2mn - m^2 \le 2mn$ such ordered pairs, and every entry changes by at most one. Hence

$$\|\mathbf{K}' - \mathbf{K}\|_F \le \sqrt{2mn}. \tag{85}$$

Centering is a contraction in Frobenius norm, so

$$\|\widetilde{\mathbf{K}}' - \widetilde{\mathbf{K}}\|_F \le \sqrt{2mn}. \tag{86}$$

For any nonzero matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\left\| \frac{\mathbf{A}}{\|\mathbf{A}\|_F} - \frac{\mathbf{B}}{\|\mathbf{B}\|_F} \right\|_F \le \frac{2\|\mathbf{A} - \mathbf{B}\|_F}{\min\{\|\mathbf{A}\|_F, \|\mathbf{B}\|_F\}}. \tag{87}$$

Applying this inequality completes the proof.

F.2 From Decision Edits to Normalized Coding Distance

Fix $\mathbf{Q}$ and let π' be obtained by changing the cluster assignments of at most $m = \eta n$ objects in π, with the same label set of size c. Assume $0 \le \eta \le 1 - 1/c$. Since active partitions have $c \ge 2$, this condition covers the perturbation levels used in our experiments and is sufficient for the finite-alphabet continuity bounds applied to both the cK-state joint distribution and the c-state marginal. The empirical joint distributions of (Y, Z) and (Y', Z) then differ in total variation by at most η. Since their Z marginal is identical, finite-alphabet entropy continuity gives

$$|H_{\mathbf{Q}}(Y' \mid Z) - H_{\mathbf{Q}}(Y \mid Z)| \le h_2(\eta) + \eta \log(cK - 1), \tag{88}$$

where $h_2(\eta) = -\eta \log \eta - (1 - \eta) \log(1 - \eta)$. Similarly,

$$|H(Y') - H(Y)| \le h_2(\eta) + \eta \log(c - 1). \tag{89}$$

If $\min\{H(Y), H(Y')\} \ge h_{\min} > 0$, then

$$\left| \frac{H_{\mathbf{Q}}(Y' \mid Z)}{H(Y')} - \frac{H_{\mathbf{Q}}(Y \mid Z)}{H(Y)} \right| \le \frac{2h_2(\eta) + \eta \log(cK - 1) + \eta \log(c - 1)}{h_{\min}}. \tag{90}$$

Equations (84) and (90) connect an observable assignment-edit rate to the stability terms in Theorem 3. Combining them gives an explicit edit-rate statement.

Corollary 2 (Stability under an assignment-edit rate)**.** *Suppose every near replica of a partition with c labels differs from it on at most $m = \eta n$ objects, where $0 \le \eta \le 1 - 1/c$. Assume the centered-kernel norms of the original and near replicas are at least $s > 0$, and their marginal entropies are at least $h_{\min} > 0$. Define*

$$B_D(\eta) = \frac{2h_2(\eta) + \eta \log(cK - 1) + \eta \log(c - 1)}{h_{\min}}. \tag{91}$$

Then, for every feasible $\mathbf{Q}$,

$$|\mathcal{J}_{\Pi'}(\mathbf{Q}) - \mathcal{J}_{\Pi}(\mathbf{Q})| \le B_D(\eta) + \frac{4\lambda n\sqrt{2\eta}}{s}. \tag{92}$$

The bound is independent of the number L of near replicas and converges to zero as $\eta \to 0$ for fixed non-degenerate partitions.

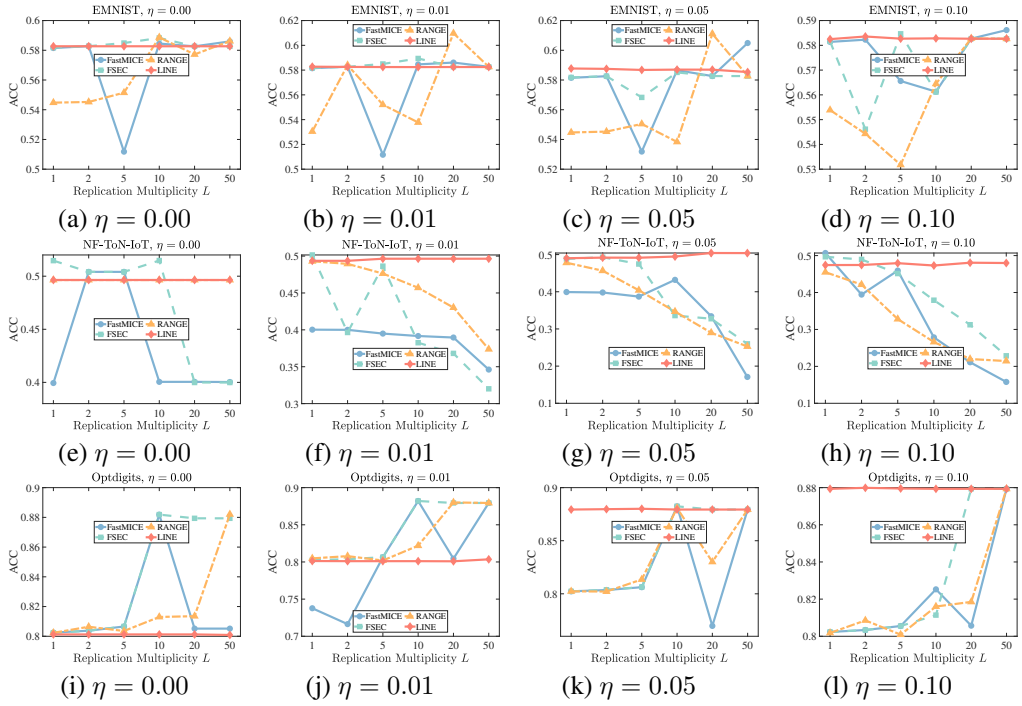
Proof. Equation (90) gives $\epsilon_D(\mathbf{Q}) \le B_D(\eta)$ uniformly in $\mathbf{Q}$. Substituting $m = \eta n$ into Eq. (84) gives $\epsilon_\phi \le 2n\sqrt{2\eta}/s$. The result follows directly from Eq. (39).

G Additional Experimental Results

G.1 Replication Robustness on Additional Datasets

Figures 4–6 provide the replication results on additional datasets using the same protocol as Section 4. Across these datasets, LINE generally shows smaller performance variation as the replication multiplicity increases. The behavior remains stable across the tested perturbation rates, while the competing methods are more sensitive to repeated or near-repeated decisions on several datasets. These results complement the main-text experiment and show that the observed multiplicity robustness is consistent across different data scales and cluster granularities.

To further examine whether simple preprocessing can resolve the multiplicity issue, we compare LINE with three competing methods equipped with exact deduplication (ED), namely FSEC+ED, FastMICE+ED, and RANGE+ED. Before consensus clustering, ED removes base partitions that induce exactly the same clustering decision, while treating label-permuted representations of the same partition as equivalent. As shown in Figure 7, when $\eta = 0$, exact replicas can be removed directly, and the resulting performance is consequently almost insensitive to L. This indicates that exact deduplication can mitigate multiplicity effects when the repeated clustering decisions are exactly identical. Once the replicas are perturbed, however, they are no longer exactly equivalent and therefore remain in the ensemble. Under $\eta \in \{0.01, 0.05, 0.10\}$, the deduplicated competing methods again exhibit noticeable variation as L increases, whereas LINE remains substantially more stable. Moreover, ED relies on an explicit preprocessing step that collapses exactly equivalent partitions

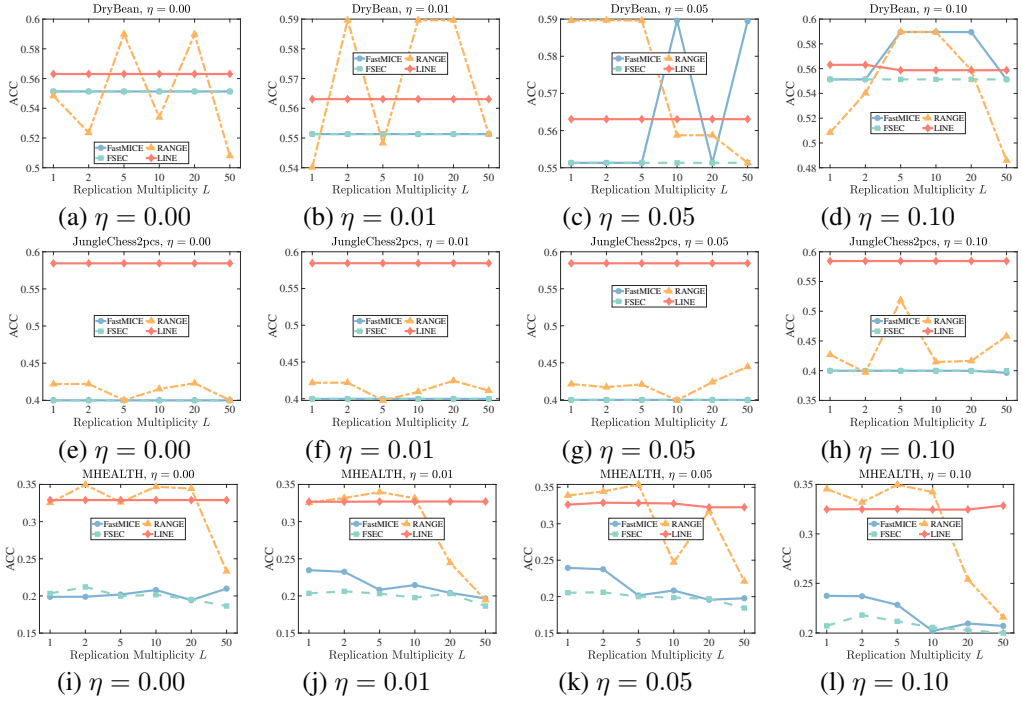


(a) $\eta = 0.00$ (b) $\eta = 0.01$ (c) $\eta = 0.05$ (d) $\eta = 0.10$

(e) $\eta = 0.00$ (f) $\eta = 0.01$ (g) $\eta = 0.05$ (h) $\eta = 0.10$

(i) $\eta = 0.00$ (j) $\eta = 0.01$ (k) $\eta = 0.05$ (l) $\eta = 0.10$

Figure 4: ACC under different replication multiplicities L and assignment perturbation rates η on EMNIST, NF-ToN-IoT, and Optdigits.



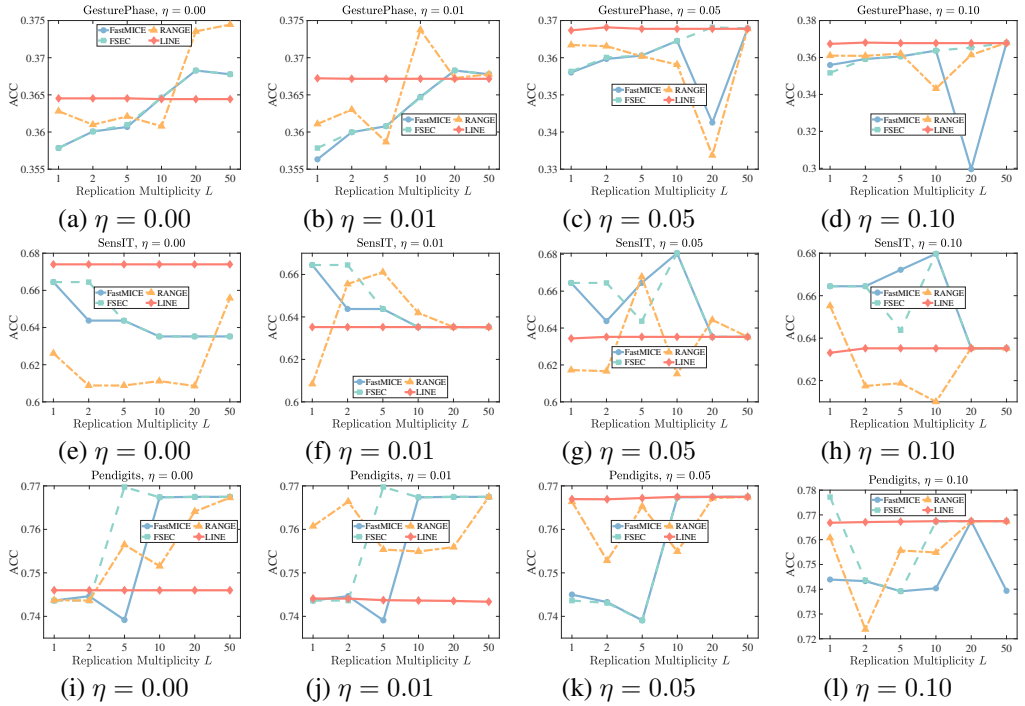
(a) $\eta = 0.00$ (b) $\eta = 0.01$ (c) $\eta = 0.05$ (d) $\eta = 0.10$

(e) $\eta = 0.00$ (f) $\eta = 0.01$ (g) $\eta = 0.05$ (h) $\eta = 0.10$

(i) $\eta = 0.00$ (j) $\eta = 0.01$ (k) $\eta = 0.05$ (l) $\eta = 0.10$

Figure 5: ACC under different replication multiplicities L and assignment perturbation rates η on DryBean, JungleChess2pcs, and MHEALTH.

before consensus learning. In contrast, LINE operates directly on the supplied ensemble without removing partition columns: equivalent decisions are retained in the input representation, while their contribution to the objective depends only on the aggregate mass of the corresponding decision-

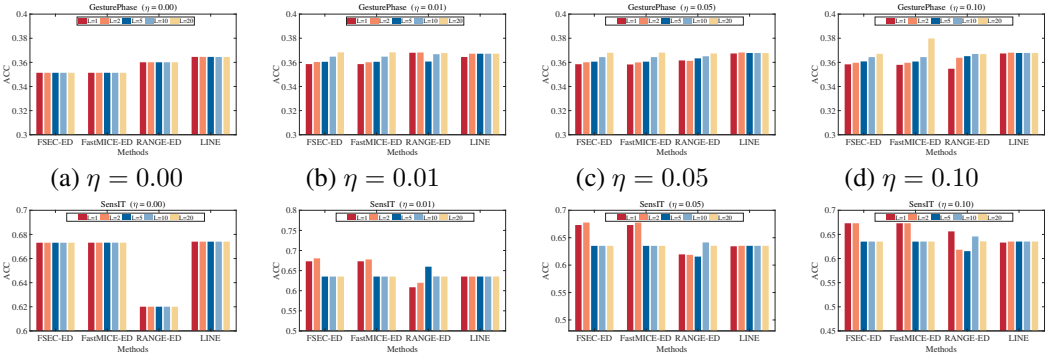


(a) $\eta = 0.00$ (b) $\eta = 0.01$ (c) $\eta = 0.05$ (d) $\eta = 0.10$

(e) $\eta = 0.00$ (f) $\eta = 0.01$ (g) $\eta = 0.05$ (h) $\eta = 0.10$

(i) $\eta = 0.00$ (j) $\eta = 0.01$ (k) $\eta = 0.05$ (l) $\eta = 0.10$

Figure 6: ACC under different replication multiplicities L and assignment perturbation rates η on GesturePhase, SensIT, and Pendigits.



(a) $\eta = 0.00$ (b) $\eta = 0.01$ (c) $\eta = 0.05$ (d) $\eta = 0.10$

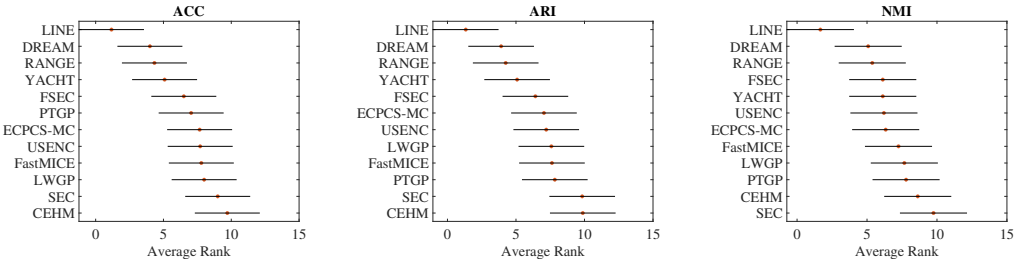
(e) $\eta = 0.00$ (f) $\eta = 0.01$ (g) $\eta = 0.05$ (h) $\eta = 0.10$

Figure 7: ACC under different replication multiplicities L after exact deduplication on GesturePhase and SensIT. ED removes decision-equivalent base partitions before consensus clustering. FSEC+ED, FastMICE+ED, and RANGE+ED are compared with LINE, which does not require explicit duplicate removal.

equivalence class. Consequently, exact multiplicity is neutralized within the consensus model rather than through hard removal. More importantly, when repeated decisions are only approximately equivalent, exact deduplication can no longer collapse them, whereas the redundancy geometry in LINE continues to characterize their dependence. These results show that exact deduplication can handle literal repetitions to some extent, but does not provide the same robustness once redundant decisions become near-repeated rather than exactly identical.

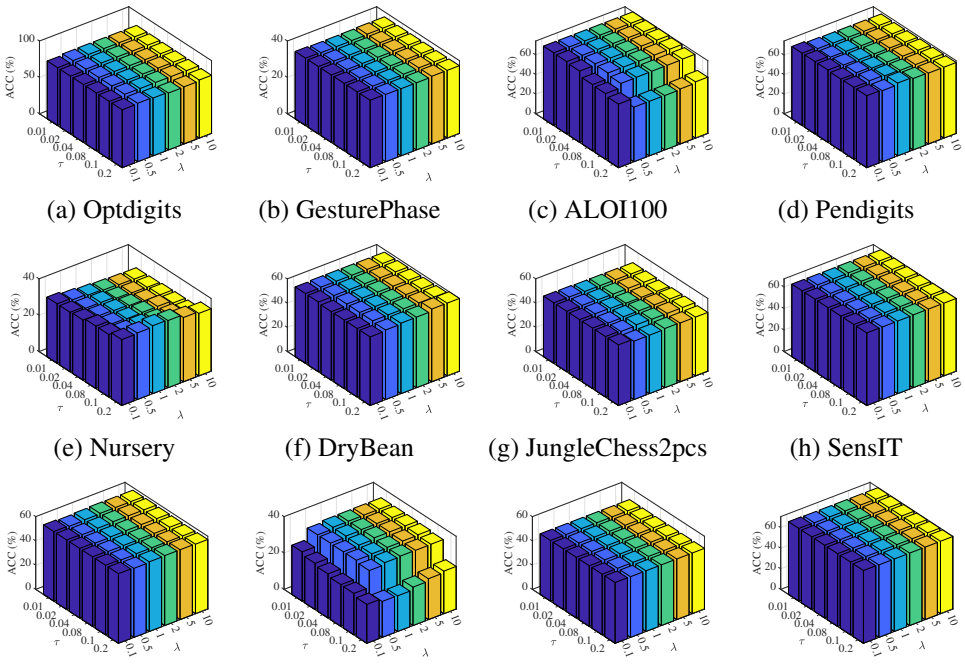
G.2 Statistical Significance Analysis.

To further assess the overall statistical differences among the compared methods, we conduct the Friedman test followed by the Nemenyi post-hoc comparison. The average rank of each method is



(a) ACC (b) ARI (c) NMI

Figure 8: Friedman test with Nemenyi post-hoc comparison. The orange dots denote the average ranks of the compared methods, and the horizontal segments indicate the corresponding critical-distance intervals. Non-overlapping intervals indicate statistically significant differences. Lower ranks indicate better performance.



(a) Optdigits (b) GesturePhase (c) ALOI100 (d) Pendigits

(e) Nursery (f) DryBean (g) JungleChess2pcs (h) SensIT

(i) EMNIST (j) MHEALTH (k) NF-ToN-IoT (l) KDDCup99

Figure 9: ACC w.r.t. τ and λ.

calculated over the evaluated datasets, where a lower rank indicates better overall performance. As shown in Figure 8, LINE achieves the best average rank among all competing methods, demonstrating its consistently favorable performance across different datasets. The critical-distance intervals further provide a statistical comparison between methods, where non-overlapping intervals indicate significant performance differences. These results complement the dataset-wise comparisons and provide additional evidence for the overall effectiveness of LINE.

G.3 Parameter Sensitivity on All Datasets

Figure 9 reports the complete sensitivity results for τ and λ on all twelve datasets. Performance remains relatively stable over broad parameter regions on most datasets, although more noticeable variations occur on ALOI100, Nursery, and MHEALTH. Overall, relatively small τ and moderate λ values tend to yield favorable performance across datasets, with $\tau \in [0.01, 0.04]$ and $\lambda \in [2, 5]$ providing a consistently effective range.