GRANULAR-BALL RELIABILITY LEARNING: ADAPTIVE SPLITTING AND RISK-AWARE CONNECTION FOR DEEP MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep multi-view clustering increasingly learns representations, sample relations, and granular structures to integrate complementary information across views. However, fixing mixed local granular balls entrenches within-granular ball errors, while incorrectly merging pure granular balls amplifies between-granular ball errors. We propose GBASC, which combines Granular-Ball Adaptive Splitting (GBAS) to refine mixed local granular balls using cross-view evidence with Granular-Ball Adaptive Connection (GBAC) to determine global assignments from reliable between-granular ball associations. Our contributions are an exact decomposition of cross-granularity pairwise errors into within-atom mixing and between-atom assignment errors, and a unified framework that jointly controls both transitions through adaptive refinement, risk-aware partition selection, and reliability-weighted representation learning. GBASC achieves state-of-the-art clustering performance on six benchmark datasets.

1 INTRODUCTION

Multi-view clustering (MVC) discovers latent categories in unlabeled data by integrating shared and complementary information (Kumar et al., 2011; Cao et al., 2015) across multiple views of the same objects (Xu et al., 2021b; Cui et al., 2023a; Tan et al., 2023; Pan & Kang, 2021; Liu et al., 2022). To capture nonlinear structures that are difficult to reveal in the original features, deep MVC uses deep networks (Zhang et al., 2019; Zhou & Shen, 2020; Wang et al., 2025a; Yuan et al., 2025) to learn view representations better suited for clustering (Xu et al., 2021a; 2023; Wu et al., 2024b; Ke et al., 2024; Yan et al., 2024; Huang et al., 2023a; 2024). Deep MVC methods optimize clustering representations through sample-relation alignment (Peng et al., 2026), granular-ball-guided learning (Su et al., 2025), and granular ball level cross-view associations, as exemplified by GFSAF (Liu et al., 2026). However, improvements in representation learning and relational consistency do not guarantee the class purity of local units or the reliability of granular ball level associations. Figure 1 and table 1 illustrates two distinct structural risks: mixed atoms can retain erroneous co-clustered pairs under hard assignment constraints, whereas incorrect links between class-pure granular balls can expose multiple cross-class sample pairs. These observations motivate separately controlling local granular ball purity and the reliability of inter granular ball associations.

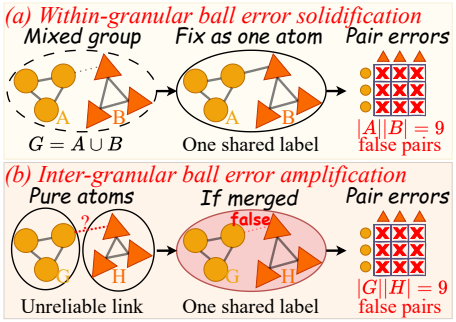


Figure 1: (a) mixed-ball fixing solidifies $|A||B|$ false pairs; (b) unreliable pure-atom link yields $|G||H|$ false pairs.

Table 1: **ORL structural audit.** (a) EBMGC-GNF: N counts mixed atoms; P counts erroneous co-clustered pairs (initial→final). (b) GFSAF: N counts erroneous cross-class links between pure granular balls; P cross-class pair exposure.

Case	N	P
(a) E-GNF (2024a)	3	$4 \rightarrow 4$
(b) GFSAF (2026)	112.7 ± 16.5	468.7 ± 122.4

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

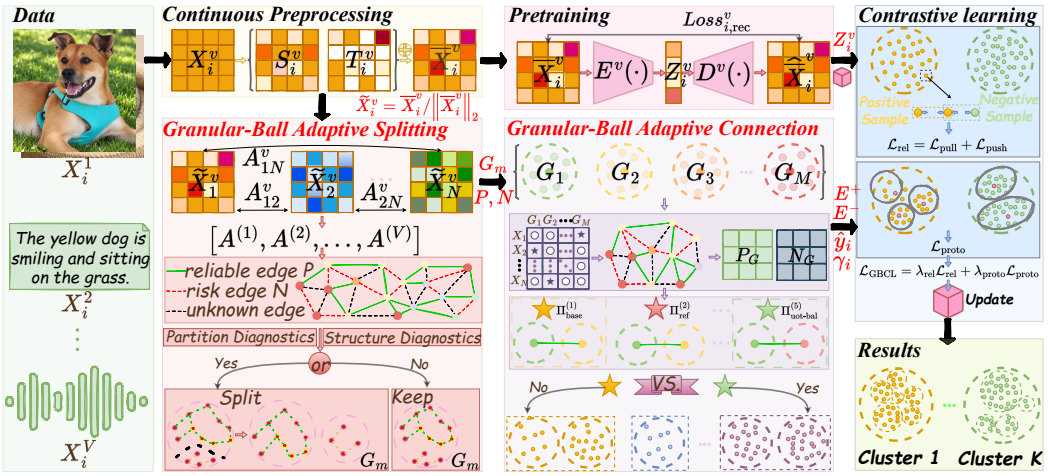


Figure 2: Overview of GBASC for reliable granularity conversion in deep multi-view clustering. GBAS refines local groups before they are fixed as atoms. GBAC then constructs a risk-aware quotient graph and determines their global assignments. Reliability-weighted structural relations guide encoder training, after which the structure is re-estimated.

We term this phenomenon "cross-granularity structural error propagation." As shown in Figure 1, an erroneous positive relation connecting class-pure granular balls G and H from different classes can cover $|G||H|$ cross-class sample pairs, but this does not imply that these pairs will necessarily belong to the same final cluster. When local granular balls remain indivisible during global assignment, two specific risks arise. **(a) Within-granular ball error entrenchment.** If A and B in an atom $G = A \cup B$ belong to different ground-truth classes, assigning this atom to any cluster produces $|A||B|$ incorrect co-clustered pairs. **(b) Between granular ball error amplification.** If class-pure atoms G and H from different classes are incorrectly assigned to the same final cluster, $|G||H|$ incorrect co-clustered pairs are produced. If this merger is irreversible, these errors cannot be eliminated in subsequent steps. Therefore, for a fixed atomic partition, within-atom mixing constitutes a lower bound on the final pairwise error, whereas between-atom assignments determine the remaining incorrect merges and separations.

To control local mixing and unreliable granular ball level associations separately, we propose *Granular-Ball Adaptive Splitting and Connection (GBASC)*, as shown in Figure 2, a deep multi-view clustering method based on granular-ball structure learning. Its core idea is to treat local granular-ball formation and global cluster assignment as two decisions requiring independent reliability assessments, and to revise intermediate granular balls through alternating structure re-estimation and representation learning. *Granular-Ball Adaptive Splitting (GBAS)* adaptively refines local granular balls using cross-view structural evidence to reduce class mixing within atoms. *Granular-Ball Adaptive Connection (GBAC)* performs global grouping according to the reliability of between-atom associations, connecting compatible local structures while suppressing incorrect merges.

Our analysis shows that atom splitting improves within-atom purity, while risk calibration suppresses cross-class connections under certain conditions. Across six benchmark tasks, ablation experiments and structural audits show that the modules address complementary sources of error and improve clustering performance on multiple datasets.

Coarse-grained clustering requires distinguishing errors caused by within-atom mixing from those caused by between-atom assignments, while separately controlling the reliability of local unit formation and global cluster assignment. To address these two problems, our contributions are twofold:

- We systematically model cross-granularity structural error propagation by exactly decomposing the final pairwise error into within-atom mixing error and between-atom assignment error, distinguishing error entrenchment within fixed granular balls from error amplification caused by incorrect granular ball mergers.

- We propose GBASC, which combines cross-view-guided atom-level refinement, risk-aware complete partition selection, and reliability-weighted representation learning to jointly control local unit formation and global cluster assignment.

2 Related Work

Deep multi-view clustering learns representations (Lin et al., 2021; Tang & Liu, 2022b; Wen et al., 2024; Ding et al., 2025) and structures (Wang et al., 2025b; Gu & Wang, 2025; Chen et al., 2025; Zhao et al., 2025) to uncover clusters (Ren et al., 2024; Cui et al., 2023b; Ren et al., 2025; Ling et al., 2023). We group methods by their primary focus: representation and sample-relation learning (Federici et al., 2020; Trosten et al., 2021; 2023; Xiao et al., 2025; Huang et al., 2020), or hierarchical and multi-granularity learning (Manduchi et al., 2023). **Representation and sample-relation learning.** MFLVC (Xu et al., 2022) separates reconstruction and cross-view consistency across feature levels, whereas CVCL (Chen et al., 2023) contrasts cluster assignments. GCFAggMVC (Yan et al., 2023) and RAV (Peng et al., 2026) model cross-sample or cross-view relations. Robust extensions address noisy views (Tang & Liu, 2022a), anomalous samples (Huang et al., 2023b), and corrupted correspondences (Xu et al., 2024; Yang et al., 2025b; Guo et al., 2024; Yang et al., 2021; Sun et al., 2025). These advances improve clustering representations, relations, and supervision (Dong et al., 2025; Wang et al., 2024; Dai et al., 2025). **Hierarchical and multi-granularity learning.** MHCN (Lin et al., 2023) learns hierarchical multi-view representations, whereas MGE (Yang et al., 2025a) combines constrained hierarchical partitions across views and granularities. GAPS (Yang et al., 2026) improves pseudo-supervision through view-quality evaluation and supervisory-granularity selection. MGBCC (Su et al., 2025) associates granular balls through within-view overlap and cross-view membership intersections to guide representation learning, then clusters fused sample embeddings. GFSAF (Liu et al., 2026) uses granular-ball representations for multi-view fusion; GBCT (Xia et al., 2025), a general clustering method, already incorporates adaptive granular-ball refinement and aggregation. These advances alone do not resolve two distinct structural risks. Mixed granular balls retained as indivisible atoms constrain subsequent assignments, while unreliable inter-ball relations can amplify errors across sample pairs through incorrect merges. GBASC links these risks to separate pairwise-error terms, addressing reliability at sample-to-atom and atom-to-cluster transitions while complementing representation, relation, and supervisory-granularity improvements.

3 Methodology

GBASC separates two structural decisions: which samples may remain indivisible, and which resulting atoms should share a cluster. We first derive their distinct error contributions, then introduce label-free criteria for splitting and connection. The selected structure supervises one guided training round before re-estimation (Figure 2). Each indivisible and finely divided granular ball will be referred to as an atom. Appendices B, C, and D specify the implementation.

3.1 Problem Setup and Structural Error

Let $\mathcal{X} = \{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$ contain N objects observed through $V \geq 2$ views, with known cluster count K. An atomic partition $\mathcal{G} = \{G_m\}_{m=1}^{M}$ restricts every member of G_m to the same predicted label c_m. To expose the cost of this restriction, let $D(A)$ count different-class pairs within A. Across disjoint sets, $D(A, B)$ and $S(A, B)$ count different- and same-class pairs, respectively. Therefore, the errors are divided into the internal atomic error $\mathcal{E}_A(\mathcal{G})$ and the inter-atomic error $\mathcal{E}_C(\mathbf{c}; \mathcal{G})$, and the inter-atomic error is further divided into the error due to error merging and the error due to error classification. The total pairwise error is

$$\mathcal{E}_{\text{pair}}(\boldsymbol{c}; \mathcal{G}) = \underbrace{\sum_{m} D(G_m)}_{\mathcal{E}_A(\mathcal{G})} + \underbrace{\sum_{m<n} \left[\mathbb{I}[c_m = c_n] D(G_m, G_n) + \mathbb{I}[c_m \neq c_n] S(G_m, G_n)\right]}_{\mathcal{E}_C(\boldsymbol{c}; \mathcal{G})}. \quad (1)$$

$\mathcal{E}_A(\mathcal{G})$ represents mixed categories within an atom, which requires atom splitting for correction. $\mathcal{E}_C(\mathbf{c}; \mathcal{G})$ denotes incorrect atom connections and must be addressed by identifying proper atom

merging. Since true labels are unavailable, the algorithm relies on sample distances, graph structures, and multi-view consistency as approximation metrics to assess connection reliability.

For these approximation metrics, let $\boldsymbol{h}_i^{(v)}$ denote a normalized encoder embedding. Self-tuning neighbor graphs (Zelnik-Manor & Perona, 2004) on each view and their concatenation give $\boldsymbol{A}^{(v)}$ and a positive connection proposal graph $\boldsymbol{P}$. A candidate edge generated from the fused representation may lack sufficient support from individual views. We therefore measure both how often individual views support it and how much their affinities vary. Let $\mathcal{O}_{ij}$ index views observing both endpoints, $c_{ij} = |\mathcal{O}_{ij}|$, and $\bar{A}_{ij} = \sum_{v\in\mathcal{O}_{ij}} A_{ij}^{(v)} / \max(c_{ij}, 1)$. Absent neighbor edges in observed views contribute zero:

$$
\begin{aligned}
s_{ij} &= \frac{\sum_{v\in\mathcal{O}_{ij}} \mathbb{I}[A_{ij}^{(v)} > 0]}{\max(c_{ij}, 1)}, & u_{ij} &= \begin{cases} 1, & c_{ij} = 0, \\ \mathrm{clip}((1 - s_{ij} + \nu_{ij})/2, 0, 1), & c_{ij} > 0, \end{cases} \\
\nu_{ij} &= \left[\frac{\sum_{v\in\mathcal{O}_{ij}} (A_{ij}^{(v)} - \bar{A}_{ij})^2}{\max(c_{ij}, 1)} \right]^{1/2}, & R_{ij} &= P_{ij} u_{ij}.
\end{aligned}
\tag{2}
$$

Low support or high dispersion increases uncertainty, which is weighted by P_{ij} to reflect its impact on likely positive pairs. When $P_{ij} = 0$, no penalty is imposed, and the pair remains unknown rather than being treated as negative or unconnectable.

3.2 Granular-Ball Adaptive Splitting

Splitting a large grain into individual atomic grains composed of single samples eliminates $\mathcal{E}_A$, but at the same time loses the grain structure. Therefore, GBAS separates the process of boundary selection from the decision on whether to accept the boundary and actually carry out the split. Ward clustering followed by hard size projection first produces M_0 size-bounded roots. For each eligible granular ball G, fused- and single-view Ward bisections, together with feasible mutual-graph spectral bisections (Shi & Malik, 1997), form a candidate pool $\Pi(G)$. Compactness alone can reward a geometric split without a supported boundary. For $\pi = \{G_L, G_R\}$, we combine compactness improvement $\bar{m}$ with neighborhood fracture $\bar{f}$ and agreement with view-specific splits $\bar{a}$. Perturbation stability $\bar{t}$ discounts fragile evidence, while child balance discourages isolating tiny granular balls:

$$
e_G(\pi) = \frac{4|G_L||G_R|}{|G|^2} (\omega_m \bar{m} + \omega_f \bar{f} + \omega_a \bar{a})[\tau_t + (1 - \tau_t)\bar{t}], \qquad \pi_G = \min_{\text{key}} \arg \max_{\pi\in\Pi(G)} e_G(\pi). \tag{3}
$$

The four scores lie in $[0, 1]$ and average over eligible observed views; $\omega_m + \omega_f + \omega_a = 1$ with nonnegative weights and $\tau_t \in [0, 1]$. Empty eligible sets give zero evidence. Canonical keys resolve ties, and only the highest-evidence candidate π_G proceeds to the local acceptance test.

A strong internal boundary makes keeping G intact less desirable, but evidence ranking does not penalize the number of resulting atoms. Set $C(G) = \binom{|G|}{2}$ and $Q_R = \max\{1, C(R)\}$ for its root R. The risk proxy $\widehat{r}(G)$ is the best split evidence, or $1 - Q(G)$ when no candidate exists. $Q(G) \in [0, 1]$ summarizes local stability. We compare keeping G with its children's frozen continuation costs:

$$
\phi_R(G) = C(G)\widehat{r}(G) + \lambda_A Q_R + \xi_A Q_R \mathbb{I}[|G| = 1], \qquad g_B(G, \pi) = \phi_R(G) - \sum_{H\in\pi} V_B(H). \tag{4}
$$

Here $C(G)$ scales risk by pair exposure, while $\lambda_A, \xi_A \geq 0$ charge for atoms and singletons. A label-free pilot computes $0 \leq V_B(G) \leq \phi_R(G)$ on the finite candidate tree. Freezing these values keeps subsequent decisions relative to the same reference.

The local cost can reject a boundary consistent with the global graph when its atom penalty outweighs local evidence. GBAS permits a second acceptance route, gated by agreement $a_G(\pi) \in [0, 1]$ with a frozen K-way spectral guide (Ng et al., 2001) $\boldsymbol{c}^{\mathrm{S},K}$. Its structural gain balances child separation against cutting supported edges, sensitivity to view deletion, and added complexity:

$$
\begin{aligned}
g_S(G, \pi) &= \left(1 + \frac{\chi}{\rho}\right) b_G(\pi) - c_P(G, \pi) - \lambda_S \rho\big[(1 - s_{\mathrm{d}}(G, \pi)) + \kappa_G(\pi)\big], \\
\Delta_A(G) &= \max\{g_B(G, \pi_G), a_G(\pi_G) g_S(G, \pi_G)\}.
\end{aligned}
\tag{5}
$$

Here b_G rewards balanced children, c_P measures cut positive mass, s_{d} measures view-deletion agreement, and κ_G penalizes complexity. The pilot ratio $\rho = M_{\mathrm{pilot}}/K$ moderates further fragmentation. χ increases split pressure when a spectral probe scores better than a Ward probe. Appendix C.1 defines these label-free quantities and the nonnegative coefficient λ_S.

Accept $G \to \pi_G$ when $\Delta_A(G) > \epsilon_A$, for tolerance $\epsilon_A > 0$. The maximum implements alternative justifications, so a structural override need not decrease local cost. If the gated structural gain strictly exceeds g_B, children become terminal locally; otherwise, recursion continues. Since the candidate pool need not contain the guide boundary, a final single pass intersects each terminal granular ball with the guide's clusters. Only proper refinements with $g_S > \epsilon_A$ are accepted. It does not repeat the local-cost test. Agreement gating is redundant because the intersection follows the guide exactly.

Theorem 1 (Atomic refinement) *Every accepted proper refinement $G \to \{H_r\}_{r=1}^{q}$ satisfies*

$$\mathcal{E}_A(\mathcal{G}_{\mathrm{before}}) - \mathcal{E}_A(\mathcal{G}_{\mathrm{after}}) = \sum_{r<t} D(H_r, H_t) \geq 0, \qquad T_{\mathrm{split}} \leq N - M_0. \tag{6}$$

This guarantee concerns refinement itself; assessing the chosen boundaries additionally requires evidence calibration. Appendix C.1 gives stronger actionwise bounds under an explicit evidence-calibration condition; Appendix F.3 gives the proof. Reduced internal mixing neither guarantees pure atoms nor resolves the remaining assignment error.

3.3 Granular-Ball Adaptive Connection

GBAC assigns whole atoms, so its graph must preserve their aggregated sample relations. For final atoms $\mathcal{G}_T = \{G_g\}_{g=1}^{M}$, let $S_{ig} = \mathbb{I}[i \in G_g]$ and $n_g = |G_g|$. We aggregate positive and risk mass separately, then attenuate links according to their risk fraction:

$$\boldsymbol{P}^Q = \boldsymbol{S}^\top \boldsymbol{P} \boldsymbol{S}, \qquad \boldsymbol{R}^Q = \boldsymbol{S}^\top \boldsymbol{R} \boldsymbol{S}, \qquad \widetilde{u}^Q_{gh} = \frac{R^Q_{gh}}{\max(P^Q_{gh}, \epsilon)}, \qquad A^Q_{gh} = P^Q_{gh} e^{-\beta \widetilde{u}^Q_{gh}}. \tag{7}$$

Positive-mass normalization makes risk quality reflect relative connection risk rather than overall quality. $\beta > 0$ controls attenuation and $\epsilon > 0$ prevents division by zero. Before attenuation, retaining diagonals of $\boldsymbol{P}^Q$ preserves sample cuts and volumes for atom-constant assignments (Appendix F.2). Repeating construction after each view deletion yields $\boldsymbol{A}^{Q,-v}$ for sensitivity assessment.

Size-weighted spectral clustering on $\boldsymbol{A}^Q$ gives a base assignment $\boldsymbol{c}_0$. Risk attenuation alone does not resolve the choice of partition. To compare sample-level structure and alternative mass allocations, we project sample spectral and sample-balanced partitions onto atoms by majority vote. Unbalanced optimal transport (UOT) refines each projection toward a mass target interpolating balanced and view-stable masses, with weight $\eta \in [0, 1]$. Thus the five slots $\boldsymbol{c}^{[0:4]}$ comprise the base, two projections, and their two UOT refinements. Projection fallback and empty-cluster safeguards preserve K nonempty clusters on the feasibility event $\mathcal{F}_C$ of finite embeddings, defined scores, and successful nonempty initializations (Appendix C.2).

These alternatives are compared by one score, rather than by their generating criteria. With $\boldsymbol{A}^{Q,-0} = \boldsymbol{A}^Q$, define

$$\begin{aligned} \mathcal{J}_C(\boldsymbol{c}) = \underbrace{\frac{1}{V+1} \sum_{v=0}^{V} \mathrm{NCut}_{\boldsymbol{A}^{Q,-v}}(\boldsymbol{c})}_{\mathrm{Cut}_Q(\boldsymbol{c})} + \mathrm{MassRisk}(\boldsymbol{c}) + \lambda_{\mathrm{stab}}[1 - \mathrm{Stab}(\boldsymbol{c})] \\ + (1-\eta)\,\mathrm{Move}(\boldsymbol{c}, \boldsymbol{c}_0). \end{aligned} \tag{8}$$

The cut term discourages breaking calibrated links across view deletions. $\mathrm{MassRisk}$ combines imbalance and deviation from the interpolated target; Stab measures agreement with view-deletion bases. The movement penalty limits reassignment relative to $\boldsymbol{c}_0$, with less resistance when η favors balancing. These terms are label-free proxies, with exact definitions in Appendix C.2.

A weighted sum can hide deterioration in one criterion behind improvement in another. We therefore check explicit trust constraints before replacing the base. Let $\boldsymbol{c}_{\mathrm{p}}$ be the lowest-scoring nonbase slot,

breaking ties by slot order, and set $b_C = 1 - s_C$, where $s_C = V^{-1} \sum_v \mathrm{clip}(\mathrm{ARI}(\boldsymbol{c}_0, \boldsymbol{c}_0^{-v}), 0, 1)$. Trust requires cut deterioration at most $b_C \max(\mathrm{Cut}_Q(\boldsymbol{c}_0), \epsilon) + \epsilon_C$, stability loss at most $b_C + \epsilon_C$, and movement at most $b_C + \epsilon_C$:

$$\widehat{\boldsymbol{c}} = \begin{cases} \boldsymbol{c}_{\mathrm{p}}, & \boldsymbol{c}_{\mathrm{p}} \neq \boldsymbol{c}_0, \quad \mathrm{Trust}(\boldsymbol{c}_{\mathrm{p}}) = 1, \quad \mathcal{J}_C(\boldsymbol{c}_{\mathrm{p}}) < \mathcal{J}_C(\boldsymbol{c}_0) - \epsilon_C, \\ \boldsymbol{c}_0, & \text{otherwise.} \end{cases} \tag{9}$$

Higher base stability tightens the admissible change. This conservative rule tests only the best-scoring alternative; failure returns the base without searching lower-ranked slots, so it is not constrained minimization over all trusted candidates.

Theorem 2 (risk-aware connection) *Each positive quotient edge satisfies $0 < A^Q_{gh} \leq P^Q_{gh}$ and $\partial A^Q_{gh} / \partial R^Q_{gh} < 0$. Suppose atoms are pure, positive quotient masses are at least ϵ, and contributing sample edges are co-observed. Under support–dispersion separation assumptions in Theorem 4 (Appendix C.2), cross-class uncertainty exceeds within-class uncertainty by a gap $\delta_u > 0$. For positive edges $e = (g, h)$ and fixed impact weights $h_e \geq 0$ with a positive within-class denominator,*

$$\frac{\sum_{e: y_g \neq y_h} h_e A^Q_e}{\sum_{e: y_g = y_h} h_e A^Q_e} \leq e^{-\beta \delta_u} \frac{\sum_{e: y_g \neq y_h} h_e P^Q_e}{\sum_{e: y_g = y_h} h_e P^Q_e}. \tag{10}$$

Here y_g is the class of pure atom G_g. Separately, on $\mathcal{F}_C$, selection satisfies $\mathcal{J}_C(\widehat{\boldsymbol{c}}) \leq \mathcal{J}_C(\boldsymbol{c}_0)$.

Theorem 2 characterizes GBAC from both edge calibration and candidate selection: risk calibration monotonically attenuates high-risk edges and, under the support–dispersion separation condition, further suppresses cross-class connections by a factor of $e^{-\beta \delta_u}$, while the conservative selection rule ensures that the final surrogate objective is no worse than that of the base assignment. These guarantees concern the connection and selection mechanisms under the stated assumptions, whereas the final semantic error $\mathcal{E}_C$ is assessed separately by the semantic audits in Appendix E.

3.4 Structural Feedback and Training

The structure depends on the representation, but treating every pseudo-label as reliable can reinforce structural errors during training. We therefore weight supervision on existing proposal edges. Let $g(i)$ index the atom of sample i, $\widehat{c}_i = \widehat{c}_{g(i)}$, and $q_i \in [0, 1]$ be its inherited atom confidence (Appendix C.2). Same- and different-label proposal edges form $\mathcal{E}^+$ and $\mathcal{E}^-$, respectively. Define

$$\begin{aligned} r_{ij} &= e^{-\beta R_{ij} / \max(P_{ij}, \epsilon)}, & q_{ij} &= \sqrt{q_i q_j}, \\ w^+_{ij} &= P_{ij} q_{ij} r_{ij} \begin{cases} 1, & g(i) = g(j), \\ \gamma_+, & g(i) \neq g(j), \end{cases} & w^-_{ij} &= P_{ij} q_{ij} s_C [\gamma_- + (1 - \gamma_-)(1 - r_{ij})]. \end{aligned} \tag{11}$$

Here $0 < \gamma_+ \leq 1$ discounts attraction across atoms because it additionally depends on connection decisions. Negative weights require pseudo-label disagreement: $1 - r_{ij}$ emphasizes questionable positive links within this set, rather than creating cannot-links from uncertainty alone; $\gamma_- \in [0, 1]$ supplies baseline repulsion.

For fused and view embeddings ($r = 0, \ldots, V$), let $d_r(i, j) = 1 - \cos(\boldsymbol{z}^{(r)}_i, \boldsymbol{z}^{(r)}_j)$. Retained observed neighbors $\mathcal{N}^{\pm}_{ir}$ provide attraction and margin-limited repulsion (Hadsell et al., 2006). To limit degree effects, normalize each sign separately:

$$\mathcal{L}_{\mathrm{rel}} = \mathrm{Avg}_{r \in \mathcal{R}_{\mathrm{act}}} \mathrm{Avg}_{i \in \mathcal{I}^{\mathrm{act}}_r} \left[\frac{\sum_{j \in \mathcal{N}^+_{ir}} w^+_{ij} d_r(i, j)}{\max(\sum_{j \in \mathcal{N}^+_{ir}} w^+_{ij}, \varepsilon)} + \frac{\sum_{j \in \mathcal{N}^-_{ir}} w^-_{ij} [m - d_r(i, j)]_+}{\max(\sum_{j \in \mathcal{N}^-_{ir}} w^-_{ij}, \varepsilon)} \right]. \tag{12}$$

Only positive-weight anchors and nonempty representations enter the averages; empty averages are zero, $m > 0$, and $\varepsilon > 0$ is a numerical floor. Weights primarily control relative neighbor contributions: away from the floor, the common factor $s_C > 0$ cancels in normalized repulsion.

Local relations do not by themselves specify global cluster targets or preserve input information. We add fixed-prototype label supervision for global targets, reconstruction for information preservation, and cross-view alignment with a variance term discouraging collapse (Bardes et al., 2022):

$$\mathcal{L} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{cv}} \mathcal{L}_{\mathrm{cv}} + \lambda_{\mathrm{rel}} \mathcal{L}_{\mathrm{rel}} + \lambda_{\mathrm{proto}} \mathcal{L}_{\mathrm{proto}}. \tag{13}$$

The nonnegative coefficients balance these roles; Appendix D defines the losses. After guided training, graphs, atoms, and assignments are rebuilt from learned embeddings. The structural guarantees apply to each fixed call and do not assert joint neural–structural convergence.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

We evaluate eight datasets with sample/view/class counts of UCI (2,000/3/10), UCI-Digits (2,000/6/10), DHA (483/2/23), ORL (400/3/40), Mfeat (2,000/6/10), NGs (500/3/5), ALOI (10,800/4/100), and BBCSport (544/2/5). We report clustering accuracy (ACC), normalized mutual information (NMI), and adjusted Rand index (ARI), using Hungarian alignment for ACC. Labels serve only post-hoc evaluation and visualization, never training or model selection. Table 2 reports the mean and standard deviation of the results from ten runs of all the algorithms. Without any special indication, all the runs were conducted under this set of conditions. Each encoder uses widths $d_v \to 512 \to 256 \to 128$, layer normalization, and dropout 0.1; positive graphs use 15 neighbors. Initial granular balls use over-clustering factor 1.5, maximum size 32, minimum split size 4, and maximum split depth 4. Adam (Kingma & Ba, 2015) uses batch size 256, learning rate 10^{-3}, and weight decay 10^{-5}. Ten warm-up epochs precede one structural refinement round and 20 guided epochs. Appendices B and D provide the remaining implementation details. In the supplementary materials, we have also provided 6 datasets for your testing to verify the accuracy of our results.

4.2 COMPARISON WITH BASELINES

Table 2 compares GBASC with nine baselines on six datasets: CVCL (Chen et al., 2023), CSOT (Zhang et al., 2024), SCM (Luo et al., 2024), SCMVC (Wu et al., 2024b), GFSAF (Liu et al., 2026), MGBCC (Su et al., 2025), BONE (Xu et al., 2026), CorreGen (Zhou et al., 2026), and RAV (Peng et al., 2026). GBASC achieves the highest mean ACC, NMI, and ARI in all 18 dataset–metric comparisons. Gains over the strongest baseline range from 3.4–13.5 ACC, 4.5–20.2 NMI, and 6.8–21.3 ARI percentage points, calculated from the displayed means. On BBCSport, ACC/ARI increase from MGBCC's 92.8/81.1 to 97.7/93.7; on ORL, they increase from CorreGen's 69.7/58.5 to 83.2/75.6. Positive margins on Mfeat and UCI extend this advantage across different clustering difficulties. The following ablation separates the contributions of local refinement and global connection calibration to this overall improvement. For more results, please refer to Appendix H.

4.3 ABLATION OF SPLITTING AND CONNECTION

Table 3 isolates GBAS (Adaptive versus Ward-M) and GBAC (Risk versus P-Connection) through a 2×2 factorial comparison. With P-Connection fixed, adaptive splitting improves DHA ACC/ARI by 7.50/9.75 points and BBCSport ACC/NMI/ARI by 3.10/8.10/9.55 points. These gains support GBAS's role in refining local units before their members receive shared assignments. With adaptive splitting fixed, risk calibration raises ALOI ACC/ARI from 90.60/87.25 to 96.85/93.85, gains of 6.25/6.60 points. GBAC therefore remains useful after local refinement, supporting separate control of inter-atom assignments. The full model leads on all twelve dataset–metric pairs, exceeding the strongest partial variant by 0.37–0.75 ACC points. Together, these results support complementary, dataset-dependent benefits from controlling both granularity transitions.

4.4 EVIDENCE FOR RELIABLE GRANULARITY CONVERSION

Figures 3–6 connect the final clustering results to the two structural controls. Figure 3 shows tighter fused-feature groups on DHA, whereas ALOI retains overlap in its two-dimensional projection. Figure 4 shows strongly diagonal assignments on UCI, Mfeat, UCI-Digits, and BBCSport, corroborating the quantitative results. Figure 5 illustrates local refinement from 35 granular balls to 105 atoms, followed by assignment to 23 groups. Retaining the base partition illustrates GBAC's fallback when no candidate regrouping is accepted. Figure 6(a, b) tests GBAS on DHA: adaptive boundaries release more cross-class pairs than geometric or random controls at comparable same-class separation. Additional releases concentrate in a few initial parents, supporting targeted removal of within-atom mixing. Figure 6(c, d) tests GBAC on ORL: full-risk ordering preferentially removes cross-class

Table 2: Comparison with nine multi-view clustering methods on six benchmarks. Entries are percentages reported as mean ± standard deviation over ten runs; higher is better. The best and second-best mean performances are boldfaced and underlined, respectively.

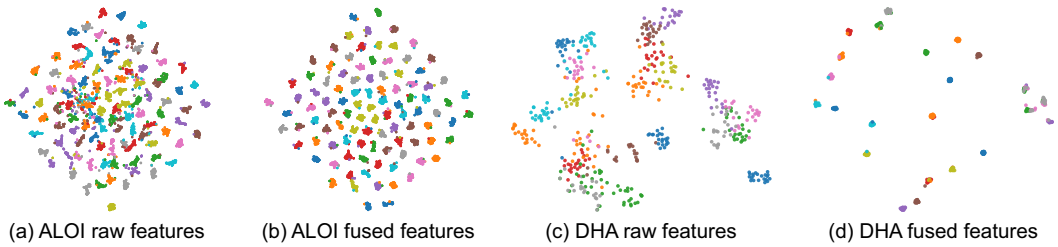
Method	BBCSport			DHA			Mfeat		
	ACC ↑	NMI ↑	ARI ↑	ACC ↑	NMI ↑	ARI ↑	ACC ↑	NMI ↑	ARI ↑
CVCL (2023)	35.0±3.4	4.3±5.5	1.4±4.3	68.0±3.5	75.5±2.1	53.3±3.9	74.2±5.8	71.1±4.7	62.3±6.9
CSOT (2024)	33.6±2.0	5.2±2.0	3.3±1.7	23.7±3.1	54.9±2.7	21.6±2.7	83.1±4.1	82.1±3.4	75.6±4.8
SCM (2024)	64.0±7.5	43.2±7.3	37.5±8.5	80.6±1.4	83.1±1.0	71.4±1.4	79.5±8.6	74.1±4.9	66.9±8.4
SCMVC (2024b)	58.8±9.3	36.8±8.1	30.7±9.7	68.6±4.1	80.7±1.8	57.1±4.0	86.7±0.5	86.7±1.3	81.9±1.2
MGBCC (2025)	92.8±2.3	85.2±3.9	81.1±4.7	57.8±13.2	78.0±2.6	48.3±11.7	58.0±0.4	67.2±0.4	54.1±0.6
GFSAF (2026)	45.1±2.5	24.6±3.4	13.1±1.2	65.3±4.4	75.0±1.3	49.6±5.1	67.5±13.0	70.1±14.1	59.1±18.3
BONE (2026)	66.4±5.0	56.9±9.9	48.7±9.8	68.2±3.4	79.5±1.6	58.9±3.4	88.2±1.2	87.5±0.8	83.3±1.0
CorreGen (2026)	27.4±1.9	3.4±1.3	1.7±1.0	77.4±2.8	83.6±0.9	69.0±2.1	86.6±7.1	84.6±3.8	77.9±8.2
RAV (2026)	55.6±6.3	32.7±5.3	27.7±5.4	59.9±4.0	71.2±2.3	45.9±4.4	92.9±3.2	89.6±2.1	87.3±3.7
GBASC	**97.7±0.1**	**93.3±0.2**	**93.7±0.1**	**86.4±0.2**	**88.9±0.2**	**79.4±0.3**	**97.6±0.0**	**94.6±0.1**	**94.7±0.1**

Method	NGs			ORL			UCI		
	ACC ↑	NMI ↑	ARI ↑	ACC ↑	NMI ↑	ARI ↑	ACC ↑	NMI ↑	ARI ↑
CVCL (2023)	51.2±4.9	37.0±5.3	26.6±5.2	47.8±1.9	70.2±1.2	33.8±2.1	79.1±5.8	77.0±3.7	69.2±6.1
CSOT (2024)	41.4±2.6	29.4±2.8	20.1±2.6	21.5±7.5	45.2±10.5	9.4±3.4	62.1±4.1	69.4±2.4	53.0±3.1
SCM (2024)	89.4±1.8	75.8±2.4	75.5±3.6	48.6±1.4	70.6±1.1	34.2±1.7	92.5±0.8	85.2±1.2	84.1±1.6
SCMVC (2024b)	76.3±8.7	57.3±6.8	55.6±8.9	40.1±3.8	64.1±2.5	25.2±3.7	81.4±0.4	80.2±0.6	72.2±0.7
MGBCC (2025)	46.2±5.4	26.4±5.0	20.4±3.7	41.6±13.6	71.4±2.3	29.4±22.5	57.6±11.5	66.7±2.1	54.1±3.7
GFSAF (2026)	23.1±2.9	5.6±5.6	0.4±0.1	33.3±8.0	58.3±6.3	16.7±7.1	41.3±9.1	51.9±20.8	34.9±13.7
BONE (2026)	69.0±9.3	60.5±10.1	53.5±11.0	59.5±3.0	78.6±1.5	47.9±3.1	80.0±5.9	81.7±2.9	73.5±5.2
CorreGen (2026)	33.0±2.7	8.2±1.4	5.5±1.2	69.7±2.7	84.1±1.6	58.5±2.9	93.9±1.5	89.2±1.8	87.2±2.8
RAV (2026)	73.0±11.1	54.2±12.3	51.3±14.5	55.1±1.4	75.1±0.7	41.4±1.8	82.0±4.8	80.3±3.2	73.0±4.9
GBASC	**98.7±0.1**	**96.0±0.3**	**96.8±0.3**	**83.2±1.1**	**91.0±0.6**	**75.6±1.4**	**97.3±0.1**	**93.7±0.1**	**94.0±0.1**

Table 3: Factorial ablation on four representative datasets. W/A denote Ward-M/Adaptive Split, and P/R denote P-Connection/Risk Connection. Entries are mean ACC, NMI, and ARI (%) over seeds 42–51; the full model is boldfaced. Protocols are frozen within each dataset.

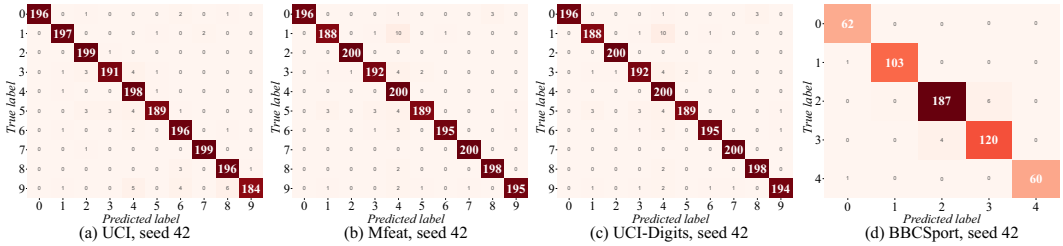
	Components				ORL			DHA			ALOI			BBCSport		
Variant	W	A	P	R	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Ward-M + P	√		√		81.95	89.84	73.58	78.20	84.95	68.90	88.55	95.10	85.30	94.10	84.60	83.40
Adaptive + P		√	√		82.85	90.63	75.02	85.70	88.25	78.65	90.60	95.55	87.25	97.20	92.70	92.95
Ward-M + Risk	√			√	82.70	90.55	74.68	79.10	85.45	69.80	96.10	96.72	93.10	94.85	86.10	85.20
Full Model		√		√	**83.22**	**91.02**	**75.64**	**86.44**	**88.92**	**79.37**	**96.85**	**97.04**	**93.85**	**97.74**	**93.25**	**93.66**

edge mass. Actual-risk attenuation reduces the cross-/same-class mass ratio as β increases, whereas shuffled risk and uniform scaling provide no comparable reduction. These audits ground the two error sources in observable structural effects; the ablations establish the corresponding benefits for final clustering.



(a) ALOI raw features (b) ALOI fused features (c) DHA raw features (d) DHA fused features

Figure 3: Raw and fused t-SNE representations on ALOI and DHA (seed 42). Panels use the same samples but separate embeddings; colors show labels only after training.



(a) UCI, seed 42 (b) Mfeat, seed 42 (c) UCI-Digits, seed 42 (d) BBCSport, seed 42

Figure 4: Confusion matrices on UCI, Mfeat, UCI-Digits, and BBCSport (seed 42). Hungarian alignment is applied only for display.

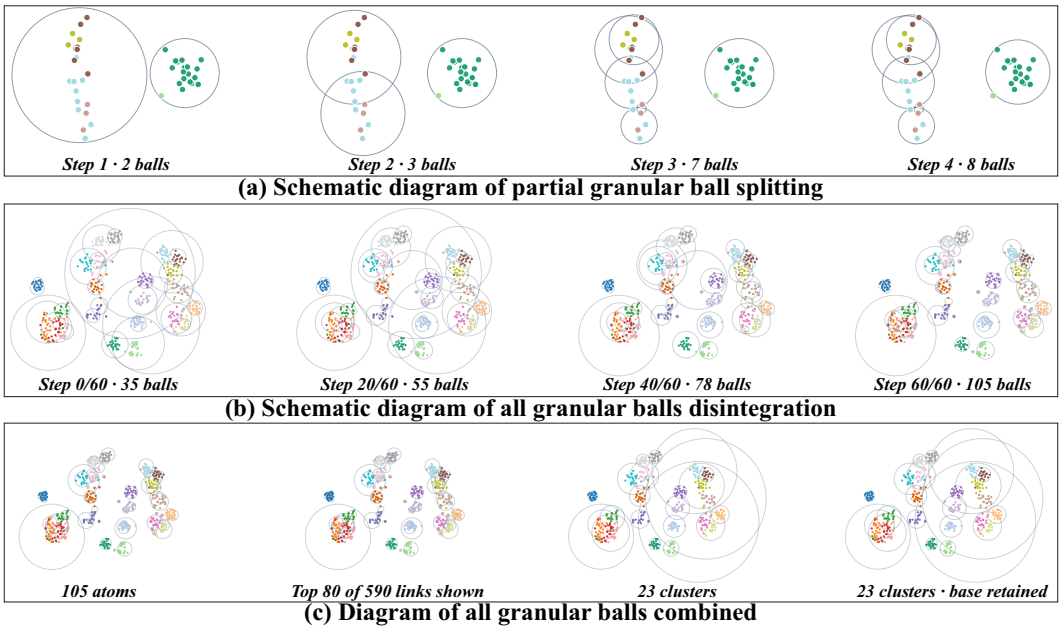


Figure 5: Splitting and connection process: (a) local refinement; (b) 35 initial balls refined into 105 atoms; (c) inter-atom links and 23-group assignment, with the base partition retained.

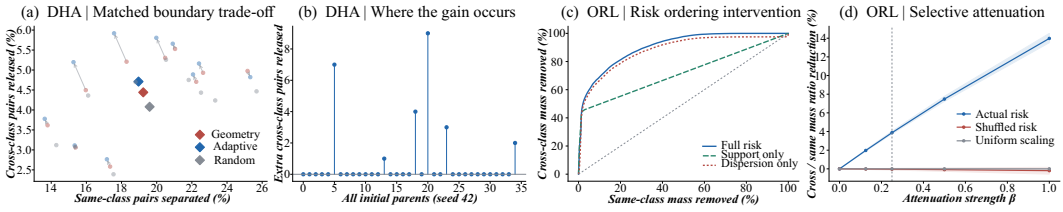


Figure 6: Module diagnostics. DHA: (a) boundary trade-off and (b) parent-wise extra cross-class pair release. ORL: (c) risk-ranked edge-mass removal and (d) selective attenuation against shuffled-risk and uniform-scaling controls.

5 Conclusion

In this work, we proposed GBASC, a granular-ball reliability learning framework for deep multi-view clustering. Multiple experiments, process visualization, and mechanism analysis verified the effectiveness of this method, and demonstrated that the granular splitting and connection mechanism showed more significant advantages on datasets with larger sample sizes. In the future, we will further extend this mechanism to incomplete multi-view (IMVC) clustering to improve the clustering performance and robustness in scenarios with missing views.

AI USE STATEMENT

We used generative AI tools to assist with experimental infrastructure, including generating scripts to run our implementation across multiple datasets and supporting the execution of comparative experiments. The proposed method and its implementation were developed and written by the authors. The manuscript and theoretical proofs were also written by the authors. AI assistance in manuscript preparation was limited to identifying formatting and spelling errors and providing feedback on the correctness of our author-written proofs. AI feedback was used as an aid to checking, not as a substitute for mathematical justification. The authors take responsibility for the final manuscript, theoretical claims, implementation, and reported experimental results.

ETHICS STATEMENT

This work studies multi-view clustering using existing benchmark datasets. Potential ethical considerations include the privacy and permitted use of face and human-action data, biases in benchmark composition, and the risk of interpreting automatically discovered clusters as definitive attributes of individuals. Our benchmark results do not establish the suitability of the method for sensitive or consequential applications involving people. Any reuse of the datasets or deployment of the method should respect applicable data-use conditions and include an assessment of privacy, fairness, and potential misuse.

REPRODUCIBILITY STATEMENT

Section 4.1 describes the experimental setup, evaluation metrics, network architecture, optimization settings, and repeated-run protocol. Appendix A provides the overall algorithm, Appendix B specifies preprocessing and numerical conventions, and Appendices C and D detail the splitting, connection, and representation-learning procedures. Appendix E describes the independent label audits, and Appendix F provides theoretical proofs and additional results. Appendices G and H report complementary mechanism experiments and comparative evaluations. These descriptions specify the procedures and assumptions underlying the reported results and support independent implementation and verification.

REFERENCES

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *ICLR*, pp. 1–23, 2022.

Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *CVPR*, pp. 586–594, 2015.

Jie Chen, Hua Mao, Wai Lok Woo, and Xi Peng. Deep multiview clustering by contrasting cluster assignments. In *ICCV*, pp. 16752–16761, 2023.

Liang Chen, Zhe Xue, Yawen Li, Meiyu Liang, Yan Wang, Anton van den Hengel, and Yuankai Qi. Medusa: A multi-scale high-order contrastive dual-diffusion approach for multi-view clustering. In *CVPR*, pp. 10295–10304, 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 *NeurIPS*, pp. 44847–44859, 2023a.

Chenhang Cui, Yazhou Ren, Jingyu Pu, Xiaorong Pu, and Lifang He. Deep multi-view subspace clustering with anchor graph. In *IJCAI*, pp. 3577–3585, 2023b.

Yuzhuo Dai, Jiaqi Jin, Zhibin Dong, Siwei Wang, Xinwang Liu, En Zhu, Xihong Yang, Xinbiao Gan, and Yu Feng. Imputation-free and alignment-free: Incomplete multi-view clustering driven by consensus semantic learning. In *CVPR*, pp. 5071–5081, 2025.

Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. In *NeurIPS*, pp. 52175–52195, 2025.

Zhibin Dong, Meng Liu, Siwei Wang, Ke Liang, Yi Zhang, Suyuan Liu, Jiaqi Jin, Xinwang Liu, and En Zhu. Enhanced then progressive fusion with view graph for multi-view clustering. In *CVPR*, pp. 15518–15527, 2025.

Marco Federici, Anjan Dutta, Patrick Forré, Nate Kushman, and Zeynep Akata. Learning robust representations via multi-view information bottleneck. In *ICLR*, pp. 1–26, 2020.

Zhibin Gu and Weili Wang. Hypergraph-enhanced contrastive learning for multi-view clustering with hyper-laplacian regularization. In *NeurIPS*, pp. 54600–54620, 2025.

Ruiming Guo, Mouxing Yang, Yijie Lin, Xi Peng, and Peng Hu. Robust contrastive multi-view clustering against dual noisy correspondence. In *NeurIPS*, pp. 121401–121421, 2024.

Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *CVPR*, pp. 1735–1742, 2006.

Haonan Huang, Guoxu Zhou, Yanghang Zheng, Yuning Qiu, Andong Wang, and Qibin Zhao. Adversarially robust deep multi-view clustering: A novel attack and defense framework. In *ICML*, pp. 20526–20558, 2024.

Weitian Huang, Sirui Yang, and Hongmin Cai. Generalized information-theoretic multi-view clustering. In *NeurIPS*, pp. 58752–58764, 2023a.

Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. Partially view-aligned clustering. In *NeurIPS*, pp. 2892–2902, 2020.

Zongmo Huang, Yazhou Ren, Xiaorong Pu, Shudong Huang, Zenglin Xu, and Lifang He. Self-supervised graph attention networks for deep weighted multi-view clustering. In *AAAI*, pp. 7936–7943, 2023b.

Guanzhou Ke, Bo Wang, Xiaoli Wang, and Shengfeng He. Rethinking multi-view representation learning via distilled disentangling. In *CVPR*, pp. 26774–26783, 2024.

Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *ICLR*, pp. 1–13, 2015.

Abhishek Kumar, Piyush Rai, and Hal Daumé, III. Co-regularized multi-view spectral clustering. In *NeurIPS*, pp. 1413–1421, 2011.

Fangfei Lin, Bing Bai, Yiwen Guo, Hao Chen, Yazhou Ren, and Zenglin Xu. MHCN: A hyperbolic neural network model for multi-view hierarchical clustering. In *ICCV*, pp. 16525–16535, 2023.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. COMPLETER: Incomplete multi-view clustering via contrastive prediction. In *CVPR*, pp. 11174–11183, 2021.

Yawen Ling, Jianpeng Chen, Yazhou Ren, Xiaorong Pu, Jie Xu, Xiaofeng Zhu, and Lifang He. Dual label-guided graph refinement for multi-view graph clustering. In *AAAI*, pp. 8791–8798, 2023.

Junjie Liu, Junlong Liu, Shaotian Yan, Rongxin Jiang, Xiang Tian, Boxuan Gu, Yaowu Chen, Chen Shen, and Jianqiang Huang. MPC: Multi-view probabilistic clustering. In *CVPR*, pp. 9509–9518, 2022.

Shuaiyu Liu, Song Wu, Jie Xu, Yazhou Ren, Yang Yang, Xiaorong Pu, and Guoyin Wang. Views attention fusion of granular-ball fuzzy representations split for improved multi-view clustering. In *AAAI*, pp. 23810–23818, 2026.

Caixuan Luo, Jie Xu, Yazhou Ren, Junbo Ma, and Xiaofeng Zhu. Simple contrastive multi-view clustering with data-level fusion. In *IJCAI*, pp. 4697–4705, 2024.

Laura Manduchi, Moritz Vandenhirtz, Alain Ryser, and Julia Vogt. Tree variational autoencoders. In *NeurIPS*, pp. 54952–54986, 2023.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *NeurIPS*, pp. 849–856, 2001.

Erlin Pan and Zhao Kang. Multi-view contrastive graph clustering. In *NeurIPS*, pp. 2148–2159, 2021.

Shuangmei Peng, Zhe Chen, Tianyang Xu, and Xiaojun Wu. Relationship alignment for view-aware multi-view clustering. In *ICLR*, pp. 147335–147347, 2026.

Yazhou Ren, Jingyu Pu, Chenhang Cui, Yan Zheng, Xinyue Chen, Xiaorong Pu, and Lifang He. Dynamic weighted graph fusion for deep multi-view clustering. In *IJCAI*, pp. 4842–4850, 2024.

Yazhou Ren, Junlong Ke, Zichen Wen, Tianyi Wu, Yang Yang, Xiaorong Pu, and Lifang He. Multi-view graph clustering via node-guided contrastive encoding. In *ICML*, pp. 51421–51435, 2025.

Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. In *CVPR*, pp. 731–737, 1997.

Peng Su, Shudong Huang, Weihong Ma, Deng Xiong, and Jiancheng Lv. Multi-view granular-ball contrastive clustering. In *AAAI*, pp. 20637–20645, 2025.

Yuan Sun, Yongxiang Li, Zhenwen Ren, Guiduo Duan, Dezhong Peng, and Peng Hu. ROLL: Robust noisy pseudo-label learning for multi-view clustering with noisy correspondence. In *CVPR*, pp. 30732–30741, 2025.

Yuze Tan, Yixi Liu, Shudong Huang, Wentao Feng, and Jiancheng Lv. Sample-level multi-view graph clustering. In *CVPR*, pp. 23966–23975, 2023.

Huayi Tang and Yong Liu. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In *CVPR*, pp. 202–211, 2022a.

Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *ICML*, pp. 21090–21110, 2022b.

Daniel J. Trosten, Sigurd Løkse, Robert Jenssen, and Michael Kampffmeyer. Reconsidering representation alignment for multi-view clustering. In *CVPR*, pp. 1255–1265, 2021.

Daniel J. Trosten, Sigurd Løkse, Robert Jenssen, and Michael C. Kampffmeyer. On the effects of self-supervision and contrastive alignment in deep multi-view clustering. In *CVPR*, pp. 23976–23985, 2023.

Fangdi Wang, Jiaqi Jin, Jingtao Hu, Suyuan Liu, Xihong Yang, Siwei Wang, Xinwang Liu, and En Zhu. Evaluate then cooperate: Shapley-based view cooperation enhancement for multi-view clustering. In *NeurIPS*, pp. 135355–135379, 2024.

Jing Wang, Songhe Feng, Kristoffer Knutsen Wickstrøm, and Michael C. Kampffmeyer. AdaptCMVC: Robust adaption to incremental views in continual multi-view clustering. In *CVPR*, pp. 10285–10294, 2025a.

Yiming Wang, Qun Li, Dongxia Chang, Jie Wen, Hua Dai, Fu Xiao, and Yao Zhao. Learning from disjoint views: A contrastive prototype matching network for fully incomplete multi-view clustering. In *NeurIPS*, pp. 18265–18284, 2025b.

Jie Wen, Shijie Deng, Waikeung Wong, Guoqing Chao, Chao Huang, Lunke Fei, and Yong Xu. Diffusion-based missing-view generation with the application on incomplete multi-view clustering. In *ICML*, pp. 52762–52778, 2024.

Danyang Wu, Zhenkun Yang, Jitao Lu, Jin Xu, Xiangmin Xu, and Feiping Nie. EBMGC-GNF: Efficient balanced multi-view graph clustering via good neighbor fusion. *TPAMI*, pp. 7878–7892, 2024a.

Song Wu, Yan Zheng, Yazhou Ren, Jing He, Xiaorong Pu, Shudong Huang, Zhifeng Hao, and Lifang He. Self-weighted contrastive fusion for deep multi-view clustering. *TMM*, pp. 9150–9162, 2024b.

Shuyin Xia, Bolun Shi, Yifan Wang, Jiang Xie, Guoyin Wang, and Xinbo Gao. GBCT: Efficient and adaptive clustering via granular-ball computing for complex data. *TNNLS*, pp. 12159–12172, 2025.

Baili Xiao, Zhibin Dong, Ke Liang, Suyuan Liu, Siwei Wang, Tianrui Liu, Xingchen Hu, En Zhu, and Xinwang Liu. EASEMVC:efficient dual selection mechanism for deep multi-view clustering. In *CVPR*, pp. 20716–20726, 2025.

Huilang Xu, Xiang-Xiang Su, Simin Chen, Guang-Yong Chen, and Xing Chen. Bridging optimization and neural networks for efficient multi-view clustering. In *AAAI*, pp. 16066–16074, 2026.

Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu. Deep embedded multi-view clustering with collaborative training. *Inf. Sci.*, pp. 279–290, 2021a.

Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, and Lifang He. Multi-VAE: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In *ICCV*, pp. 9234–9243, 2021b.

Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *CVPR*, pp. 16051–16060, 2022.

Jie Xu, Yazhou Ren, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang, Xiaorong Pu, Philip S. Yu, and Lifang He. Self-supervised discriminative feature learning for deep multi-view clustering. *TKDE*, pp. 7470–7482, 2023.

Jie Xu, Yazhou Ren, Xiaolong Wang, Lei Feng, Zheng Zhang, Gang Niu, and Xiaofeng Zhu. Investigating and mitigating the side effects of noisy views for self-supervised clustering algorithms in practical multi-view scenarios. In *CVPR*, pp. 22957–22966, 2024.

Weiqing Yan, Yuanyang Zhang, Chenlei Lv, Chang Tang, Guanghui Yue, Liang Liao, and Weisi Lin. GCFAgg: Global and cross-view feature aggregation for multi-view clustering. In *CVPR*, pp. 19863–19872, 2023.

Xiaoqiang Yan, Zhixiang Jin, Fengshou Han, and Yangdong Ye. Differentiable information bottleneck for deterministic multi-view clustering. In *CVPR*, pp. 27435–27444, 2024.

Jie Yang, Wei Chen, Feng Liu, Peng Zhou, Zhongli Wang, Xinyan Liang, and Bingbing Jiang. Multi-view clustering via multi-granularity ensemble. In *IJCAI*, pp. 6794–6802, 2025a.

Jie Yang, Cheng-You Lu, Zhongli Wang, Hsiang-Ting Chen, Guang-Kui Xu, Chenglong Zhang, Shuting Dong, Xinyan Liang, and Bingbing Jiang. Multi-view clustering with granularity-aware pseudo supervision. In *AAAI*, pp. 27538–27546, 2026.

Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *CVPR*, pp. 1134–1143, 2021.

Xihong Yang, Siwei Wang, Fangdi Wang, Jiaqi Jin, Suyuan Liu, Yue Liu, En Zhu, Xinwang Liu, and Yueming Jin. Automatically identify and rectify: Robust deep contrastive multi-view clustering in noisy scenarios. In *ICML*, pp. 71383–71404, 2025b.

Honglin Yuan, Xingfeng Li, Jian Dai, Xiaojian You, Yuan Sun, and Zhenwen Ren. Deep streaming view clustering. In *ICML*, pp. 73526–73540, 2025.

Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In *NeurIPS*, pp. 1601–1608, 2004.

Changqing Zhang, Yeqing Liu, and Huazhu Fu. AE2-Nets: Autoencoder in autoencoder networks. In *CVPR*, pp. 2577–2585, 2019.

Qian Zhang, Lin Zhang, Ran Song, Runmin Cong, Yonghuai Liu, and Wei Zhang. Learning common semantics via optimal transport for contrastive multi-view clustering. *TIP*, pp. 4501–4515, 2024.

Bowen Zhao, Qianqian Wang, Zhengming Ding, and Quanxue Gao. Attribute-missing multi-view graph clustering. In *CVPR*, pp. 25832–25841, 2025.

Haochen Zhou, Guofeng Ding, Mouxing Yang, Peng Hu, Yijie Lin, and Xi Peng. Uncover underlying correspondence for robust multi-view clustering. In *ICLR*, pp. 86995–87014, 2026.

Runwu Zhou and Yi-Dong Shen. End-to-end adversarial-attention network for multi-modal clustering. In *CVPR*, pp. 14619–14628, 2020.

We thank the reviewers for taking the time to read our appendices. They cover the framework of the proposed algorithm (Appendix A), implementation and numerical conventions (Appendix B), complete splitting and connection algorithms (Appendix C), structural feedback and training details (Appendix D), independent label audits (Appendix E), theoretical proofs and strengthened results (Appendix F), mechanistic evaluation of splitting and connection (Appendix G), and additional comparative experiments (Appendix H).

A Framework of the Proposed Algorithm

Algorithm 1 summarizes GBASC under the default experimental configuration. GBAS refines local granular balls through evidence-guided splitting, and GBAC selects a global assignment through risk-aware connection. The resulting partition supplies reliability-weighted supervision for one guided training round. The structure is then rebuilt from the learned embeddings to produce the final clustering assignments.

Algorithm 1 Granular-Ball Adaptive Splitting and Connection (GBASC)

Input: Multi-view data $\mathcal{X} = \{\mathbf{X}^{(v)}\}_{v=1}^{V}$, observation mask $\mathbf{O}$, cluster count K, and default parameters; $T_{\mathrm{w}} = 10$ warm-up epochs and $T_{\mathrm{g}} = 20$ guided epochs.
Output: Final sample-level clustering assignments $\widehat{\mathbf{y}}$.

Preprocess the views and initialize the encoders and decoders.
Run T_{w} warm-up epochs under the default protocol.
Set the initial structural representations $\mathcal{H}$ according to this protocol.
for $t = 0, 1$ **do**
    Build view graphs and proposal graph $\mathbf{P}$; compute risk matrix $\mathbf{R}$ using Eq. (2).
    GBAS: adaptive splitting
    Initialize size-bounded roots by Ward clustering and hard size projection.
    Retain the highest-evidence split at each eligible node using Eq. (3).
    Run the base pilot and freeze V_B, ρ, and the structural probes.
    Recursively apply the $\Delta_A(G) > \epsilon_A$ gate with the local termination rule in Appendix C.1.
    Apply one terminal guide pass; accept proper refinements with $g_S > \epsilon_A$.
    Record the final atoms $\mathcal{G}$, membership matrix $\mathbf{S}$, and sample confidences q_i.
    GBAC: risk-aware connection
    Aggregate $\mathbf{P}^Q = \mathbf{S}^\top \mathbf{P}\mathbf{S}$ and $\mathbf{R}^Q = \mathbf{S}^\top \mathbf{R}\mathbf{S}$, retaining their diagonals.
    Construct the calibrated graph $\mathbf{A}^Q$ using Eq. (7).
    Obtain the size-weighted spectral base $\mathbf{c}_0$ and view-deletion stability s_C.
    Project the sample spectral and balanced partitions onto atoms; form their two UOT refinements.
    Choose the lowest-$\mathcal{J}_C$ nonbase candidate $\mathbf{c}_{\mathrm{p}}$, breaking ties by slot order.
    if the acceptance test in Eq. (9) holds **then**
        $\widehat{\mathbf{c}} \leftarrow \mathbf{c}_{\mathrm{p}}$.
    else
        $\widehat{\mathbf{c}} \leftarrow \mathbf{c}_0$.
    end if
    Assign $\widehat{y}_i \leftarrow \widehat{c}_{g(i)}$ for every sample, where $g(i)$ is its atom index.
    if $t = 0$ **then**
        Construct weighted relations using Eq. (11) and initialize fixed prototypes.
        for $e = 1, \ldots, T_{\mathrm{g}}$ **do**
            Freeze the full-dataset representation memory for the current epoch.
            Update the encoders and decoders over mini-batches using Eq. (13).
            Refresh the representation memory after the epoch.
        end for
        Recompute $\mathcal{H}$ from the learned embeddings.
    end if
end for
return $\widehat{\mathbf{y}}$.

B Implementation and Numerical Conventions

B.1 Inputs, Preprocessing, and Deterministic Conventions

Let $\mathcal{X} = \{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$ contain N objects, with $\boldsymbol{X}^{(v)} \in \mathbb{R}^{N\times d_v}$ and observation mask $\boldsymbol{O} \in \{0,1\}^{N\times V}$, where $O_{iv} = 1$ means that object i is observed in view v. We assume $V \geq 2$, $2 \leq K < N$, at least one observed view for every object, and at least one observed object in every view. The cluster count K is known, but class labels are unavailable to the algorithm. Throughout, we use $\epsilon = 10^{-12}$, $[x]_+ = \max\{x, 0\}$, and $\sigma(t) = (1 + e^{-t})^{-1}$.

We preprocess every view with the same continuous rule,

$$\widetilde{\boldsymbol{X}}^{(v)} = (1 - \alpha_v)\,\mathrm{ZS}(\boldsymbol{X}^{(v)}) + \alpha_v\,\mathrm{TL}(\boldsymbol{X}^{(v)}), \qquad \alpha_v = \sigma(40(s_v - 0.8))\,e^{-4e_v},$$

where s_v is the zero fraction, e_v is the fraction of squared energy in negative entries, ZS standardizes each feature, and TL applies TF–IDF reweighting followed by row ℓ_2 normalization. All statistics and both transformations use only rows with $O_{iv} = 1$. We set $e_v = 0$ when the total squared energy is zero; TL clips negative entries to zero before reweighting. Unobserved output rows, zero-variance standardized coordinates, and zero TF–IDF rows are returned as zero. This rule gives more TF–IDF weight to sparse nonnegative views and more standardization weight to dense or signed views without a separate preprocessing choice for each view.

Each observed view is passed through a view-specific encoder E_v with dimensions $d_v \to 512 \to 256 \to 128$. The normalized view representations and their fused representation are

$$\boldsymbol{z}_i^{(v)} = \frac{E_v(\widetilde{\boldsymbol{x}}_i^{(v)})}{\max(\|E_v(\widetilde{\boldsymbol{x}}_i^{(v)})\|_2, \epsilon)}, \qquad \bar{\boldsymbol{z}}_i = \mathrm{norm}_2\left(\frac{\sum_v O_{iv}\boldsymbol{z}_i^{(v)}}{\max(1, \sum_v O_{iv})}\right).$$

We use $\mathrm{norm}_2(\mathbf{0}) = \mathbf{0}$ and assign zero cosine similarity whenever either vector is zero. The structural operator can use either the preprocessed inputs or the learned embeddings; we denote the current view representation by $\boldsymbol{h}_i^{(v)}$. During relation-graph construction, distance ties follow the fixed brute-force nearest-neighbor backend on the caller's row order. Atomic routines use the content-canonical order described next. These conventions also define preprocessing, normalization, and neighborhoods on degenerate inputs.

Before a structural call, we normalize observed rows in each view to unit ℓ_2 norm and zero unobserved rows. For each view, we form a 128-bit BLAKE2b digest from its feature dimension and the sorted multiset of records containing the observation bit and contiguous normalized row bytes. Views are stably sorted by this digest, with their original indices resolving digest collisions, and the radially mapped view blocks are concatenated in that order. A second 128-bit BLAKE2b digest is computed for each sample from its concatenated row and correspondingly ordered mask row. Samples are stably sorted by this digest, with their original indices resolving collisions. Every atomic routine that is not permutation equivariant uses this order and maps its output back to the caller's sample order. The resulting atoms are numbered in increasing order of the pair consisting of the minimum original member index and atom size. The digests serve only as deterministic content identifiers; the analysis uses neither statistical evidence from them nor a cryptographic assumption.

We identify partitions up to label permutation. To compare two labeled vectors, we pass their ordered contingency matrix to the fixed maximum-weight linear-sum-assignment backend. For a nonunique optimum, its deterministic return with rows and columns ordered as $0, \ldots, K-1$ defines the alignment. A hard label obtained by $\arg\max$ or $\arg\min$ takes the smallest numeric label on a tie, and an objective tie within an indexed candidate list takes the smallest slot index. Ward, spectral, QR, SVD, and K-means calls use the fixed backend stated below. Randomized calls use the reported seed and restart order; a backend tie takes the first returned solution under the canonical order. Thus every partition and every subsequent content-derived rank is single valued.

B.2 Sample Graphs and Edge Risk

The reported neighborhood parameter is $k^+ = 15$. For $n_v = \sum_i O_{iv}$, view v uses $k_v^+ = \min\{15, n_v - 1\}$ if $n_v \geq 2$ and the empty graph otherwise, while the fused graph uses $k_P = \min\{15, N - 1\}$. The same rule is applied to every leave-one-view-out graph after removing the

indicated view. No validation-label or data-dependent search over k^+ is performed. We construct a symmetric self-tuning k_v^+-nearest-neighbor graph in each view, assigning a candidate neighbor pair the affinity

$$A_{ij}^{(v)} = \exp\left(-\frac{(d_{ij}^{(v)})^2}{\max(\sigma_i^{(v)}\sigma_j^{(v)}, \epsilon)}\right),$$

where $d_{ij}^{(v)}$ is cosine distance and $\sigma_i^{(v)}$ is the distance from i to its farthest retained neighbor. A zero scale is replaced by the median positive retained scale in that view, or by one if no positive scale exists. Directed neighborhoods are symmetrized by their maximum affinity. Applying the same construction to the concatenation of row-normalized views gives the positive connection proposal graph $\boldsymbol{P} \in \mathbb{R}_+^{N\times N}$.

For an edge with $P_{ij} > 0$, let $\mathcal{O}_{ij}$ contain the views in which both endpoints are observed and write $c_{ij} = |\mathcal{O}_{ij}|$. A missing neighbor edge in an observed view contributes zero affinity. We measure cross-view support and affinity variation, and assign the corresponding risk mass, by

$$\begin{aligned}
s_{ij} &= \frac{1}{\max(c_{ij}, 1)} \sum_{v\in\mathcal{O}_{ij}} \mathbb{I}[A_{ij}^{(v)} > 0], \\
\bar{A}_{ij} &= \frac{1}{\max(c_{ij}, 1)} \sum_{v\in\mathcal{O}_{ij}} A_{ij}^{(v)}, \\
\nu_{ij} &= \left[\frac{1}{\max(c_{ij}, 1)} \sum_{v\in\mathcal{O}_{ij}} (A_{ij}^{(v)} - \bar{A}_{ij})^2\right]^{1/2}, \\
u_{ij} &= \begin{cases} \mathrm{clip}\left(\dfrac{1 - s_{ij} + \nu_{ij}}{2}, 0, 1\right), & c_{ij} > 0, \\ 1, & c_{ij} = 0, \end{cases} \\
R_{ij} &= P_{ij} u_{ij}.
\end{aligned}$$

The score is nonincreasing in support and nondecreasing in affinity variation, and clipping gives $0 \le R_{ij} \le P_{ij}$. For $P_{ij} = 0$, we set $R_{ij} = 0$ and leave the pair unknown. Thus $\boldsymbol{R}$ measures uncertainty on existing positive edges; it does not encode cannot-link relations.

B.3 Graph Measures and Structural Scores

For the graph measures used below, fix $L \ge 1$, labels $\boldsymbol{c} \in [K]^L$, and positive node weights $w = (w_1, \ldots, w_L)^\top \in \mathbb{R}_{>0}^L$. Let $C_k^{(L)}(\boldsymbol{c}) = \{r : c_r = k\}$, $\mathcal{K}(\boldsymbol{c}) = \{k : C_k^{(L)}(\boldsymbol{c}) \neq \emptyset\}$, and $W(A, B) = \sum_{i\in A, j\in B} W_{ij}$. Since $L > 0$, at least one cluster is nonempty. For a nonnegative $L \times L$ graph $\boldsymbol{W}$, define

$$\begin{aligned}
\mathrm{cut}_{\boldsymbol{W}}(C, \bar{C}) &= \sum_{i\in C, j\notin C} W_{ij}, \\
\mathrm{vol}_{\boldsymbol{W}}(C) &= \sum_{i\in C, j} W_{ij}, \\
\ell_{\boldsymbol{W}}(C) &= \begin{cases} 1, & \mathrm{vol}_{\boldsymbol{W}}(C) \le \epsilon, \\ \mathrm{clip}\left(\dfrac{\mathrm{cut}_{\boldsymbol{W}}(C, \bar{C})}{\mathrm{vol}_{\boldsymbol{W}}(C)}, 0, 1\right), & \text{otherwise}, \end{cases} \\
\mathrm{NCut}_{\boldsymbol{W}}(\boldsymbol{c}) &= \frac{1}{|\mathcal{K}(\boldsymbol{c})|} \sum_{k\in\mathcal{K}(\boldsymbol{c})} \ell_{\boldsymbol{W}}(C_k^{(L)}(\boldsymbol{c})).
\end{aligned}$$

Cluster balance uses the weighted cluster proportions,

$$p_k^{(w)}(\boldsymbol{c}) = \frac{\sum_{r\in C_k^{(L)}(\boldsymbol{c})} w_r}{\sum_{r=1}^L w_r}, \qquad \mathrm{Bal}_w(\boldsymbol{c}) = -\frac{\sum_{k: p_k^{(w)}(\boldsymbol{c})>0} p_k^{(w)}(\boldsymbol{c}) \log p_k^{(w)}(\boldsymbol{c})}{\log K},$$

with $\mathrm{Bal}(\boldsymbol{c}) := \mathrm{Bal}_w(\boldsymbol{c})$ for unit node weights. On the sample graph ($L = N$), the within-cluster risk fraction is

$$\mathrm{RiskIn}_{\boldsymbol{R}}(\boldsymbol{c}) = \begin{cases} 0, & R([N],[N]) \le \epsilon, \\ \mathrm{clip}\left(\dfrac{\sum_{k\in\mathcal{K}(\boldsymbol{c})} R(C_k^{(N)}(\boldsymbol{c}), C_k^{(N)}(\boldsymbol{c}))}{R([N],[N])}, 0, 1 \right), & \text{otherwise.} \end{cases}$$

All these quantities are evaluated without ground-truth labels.

B.4 Pairwise Structural-Error Decomposition

To analyze structural-error propagation across granularities, let $y_i \in \mathcal{Y}$ denote the ground-truth label of object i. For a set A and disjoint sets A, B, write

$$D(A) = \sum_{\substack{i<j \\ i,j\in A}} \mathbb{I}[y_i \neq y_j],$$

$$D(A,B) = \sum_{i\in A}\sum_{j\in B} \mathbb{I}[y_i \neq y_j], \qquad S(A,B) = \sum_{i\in A}\sum_{j\in B} \mathbb{I}[y_i = y_j].$$

Suppose the connection stage preserves the atoms in the sample partition $\mathcal{G} = \{G_m\}_{m=1}^M$ and assigns label $c_m \in [K]$ to G_m. We abbreviate $D_{mm} = D(G_m)$, $D_{mn} = D(G_m, G_n)$, and $S_{mn} = S(G_m, G_n)$.

Proposition 1 (Exact structural-error decomposition) *For the atom-constant prediction $\widehat{y}_i = c_m$ when $i \in G_m$, the pairwise clustering disagreement satisfies*

$$\mathcal{E}_{\mathrm{pair}} = \underbrace{\sum_{m=1}^{M} D_{mm}}_{\mathcal{E}_{\mathrm{atom}}} + \underbrace{\sum_{m<n} \mathbb{I}[c_m = c_n] D_{mn} + \sum_{m<n} \mathbb{I}[c_m \neq c_n] S_{mn}}_{\mathcal{E}_{\mathrm{connect}}}.$$

Consequently, $\mathcal{E}_{\mathrm{pair}} \ge \mathcal{E}_{\mathrm{atom}}$ under every connection rule that preserves atomic integrity. Splitting an atom G into nonempty children $\{H_r\}_{r=1}^q$ changes its irreparable term by

$$D(G) - \sum_{r=1}^{q} D(H_r) = \sum_{r<t} \sum_{i\in H_r} \sum_{j\in H_t} \mathbb{I}[y_i \neq y_j] \ge 0.$$

Joining two pure atoms from different classes introduces exactly $|G_m||G_n|$ false-positive co-clustered pairs.

Proposition 1 separates errors inside atoms from errors in their connections. This motivates keeping local granular balls revisable through atomic splitting before calibrating inter granular ball connections, without asserting that the two-stage order is necessary or unique. The proof is given in Appendix F.1.

C Complete Splitting and Connection Algorithms

C.1 Complete Evidence-Guided Atomic Splitting

C.1.1 Initial Granules and Size Constraints

Atomic splitting forms the indivisible units used by the connection stage. We request an initial Ward partition with

$$M_{\mathrm{init}} = \min\{N, \max(\lceil 1.5K \rceil, \lceil N/B_{\max} \rceil)\}, \qquad B_{\max} = 32,$$

and apply the hard size projection below. Local recursive candidates are generated only when $|G| \ge 4$ and the depth is smaller than four. Ward geometry and atomic evidence use the radially mapped representation

$$T_\lambda(\boldsymbol{h}) = \boldsymbol{h}\left[(1-\lambda_{\mathrm{r}}) + \frac{\lambda_{\mathrm{r}}}{\max(\|\boldsymbol{h}\|_2, \epsilon)}\right], \qquad \lambda_{\mathrm{r}} = \frac{D_{\mathrm{H}}^2}{D_{\mathrm{H}}^2 + N^2}, \quad D_{\mathrm{H}} = \sum_v \dim(\boldsymbol{h}_i^{(v)}).$$

The map limits the influence of high-dimensional row norms. Its dimension D_{H} is recomputed at each structural stage: $\sum_v d_v$ for inputs and $128V$ for learned embeddings. Write $\check{\boldsymbol{h}}_i^{(v)} = T_\lambda(\boldsymbol{h}_i^{(v)})$ for the zero-masked result. Let $\check{\boldsymbol{H}}_U$ contain the concatenated radially mapped rows indexed by U in content-canonical order, and let $\mathrm{Ward}_q(\check{\boldsymbol{H}}_U)$ denote the nonempty cells returned by the fixed Ward backend with q requested cells. For $B = B_{\max}$, define the hard projection recursively by

$$\mathcal{H}_B(U) = \begin{cases} \{U\}, & |U| \le B, \\ \displaystyle\bigcup_{H \in \mathrm{Ward}_{\lceil |U|/B \rceil}(\check{\boldsymbol{H}}_U)} \mathcal{H}_B(H), & |U| > B, \end{cases}$$

$$\mathcal{G}_0 = \mathrm{HardProj}_B\big(\mathrm{Ward}_{M_{\mathrm{init}}}(\check{\boldsymbol{H}}_{[N]})\big) := \bigcup_{U \in \mathrm{Ward}_{M_{\mathrm{init}}}(\check{\boldsymbol{H}}_{[N]})} \mathcal{H}_B(U), \qquad M_0 = |\mathcal{G}_0|.$$

When $|U| > B$, Ward is asked for at least two nonempty cells. Each retained cell is therefore a proper subset of U, so its integer size decreases until all roots in $\mathcal{G}_0$ have size at most $B_{\max}$. This feasibility projection is completed before any objective-driven split and is excluded from those decisions.

C.1.2 Candidate Evidence and Ranking

For a splittable granular ball G, the candidate pool contains a fused-view Ward bisection, single-view Ward bisections fitted to the zero-masked blocks, and, when feasible, a spectral bisection of the sum of the per-view mutual-neighbor graphs. Duplicate partitions are removed. Consider a candidate $\pi = \{G_L, G_R\}$. Write $G_v = \{i \in G : O_{iv} = 1\}$ and $G_{L,v} = G_L \cap G_v$, with $G_{R,v}$ defined analogously. A view is eligible when $|G_v| \ge 4$ and $\min(|G_{L,v}|, |G_{R,v}|) \ge 2$; the eligible set is $\mathcal{V}_G(\pi)$. Let $\mathcal{E}_v(A)$ be the edges induced by A from the unordered mutual-neighbor graph of view v, whose effective global count is $k_v^+ = \min\{15, n_v - 1\}$, and let $\mathcal{E}_v^\times(G, \pi) = \{\{i, j\} \in \mathcal{E}_v(G_v) : |\{i, j\} \cap G_{L,v}| = 1\}$. Define

$$\begin{aligned}
\mathrm{SSE}_v(A) &= \sum_{i \in A} \|\check{\boldsymbol{h}}_i^{(v)} - \bar{\check{\boldsymbol{h}}}_A^{(v)}\|_2^2, \\
m_v &= \mathrm{clip}\left(1 - \frac{\mathrm{SSE}_v(G_{L,v}) + \mathrm{SSE}_v(G_{R,v})}{\max(\mathrm{SSE}_v(G_v), \epsilon)}, 0, 1\right), \\
x_v &= \begin{cases} |\mathcal{E}_v^\times(G, \pi)| / |\mathcal{E}_v(G_v)|, & \mathcal{E}_v(G_v) \neq \emptyset, \\ 0, & \mathcal{E}_v(G_v) = \emptyset, \end{cases} \\
x_v^0 &= \frac{2|G_{L,v}||G_{R,v}|}{|G_v|(|G_v| - 1)}, \\
f_v &= \begin{cases} \mathrm{clip}\left(1 - \dfrac{x_v}{x_v^0}, 0, 1\right), & \mathcal{E}_v(G_v) \neq \emptyset \text{ and } x_v^0 > \epsilon, \\ 0, & \text{otherwise}, \end{cases} \\
a_v &= \mathrm{clip}\Big(\mathrm{ARI}(\pi|_{G_v}, \mathrm{Ward}_2(\check{\boldsymbol{h}}_{G_v}^{(v)})), 0, 1\Big).
\end{aligned}$$

Rebuilding the mutual graph under three deterministic perturbations with ℓ_∞ magnitude at most 0.03 gives fracture scores $f_v^{(1)}, f_v^{(2)}, f_v^{(3)}$ and

$$\begin{aligned}
t_v &= \mathrm{clip}\left(1 - \frac{1}{3}\sum_{b=1}^{3} |f_v^{(b)} - f_v|, 0, 1\right), \\
\bar{z} &= \begin{cases} \dfrac{\sum_{v \in \mathcal{V}_G(\pi)} |G_v| z_v}{\sum_{v \in \mathcal{V}_G(\pi)} |G_v|}, & \mathcal{V}_G(\pi) \neq \emptyset, \\ 0, & \mathcal{V}_G(\pi) = \emptyset, \end{cases} \qquad z \in \{m, f, a, t\}.
\end{aligned}$$

One eligible view contributes its own evidence, while an empty eligible set gives four zero aggregates. Multiple views are averaged without separate per-view thresholds, so acceptance need not imply unanimous or majority support. The split score is

$$e_G(\pi) = \frac{4|G_L||G_R|}{|G|^2}\left(0.45\bar{m} + 0.30\bar{f} + 0.25\bar{a}\right)\left(0.5 + 0.5\bar{t}\right) \in [0, 1].$$

The balance factor suppresses extremely imbalanced proposals. For an unordered bipartition $\pi = \{G_L, G_R\}$, let $\mathrm{key}_G(\pi)$ be the lexicographically ordered pair of its two lexicographically sorted member tuples in content-canonical indices. Proposals with the same key are deduplicated while retaining all source tags. The remaining candidates are ordered lexicographically by $(-e_G(\pi), \mathrm{key}_G(\pi))$; hence evidence is ranked in descending order and exact ties use the canonical key. Only the rank-zero proposal π_G is recursively evaluated. For the equivalent objective formulation below, define the zero-based rank

$$r_G(\pi) = \#\left\{\pi' : (-e_G(\pi'), \mathrm{key}_G(\pi')) <_{\mathrm{lex}} (-e_G(\pi), \mathrm{key}_G(\pi))\right\}.$$

Appendix C.3 gives the exact perturbation, atomic-quality, consensus-edge, and empty-graph conventions used throughout this stage.

C.1.3 Frozen Pilot and Continuation Costs

For each feasible root R, write $C(G) = \binom{|G|}{2}$ and $Q_R = \max\{1, C(R)\}$. Let q_{cov}, q_{ov}, q_{cmp}, q_{pert}, and q_{cv} denote the normalized mutual-neighbor coverage, neighborhood overlap, compactness, perturbation stability, and cross-view stability of G, respectively; their reproducible definitions are given in Appendix C.3. The label-free stability score is

$$Q(G) = \frac{q_{\mathrm{cov}} + q_{\mathrm{ov}} + q_{\mathrm{cmp}} + q_{\mathrm{pert}} + q_{\mathrm{cv}}}{5} \in [0, 1].$$

A singleton is assigned $Q(G) = 0$ because it is a safe fallback rather than positive stability evidence. We then define

$$\widehat{r}(G) = \begin{cases} e_G(\pi_G), & \text{if a rank-zero candidate } \pi_G \text{ exists,} \\ 1 - Q(G), & \text{otherwise,} \end{cases}$$
$$\phi_R(G) = C(G)\widehat{r}(G) + \lambda_A Q_R + \xi_A Q_R \mathbb{I}[|G| = 1],$$

where the GBASC uses $\lambda_A = \xi_A = 0.10$. Within the frozen base objective, every additional leaf incurs a root-normalized complexity cost and every singleton incurs an extra penalty.

For analysis, expand the potential candidate tree of each feasible root by applying the deterministic generators to every eligible child up to depth four. Bounded depth, finite branching, and nonempty proper children make this tree finite. Each node retains its root and depth index, including branches evaluated lazily in the implementation; we suppress these indices when unambiguous. All candidate keys, evidence scores, and ranks are fixed before accept/reject decisions, so this expansion leaves the executed branch unchanged.

A label-free pilot is run once before final decisions. For the rank-zero proposal π_G, it computes

$$V_B(G) = \begin{cases} \sum_{H \in \pi_G} V_B(H), & \phi_R(G) - \sum_{H \in \pi_G} V_B(H) > \epsilon_A, \\ \phi_R(G), & \text{otherwise,} \end{cases} \qquad \epsilon_A = 10^{-12},$$

with $V_B(G) = \phi_R(G)$ when no candidate exists. Its values and leaf count M_{pilot} are then frozen, giving $\rho = M_{\mathrm{pilot}}/K$. The recursion itself uses neither ρ nor the subsequent structural gain.

Lemma 1 (Frozen-pilot range) *For every depth-indexed node G in the finite potential pilot tree rooted at R,*

$$0 \le V_B(G) \le \phi_R(G).$$

If the pilot accepts the rank-zero proposal at G, then $V_B(G) < \phi_R(G) - \epsilon_A$.

For any conceptual candidate π, define its prospective frozen-base gain by

$$g_B(G, \pi) = \phi_R(G) - \sum_{H \in \pi} V_B(H).$$

Only π_G requires recursive evaluation, as shown by the bounds below. The frozen continuation prevents later changes to a parent from generating new evidence that rescues a rejected decision.

C.1.4 Structural Guidance and Split Acceptance

Let $\mathcal{P}_{K,L} = \{\boldsymbol{c} \in [K]^L : \mathcal{K}(\boldsymbol{c}) = [K]\}$ contain the assignments with all K granular balls nonempty. In addition to the M_{init}-way initialization, compute the two K-way probes

$$\boldsymbol{c}^{\mathrm{W},K} = \mathrm{Ward}_K(\check{\boldsymbol{H}}) \in \mathcal{P}_{K,N}, \qquad \boldsymbol{c}^{\mathrm{S},K} = \mathrm{Spectral}_K(\boldsymbol{P}) \in [K]^N,$$

where $\check{\boldsymbol{H}}$ concatenates the radially mapped blocks and Ward uses the fixed deterministic backend. Here K is the requested granular ball count, independent of M_{init}. On a degenerate embedding, the spectral probe may have fewer than K active labels; the graph and structural measures in the graph-measure conventions and the within-cluster risk definition remain defined. Both probes and every leave-one-view-out spectral probe are computed once and frozen before local recursion. Their structural criterion and the resulting split pressure are

$$\mathcal{J}_{\mathrm{str}}(\boldsymbol{c}) = \frac{1}{2}\left[\mathrm{NCut}_{\boldsymbol{P}}(\boldsymbol{c}) + \mathrm{RiskIn}_{\boldsymbol{R}}(\boldsymbol{c}) + 1 - \mathrm{Bal}(\boldsymbol{c})\right],$$
$$\chi = 1 + 400\left[1 - \exp\Big(-\max\Big\{0, \sqrt{N/K}\left(\mathcal{J}_{\mathrm{str}}(\boldsymbol{c}^{\mathrm{W},K}) - \mathcal{J}_{\mathrm{str}}(\boldsymbol{c}^{\mathrm{S},K})\right)\Big\}\Big)\right].$$

For any candidate π of G, define its smallest-child balance $b_G(\pi) = 1 - \max_{H\in\pi} |H|/|G|$, its positive-edge cut fraction

$$c_P(G,\pi) = \frac{P(G,G) - \sum_{H\in\pi} P(H,H)}{\max(P(G,G),\epsilon)},$$

its mean leave-one-view-out agreement $s_{\mathrm{d}}(G,\pi) = V^{-1}\sum_{v=1}^{V} \mathrm{clip}(\mathrm{ARI}(\pi, \boldsymbol{c}^{\mathrm{S},-v}|_G), 0, 1)$, and its complexity $\kappa_G(\pi) = (|\pi| - 1)\log(1 + |G|)/|G|$. Here, $\boldsymbol{c}^{\mathrm{S},-v} \in [K]^N$ is the spectral guide after removing view v, and $P(A,B) = \sum_{i\in A, j\in B} P_{ij}$. The normalized structural gain is

$$g_S(G,\pi) = \left(1 + \frac{\chi}{\rho}\right) b_G(\pi) - c_P(G,\pi) - 0.015\rho\left[(1 - s_{\mathrm{d}}(G,\pi)) + \kappa_G(\pi)\right].$$

For local candidates, we retain this certificate only when its boundary agrees with the global guide:

$$a_G(\pi) = \left[\mathrm{clip}\left(\frac{\mathrm{ARI}(\pi, \boldsymbol{c}^{\mathrm{S},K}|_G) - \tau_A}{1 - \tau_A}, 0, 1\right)\right]^p, \quad \tau_A = 1 - 10^{-9}, \quad p = 1.$$

To express the rank-zero rule as an objective, consider a nonempty candidate pool and set $\epsilon_{\mathrm{mach}} = 2^{-52}$. The child-cost bound, unpenalized gain, and rank penalty are defined together as

$$\begin{aligned}
B_R(\pi) &= \sum_{H\in\pi} C(H) + |\pi|\lambda_A Q_R + \xi_A Q_R \sum_{H\in\pi} \mathbb{I}[|H| = 1],\\
\bar{\Delta}_A(G,\pi) &= \max\{g_B(G,\pi), a_G(\pi) g_S(G,\pi)\},\\
\ell_G(\pi) &= \max\{\phi_R(G) - B_R(\pi), a_G(\pi) g_S(G,\pi)\},\\
u_G(\pi) &= \max\{\phi_R(G), a_G(\pi) g_S(G,\pi)\},\\
L_G &= \min_{\pi} \ell_G(\pi), \qquad U_G = \max_{\pi} u_G(\pi).
\end{aligned}$$

Using these bounds, set the rank penalty and candidate objective to

$$\begin{aligned}
\delta_{\mathrm{tie}} &= 64\epsilon_{\mathrm{mach}}(1 + |\phi_R(G)| + |L_G| + |U_G|) > 0,\\
d_G &= U_G - L_G + \delta_{\mathrm{tie}}, \qquad \omega_G(\pi) = r_G(\pi) d_G,\\
\mathcal{J}_A(G \to \pi) &= \phi_R(G) - \bar{\Delta}_A(G,\pi) + \omega_G(\pi).
\end{aligned}$$

By Lemma 1 and $\widehat{r}(H) \in [0,1]$, $0 \le V_B(H) \le \phi_R(H) \le C(H) + \lambda_A Q_R + \xi_A Q_R \mathbb{I}[|H| = 1]$. Summing over children bounds the base gain, and taking its maximum with the structural gain gives

$$\begin{aligned}
0 \le \sum_{H\in\pi} V_B(H) &\le B_R(\pi),\\
\ell_G(\pi) \le \bar{\Delta}_A(G,\pi) &\le u_G(\pi),\\
\phi_R(G) - u_G(\pi) + \omega_G(\pi) \le \mathcal{J}_A(G \to \pi) &\le \phi_R(G) - \ell_G(\pi) + \omega_G(\pi).
\end{aligned}$$

For $r_G(\pi) \geq 1$, the same bounds imply

$$\begin{aligned}\mathcal{J}_A(G \to \pi) &\geq \phi_R(G) - U_G + r_G(\pi) d_G \\ &\geq \phi_R(G) - L_G + \delta_{\text{tie}} \\ &> \phi_R(G) - L_G \geq \mathcal{J}_A(G \to \pi_G).\end{aligned}$$

Thus every positive-rank proposal can be pruned without recursive evaluation. This is an exact-real-arithmetic comparison; the deterministic tie margin supplies no general floating-point guarantee without a forward-error analysis. Only π_G requires the frozen-base gain, and its unified gain and objective reduce to

$$\begin{aligned}\Delta_A(G) = \bar{\Delta}_A(G, \pi_G) &= \max\{g_B(G, \pi_G), a_G(\pi_G) g_S(G, \pi_G)\}, \\ \mathcal{J}_A(G \to \pi_G) &= \phi_R(G) - \Delta_A(G).\end{aligned}$$

The split is accepted exactly when $\Delta_A(G) > \epsilon_A$. Either the base continuation or the structural certificate can therefore justify acceptance. The leaf and singleton costs act through g_B; a structural decision may override a negative base gain and need not decrease that leaf cost. If the structural certificate strictly dominates g_B, its children are terminal within the local recursion. Otherwise, the same rule is applied recursively. After local recursion, the global guide partitions each remaining atom into nonempty cells,

$$\pi_G^{\text{term}} = \left\{ G \cap C_k^{(N)}(\boldsymbol{c}^{\mathrm{S},K}) : k \in \mathcal{K}(\boldsymbol{c}^{\mathrm{S},K}),\ G \cap C_k^{(N)}(\boldsymbol{c}^{\mathrm{S},K}) \neq \emptyset \right\}.$$

When π_G^{term} is a proper refinement, it is accepted exactly when

$$g_S(G, \pi_G^{\text{term}}) > \epsilon_A.$$

Since π_G^{term} and $\boldsymbol{c}^{\mathrm{S},K}|_G$ encode the same equivalence relation, their adjusted Rand index is one and $a_G(\pi_G^{\text{term}}) = 1$. This weight supplies no independent terminal evidence and is omitted from the terminal gate. The terminal guide makes a separate, single pass using neither e_G, $Q(G)$, g_B, nor the child-cost test. All accepted children are nonempty. Parents are stored for audit and recoverability, but accepted splits are not revoked within the same structural call.

C.1.5 Refinement Guarantees

We next separate properties that hold for every refinement from a semantic statement requiring independently quantified evidence calibration. For an atom partition $\mathcal{G}$, define

$$\mathcal{E}_A(\mathcal{G}) = \sum_{G \in \mathcal{G}} D(G), \qquad \mathrm{Pur}_{\mathrm{maj}}(\mathcal{G}) = \frac{1}{N} \sum_{G \in \mathcal{G}} \max_{c \in \mathcal{Y}} |\{i \in G : y_i = c\}|.$$

For a feasible root R, let $Q_R = \max\{1, \binom{|R|}{2}\}$ and define the root-normalized pair purity of a refinement tree $\mathcal{T}_R$ by

$$\mathrm{Pur}_{2,R}(\mathcal{T}_R) = 1 - \frac{1}{Q_R} \sum_{L \in \mathrm{leaves}(\mathcal{T}_R)} D(L).$$

Let $\mathcal{A}_{\mathrm{loc}}$ be the finite family of potential local binary node–proposal pairs fixed before unified decisions. For $a = (G, \pi)$ with $\pi = \{G_L, G_R\}$, set

$$B_a = D(G_L, G_R), \qquad \mu_a = \frac{4B_a}{|G|^2}, \qquad e_a = e_G(\pi).$$

The normalization matches the balance factor in the split-evidence definition. For predetermined nonnegative radii $\{\eta_a\}_{a \in \mathcal{A}_{\mathrm{loc}}}$, define the simultaneous event

$$\mathcal{E}_A^{\mathrm{cal}}(\boldsymbol{\eta}) = \{\mu_a \geq e_a - \eta_a \text{ for every } a \in \mathcal{A}_{\mathrm{loc}}\}.$$

If $\mathcal{A}_{\mathrm{loc}} = \emptyset$, then $\boldsymbol{\eta}$ is the empty family and we adopt the convention $\mathcal{E}_A^{\mathrm{cal}}(\boldsymbol{\eta}) = \Omega$. The calibrated actionwise conclusions then have no instances and hold vacuously. The radii must either be fixed without labels and independently of the selected action, or be generated by the prespecified independent audit protocol below; they cannot be freely chosen after inspecting the unknown labels.

Theorem 3 (Strengthened atomic-refinement certificate) *Condition on the fixed ground-truth labels used for analysis, representations, graphs, potential candidate family, and pilot values. Suppose an accepted proper refinement replaces the current atom G by nonempty children $\pi = \{H_r\}_{r=1}^{q}$, and define $B(G,\pi) = \sum_{r<t} D(H_r, H_t)$. Then*

$$\mathcal{E}_A(\mathcal{G}_{\text{before}}) - \mathcal{E}_A(\mathcal{G}_{\text{after}}) = B(G,\pi) \geq 0,$$
$$\text{Pur}_{\text{maj}}(\mathcal{G}_{\text{after}}) \geq \text{Pur}_{\text{maj}}(\mathcal{G}_{\text{before}}).$$

The first inequality is strict exactly when the boundary separates at least one different-class pair. If this accepted refinement is a post-projection local binary action $a = (G, \{G_L, G_R\}) \in \mathcal{A}_{\text{loc}}$ belonging to its unique feasible root R, let its immediate tree states satisfy

$$\text{leaves}(\mathcal{T}_{\text{after}}) = \big(\text{leaves}(\mathcal{T}_{\text{before}}) \setminus \{G\}\big) \cup \{G_L, G_R\}.$$

On $\mathcal{E}_A^{\text{cal}}(\boldsymbol{\eta})$,

$$B_a \geq \frac{|G|^2}{4}[e_a - \eta_a]_+, \qquad \text{Pur}_{2,R}(\mathcal{T}_{\text{after}}) - \text{Pur}_{2,R}(\mathcal{T}_{\text{before}}) = \frac{B_a}{Q_R} \geq \frac{|G|^2}{4Q_R}[e_a - \eta_a]_+.$$

If $s_a = \mathbb{I}[|G_L| = 1] + \mathbb{I}[|G_R| = 1]$, the exact reduction of the semantic leaf cost

$$\Phi_R^\star(\mathcal{T}_R) = \sum_{L \in \text{leaves}(\mathcal{T}_R)} \big[D(L) + \lambda_A Q_R + \xi_A Q_R \mathbb{I}[|L| = 1]\big]$$

caused by this immediate split is

$$\Delta_a^\star = B_a - \lambda_A Q_R - \xi_A Q_R s_a \geq \frac{|G|^2}{4}[e_a - \eta_a]_+ - Q_R(\lambda_A + \xi_A s_a).$$

Hence a positive right-hand side is a sufficient, noncircular certificate of strict semantic-cost reduction. Separately, the frozen base pilot obeys $0 \leq V_B(G) \leq \phi_R(G)$, with $V_B(G) < \phi_R(G) - \epsilon_A$ whenever that pilot accepts at G. Finally, within each input-space or learned-space structural call, starting from M_0 feasible roots, the local and terminal guided stages together accept at most $N - M_0$ proper splits and terminate.

The atomic-error and majority-purity bounds apply to every proper refinement, including hard size projection and multiway terminal guidance. The calibrated purity and immediate semantic-cost bounds apply only to post-projection local binary actions in $\mathcal{A}_{\text{loc}}$. They require the stated calibration event and margin: splitting a pure atom can leave $\mathcal{E}_A$ unchanged, and a structural decision can override the base complexity cost. Acceptance alone therefore certifies neither a decrease of $\Phi_R^\star$ nor lower total clustering error. Appendix F.3 proves these claims and gives an independent label audit that certifies the calibration event uniformly after label-free decisions are frozen. Because the audit queries true pair labels, its certificate is post hoc and does not supply an unsupervised run-time guarantee for e_G.

C.2 Risk-Aware Quotient-Graph Granular-Ball Connection

C.2.1 Quotient Graphs and Atom Confidence

Because $K < N$ and the initial granule count gives $M_{\text{init}} > K$, splitting yields a final atomization $\mathcal{G}_T = \{G_g\}_{g=1}^{M_T}$ with $M_T \geq M_0 > K$, while a semantic class may contain several such granular balls. For partitions of $[N]$, write $\mathcal{H} \preceq \mathcal{G}$ when every cell of $\mathcal{H}$ is contained in a unique cell of $\mathcal{G}$. Because all post-projection operations are refinements, $\mathcal{G}_T \preceq \mathcal{G}_0$. In this subsection set $M = M_T$. We connect all atoms through complete K-way partitions rather than greedily merging one pair at a time. Let $\boldsymbol{S} \in \{0,1\}^{N \times M}$ be the membership matrix, where $S_{ig} = 1$ iff $i \in G_g$, and let $n_g = |G_g|$. We aggregate positive and risk mass as

$$\boldsymbol{P}^Q = \boldsymbol{S}^\top \boldsymbol{P} \boldsymbol{S}, \qquad \boldsymbol{R}^Q = \boldsymbol{S}^\top \boldsymbol{R} \boldsymbol{S}.$$

Write $g(i) = g$ for the unique atom satisfying $i \in G_g$. The diagonal of $\boldsymbol{P}^Q$ is retained: it stores within-atom affinity needed to reproduce sample-graph volumes. We also define the atom confidence

$$q_g^{\text{A}} = \begin{cases} 1, & |G_g| = 1, \\ \dfrac{P(G_g, G_g)}{\max(P(G_g, G_g) + R(G_g, G_g), \epsilon)}, & \text{otherwise.} \end{cases}$$

The singleton value $Q(G) = 0$ prevents the keep objective from rewarding fragmentation. In contrast, $q_g^{\mathrm{A}} = 1$ preserves feedback from a singleton, which has no contradictory within-atom pair. These two quantities serve different purposes and are not interchangeable.

Lemma 2 (Exact positive-graph quotient) *For every partition that is constant within each atom, the cut, volume, and normalized-cut values computed on $\boldsymbol{P}$ are exactly equal to those computed on $\boldsymbol{P}^Q$. Thus, $\boldsymbol{P}^Q$ is an exact compression of $\boldsymbol{P}$ under the atomic-integrity constraint, rather than a center-based approximation.*

C.2.2 Risk Calibration and the Base Partition

The proof is given in Appendix F.2. Reliability calibration is deliberately performed after quotient aggregation. With $d_\epsilon(p) = \max(p, \epsilon)$, define

$$\widetilde{u}_{gh}^Q = \frac{R_{gh}^Q}{d_\epsilon(P_{gh}^Q)}, \qquad A_{gh}^Q = P_{gh}^Q \exp(-\beta \widetilde{u}_{gh}^Q), \qquad \beta = 0.25.$$

Because $0 \le R_{gh}^Q \le P_{gh}^Q$, every positive quotient edge satisfies $0 \le \widetilde{u}_{gh}^Q \le 1$ and therefore $e^{-0.25} \le A_{gh}^Q / P_{gh}^Q \le 1$. The mechanism is thus a bounded relative reweighting, not a claim that every risky edge is deleted. At sample level, set

$$\widetilde{u}_{ij} = \frac{R_{ij}}{d_\epsilon(P_{ij})}, \qquad A_{ij} = P_{ij} \exp(-\beta \widetilde{u}_{ij}),$$

which defines the risk-aware sample graph $\boldsymbol{A}$ used below. The same construction on each leave-one-view-out graph yields $\{\boldsymbol{A}^{Q,-v}\}_{v=1}^V$. We make the spectral operators explicit. For a symmetric nonnegative $L \times L$ graph $\boldsymbol{W}$, let

$$\mathrm{Emb}_K(\boldsymbol{W}) = \mathrm{rowNorm}\Big(\mathrm{eig}^{\uparrow K}(\boldsymbol{D}_{\boldsymbol{W}}^{-1/2} \boldsymbol{W} \boldsymbol{D}_{\boldsymbol{W}}^{-1/2})\Big), \qquad (D_{\boldsymbol{W}})_{ii} = d_i = \sum_j W_{ij},$$

where $(D_{\boldsymbol{W}}^{-1/2})_{ii} = d_i^{-1/2}$ for $d_i > \epsilon$ and is zero otherwise, and $\mathrm{eig}^{\uparrow K}$ returns eigenvectors for the K largest eigenvalues. The fixed implementation uses a dense symmetric eigensolver for quotient graphs and an ARPACK spectral backend for sample graphs; all statements condition on the finite embeddings returned by these fixed backends and do not assert a unique basis in a repeated eigenspace. Let $\mathrm{WKM}_K(\boldsymbol{E}; \boldsymbol{w})$ denote K-means on the rows of $\boldsymbol{E}$ with positive sample weights $\boldsymbol{w}$, a fixed seed, backend, and restart order. With $\boldsymbol{E}_Q = \mathrm{Emb}_K(\boldsymbol{A}^Q)$, the base atom partition is

$$\boldsymbol{c}_0 = \mathrm{WKM}_K(\boldsymbol{E}_Q; n), \qquad n = (n_1, \ldots, n_M)^\top.$$

The leave-one-view-out bases $\boldsymbol{c}_0^{-v}$ are obtained by the same operator on $\boldsymbol{A}^{Q,-v}$ and are deterministically aligned to $\boldsymbol{c}_0$.

C.2.3 Projected Candidate Partitions

Let $\boldsymbol{E}_S = \mathrm{Emb}_K(\boldsymbol{A})$. The sample reference is the deterministic cluster-QR discretization of this embedding. Specifically, pivoted QR of $\boldsymbol{E}_S^\top$ returns the first K row indices $\boldsymbol{p} = (p_1, \ldots, p_K)$; if $\boldsymbol{E}_S[\boldsymbol{p}, :]^\top = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the SVD returned by the fixed numerical backend, then

$$c_{\mathrm{ref},i} = \min \arg\max_{k \in [K]} \left| [\boldsymbol{E}_S \boldsymbol{U} \boldsymbol{V}^\top]_{ik} \right|.$$

The sample-balanced probe $\boldsymbol{c}_{\mathrm{bal}}$ is obtained from $\boldsymbol{c}_{\mathrm{ref}}$ by a deterministic capacity refinement. Let $b_k = \lfloor N/K \rfloor + \mathbb{I}[k \le N - K\lfloor N/K \rfloor]$. At each of at most ten updates, nonempty current labels define centers and squared distances D_{ik}^S in $\boldsymbol{E}_S$. With $D_{i,(2)}^S$ the second-smallest distance, set

$$\tau_S = \max\Big(\epsilon, 0.1 \,\mathrm{median}_i\, D_{i,(2)}^S\Big), \qquad H_{ik} = \exp\left[-\frac{D_{ik}^S - \min_\ell D_{i\ell}^S}{\tau_S}\right] + 10^{-300}.$$

Fifty alternating row normalizations and column scalings to the common fractional column target N/K are followed by one row normalization. Entries are then visited in nonincreasing order, and

each unassigned sample is allocated to a cluster with remaining capacity; exact ties follow the fixed array-sort backend order. This produces exactly the capacities $\{b_k\}$ and the iteration stops when labels no longer change. If the initial cluster-QR labels omit a cluster, the seed-fixed K-means labels on $\boldsymbol{E}_S$ are used for the first center update.

For a sample partition $\boldsymbol{c}$, first form the raw majority labels

$$\ell_g(\boldsymbol{c}) = \min \arg\max_{k\in[K]} \sum_{i\in G_g} \mathbb{I}[c_i = k].$$

When $\{\ell_g(\boldsymbol{c}) : g \in [M]\} = [K]$, let $\sigma_{\boldsymbol{c}}$ be the maximizing permutation returned under the fixed alignment convention from the objective $\sum_g n_g \mathbb{I}[c_{0g} = \sigma(\ell_g(\boldsymbol{c}))]$. The hard sample-to-atom projection is

$$[\Pi_{\mathcal{G}}(\boldsymbol{c})]_g = \begin{cases} \sigma_{\boldsymbol{c}}(\ell_g(\boldsymbol{c})), & \{\ell_h(\boldsymbol{c}) : h \in [M]\} = [K], \\ c_{0g}, & \text{otherwise.} \end{cases}$$

Omitting any cluster therefore triggers whole-vector fallback. Otherwise, the projected labels are aligned to $\boldsymbol{c}_0$ by atom-size-weighted maximum overlap. The ordered family $\mathcal{C}_5^{\mathrm{ord}} = (\boldsymbol{c}^{[0]}, \ldots, \boldsymbol{c}^{[4]})$ starts with the base $\boldsymbol{c}^{[0]} = \boldsymbol{c}_0$, followed by $\boldsymbol{c}^{[1]} = \Pi_{\mathcal{G}}(\boldsymbol{c}_{\mathrm{ref}})$ and $\boldsymbol{c}^{[2]} = \Pi_{\mathcal{G}}(\boldsymbol{c}_{\mathrm{bal}})$. The remaining slots $\boldsymbol{c}^{[3]}$ and $\boldsymbol{c}^{[4]}$ are the unbalanced optimal-transport (UOT) refinements initialized by $\boldsymbol{c}^{[1]}$ and $\boldsymbol{c}^{[2]}$, respectively. The five slots need not be distinct.

We condition all subsequent connection definitions and guarantees on $\mathcal{F}_C$: the embeddings $\boldsymbol{E}_Q$, $\boldsymbol{E}_S$, and all leave-one-view-out embeddings are finite; every center-initializing K-means call returns K nonempty granular balls; $\boldsymbol{c}_0 \in \mathcal{P}_{K,M}$; and both UOT calls return finite couplings and scores. Projection fallback and reversion of empty-cluster updates then ensure that all five slots belong to $\mathcal{P}_{K,M}$. No algorithmic or theoretical claim is made outside this event.

C.2.4 Unbalanced Optimal-Transport Refinement

The common UOT target interpolates between balanced mass and masses stable under view deletion. Using the reference and balanced probes on $\boldsymbol{A}$, compute

$$\mathcal{J}_{\mathrm{s}}(\boldsymbol{c}) = \tfrac{1}{2}\big[\,\mathrm{NCut}_{\boldsymbol{A}}(\boldsymbol{c}) + 1 - \mathrm{Bal}(\boldsymbol{c})\big],$$
$$\eta = \frac{1}{2}\left[1 + \tanh\left(\frac{1}{2}\sqrt{NK}\,\frac{\mathcal{J}_{\mathrm{s}}(\boldsymbol{c}_{\mathrm{ref}}) - \mathcal{J}_{\mathrm{s}}(\boldsymbol{c}_{\mathrm{bal}})}{\max(\mathcal{J}_{\mathrm{s}}(\boldsymbol{c}_{\mathrm{ref}}), \epsilon)}\right)\right].$$

Let $m_k(\boldsymbol{c}) = \sum_{g:c_g=k} n_g$. After aligning the leave-one-view-out base partitions $\{\boldsymbol{c}_0^{-v}\}_{v=1}^V$ to $\boldsymbol{c}_0$, form

$$\bar{t}_{\mathrm{stab},k} = \frac{m_k(\boldsymbol{c}_0) + \sum_{v=1}^V m_k(\boldsymbol{c}_0^{-v})}{V+1},$$
$$t'_{\mathrm{stab},k} = \max(\bar{t}_{\mathrm{stab},k}, N\times 10^{-9}), \qquad \boldsymbol{t}_{\mathrm{stab}} = \frac{N\boldsymbol{t}'_{\mathrm{stab}}}{\sum_k t'_{\mathrm{stab},k}}.$$

The common UOT target is

$$\boldsymbol{t} = \eta(N/K)\mathbf{1} + (1-\eta)\boldsymbol{t}_{\mathrm{stab}}.$$

It is normalized once more to total mass N in floating-point arithmetic. Let $\boldsymbol{e}_g$ be row g of $\boldsymbol{E}_Q$. At the start of each outer assignment update, the current nonempty assignment gives the sample-mass-weighted centers and distances

$$\boldsymbol{\mu}_k = \frac{\sum_{g:c_g=k} n_g \boldsymbol{e}_g}{\sum_{g:c_g=k} n_g}, \qquad D_{gk} = \|\boldsymbol{e}_g - \boldsymbol{\mu}_k\|_2^2,$$

and $D_{g,(2)}$ denotes the second-smallest center distance. Define

$$s_{\text{view}} = \frac{1}{V}\sum_{v=1}^{V} \text{clip}(\text{ARI}(\boldsymbol{c}_0, \boldsymbol{c}_0^{-v}), 0, 1),$$
$$\gamma_g = q_g^{\text{A}} \frac{1}{V}\sum_{v=1}^{V} \mathbb{I}[c_{0g} = c_{0g}^{-v}],$$
$$\overline{D}_2 = \text{median}_g\, D_{g,(2)}, \qquad m_g = \max(D_{g,(2)}, 10^{-3}\overline{D}_2),$$
$$C_{gk} = D_{gk} - 0.05(1-\eta)\gamma_g\, s_{\text{view}} m_g \mathbb{I}[k = c_g^{\text{init}}], \qquad \tau_T = \max(\epsilon, 0.02\overline{D}_2).$$

The discount applies to the initial assignment through its view stability and atom confidence. With masses $n = (n_g)_g$, target $\boldsymbol{t}$, and row-shifted kernel $K_{T,gk} = \exp[-(C_{gk} - \min_\ell C_{g\ell})/\tau_T] + 10^{-300}$, initialize $v_k^{(0)} = 1$ at each outer update and alternate

$$u_g \leftarrow \frac{n_g}{\max(\sum_k K_{T,gk} v_k, 10^{-300})},$$
$$v_k \leftarrow \left(\frac{t_k}{\max(\sum_g K_{T,gk} u_g, 10^{-300})}\right)^{\rho_T},$$
$$\Gamma_{gk} = u_g K_{T,gk} v_k / n_g,$$

where $\rho_T = \text{clip}(s_{\text{view}} + \eta(1 - s_{\text{view}}), 0.05, 0.999999)$. Atom labels are updated by $\min \arg\max_k \Gamma_{gk}$. We use at most 15 assignment updates and 100 scaling updates per assignment step; an empty-cluster update reverts to the preceding nonempty assignment. Among the generated assignments $\{\boldsymbol{c}_T^{(\ell)}\}$, the UOT slot returns the earliest minimizer of the following score, with MassErr defined in the connection-term definitions:

$$\mathcal{J}_T(\boldsymbol{c}_T^{(\ell)}) = 2.5\eta[1 - \text{Bal}_n(\boldsymbol{c}_T^{(\ell)})] + (1-\eta)\,\text{MassErr}(\boldsymbol{c}_T^{(\ell)}, \boldsymbol{t}).$$

C.2.5 Candidate Scoring and Safe Selection

For an atom partition $\boldsymbol{c}$, let $p_k(\boldsymbol{c}) = N^{-1}\sum_{g:c_g=k} n_g = p_k^{(n)}(\boldsymbol{c})$, consistent with the graph-measure conventions. Define

$$\text{MassErr}(\boldsymbol{c}, \boldsymbol{t}) = \frac{1}{2N}\sum_k \left|\sum_{g:c_g=k} n_g - t_k\right|,$$
$$\text{Stab}(\boldsymbol{c}) = \frac{1}{V}\sum_{v=1}^{V} \text{ARI}(\boldsymbol{c}, \boldsymbol{c}_0^{-v}),$$
$$\text{Move}(\boldsymbol{c}, \boldsymbol{c}_0) = \frac{1}{N}\sum_g n_g \mathbb{I}[c_g \neq c_{0g}],$$

where labels are aligned to $\boldsymbol{c}_0$ before computing movement. Let $\boldsymbol{A}^{Q,-0} = \boldsymbol{A}^Q$ and define

$$\text{Cut}_Q(\boldsymbol{c}) = \frac{1}{V+1}\sum_{v=0}^{V} \text{NCut}_{\boldsymbol{A}^{Q,-v}}(\boldsymbol{c}),$$
$$\text{MassRisk}(\boldsymbol{c}) = 2.5\eta[1 - \text{Bal}_n(\boldsymbol{c})] + (1-\eta)\,\text{MassErr}(\boldsymbol{c}, \boldsymbol{t}).$$

All five indexed candidates are evaluated by the single objective

$$\mathcal{J}_C(\boldsymbol{c}) = \text{Cut}_Q(\boldsymbol{c}) + \text{MassRisk}(\boldsymbol{c}) + 0.2[1 - \text{Stab}(\boldsymbol{c})] + (1-\eta)\,\text{Move}(\boldsymbol{c}, \boldsymbol{c}_0).$$

Risk exposure is reported as a diagnostic; risk already enters the objective through graph calibration. Choose the best nonbase slot before applying the trust check,

$$j_{\text{p}} = \min \arg \min_{j \in \{1,2,3,4\}} \mathcal{J}_C(\boldsymbol{c}^{[j]}), \qquad \boldsymbol{c}_{\text{p}} = \boldsymbol{c}^{[j_{\text{p}}]}, \qquad b_C = \text{clip}(1 - s_{\text{view}}, 0, 1),$$

where the outer minimum resolves objective ties by slot order. The evidence-budget check $\mathrm{Trust}(\boldsymbol{c}_{\mathrm{p}}) = 1$ holds exactly when

$$\begin{aligned}\mathrm{Cut}_Q(\boldsymbol{c}_{\mathrm{p}}) - \mathrm{Cut}_Q(\boldsymbol{c}_0) &\le b_C \max(\mathrm{Cut}_Q(\boldsymbol{c}_0), \epsilon) + \epsilon_C,\\ \mathrm{Stab}(\boldsymbol{c}_0) - \mathrm{Stab}(\boldsymbol{c}_{\mathrm{p}}) &\le b_C + \epsilon_C,\\ \mathrm{Move}(\boldsymbol{c}_{\mathrm{p}}, \boldsymbol{c}_0) &\le b_C + \epsilon_C.\end{aligned}$$

The same three inequalities, with an arbitrary candidate $\boldsymbol{c}$ in place of $\boldsymbol{c}_{\mathrm{p}}$, define $\mathrm{Trust}(\boldsymbol{c})$ when it is used below. We select

$$\widehat{\boldsymbol{c}} = \begin{cases}\boldsymbol{c}_{\mathrm{p}}, & \boldsymbol{c}_{\mathrm{p}} \neq \boldsymbol{c}_0 \text{ and } \mathrm{Trust}(\boldsymbol{c}_{\mathrm{p}}) = 1 \text{ and } \mathcal{J}_C(\boldsymbol{c}_{\mathrm{p}}) < \mathcal{J}_C(\boldsymbol{c}_0) - \epsilon_C,\\ \boldsymbol{c}_0, & \text{otherwise,}\end{cases} \qquad \epsilon_C = 10^{-12}.$$

C.2.6 Semantic Connection Guarantees

To connect the label-free objective to the semantic error used only in analysis, let $\mathcal{H} = \{H_r\}_{r=1}^{L}$ be any atomization and $\boldsymbol{c} \in \mathcal{P}_{K,L}$. Define its normalized inter-atomic loss by

$$\mathcal{L}_C(\boldsymbol{c}; \mathcal{H}) = \binom{N}{2}^{-1} \sum_{r<s} \left[\mathbb{I}[c_r = c_s] D(H_r, H_s) + \mathbb{I}[c_r \neq c_s] S(H_r, H_s)\right].$$

When the current final atomization $\mathcal{G}_T$ is fixed, we abbreviate $\mathcal{L}_C(\boldsymbol{c}) := \mathcal{L}_C(\boldsymbol{c}; \mathcal{G}_T)$. Likewise, $\mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}; \mathcal{H})$ denotes the pairwise disagreement in Proposition 1 for the sample labels induced by assigning label c_r to every member of H_r.

Let $\mathcal{E}_Q^+ = \{(g,h) : g < h,\ P_{gh}^Q > 0\}$. All edge-indexed sums and expectations below range over $\mathcal{E}_Q^+$. For $e = (g,h) \in \mathcal{E}_Q^+$, write

$$\begin{aligned}P_e^Q = P_{gh}^Q, \quad R_e^Q = R_{gh}^Q, \qquad A_e^Q = A_{gh}^Q,\\ U_e = \widetilde{u}_{gh}^Q, \qquad \zeta_e = e^{-\beta U_e}, \quad Z_e = \frac{D_{gh}}{n_g n_h} \in [0,1].\end{aligned}$$

Here Z_e is the fraction of cross-atom pairs having different labels; for pure atoms it is exactly zero for a same-class edge and one for a cross-class edge. Fix any nonnegative impact weights h_e with $\sum_e h_e P_e^Q > 0$ and define

$$\pi_e = \frac{h_e P_e^Q}{\sum_f h_f P_f^Q}, \qquad \Theta_0 = \mathbb{E}_\pi[Z_E],$$
$$\Theta_\beta = \frac{\mathbb{E}_\pi[\zeta_E Z_E]}{\mathbb{E}_\pi[\zeta_E]}.$$

where $E \sim \pi$ denotes an edge drawn according to the displayed mass function. Choosing $h_e = 1$ measures affinity-weighted error exposure, whereas $h_e = n_g n_h$ additionally emphasizes the Cartesian-product amplification of connecting large atoms.

Theorem 4 (Strengthened risk-attenuation and connection certificate) *Condition on the fixed ground-truth labels used for analysis, atoms, and nonnegative graphs satisfying $P_{ij} = 0 \Rightarrow R_{ij} = 0$. For the selection claim, also condition on $\mathcal{F}_C$ and the resulting indexed family $\mathcal{C}_5^{\mathrm{ord}} \in (\mathcal{P}_{K,M})^5$. For every atom pair with $P_{gh}^Q > 0$, attenuation and its sensitivity to risk satisfy*

$$0 < A_{gh}^Q \le P_{gh}^Q, \qquad 0 \le P_{gh}^Q - A_{gh}^Q \le \beta R_{gh}^Q,$$
$$\frac{\partial A_{gh}^Q}{\partial R_{gh}^Q} = -\beta \frac{P_{gh}^Q}{d_\epsilon(P_{gh}^Q)} e^{-\beta \widetilde{u}_{gh}^Q} < 0.$$

For positive edges b and g with $U_b - U_g \ge \delta \ge 0$, $A_b^Q / A_g^Q \le e^{-\beta\delta} P_b^Q / P_g^Q$. Calibration after aggregation also satisfies the conservative bound

$$A_{gh}^Q \le \sum_{i \in G_g, j \in G_h : P_{ij} > 0} P_{ij} e^{-\beta R_{ij} / d_\epsilon(P_{ij})}.$$

The absolute pollution-weighted bridge mass and the change in normalized exposure obey

$$\sum_e h_e A_e^Q Z_e \le \sum_e h_e P_e^Q Z_e,$$
$$\Theta_\beta - \Theta_0 = \frac{\mathrm{Cov}_\pi(\zeta_E, Z_E)}{\mathbb{E}_\pi[\zeta_E]}.$$

Thus $\Theta_\beta \le \Theta_0$ holds exactly when $\mathrm{Cov}_\pi(\zeta_E, Z_E) \le 0$. A sufficient condition is that $u \mapsto \mathbb{E}_\pi[Z_E \mid U_E = u]$ is nondecreasing. This order-calibration condition is not necessary and is not implied by monotone attenuation alone.

A concrete sufficient condition follows from the implemented view statistics. Suppose $\boldsymbol{R}$ is constructed by the cross-view risk definition, every atom incident to $\mathcal{E}_Q^+$ is pure with unique label y_g, every positive quotient edge satisfies $P_e^Q \ge \epsilon$, and every contributing positive sample edge has at least one co-observed view, $c_{ij} \ge 1$. Let s_+, s_-, ν_+, ν_- satisfy

$$s_{ij} \ge s_+, \quad \nu_{ij} \le \nu_+ \qquad \text{on positive same-class sample edges},$$
$$s_{ij} \le s_-, \quad \nu_{ij} \ge \nu_- \qquad \text{on positive cross-class sample edges}.$$

With $u_+ = \mathrm{clip}((1 - s_+ + \nu_+)/2, 0, 1)$, $u_- = \mathrm{clip}((1 - s_- + \nu_-)/2, 0, 1)$, and $\delta_u = u_- - u_+ > 0$, any nonnegative impact weights with $\sum_{e: y_g = y_h} h_e P_e^Q > 0$ satisfy

$$\frac{\sum_{e:\, y_g \ne y_h} h_e A_e^Q}{\sum_{e:\, y_g = y_h} h_e A_e^Q} \le e^{-\beta \delta_u} \frac{\sum_{e:\, y_g \ne y_h} h_e P_e^Q}{\sum_{e:\, y_g = y_h} h_e P_e^Q}.$$

Hence purity and support–dispersion separation bound the calibrated cross-class-to-within-class exposure ratio on $\mathcal{E}_Q^+$ by the original ratio multiplied by $e^{-\beta\delta_u}$.

On $\mathcal{F}_C$, the selected partition satisfies $\mathcal{J}_C(\widehat{\boldsymbol{c}}) \le \mathcal{J}_C(\boldsymbol{c}_0)$. An accepted nonbase proposal satisfies the strict margin in the selection rule and attains $\mathcal{J}_C(\widehat{\boldsymbol{c}}) = \min_{0 \le j \le 4} \mathcal{J}_C(\boldsymbol{c}^{[j]})$.

Attenuation is a property of the graph transformation, and the fallback protects the fixed surrogate on $\mathcal{F}_C$. Their semantic interpretation requires the stated calibration assumptions: a consistently supported false edge need not receive additional suppression. The exposure ratio bound applies to pure atoms and positive quotient edges with the specified risk ordering. It does not cover atom pairs absent from $\mathcal{E}_Q^+$ that are later assigned to the same cluster. Neither that bound nor surrogate-safe fallback alone guarantees lower $\mathcal{L}_C$ or higher clustering accuracy.

C.3 Atomic Evidence Definitions

This subsection closes the definitions used by the split evidence and atomic quality. The displays below form a definition block, not a chain of probabilistic or semantic implications. In each current radially mapped view block, let $n_v = \sum_i O_{iv}$ and

$$k_v = \begin{cases} 0, & n_v \le 1, \\ \min\{15, n_v - 1\}, & n_v \ge 2. \end{cases}$$

Let $\mathcal{N}_v^{\rightarrow}(i)$ be the directed Euclidean k_v-nearest-neighbor set among observed samples; the graph is empty when $n_v < 2$. Let

$$\mathcal{E}_v(A) = \big\{\{i, j\} \subseteq A : j \in \mathcal{N}_v^{\rightarrow}(i),\ i \in \mathcal{N}_v^{\rightarrow}(j)\big\}$$

be the unordered mutual edges induced by A. For endpoints co-observed in at least one view, the consensus graph retains an edge whenever its number of supporting per-view mutual graphs divided by its positive number of co-observed views is at least 0.5; denote its induced edge set by $\mathcal{E}_\cap(A)$.

For any graph H on A, define

$$\mathrm{cov}_H(A) = \begin{cases} 1, & |A| \le 1, \\ |A|^{-1} \sum_{i \in A} \mathbb{I}[\mathcal{N}_H(i) \cap A \neq \emptyset], & |A| > 1, \end{cases}$$

$$\mathrm{Jac}(B, C) = \begin{cases} |B \cap C| / |B \cup C|, & B \cup C \neq \emptyset, \\ 1, & B \cup C = \emptyset, \end{cases}$$

$$\mathrm{ov}_v(A) = \begin{cases} |\mathcal{E}_v(A)|^{-1} \displaystyle\sum_{\{i,j\} \in \mathcal{E}_v(A)} \mathrm{Jac}(\mathcal{N}_v^{\rightarrow}(i) \setminus \{j\}, \mathcal{N}_v^{\rightarrow}(j) \setminus \{i\}), & \mathcal{E}_v(A) \neq \emptyset, \\ 0, & \mathcal{E}_v(A) = \emptyset. \end{cases}$$

We use $\mathrm{cov}_v(A)$ and $\mathrm{cov}_\cap(A)$ as shorthand for $\mathrm{cov}_H(A)$ when H is, respectively, the view-v mutual graph and the consensus mutual graph. The consensus overlap $\mathrm{ov}_\cap(A)$ is defined by the same last display using consensus mutual neighborhoods for both neighbor sets.

For $A_v = \{i \in A : O_{iv} = 1\}$, define

$$\bar{\check{\boldsymbol{h}}}_{A_v}^{(v)} = |A_v|^{-1} \sum_{i \in A_v} \check{\boldsymbol{h}}_i^{(v)}, \qquad r_v(A) = \left\{ |A_v|^{-1} \sum_{i \in A_v} \|\check{\boldsymbol{h}}_i^{(v)} - \bar{\check{\boldsymbol{h}}}_{A_v}^{(v)}\|_2^2 \right\}^{1/2}$$

only when $|A_v| \ge 1$, and set $r_v(A) = 0$ when $A_v = \emptyset$. For each member of a set with $|A_v| \ge 2$, take the first retained neighbor distance strictly larger than 10^{-12}, or 10^{-6} if all retained distances are zero; let $\sigma_v(A)$ be the median of the resulting finite positive scales. The compactness is

$$\mathrm{cmp}_v(A) = \begin{cases} 1, & |A_v| \le 1, \\ [1 + r_v(A) / \max(\sigma_v(A), \epsilon)]^{-1}, & \sigma_v(A) \text{ exists}, \\ 0, & \text{otherwise}. \end{cases}$$

For $b \in \{1, 2, 3\}$, an observed coordinate in the content-canonicalized, radially mapped block, using one-based row and coordinate indices, is perturbed by

$$\delta_{i\ell}^{(b)} = 0.03 \sin(2\pi[0.754877666 i + 0.569840296 \ell + 0.438579021 b]), \qquad |\delta_{i\ell}^{(b)}| \le 0.03.$$

Let $\mathcal{E}_v^{(b)}(A)$ and $\mathrm{cov}_v^{(b)}(A)$ be obtained after rebuilding the view graph. With edge-set Jaccard equal to one when both sets are empty, define

$$\mathrm{pert}_v(A) = \frac{1}{3} \sum_{b=1}^{3} \left[\frac{1}{2} \mathrm{Jac}(\mathcal{E}_v(A), \mathcal{E}_v^{(b)}(A)) + \frac{1}{2} \{1 - |\mathrm{cov}_v(A) - \mathrm{cov}_v^{(b)}(A)|\} \right].$$

For a nonsingleton A, a view enters the quality average when $|A_v| \ge 2$; write this set as $\mathcal{V}_A^Q$, and interpret an average over an empty set as zero. The first four quality components are

$$q_{\mathrm{cov}}(A) = \tfrac{1}{2} \mathrm{Avg}_{v \in \mathcal{V}_A^Q} \mathrm{cov}_v(A_v) + \tfrac{1}{2} \mathrm{cov}_\cap(A),$$

$$q_{\mathrm{ov}}(A) = \tfrac{1}{2} \mathrm{Avg}_{v \in \mathcal{V}_A^Q} \mathrm{ov}_v(A_v) + \tfrac{1}{2} \mathrm{ov}_\cap(A),$$

$$q_{\mathrm{cmp}}(A) = \mathrm{Avg}_{v \in \mathcal{V}_A^Q} \mathrm{cmp}_v(A_v), \qquad q_{\mathrm{pert}}(A) = \mathrm{Avg}_{v \in \mathcal{V}_A^Q} \mathrm{pert}_v(A_v).$$

For each view pair $v < w$, let $A_{vw} = A_v \cap A_w$. The pair is valid when $|A_{vw}| \ge 2$ and at least one of $\mathcal{E}_v(A_{vw})$ and $\mathcal{E}_w(A_{vw})$ is nonempty, and its score is

$$s_{vw}(A) = \tfrac{1}{2} \mathrm{Jac}(\mathcal{E}_v(A_{vw}), \mathcal{E}_w(A_{vw})) + \tfrac{1}{2} [1 - |\mathrm{cov}_v(A_{vw}) - \mathrm{cov}_w(A_{vw})|].$$

If $|\mathcal{V}_A^Q| < 2$, set $q_{\mathrm{cv}}(A) = q_{\mathrm{pert}}(A)$; otherwise, average $s_{vw}(A)$ over valid view pairs, using the neutral value 0.5 when none is valid. Each component is clipped to $[0, 1]$ before the atomic-quality definition is evaluated. For a singleton, the atomic-quality definition overrides these component conventions and sets $Q(A) = 0$. For split evidence, the per-view evidence definition sets $f_v = 0$ on an empty mutual graph, and the evidence aggregation rule uses the same three rebuilt graphs. These conventions make every evidence value defined under arbitrary missingness satisfying Section B.1. In particular, coverage, overlap, compactness, perturbation stability, cross-view stability, and every clipped quality component lie in $[0, 1]$. The three perturbations defined above are deterministic sensitivity probes; they are not the independent random draws used by the label audits below.

D Structural Feedback and Training Details

D.1 Structural Supervision

The selected structural partition supplies weighted neighbor relations for representation learning. For sample i, let $g(i)$ be its atom index and $q_i = q^{\mathrm{A}}_{g(i)}$ its inherited atom confidence. Write $s_C = V^{-1}\sum_v \mathrm{clip}(\mathrm{ARI}(\boldsymbol{c}_0, \boldsymbol{c}_0^{-v}), 0, 1)$ for the leave-one-view-out stability recorded by the connection stage. Given the available history $\{\widehat{\boldsymbol{c}}^{(r)}_{\mathrm{A}}\}_r \subseteq [K]^M$ of selected atom partitions, define $\widehat{c}^{(r)}_i = \widehat{c}^{(r)}_{\mathrm{A},g(i)}$ and let $\widehat{c}_i$ denote the current pre-update induced label. We retain positive-graph edges whose endpoints consistently share or differ in label,

$$\mathcal{E}^+ = \{(i,j) : P_{ij} > 0,\ \widehat{c}^{(r)}_i = \widehat{c}^{(r)}_j \text{ for every } r\},$$
$$\mathcal{E}^- = \{(i,j) : P_{ij} > 0,\ \widehat{c}^{(r)}_i \neq \widehat{c}^{(r)}_j \text{ for every } r\}.$$

The reported protocol uses one guided refinement round, so these sets are determined by the current pre-update partition. Their reliability comes from atom confidence, edge risk, and view stability. The history-based definition also permits multiple rounds, in which conflicting decisions exclude an edge from both sets. Pairs outside $\boldsymbol{P}$ remain unknown.

D.2 Reliability Weights and Relation Loss

For retained edges, combine endpoint confidence with risk attenuation and assign the corresponding pull and repel weights,

$$q_{ij} = \sqrt{q_i q_j}, \qquad r_{ij} = e^{-\beta R_{ij}/\max(P_{ij},\epsilon)},$$
$$w^+_{ij} = P_{ij} q_{ij} r_{ij} \begin{cases} 1, & g(i) = g(j), \\ 0.75, & g(i) \neq g(j), \end{cases} \quad (i,j) \in \mathcal{E}^+,$$
$$w^-_{ij} = P_{ij} q_{ij} s_C [0.5 + 0.5(1 - r_{ij})], \quad (i,j) \in \mathcal{E}^-.$$

For each anchor, retain at most the 16 highest-weight pull neighbors and 32 highest-weight repel neighbors. Set $\boldsymbol{z}^{(0)}_i = \bar{\boldsymbol{z}}_i$, $O_{i0} = 1$, and $d_r(i,j) = 1 - \cos(\boldsymbol{z}^{(r)}_i, \boldsymbol{z}^{(r)}_j)$ for $0 \le r \le V$. After filtering the retained neighbors for observation in representation r, denote the resulting sets by $\mathcal{N}^+_{ir}$ and $\mathcal{N}^-_{ir}$. Each Avg below includes only the eligible nonempty sets specified in its index and equals zero when none exists. With margin $m = 0.8$, the relation loss is

$$\ell^+_{ir} = \frac{\sum_{j\in\mathcal{N}^+_{ir}} w^+_{ij} d_r(i,j)}{\max(\sum_{j\in\mathcal{N}^+_{ir}} w^+_{ij}, 10^{-8})},$$
$$\ell^-_{ir} = \frac{\sum_{j\in\mathcal{N}^-_{ir}} w^-_{ij} [m - d_r(i,j)]_+}{\max(\sum_{j\in\mathcal{N}^-_{ir}} w^-_{ij}, 10^{-8})},$$
$$\mathcal{I}^{\mathrm{act}}_r = \left\{ i : O_{ir} = 1,\ \sum_{j\in\mathcal{N}^+_{ir}} w^+_{ij} + \sum_{j\in\mathcal{N}^-_{ir}} w^-_{ij} > 0 \right\},$$
$$\mathcal{L}_{\mathrm{rel}} = \mathrm{Avg}_{r:\,|\mathcal{I}^{\mathrm{act}}_r|>0} \frac{1}{|\mathcal{I}^{\mathrm{act}}_r|} \sum_{i\in\mathcal{I}^{\mathrm{act}}_r} (\ell^+_{ir} + \ell^-_{ir}).$$

This first averages over active anchors within each representation, then over eligible fused and view representations. Neighbor endpoints come from a stop-gradient, full-dataset representation memory, frozen within each epoch and refreshed after that epoch.

D.3 Representation Objectives and Structure Re-estimation

For reconstruction and cross-view alignment, let $D_v : \mathbb{R}^{128} \to \mathbb{R}^{d_v}$ be the decoder paired with E_v, $\mathcal{I}_{vw} = \{i : O_{iv}O_{iw} = 1\}$, and $d_z = 128$. The two losses are

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{V}\sum_v \frac{\sum_i O_{iv}\|D_v(E_v(\widetilde{\boldsymbol{x}}_i^{(v)})) - \widetilde{\boldsymbol{x}}_i^{(v)}\|_2^2/d_v}{\max(\sum_i O_{iv}, 1)},$$

$$\mathcal{L}_{\mathrm{cv}} = \mathrm{Avg}_{v<w:\,|\mathcal{I}_{vw}|>0}\left[1 - \frac{1}{|\mathcal{I}_{vw}|}\sum_{i\in\mathcal{I}_{vw}} \cos(\boldsymbol{z}_i^{(v)}, \boldsymbol{z}_i^{(w)})\right]$$

$$+ \mathrm{Avg}_{v:\,\sum_i O_{iv}\geq 2}\frac{1}{d_z}\sum_{\ell=1}^{d_z}\frac{[\gamma - \sqrt{\mathrm{Var}_{i:O_{iv}=1}(z_{i\ell}^{(v)}) + 10^{-4}}]_+}{\gamma}, \qquad \gamma = d_z^{-1/2}.$$

The alignment term compares co-observed views, while its variance term uses only views with at least two observed samples. For prototype supervision, set $\widetilde{q}_i = q_i(0.5 + 0.5s_C)$. The normalized prototypes $\{\boldsymbol{p}_k\}_{k=1}^K$ are constructed once from a seed-fixed Gaussian matrix by deterministic sign-corrected QR factorization and remain fixed throughout guided training. Using the pre-update labels gives

$$\mathcal{L}_{\mathrm{proto}} = -\,\mathrm{Avg}_{0\leq r\leq V:\,\sum_i O_{ir}\widetilde{q}_i>0}\frac{\sum_i O_{ir}\widetilde{q}_i \log\dfrac{\exp(\cos(\boldsymbol{z}_i^{(r)}, \boldsymbol{p}_{\widehat{c}_i})/\tau)}{\sum_{k=1}^K \exp(\cos(\boldsymbol{z}_i^{(r)}, \boldsymbol{p}_k)/\tau)}}{\max(\sum_i O_{ir}\widetilde{q}_i, 10^{-8})}, \qquad \tau = 0.1.$$

The complete encoder–decoder objective combines these terms as

$$\mathcal{L} = \mathcal{L}_{\mathrm{rec}} + 0.5\mathcal{L}_{\mathrm{cv}} + 0.5\mathcal{L}_{\mathrm{rel}} + \mathcal{L}_{\mathrm{proto}}.$$

After encoder optimization, we re-evaluate the graph construction and the splitting and connection rules on the learned embeddings to obtain the final prediction. Theorems 1 and 2 condition on the representations and graphs of each structural update. They do not assert that neural training and structural updates jointly minimize a single global objective.

E Independent Label Audits

E.1 Connection Selection Audit

The gap between the label-free surrogate and semantic error can be assessed without assuming an unknown uniform calibration constant. Fix the dataset, representations, graphs, atoms, candidates, and selected output before accessing any audit labels. For $N \geq 2$, draw unordered sample pairs independently and uniformly with replacement from all $\binom{N}{2}$ pairs. For an atom assignment $\boldsymbol{c}$, define

$$\ell_{\boldsymbol{c}}(i,j) = \mathbb{I}[g(i) \neq g(j)]\left\{\mathbb{I}[c_{g(i)} = c_{g(j)}]\mathbb{I}[y_i \neq y_j] + \mathbb{I}[c_{g(i)} \neq c_{g(j)}]\mathbb{I}[y_i = y_j]\right\}.$$

Choose $m_C \geq 1$ without audit labels and fix $\delta \in (0,1)$ before drawing the pairs. Using the same m_C pairs for the four nonbase slots, let

$$\widehat{\mathcal{L}}_C(\boldsymbol{c}) = \frac{1}{m_C}\sum_{t=1}^{m_C}\ell_{\boldsymbol{c}}(I_t, J_t),$$

$$\widehat{d}_j = \widehat{\mathcal{L}}_C(\boldsymbol{c}^{[j]}) - \widehat{\mathcal{L}}_C(\boldsymbol{c}_0), \qquad r_C(\delta) = \sqrt{\frac{2\log(8/\delta)}{m_C}}.$$

Audit labels are never used to construct graphs, tune parameters, generate candidates, or choose the output.

Corollary 1 (Audit-certified semantic connection change) *Under this audit protocol, with probability at least $1-\delta$, simultaneously for the four nonbase slots,*

$$\left|\mathcal{L}_C(\boldsymbol{c}^{[j]}) - \mathcal{L}_C(\boldsymbol{c}_0) - \widehat{d}_j\right| \leq r_C(\delta).$$

If a nonbase proposal is selected, let $j_{\mathrm{sel}} = j_{\mathrm{p}}$. Then $\widehat{d}_{j_{\mathrm{sel}}} + r_C(\delta) < 0$ certifies a strict decrease in both connection loss and normalized total pairwise error relative to $\mathbf{c}_0$ on the common final atoms.

Proof. For each slot, set $X_t^{(j)} = \ell_{\boldsymbol{c}^{[j]}}(I_t, J_t) - \ell_{\boldsymbol{c}_0}(I_t, J_t)$. The pair losses lie in $[0,1]$, so $X_t^{(j)} \in [-1,1]$; uniform sampling gives $d_j := \mathbb{E}[X_t^{(j)}] = \mathcal{L}_C(\boldsymbol{c}^{[j]}) - \mathcal{L}_C(\boldsymbol{c}_0)$. Two-sided Hoeffding with range length two and a union bound therefore yield

$$\Pr\left\{\max_{1\le j\le 4} |\widehat{d}_j - d_j| > r_C(\delta)\right\} \le \sum_{j=1}^{4} 2\exp\left[-\frac{m_C r_C(\delta)^2}{2}\right] = 8\exp[-\log(8/\delta)] = \delta.$$

The union bound does not require independence between slots. On the simultaneous event, the selected slot satisfies $\mathcal{L}_C(\widehat{\boldsymbol{c}}) - \mathcal{L}_C(\boldsymbol{c}_0) \le \widehat{d}_{j_{\mathrm{sel}}} + r_C(\delta)$. Proposition 1 gives

$$\frac{\mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}; \mathcal{G}_T)}{\binom{N}{2}} = \frac{\mathcal{E}_A(\mathcal{G}_T)}{\binom{N}{2}} + \mathcal{L}_C(\boldsymbol{c}; \mathcal{G}_T),$$
$$\frac{\mathcal{E}_{\mathrm{pair}}(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}_0; \mathcal{G}_T)}{\binom{N}{2}} = \mathcal{L}_C(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{L}_C(\boldsymbol{c}_0; \mathcal{G}_T).$$

Thus, a negative observed upper bound proves both strict decreases. ∎

The observed difference-calibration bias is simply the identity

$$\widehat{b}_j = \widehat{d}_j - [\mathcal{J}_C(\boldsymbol{c}^{[j]}) - \mathcal{J}_C(\boldsymbol{c}_0)].$$

Substitution rewrites the same certificate as $\mathcal{J}_C(\widehat{\boldsymbol{c}}) - \mathcal{J}_C(\boldsymbol{c}_0) + \widehat{b}_{j_{\mathrm{sel}}} + r_C(\delta) < 0$; it does not provide an additional calibration guarantee for the surrogate.

E.2 End-to-end audit against a pre-refinement baseline.

Let $\mathcal{G}_0 = \{R_m\}_{m=1}^{M_0}$ be the feasible roots after hard size projection, and let $\mathcal{G}_T = \{G_g\}_{g=1}^{M_T} \preceq \mathcal{G}_0$ be the final atoms in the same structural call. For this optional diagnostic, apply the same deterministic base quotient operator in the quotient construction above directly to $\mathcal{G}_0$ to obtain $\boldsymbol{c}_{\mathrm{pre}} \in \mathcal{P}_{K,M_0}$, conditioning on a finite embedding and K nonempty granular balls. This label-free comparator is fixed before the audit and is never used to generate, rank, accept, or select split or connection candidates. Each final atom belongs to a unique root $R_{p(g)}$, so we lift the baseline by

$$[\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}]_g = [\boldsymbol{c}_{\mathrm{pre}}]_{p(g)}, \qquad g \in [M_T].$$

Condition on the fixed dataset, representations, graphs, $\mathcal{G}_0$, $\mathcal{G}_T$, $\boldsymbol{c}_{\mathrm{pre}}$, and $\widehat{\boldsymbol{c}}$ before accessing audit labels. Choose $m_E \ge 1$ using label-free information only, fix $\delta \in (0,1)$, and draw a fresh set of m_E pairs independently and uniformly with replacement from all $\binom{N}{2}$ pairs. Using the final-atom loss defined above, let

$$\widehat{d}_E = \frac{1}{m_E}\sum_{t=1}^{m_E}\left[\ell_{\widehat{\boldsymbol{c}}}(I_t, J_t) - \ell_{\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}}(I_t, J_t)\right],$$
$$r_E(\delta) = \sqrt{\frac{2\log(1/\delta)}{m_E}}.$$

Corollary 2 (Audit-certified refinement–connection improvement) *Under this audit protocol, the complete structural-stage change satisfies the deterministic identity*

$$\frac{\mathcal{E}_{\mathrm{pair}}(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}_{\mathrm{pre}}; \mathcal{G}_0)}{\binom{N}{2}} = \mathcal{L}_C(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{L}_C(\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}; \mathcal{G}_T).$$

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

$$\mathcal{L}_C(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{L}_C(\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}; \mathcal{G}_T) \le \widehat{d}_E + r_E(\delta).$$

Consequently, a negative observed upper bound certifies

$$\mathcal{E}_{\mathrm{pair}}(\widehat{\boldsymbol{c}}; \mathcal{G}_T) < \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}_{\mathrm{pre}}; \mathcal{G}_0).$$

Proof. The lift assigns each sample its original root's baseline label and hence preserves the baseline sample partition:

$$\mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}^{\uparrow}_{\mathrm{pre}}; \mathcal{G}_T) = \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}_{\mathrm{pre}}; \mathcal{G}_0).$$

Comparing both predictions on $\mathcal{G}_T$ now cancels the common atomic term by Proposition 1, proving the deterministic identity. Pairs crossing final children of the same root remain in the lifted baseline's connection term, so the refinement-induced change is retained. Set $X_t = \ell_{\widehat{\boldsymbol{c}}}(I_t, J_t) - \ell_{\boldsymbol{c}^{\uparrow}_{\mathrm{pre}}}(I_t, J_t)$. Then $X_t \in [-1, 1]$, its empirical mean is $\widehat{d}_E$, and its expectation is the connection-loss difference in the identity. One-sided Hoeffding gives

$$\begin{aligned}\Pr\Big\{\mathbb{E}[X_t] > \widehat{d}_E + r_E(\delta)\Big\} &\le \exp\left[-\frac{m_E r_E(\delta)^2}{2}\right] \\ &= \exp[-\log(1/\delta)] = \delta.\end{aligned}$$

Combining this event with the identity proves the claim. ∎

Appendix F.4 gives the connection proof and further details for both audits. The probabilities here come only from independent post-selection sampling on the fixed dataset; neither audit gives a generalization guarantee for unseen data.

F Theoretical Proofs and Strengthened Results

F.1 Proof of the Structural-Error Decomposition

Proof of Proposition 1. Fix an atomization $\mathcal{G} = \{G_m\}_{m=1}^{M}$ and an atom-label vector $\boldsymbol{c}$, so that $\widehat{y}_i = c_m$ for $i \in G_m$. Within an atom, every pair receives the same prediction. Between two atoms, the prediction agrees exactly when their assigned labels agree. Partitioning all unordered sample pairs accordingly gives

$$\begin{aligned}\mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}; \mathcal{G}) &= \sum_{m=1}^{M} \sum_{\substack{i<j \\ i,j \in G_m}} \mathbb{I}[y_i \neq y_j] + \sum_{m<n} \sum_{\substack{i \in G_m \\ j \in G_n}} |\mathbb{I}[c_m = c_n] - \mathbb{I}[y_i = y_j]| \\ &= \sum_{m=1}^{M} D(G_m) + \sum_{m<n} [\mathbb{I}[c_m = c_n] D(G_m, G_n) + \mathbb{I}[c_m \neq c_n] S(G_m, G_n)] \\ &= \mathcal{E}_A(\mathcal{G}) + \binom{N}{2} \mathcal{L}_C(\boldsymbol{c}; \mathcal{G}) \ge \mathcal{E}_A(\mathcal{G}).\end{aligned}$$

This proves the decomposition and its lower bound. The refinement identity follows from the pair-count argument in the proof of Theorems 1 and 3 below.

Changing the co-clustering indicator of (G_m, G_n) from different to same changes that pair block's contribution by $D(G_m, G_n) - S(G_m, G_n)$. For pure atoms from different classes, this difference is $|G_m||G_n|$. It counts the errors introduced within their pair block; a larger reassignment may also change relations with other atoms. ∎

F.2 Proof of Exact Quotient Preservation

Proof of Lemma 2. Let $\boldsymbol{h}_k \in \{0, 1\}^M$ indicate the atoms in cluster k, and let $\boldsymbol{u}_k = \boldsymbol{S}\boldsymbol{h}_k$ indicate its samples. Since every sample belongs to exactly one atom, $\boldsymbol{S}\mathbf{1}_M = \mathbf{1}_N$. Substituting $\boldsymbol{P}^Q = \boldsymbol{S}^\top \boldsymbol{P} \boldsymbol{S}$ into the cut and volume definitions gives

$$\begin{aligned}\mathrm{cut}_{\boldsymbol{P}}(k, \bar{k}) &= \boldsymbol{u}_k^\top \boldsymbol{P}(\mathbf{1}_N - \boldsymbol{u}_k) \\ &= \boldsymbol{h}_k^\top \boldsymbol{S}^\top \boldsymbol{P} \boldsymbol{S}(\mathbf{1}_M - \boldsymbol{h}_k) \\ &= \boldsymbol{h}_k^\top \boldsymbol{P}^Q(\mathbf{1}_M - \boldsymbol{h}_k) = \mathrm{cut}_{\boldsymbol{P}^Q}(k, \bar{k}), \\ \mathrm{vol}_{\boldsymbol{P}}(k) &= \boldsymbol{u}_k^\top \boldsymbol{P} \mathbf{1}_N \\ &= \boldsymbol{h}_k^\top \boldsymbol{S}^\top \boldsymbol{P} \boldsymbol{S} \mathbf{1}_M \\ &= \boldsymbol{h}_k^\top \boldsymbol{P}^Q \mathbf{1}_M = \mathrm{vol}_{\boldsymbol{P}^Q}(k).\end{aligned}$$

The volume threshold, clipped cut-to-volume ratio and set of active clusters therefore agree on both graphs. Averaging over these clusters yields $\mathrm{NCut}_{\boldsymbol{P}}(\boldsymbol{c}) = \mathrm{NCut}_{\boldsymbol{P}^Q}(\boldsymbol{c})$. The diagonal masses P^Q_{gg} contribute to volume, although they vanish from the cut because $h_{gk}(1 - h_{gk}) = 0$, and must therefore be retained. The equality concerns atom-constant partitions and does not identify the spectral embeddings or unconstrained optima of the two graphs. ∎

F.3 Proof of Atomic Refinement and Its Semantic Certificate

We first record that the potential candidate tree is finite: every proposed split has nonempty disjoint children, the minimum splittable size is four, and the recursion has a fixed maximum depth. The definitions are acyclic. Fixed representations and masks determine the transformed views, $\boldsymbol{P}$, and $\boldsymbol{R}$; these determine the potential candidates, their evidence, and the K-way probes; evidence ranking fixes π_G and ϕ_R; the pilot then fixes V_B, M_{pilot}, and ρ; only afterward are g_B, g_S, and the unified decisions evaluated. The pilot does not depend on ρ or g_S.

Proof of Lemma 1. Fix a feasible root R and use backward induction on its finite pilot tree. A node without a proposal has $V_B(G) = \phi_R(G) \geq 0$. At any other node, the induction hypothesis makes the child sum $S_G = \sum_{H \in \pi_G} V_B(H)$ nonnegative. If the pilot accepts, its acceptance condition gives

$$0 \leq V_B(G) = S_G < \phi_R(G) - \epsilon_A < \phi_R(G).$$

Otherwise, it sets $V_B(G) = \phi_R(G)$. Both branches thus satisfy $0 \leq V_B(G) \leq \phi_R(G)$, and the accepted branch satisfies the claimed strict inequality. Induction over the finite tree completes the proof. ∎

The quantities used below have different roles. The true-label quantity $\mathcal{E}_A$ measures atom-internal pair error; ϕ_R, V_B, and $\mathcal{J}_A$ are label-free algorithmic scores; and $\Phi^\star_R$ is a true-label semantic leaf cost used only for analysis. The range just proved is a property of the frozen pilot. The rank bounds in the candidate-objective bounds and rank margin justify pruning positive-rank proposals under the defined conceptual $\mathcal{J}_A$ in exact arithmetic; they do not imply descent of $\mathcal{E}_A$ or $\Phi^\star_R$. In particular, the structural-dominant branch may override a negative base gain.

Proof of Theorems 1 and 3. Fix one structural call and suppose that a proper refinement replaces G by nonempty, disjoint children $\{H_r\}_{r=1}^q$. Every different-class pair in G lies either within one child or between two children. All other atoms remain unchanged, so

$$\begin{aligned}
\mathcal{E}_A(\mathcal{G}_{\mathrm{before}}) - \mathcal{E}_A(\mathcal{G}_{\mathrm{after}}) &= D(G) - \sum_{r=1}^{q} D(H_r) \\
&= \sum_{r<t} \sum_{i \in H_r} \sum_{j \in H_t} \mathbb{I}[y_i \neq y_j] \\
&= \sum_{r<t} D(H_r, H_t) \geq 0.
\end{aligned}$$

The decrease is strict exactly when the boundary separates a different-class pair. To establish majority-purity monotonicity, let $n_{rc} = |\{i \in H_r : y_i = c\}|$ and choose $c^\star \in \arg\max_c \sum_r n_{rc}$. Then

$$\begin{aligned}
N[\mathrm{Pur}_{\mathrm{maj}}(\mathcal{G}_{\mathrm{after}}) - \mathrm{Pur}_{\mathrm{maj}}(\mathcal{G}_{\mathrm{before}})] &= \sum_{r=1}^{q} \max_c n_{rc} - \sum_{r=1}^{q} n_{rc^\star} \\
&\geq \sum_{r=1}^{q} n_{rc^\star} - \sum_{r=1}^{q} n_{rc^\star} = 0.
\end{aligned}$$

These two monotonicities need not be strict simultaneously.

For a specified post-projection local binary action $a = (G, \{G_L, G_R\}) \in \mathcal{A}_{\mathrm{loc}}$, the calibration event gives $\mu_a \geq e_a - \eta_a$. Since $\mu_a \geq 0$, we obtain

$$\begin{aligned} B_a &= D(G) - D(G_L) - D(G_R) \\ &= D(G_L, G_R) = \frac{|G|^2}{4}\mu_a \\ &\geq \frac{|G|^2}{4}[e_a - \eta_a]_+. \end{aligned}$$

Dividing by the positive root normalization Q_R gives the pair-purity gain B_a/Q_R and its stated lower bound.

The parent satisfies $|G| \geq 4$, so it has no singleton penalty. The immediate split adds one leaf and s_a singleton children, while all other leaf contributions to $\Phi_R^\star$ cancel. Therefore,

$$\begin{aligned} \Delta_a^\star &= \Phi_R^\star(\mathcal{T}_{\mathrm{before}}) - \Phi_R^\star(\mathcal{T}_{\mathrm{after}}) \\ &= \big[D(G) + \lambda_A Q_R\big] - \big[D(G_L) + D(G_R) + 2\lambda_A Q_R + \xi_A Q_R s_a\big] \\ &= B_a - Q_R(\lambda_A + \xi_A s_a) \\ &\geq \frac{|G|^2}{4}[e_a - \eta_a]_+ - Q_R(\lambda_A + \xi_A s_a). \end{aligned}$$

A positive right-hand side certifies strict semantic-cost reduction for this local binary action. This conclusion requires calibration and does not follow from the acceptance gate, whose structural branch can override a negative base gain.

To compare total errors, give every child its parent's predicted label and denote the resulting assignment by $\boldsymbol{c}^\uparrow$. The sample predictions remain unchanged. Applying the decomposition already proved in Proposition 1 gives

$$\begin{aligned} 0 &= \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}^\uparrow; \mathcal{G}_{\mathrm{after}}) - \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}; \mathcal{G}_{\mathrm{before}}) \\ &= -\sum_{r<t} D(H_r, H_t) + \binom{N}{2}\left[\mathcal{L}_C(\boldsymbol{c}^\uparrow; \mathcal{G}_{\mathrm{after}}) - \mathcal{L}_C(\boldsymbol{c}; \mathcal{G}_{\mathrm{before}})\right]. \end{aligned}$$

Thus, reducing the atomic term alone does not reduce total pairwise error.

Finally, let M_t be the leaf count after t accepted refinements, starting from M_0 feasible roots. Each refinement replaces one leaf by $q_s \geq 2$ nonempty children, and accepted parents are not restored within the call. Hence,

$$M_0 + t \leq M_0 + \sum_{s=1}^{t}(q_s - 1) = M_t \leq N.$$

It follows that $t \leq N - M_0$, covering both local and terminal-guide refinements within each input-space or learned-space structural call. ∎

Independent certification of the simultaneous calibration event. Freeze all graphs, candidates, evidence scores and label-free decisions, and fix an audit family $\mathcal{A}_{\mathrm{aud}} \subseteq \mathcal{A}_{\mathrm{loc}}$. If the family is empty, its simultaneous event is Ω and no samples are drawn. Otherwise, let $L = |\mathcal{A}_{\mathrm{aud}}| \geq 1$ and choose $\delta \in (0,1)$ and $m_a \geq 1$ without audit labels. For each action, draw m_a pairs independently and uniformly with replacement from $G_L \times G_R$, and write $X_{at} = \mathbb{I}[y_{I_{at}} \neq y_{J_{at}}]$. Define

$$\alpha_a = \frac{4|G_L||G_R|}{|G|^2}, \qquad \widehat{\mu}_a = \frac{\alpha_a}{m_a}\sum_{t=1}^{m_a} X_{at}, \qquad r_a(\delta) = \alpha_a\sqrt{\frac{\log(L/\delta)}{2m_a}}.$$

Since $\mu_a = \alpha_a \mathbb{E}[X_{at}]$ and $X_{at} \in [0,1]$, one-sided Hoeffding followed by a union bound gives

$$\begin{aligned} \Pr\{\exists a \in \mathcal{A}_{\mathrm{aud}} : \mu_a < \widehat{\mu}_a - r_a(\delta)\} &\leq \sum_{a \in \mathcal{A}_{\mathrm{aud}}} \exp\left[-2m_a\left(\frac{r_a(\delta)}{\alpha_a}\right)^2\right] \\ &= \sum_{a \in \mathcal{A}_{\mathrm{aud}}} \frac{\delta}{L} = \delta. \end{aligned}$$

Set $\eta_a^{\mathrm{cert}}(\delta) = [e_a - \widehat{\mu}_a + r_a(\delta)]_+$. On the resulting simultaneous event, we have

$$\begin{aligned} e_a - \eta_a^{\mathrm{cert}}(\delta) &= e_a - [e_a - \widehat{\mu}_a + r_a(\delta)]_+ \\ &= \min\{e_a, \widehat{\mu}_a - r_a(\delta)\} \\ &\le \widehat{\mu}_a - r_a(\delta) \le \mu_a. \end{aligned}$$

Taking $\mathcal{A}_{\mathrm{aud}} = \mathcal{A}_{\mathrm{loc}}$ therefore certifies the full calibration event. Auditing only the fixed accepted local actions certifies the actionwise bounds for those actions alone. All probabilities condition on the fixed dataset labels and label-free outputs. The randomness comes from the audit draws, rather than the deterministic perturbations, and audit labels are never used for splitting.

F.4 Proof of Risk Attenuation, Bridge Suppression, and Fallback

Three risk quantities are used and are not interchangeable. The raw cross-view risk $u_{ij} \in [0,1]$ enters the implemented relation $R_{ij} = P_{ij}u_{ij}$; the stabilized sample risk is $\widetilde{u}_{ij} = R_{ij}/d_\epsilon(P_{ij})$; and the quotient risk is $U_e = R_e^Q/d_\epsilon(P_e^Q)$. The identity expressing U_e as a P_{ij}-weighted average of the raw u_{ij} is used below only under $P_e^Q \ge \epsilon$.

Proof of Theorems 2 and 4. Let $\beta > 0$, $\epsilon > 0$ and $d_\epsilon(p) = \max(p, \epsilon)$. For a positive quotient edge e, write $p = P_e^Q$, $r = R_e^Q$ and $U_e = r/d_\epsilon(p)$. Nonnegativity gives $0 < A_e^Q = pe^{-\beta U_e} \le p$. Holding p, β and ϵ fixed, we have

$$\frac{\partial A_e^Q}{\partial r} = -\frac{\beta p}{d_\epsilon(p)} e^{-\beta r/d_\epsilon(p)} < 0,$$

$$0 \le p - A_e^Q = p\left(1 - e^{-\beta r/d_\epsilon(p)}\right) \le \frac{\beta p r}{d_\epsilon(p)} \le \beta r,$$

where the second line uses $1 - e^{-x} \le x$ for $x \ge 0$. For positive edges b, g with $U_b - U_g \ge \delta \ge 0$, the same definition yields

$$\frac{A_b^Q}{A_g^Q} = \frac{P_b^Q}{P_g^Q} e^{-\beta(U_b - U_g)} \le e^{-\beta\delta} \frac{P_b^Q}{P_g^Q}.$$

To compare calibration before and after aggregation, fix an atom pair with $q = P_{gh}^Q > 0$. Let $\mathcal{I}_{gh} = \{(i,j) : i \in G_g, j \in G_h, P_{ij} > 0\}$, $w_{ij} = P_{ij}/q$ and $\widetilde{u}_{ij} = R_{ij}/d_\epsilon(P_{ij})$. These weights sum to one, and $P_{ij} \le q$. Since $p/d_\epsilon(p)$ is nondecreasing for $p \ge 0$, and $P_{ij} = 0$ implies $R_{ij} = 0$, we obtain

$$\begin{aligned} \sum_{(i,j)\in\mathcal{I}_{gh}} w_{ij}\widetilde{u}_{ij} &= \sum_{(i,j)\in\mathcal{I}_{gh}} \frac{R_{ij}}{q} \frac{P_{ij}}{d_\epsilon(P_{ij})} \\ &\le \sum_{(i,j)\in\mathcal{I}_{gh}} \frac{R_{ij}}{d_\epsilon(q)} \\ &= \frac{R_{gh}^Q}{d_\epsilon(q)} = \widetilde{u}_{gh}^Q. \end{aligned}$$

The function $u \mapsto e^{-\beta u}$ is decreasing and convex. Applying monotonicity and then Jensen's inequality gives

$$\begin{aligned} A_{gh}^Q &= qe^{-\beta\widetilde{u}_{gh}^Q} \\ &\le q \exp\left[-\beta \sum_{(i,j)\in\mathcal{I}_{gh}} w_{ij}\widetilde{u}_{ij}\right] \\ &\le q \sum_{(i,j)\in\mathcal{I}_{gh}} w_{ij} e^{-\beta\widetilde{u}_{ij}} \\ &= \sum_{(i,j)\in\mathcal{I}_{gh}} P_{ij} e^{-\beta R_{ij}/d_\epsilon(P_{ij})}. \end{aligned}$$

This proves the aggregation inequality, including the regime $q < \epsilon$.

For fixed nonnegative impact weights h_e, the relation $A_e^Q = P_e^Q \zeta_e$ with $0 < \zeta_e \leq 1$ also gives

$$\sum_e h_e A_e^Q Z_e = \sum_e h_e P_e^Q \zeta_e Z_e \leq \sum_e h_e P_e^Q Z_e.$$

Here $Z_e = D_{gh}/(n_g n_h)$ measures different-class pairs across the entire atom pair, including pairs without a positive sample edge. Using the weights π_e from the exposure definition, whose normalizing mass is assumed positive, we have

$$\begin{aligned}\Theta_\beta - \Theta_0 &= \frac{\mathbb{E}_\pi[\zeta_E Z_E]}{\mathbb{E}_\pi[\zeta_E]} - \mathbb{E}_\pi[Z_E] \\ &= \frac{\mathbb{E}_\pi[\zeta_E Z_E] - \mathbb{E}_\pi[\zeta_E]\mathbb{E}_\pi[Z_E]}{\mathbb{E}_\pi[\zeta_E]} \\ &= \frac{\mathrm{Cov}_\pi(\zeta_E, Z_E)}{\mathbb{E}_\pi[\zeta_E]}.\end{aligned}$$

The denominator is positive, so nonpositive covariance is necessary and sufficient for normalized exposure not to increase. To establish the stated sufficient condition, let $m(u) = \mathbb{E}_\pi[Z_E \mid U_E = u]$ and let U' be an independent copy of U_E. Conditioning on U_E yields

$$\begin{aligned}2\,\mathrm{Cov}_\pi(\zeta_E, Z_E) &= 2\,\mathrm{Cov}_\pi(e^{-\beta U_E}, m(U_E)) \\ &= \mathbb{E}\Big[(e^{-\beta U_E} - e^{-\beta U'})(m(U_E) - m(U'))\Big] \\ &\leq 0\end{aligned}$$

whenever m is nondecreasing on the risk support. The last inequality holds because the factors have opposite signs, proving $\Theta_\beta \leq \Theta_0$ under the calibration condition.

Now impose the purity, co-observation and support–dispersion assumptions of the theorem. Substitution into the raw risk definition gives $u_{ij} \leq u_+$ on positive same-class edges and $u_{ij} \geq u_-$ on positive cross-class edges. For $e = (g, h)$, the assumptions $P_e^Q \geq \epsilon$ and $R_{ij} = P_{ij} u_{ij}$ give

$$\begin{aligned}U_e &= \frac{R_e^Q}{P_e^Q} \\ &= \frac{\sum_{i \in G_g, j \in G_h} P_{ij} u_{ij}}{\sum_{i \in G_g, j \in G_h} P_{ij}}.\end{aligned}$$

Purity ensures that all contributing sample edges share their atom pair's class relation. The weighted average therefore satisfies $U_e \leq u_+$ for same-class atom pairs and $U_e \geq u_-$ for cross-class pairs. Since $u_+ < u_-$ and Z_e is binary for pure atoms, its conditional mean is nondecreasing on the weighted risk support. Moreover,

$$\begin{aligned}U_e \leq u_+ \quad &\Longrightarrow \quad A_e^Q \geq e^{-\beta u_+} P_e^Q, && y_g = y_h, \\ U_e \geq u_- \quad &\Longrightarrow \quad A_e^Q \leq e^{-\beta u_-} P_e^Q, && y_g \neq y_h.\end{aligned}$$

Multiplying by the fixed $h_e \geq 0$, summing over each class of edges and using the positive within-class denominator gives

$$\begin{aligned}\frac{\sum_{e: y_g \neq y_h} h_e A_e^Q}{\sum_{e: y_g = y_h} h_e A_e^Q} &\leq \frac{e^{-\beta u_-} \sum_{e: y_g \neq y_h} h_e P_e^Q}{e^{-\beta u_+} \sum_{e: y_g = y_h} h_e P_e^Q} \\ &= e^{-\beta \delta_u} \frac{\sum_{e: y_g \neq y_h} h_e P_e^Q}{\sum_{e: y_g = y_h} h_e P_e^Q}.\end{aligned}$$

If the original cross-class mass is positive, the ratio strictly decreases; otherwise, both cross-class masses remain zero.

Finally, condition on $\mathcal{F}_C$, so all five candidate slots and scores are defined. The selected proposal minimizes the four nonbase scores, but is accepted only if it passes the trust checks and improves $\mathcal{J}_C$

by more than ϵ_C. Otherwise, the rule returns $\boldsymbol{c}_0$. Thus $\mathcal{J}_C(\widehat{\boldsymbol{c}}) \le \mathcal{J}_C(\boldsymbol{c}_0)$, and an accepted nonbase proposal minimizes all five indexed scores. Under fallback, the base need not minimize those scores. This comparison uses the base already obtained from the calibrated graph; the guarantees concern affinity exposure and the fixed surrogate, not true connection loss. ∎

Proof of Corollary 1. Fix the graph, atoms, base, four nonbase slots and selection rule before drawing audit pairs. For each slot j, define $X_t^{(j)} = \ell_{\boldsymbol{c}^{[j]}}(I_t, J_t) - \ell_{\boldsymbol{c}_0}(I_t, J_t)$. Independent uniform sampling with replacement over all $\binom{N}{2}$ pairs gives

$$X_t^{(j)} \in [-1, 1], \qquad d_j := \mathbb{E}[X_t^{(j)}] = \mathcal{L}_C(\boldsymbol{c}^{[j]}) - \mathcal{L}_C(\boldsymbol{c}_0), \qquad \widehat{d}_j = \frac{1}{m_C} \sum_{t=1}^{m_C} X_t^{(j)}.$$

Since the range length is two, two-sided Hoeffding and a union bound yield

$$\begin{aligned} \Pr\left\{ \max_{1 \le j \le 4} |\widehat{d}_j - d_j| > r_C(\delta) \right\} &\le \sum_{j=1}^{4} 2 \exp\left[-\frac{m_C r_C(\delta)^2}{2} \right] \\ &= 8 \exp\left[-\frac{m_C}{2} \frac{2 \log(8/\delta)}{m_C} \right] = \delta. \end{aligned}$$

The same pair draws may serve all four slots, since the union bound does not require independence between slots. On the simultaneous event, a selected nonbase slot satisfying the observed certificate obeys

$$\begin{aligned} \mathcal{L}_C(\widehat{\boldsymbol{c}}) - \mathcal{L}_C(\boldsymbol{c}_0) &= d_{j_{\mathrm{sel}}} \\ &\le \widehat{d}_{j_{\mathrm{sel}}} + r_C(\delta) < 0. \end{aligned}$$

Both assignments use the same atoms, so their common atomic term cancels in Proposition 1. Their normalized total pairwise error therefore has the same decrease. Finally, substituting $\widehat{d}_j = \mathcal{J}_C(\boldsymbol{c}^{[j]}) - \mathcal{J}_C(\boldsymbol{c}_0) + \widehat{b}_j$ proves the equivalent certificate in terms of the observed bias. This substitution is an identity and does not independently calibrate the surrogate. ∎

Proof of Corollary 2. Since $\mathcal{G}_T \preceq \mathcal{G}_0$, each final atom belongs to a unique root. Let $g_T(i)$ and $g_0(i)$ identify the final atom and root containing sample i. By the definition of the lifted baseline,

$$[\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}]_{g_T(i)} = [\boldsymbol{c}_{\mathrm{pre}}]_{p(g_T(i))} = [\boldsymbol{c}_{\mathrm{pre}}]_{g_0(i)}.$$

The lifted and original baselines thus induce exactly the same sample partition. Both predictions can now be compared on $\mathcal{G}_T$, where their common atomic term cancels:

$$\begin{aligned} &\frac{\mathcal{E}_{\mathrm{pair}}(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}_{\mathrm{pre}}; \mathcal{G}_0)}{\binom{N}{2}} \\ &= \frac{\mathcal{E}_{\mathrm{pair}}(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}; \mathcal{G}_T)}{\binom{N}{2}} \\ &= \mathcal{L}_C(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{L}_C(\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}; \mathcal{G}_T). \end{aligned}$$

Pairs separated into different children of one root remain included in the lifted baseline's connection term, so no refinement contribution has been discarded.

Fix the compared objects, m_E and δ without audit labels, and draw the fresh pairs independently and uniformly with replacement. Set $X_t = \ell_{\widehat{\boldsymbol{c}}}(I_t, J_t) - \ell_{\boldsymbol{c}_{\mathrm{pre}}^{\uparrow}}(I_t, J_t)$. Its expectation is the displayed loss difference, its empirical mean is $\widehat{d}_E$, and $X_t \in [-1, 1]$. One-sided Hoeffding therefore gives

$$\begin{aligned} \Pr\left\{ \mathbb{E}[X_t] > \widehat{d}_E + r_E(\delta) \right\} &\le \exp\left[-\frac{m_E r_E(\delta)^2}{2} \right] \\ &= \exp\left[-\frac{m_E}{2} \frac{2 \log(1/\delta)}{m_E} \right] = \delta. \end{aligned}$$

Combining this event with the common-atomization identity yields, with probability at least $1 - \delta$,

$$\frac{\mathcal{E}_{\mathrm{pair}}(\widehat{\boldsymbol{c}}; \mathcal{G}_T) - \mathcal{E}_{\mathrm{pair}}(\boldsymbol{c}_{\mathrm{pre}}; \mathcal{G}_0)}{\binom{N}{2}} \le \widehat{d}_E + r_E(\delta).$$

A negative observed upper bound proves the claimed strict improvement over the pre-refinement baseline. ∎

F.5 Scope of the Guarantees

General deterministic structural facts. Once labels used only for analysis and all label-free objects are fixed, Proposition 1 is an exact identity. Every proper refinement makes $\mathcal{E}_A$ nonincreasing and majority purity nondecreasing, while finite termination follows independently from strict leaf-count growth. These two refinement monotonicities hold for any proper refinement and therefore do not prove that the proposed evidence ranking or acceptance gate is superior to another rule. For atom-constant partitions, $\boldsymbol{P}^{Q} = \boldsymbol{S}^{\top}\boldsymbol{P}\boldsymbol{S}$ preserves cut, volume, and the paper's clipped active-cluster normalized cut exactly. It does not preserve spectral eigenvectors, embeddings, K-means outputs, or unconstrained optima.

Deterministic mechanism properties. For fixed positive quotient mass, risk calibration changes the baseline $A_e^Q(0) = P_e^Q$ to $A_e^Q(\beta) \le P_e^Q$, with the derivative, absolute gap, relative attenuation, and aggregation inequality proved above. Absolute pollution-weighted exposure cannot increase, but this fact alone does not order its normalized version. On $\mathcal{F}_C$, fallback guarantees only $\mathcal{J}_C(\widehat{\boldsymbol{c}}) \le \mathcal{J}_C(\boldsymbol{c}_0)$. Because $\boldsymbol{c}_0$ is itself computed on the calibrated graph, the comparison $\boldsymbol{c}_0 \mapsto \widehat{\boldsymbol{c}}$ evaluates candidate selection rather than the change $\beta = 0 \mapsto \beta > 0$.

Conditional semantic bridges. On $\mathcal{E}_A^{\mathrm{cal}}$, a specified post-projection local binary action strictly reduces $\Phi_R^{\star}$ only when the final lower bound in the semantic-gain bound is positive. This certificate does not cover hard projection or terminal-guide refinement. Normalized pollution is nonincreasing exactly when the covariance in the covariance identity is nonpositive; monotonic conditional error in the quotient risk is one sufficient condition. Under pure atoms, co-observation, positive quotient-mass thresholds, support–dispersion separation with $\delta_u > 0$, fixed nonnegative impacts, and a positive within-class denominator, the bridge-ratio bound controls the cross-class-to-within-class exposure ratio on existing positive quotient edges. It does not by itself control $\mathcal{L}_C$.

Independent-label audit certificates and exclusions. The local-action audit in Appendix F.3 and Corollaries 1 and 2 are post-selection finite-sample certificates on one fixed dataset after all audited objects have been frozen. The end-to-end audit compares the complete refinement–connection output with a label-free pre-refinement baseline within one structural call and one representation space. If audits share a label budget, their confidence parameters must be allocated to control the desired familywise failure probability. Without their stated conditions, the theory does not claim that the evidence gate dominates other splits, that every accepted split reduces total true error, that risk scores unconditionally identify false cross-class edges, or that surrogate fallback lowers $\mathcal{L}_C$. It provides neither unseen-data generalization nor monotonic clustering accuracy, global optimality over all K-partitions, neural feedback convergence, or joint convergence of the neural and discrete updates.

G Mechanistic Evaluation of Splitting and Connection

G.1 Datasets

Table 4 describes the local dataset versions used in these experiments. A view denotes a feature representation of the same samples. The collection covers human actions, faces, handwritten digits, objects, biological data, scientific documents, and clothing images.

G.2 Experimental Protocol and Figure Conventions

All experiments use the default experimental configuration. We examine three successive decisions: where GBAS places splitting boundaries, how GBAC converts the resulting freedom to reassign atoms into net error correction, and how its partition selection improves on base merging.

Each comparison uses ten paired runs with random seeds 42–51. The horizontal axis, *Random seed / paired mean*, shows these runs with selected seed labels, followed by their mean gain at *Mean*. All vertical values are gains in percentage points (pp): positive values indicate improvement and negative values indicate deterioration. Green and orange-red bars denote positive and negative gains, respectively; dots on the zero line denote zero gains. A blue diamond and its error bar show

Table 4: Dataset versions used in the mechanism experiments. Counts refer to the local experimental data.

Dataset	Samples	Classes	Views	Description
DHA	483	23	2	Human actions represented by two feature sets.
ORL	400	40	3	Face images with variations in illumination and expression.
Mfeat	2,000	10	6	Handwritten digits described by six feature sets.
Animal	11,673	20	4	An Animals with Attributes image subset with four feature sets.
Caltech101-7	1,474	7	6	Seven object categories from Caltech101 with six visual feature sets.
COIL20	1,440	20	3	Object images represented by grayscale, LBP, and Gabor features.
MNIST-10k	10,000	10	3	Handwritten digits 0–9 with three feature representations.
NUS-WIDE	2,400	12	5	Twelve animal categories, with 200 images per category and five visual feature sets.
Prokaryotic	551	4	3	Prokaryotic organisms represented by gene composition, proteome composition, and text.
ALOI	10,800	100	4	Object images under varying capture conditions, with four color and texture feature sets.
100Leaves	1,600	100	3	Sixteen leaf images per species, described by shape, fine-scale margin, and texture features.
CiteSeer	3,312	6	2	Scientific documents represented by content and citation information.
Cora	2,708	7	4	Scientific documents represented by content and different citation relations.
Multi-FMNIST	10,000	10	3	A multi-view Fashion-MNIST subset with ten clothing categories and three image representations.

the paired mean and an unadjusted 95% bootstrap confidence interval obtained from 2,000 resamples of the ten paired differences. An annotation such as "9/10 positive" counts runs with strictly positive gains. Each panel uses its own vertical scale, so cross-dataset comparisons use the numerical values.

G.3 Boundary Quality at Matched Granularity

Experimental design. For each original granular ball, we match the multiset of child-block sizes between evidence-guided and geometric splitting. Let B be the number of different-class sample pairs separated into distinct child blocks, divided by the total number of sample pairs within the original granular balls included in the comparison. We measure

$$G_{\mathrm{boundary}} = 100\,(B_{\mathrm{evidence}} - B_{\mathrm{geometry}})\,. \tag{27}$$

Matching child counts and sizes isolates the effect of boundary placement at a fixed granularity.

Results and module roles. Figure 7 shows positive mean gains on all nine displayed datasets. ORL, DHA, and Prokaryotic gain 0.6482, 0.2698, and 0.2566 pp, respectively; Animal and MNIST-10k improve in all ten runs. These matched-size comparisons support GBAS's ability to place boundaries that better separate different classes. By reducing the different-class pairs retained within indivisible children, GBAS provides better-separated local units for GBAC's subsequent assignment decisions. This experiment establishes the local boundary contribution of splitting; the following experiment evaluates how connection uses the resulting units.

G.4 Net Error Correction after Reconnection

Experimental design. We compare two assignments on the same split child blocks. In the inherited assignment, every child retains its original granular ball's cluster membership, preserving the original sample partition. In the selected assignment, GBAC reconnects the children and applies

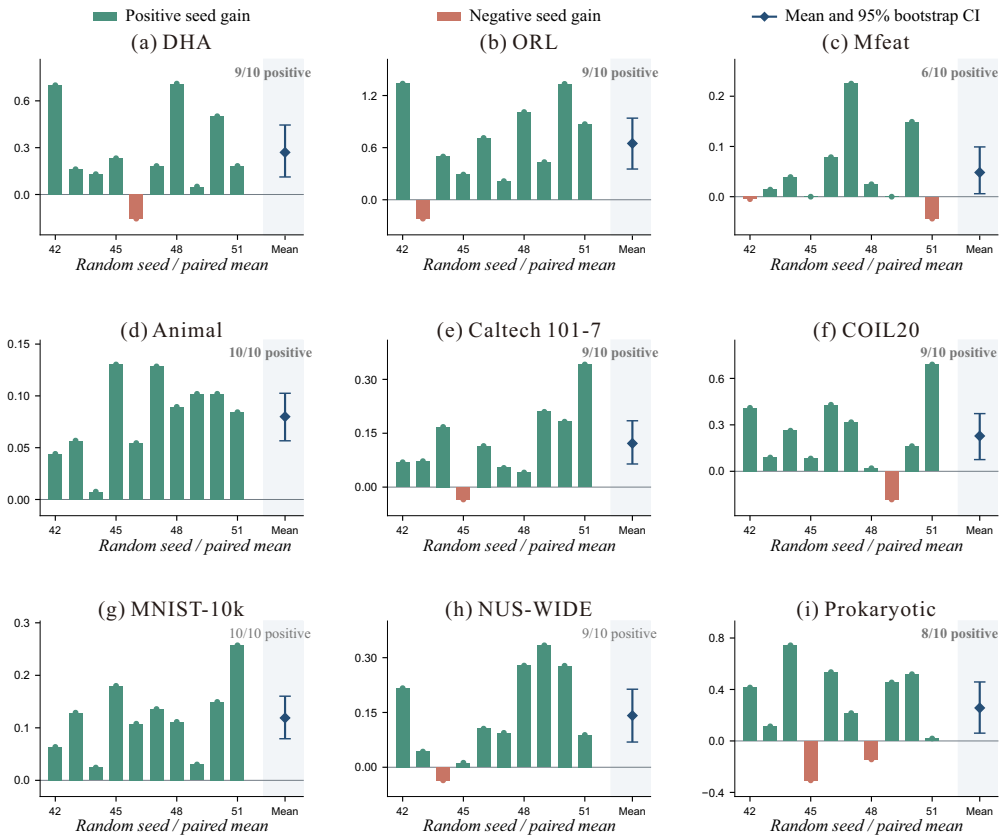


Figure 7: **Boundary quality at matched child-block sizes.** The vertical metric is *Cross-class pair separation gain (pp)*, $100(B_{\text{evidence}} - B_{\text{geometry}})$, normalized by all sample pairs within the compared original granular balls. Panels show ten paired runs and their mean with an unadjusted 95% bootstrap confidence interval. Positive values indicate that evidence-guided splitting separates more different-class pairs at the same granularity.

its partition selection rule. For a partition with sample assignments $\widehat{y}_i$ and evaluation labels y_i, the sample-pair error rate is

$$E = \frac{\sum_{i<j} \left(\mathbb{I}[y_i \neq y_j,\, \widehat{y}_i = \widehat{y}_j] + \mathbb{I}[y_i = y_j,\, \widehat{y}_i \neq \widehat{y}_j] \right)}{\binom{N}{2}}. \tag{28}$$

This rate jointly measures erroneous merges of different-class samples and erroneous separations of same-class samples over all sample pairs. The ground-truth labels are used for evaluation after the compared partitions have been obtained. We measure the reconnection gain as

$$G_{\text{reconnect}} = 100 \left(E_{\text{inherit}} - E_{\text{selected}} \right). \tag{29}$$

A positive gain means that corrected pairwise errors outweigh newly introduced errors in the selected partition.

Results and module roles. As shown in Figure 8, the mean pairwise error rate decreases by 0.0903 pp on ALOI, 0.0037 pp on 100Leaves, 1.8665 pp on CiteSeer, and 0.2377 pp on Multi-FMNIST. ALOI improves in all ten runs, showing consistent net correction. CiteSeer has the largest mean gain and greater variation across seeds, while 100Leaves exhibits a smaller mean improvement. The results connect the roles of the two modules: GBAS makes parts of an original granular ball independently assignable, and GBAC uses this freedom to reduce the combined cost of erroneous merges and separations. Together, these results show that the splitting–connection transition turns the reassignment freedom created by GBAS into net pairwise error correction, linking local structural refinement to improved global assignments.

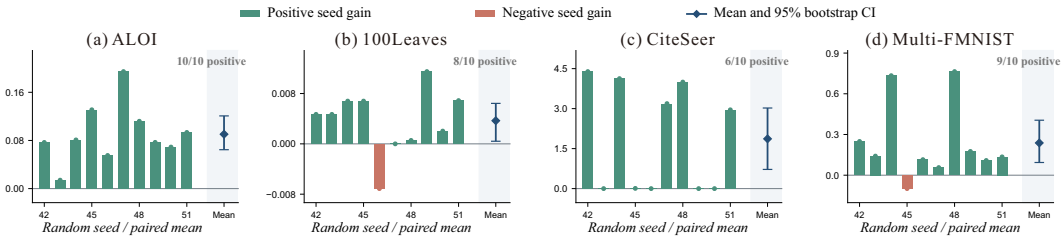


Figure 8: **Net error correction through reconnection.** The vertical metric is *Pairwise error reduction after reconnection (pp)*, $100(E_{\text{inherit}} - E_{\text{selected}})$. Both error rates count erroneous merges and separations over all sample pairs. The inherited and selected assignments use the same split child blocks. Blue diamonds and error bars show paired means and unadjusted 95% bootstrap confidence intervals over ten runs.

G.5 Additional Value of Connection Selection

Experimental design. We hold the split child blocks fixed and compare GBAC's base merging assignment $\boldsymbol{c}_0$ with its selected assignment $\widehat{\boldsymbol{c}}$. The base is the size-weighted spectral partition of the calibrated quotient graph; the selected assignment is the output of the candidate comparison and trust checks in Equation 9. Their gain is

$$G_{\text{selection}} = 100\,(E_{\text{base}} - E_{\text{selected}})\,. \tag{30}$$

This comparison measures the extra pairwise error reduction attributable to partition selection on the shared split structure.

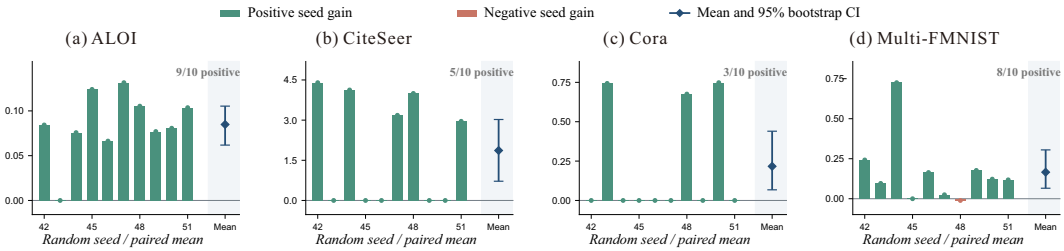


Figure 9: **Connection selection relative to base merging.** The vertical metric is *Pairwise error reduction over base merging (pp)*, $100(E_{\text{base}} - E_{\text{selected}})$, with each error rate normalized by all sample pairs. Both assignments use identical split child blocks. Positive, zero, and negative values denote improved, unchanged, and degraded pairwise error, respectively. Blue diamonds and error bars summarize the ten paired runs by their means and unadjusted 95% bootstrap confidence intervals.

Results and module roles. Figure 9 shows mean error reductions of 0.0848 pp on ALOI, 1.8658 pp on CiteSeer, 0.2167 pp on Cora, and 0.1658 pp on Multi-FMNIST. The mean gain is positive on all four datasets, although its seed-wise distribution differs across datasets: The mean gain is positive on all four datasets, although its seed-wise distribution differs across datasets: ALOI and Multi-FMNIST show positive gains in nine and eight runs, whereas the larger mean improvements on CiteSeer and Cora arise from a smaller subset of runs. This pattern is consistent with the role of selection as an additional correction step rather than the sole source of connection quality. This pattern is consistent with the role of selection as an additional correction step rather than the sole source of connection quality. With GBAS's output fixed, these comparisons identify an additional contribution from GBAC's choice among candidate partitions. GBAS defines the units available for reassignment, while GBAC's selection mechanism uses those units to improve the global grouping beyond the base connection on the displayed datasets.

G.6 Implications for GBASC

The three experiments trace complementary contributions across the two granularity transitions. At matched child-block sizes, GBAS improves the separation of different-class pairs and supplies better resolved local units. GBAC then uses these units to obtain net pairwise error reductions, with candidate selection providing further average gains over base merging. Together, the displayed results provide mechanism evidence for GBASC's joint control of local mixing and global assignment errors through adaptive splitting and risk-aware connection.

H Additional Comparative Experiments

Table 5: Additional comparisons on twelve benchmark configurations, part I. Entries are percentages reported as mean ± sample standard deviation. Higher values are better; the highest mean for each configuration and metric is boldfaced.

Method	CiteSeer				HW			
	ACC ↑	NMI ↑	ARI ↑	Purity ↑	ACC ↑	NMI ↑	ARI ↑	Purity ↑
AEMVC	23.2±1.1	2.9±1.0	0.7±0.4	25.3±1.5	87.3±5.4	91.5±2.3	83.0±5.8	89.6±3.4
BONE	51.1±2.6	28.1±2.7	26.5±3.1	53.9±3.1	88.4±1.3	88.1±1.0	83.8±0.8	88.4±1.3
CSOT	26.1±3.1	4.3±2.1	2.5±1.2	27.8±2.1	80.7±5.2	81.5±3.2	73.7±5.2	81.9±4.5
CVCL	36.4±3.1	15.0±1.8	11.5±1.9	40.3±3.2	73.7±2.7	70.4±1.0	61.1±1.9	76.3±1.0
DealMVC	36.0±4.9	15.4±4.5	11.3±3.7	40.7±4.4	73.9±5.6	73.1±2.8	63.1±5.0	74.1±5.7
DMAC	19.6±0.5	0.4±0.1	0.2±0.1	22.5±0.3	24.3±1.8	15.0±2.8	6.8±1.7	26.7±2.2
GCFAggMVC	25.9±1.5	5.7±1.2	2.9±1.5	29.0±1.8	78.9±5.1	78.7±3.4	68.4±6.4	79.6±4.0
SCM	58.7±1.0	30.8±0.9	30.5±1.6	**61.2±0.8**	82.0±2.9	74.3±2.3	67.3±3.6	82.0±2.9
SCMVC	47.9±4.3	25.0±2.5	20.4±3.4	51.1±3.7	86.1±0.7	86.3±0.8	81.0±1.3	86.1±0.7
GBASC (Ours)	**59.1±0.9**	**32.5±0.8**	**31.6±0.8**	60.9±0.9	**97.6±0.0**	**94.6±0.1**	**94.7±0.1**	**97.6±0.0**

Method	ProteinFold				Yale			
	ACC ↑	NMI ↑	ARI ↑	Purity ↑	ACC ↑	NMI ↑	ARI ↑	Purity ↑
AEMVC	26.7±3.6	31.4±4.8	5.6±1.9	30.2±4.0	60.8±4.5	63.7±5.7	38.2±7.5	62.2±4.7
BONE	22.2±1.7	23.8±4.8	6.4±3.0	24.2±2.6	53.2±7.9	58.4±5.3	34.2±7.6	54.2±6.9
CSOT	24.5±1.3	28.6±1.0	10.6±0.9	26.3±0.6	21.0±5.2	24.0±6.2	6.3±2.5	21.1±5.1
CVCL	21.8±1.1	25.7±3.3	6.0±2.4	24.1±2.0	47.8±1.7	52.9±1.2	26.5±1.1	48.8±1.8
DealMVC	26.7±0.8	33.1±0.9	12.9±0.4	30.3±0.7	33.9±3.5	38.9±4.7	12.3±3.1	36.4±3.6
DMAC	22.9±1.7	29.0±1.3	4.9±0.8	28.2±1.6	28.5±1.8	32.6±1.6	6.9±1.3	30.7±0.9
GCFAggMVC	20.4±2.8	29.3±1.4	5.8±1.3	26.4±2.1	31.4±8.3	38.2±6.3	11.6±6.1	33.5±7.7
SCM	31.7±2.2	42.8±1.8	16.4±1.7	38.6±2.0	51.2±2.7	54.6±1.7	27.2±2.4	52.6±2.2
SCMVC	31.2±1.2	42.3±1.0	**16.8±1.1**	38.0±1.8	48.2±7.0	51.0±5.5	25.4±6.6	49.3±6.7
GBASC (Ours)	**33.5±0.6**	**44.4±0.5**	16.1±0.8	**41.8±0.8**	**67.2±0.4**	**66.2±0.5**	**46.0±0.7**	**67.2±0.4**

Method	NGs				UCI_Digits			
	ACC ↑	NMI ↑	ARI ↑	Purity ↑	ACC ↑	NMI ↑	ARI ↑	Purity ↑
AEMVC	28.4±2.0	6.6±4.2	3.1±2.4	28.9±2.0	92.1±6.9	93.8±2.6	89.4±6.8	93.7±4.8
BONE	69.0±9.3	60.5±10.1	53.5±11.0	69.6±8.7	88.4±2.7	87.4±1.9	83.3±3.1	88.4±2.7
CSOT	41.4±2.6	29.4±2.8	20.1±2.6	41.7±2.4	82.6±4.2	81.5±2.4	74.2±4.2	82.9±3.7
CVCL	51.2±4.9	37.0±5.3	26.5±5.2	55.1±4.8	74.5±4.0	71.1±3.3	61.9±5.2	75.9±4.0
DealMVC	71.8±10.0	52.3±10.5	48.5±12.1	72.1±9.6	73.2±4.7	72.1±2.8	61.4±4.5	73.4±4.6
DMAC	24.0±1.5	1.0±0.4	0.0±0.4	24.6±1.6	26.7±4.1	18.3±5.5	8.7±4.0	29.2±4.8
GCFAggMVC	52.0±8.2	32.3±7.4	25.7±8.2	53.7±7.4	80.5±2.2	79.0±1.9	69.5±3.8	80.5±2.2
SCM	89.4±1.8	75.8±2.4	75.5±3.6	89.4±1.8	82.3±6.3	74.6±4.7	68.3±7.7	82.3±6.3
SCMVC	76.3±8.7	57.3±6.8	55.6±8.9	77.2±7.0	86.2±0.8	86.2±1.3	81.2±1.2	86.4±0.5
GBASC (Ours)	**98.7±0.1**	**96.0±0.3**	**96.8±0.3**	**98.7±0.1**	**97.6±0.0**	**94.6±0.0**	**94.7±0.0**	**97.6±0.0**

Datasets. We extend the comparison to twelve benchmark configurations spanning document clustering, handwritten digits, faces, human actions, objects, and protein folds. CiteSeer contains scientific publications represented by content and citation information, NGs contains newsgroup documents, and BBCSport contains sports news articles. HW, UCI_Digits, Mfeat, and UCI are variants of the same handwritten-digit collection, using six-view or three-view feature configurations. Yale and ORL provide face images, DHA contains human actions, ALOI comprises object images captured under varying conditions, and ProteinFold provides multiple representations for protein-fold categorization.

Table 6: Additional comparisons on twelve benchmark configurations, part II. Entries are percentages reported as mean ± sample standard deviation. Higher values are better; the highest mean for each configuration and metric is boldfaced.

Method	DHA				ORL			
	ACC ↑	NMI ↑	ARI ↑	Purity ↑	ACC ↑	NMI ↑	ARI ↑	Purity ↑
AEMVC	59.6±2.4	70.7±2.3	44.7±3.9	62.4±2.2	66.7±2.4	81.5±1.5	50.7±3.6	71.4±2.0
BONE	68.2±3.4	79.5±1.6	58.9±3.4	69.5±3.1	59.5±3.0	78.6±1.5	47.9±3.1	62.2±2.9
CSOT	23.7±3.1	54.9±2.7	21.6±2.7	23.7±3.2	21.5±7.5	45.2±10.5	9.4±3.4	23.0±8.4
CVCL	68.0±3.5	75.5±2.1	53.3±3.9	69.0±3.4	47.8±1.9	70.2±1.2	33.8±2.1	50.9±2.2
DealMVC	60.4±3.0	79.0±1.1	52.8±2.4	64.9±2.1	29.5±2.8	54.4±2.5	13.2±2.5	31.4±2.9
DMAC	45.1±6.7	55.2±6.4	27.2±6.8	47.5±6.5	39.3±3.3	60.2±2.8	19.8±3.4	42.4±3.4
GCFAggMVC	76.6±4.0	83.2±1.3	66.9±2.8	78.0±3.1	35.1±1.6	60.6±1.5	19.3±2.2	38.4±1.5
SCM	80.6±1.4	83.1±1.0	71.4±1.4	83.3±1.3	48.6±1.4	70.6±1.1	34.2±1.7	51.7±1.6
SCMVC	68.6±4.1	80.7±1.8	57.1±4.0	71.4±3.4	40.1±3.8	64.1±2.5	25.2±3.7	43.4±3.3
GBASC (Ours)	**86.4±0.2**	**88.9±0.2**	**79.4±0.3**	**87.2±0.2**	**83.2±1.1**	**91.0±0.6**	**75.6±1.4**	**84.4±1.0**
Method	Mfeat				UCI			
	ACC ↑	NMI ↑	ARI ↑	Purity ↑	ACC ↑	NMI ↑	ARI ↑	Purity ↑
AEMVC	92.0±7.8	93.7±3.6	89.1±9.1	93.7±5.7	85.7±4.4	88.6±1.7	80.7±4.2	88.0±2.7
BONE	88.2±1.2	87.5±0.8	83.3±1.0	88.2±1.1	80.0±5.9	81.7±2.9	73.5±5.2	81.1±5.0
CSOT	83.1±4.1	82.1±3.4	75.6±4.8	83.4±3.3	62.1±4.1	69.4±2.4	53.0±3.1	62.7±4.2
CVCL	74.2±5.8	71.1±4.7	62.3±6.9	76.7±4.6	79.1±5.8	77.0±3.7	69.2±6.1	79.7±4.6
DealMVC	72.2±2.7	71.6±1.4	60.1±2.7	72.7±2.1	69.0±3.4	69.3±2.1	57.1±3.4	69.4±3.4
DMAC	25.8±2.4	17.3±3.2	8.0±2.2	28.6±2.6	29.5±4.2	20.0±5.0	10.7±3.8	31.9±4.1
GCFAggMVC	80.6±3.1	79.2±2.8	69.7±5.4	80.6±3.1	73.3±3.2	71.7±2.1	59.5±3.6	73.5±2.8
SCM	79.5±8.6	74.1±4.9	66.9±8.4	80.2±7.1	92.5±0.8	85.2±1.2	84.1±1.6	92.5±0.8
SCMVC	86.7±0.5	86.7±1.3	81.9±1.2	86.7±0.5	81.4±0.4	80.1±0.6	72.2±0.7	81.4±0.4
GBASC (Ours)	**97.6±0.0**	**94.6±0.1**	**94.7±0.1**	**97.6±0.0**	**97.3±0.1**	**93.7±0.1**	**94.0±0.1**	**97.3±0.1**
Method	BBCSport				ALOI			
	ACC ↑	NMI ↑	ARI ↑	Purity ↑	ACC ↑	NMI ↑	ARI ↑	Purity ↑
AEMVC	30.5±1.8	4.4±1.7	1.2±1.2	37.2±1.6	73.8±1.8	86.8±0.8	39.7±6.6	77.5±1.3
BONE	66.4±5.0	56.9±9.9	48.7±9.8	75.2±6.5	33.3±14.5	58.7±20.3	22.4±13.3	34.7±15.4
CSOT	33.6±2.0	5.2±2.0	3.3±1.7	37.6±2.0	47.5±2.2	77.2±0.9	38.3±2.6	49.7±2.0
CVCL	35.0±3.4	4.3±5.5	1.4±4.3	37.5±3.9	81.0±2.6	89.3±0.8	75.3±2.1	82.1±2.2
DealMVC	49.7±5.3	29.2±5.9	20.5±4.6	53.8±5.1	54.7±1.0	81.1±0.8	50.4±2.0	57.7±1.3
DMAC	25.1±1.2	1.1±0.5	0.1±0.4	35.6±0.3	55.8±2.4	71.9±2.0	39.1±2.0	58.7±2.6
GCFAggMVC	43.0±5.0	21.3±5.4	14.3±5.5	48.5±6.1	86.4±1.0	93.8±0.2	83.2±0.8	87.5±0.9
SCM	64.0±7.5	43.2±7.3	37.5±8.5	69.2±6.1	59.2±2.2	75.4±1.3	48.4±2.2	61.2±2.3
SCMVC	58.8±9.3	36.7±8.1	30.7±9.7	63.3±8.3	80.7±1.2	91.0±0.3	76.1±1.0	82.5±1.1
GBASC (Ours)	**97.7±0.1**	**93.3±0.2**	**93.7±0.1**	**97.9±0.0**	**96.9±2.4**	**97.0±0.5**	**93.9±2.3**	**97.8±1.9**

Compared methods and evaluation. We compare GBASC with nine multi-view clustering baselines: AEMVC, BONE, CSOT, CVCL, DealMVC, DMAC, GCFAggMVC, SCM, and SCMVC. GBASC combines evidence-guided atomic splitting, risk-aware inter-atom connection, and struc-

tural feedback to representation learning. We evaluate clustering accuracy (ACC), normalized mutual information (NMI), adjusted Rand index (ARI), and cluster purity; higher values indicate better performance for all four metrics. Results are reported on a percentage scale as mean $\pm$ sample standard deviation, rounded to one decimal place. A displayed standard deviation of 0.0 may result from this rounding.

Comparative results. Tables 5 and 6 report the complete comparison. GBASC achieves the highest mean ACC and NMI on all twelve configurations and the highest mean ARI and purity on eleven configurations each, leading in 46 of the 48 configuration–metric comparisons. It leads on all four metrics in ten configurations. Comparing each GBASC result with the strongest baseline for the same configuration and metric gives average gains of 8.8, 6.7, 11.9, and 7.2 percentage points in ACC, NMI, ARI, and purity, respectively, calculated from the displayed means and averaged over the twelve reported configurations.

The advantage is consistent across metrics and dataset types rather than being concentrated in a particular evaluation criterion or benchmark. On BBCSport, GBASC reaches 97.7% ACC and 93.7% ARI, exceeding the strongest baselines by 31.3 and 45.0 percentage points. On ORL, the corresponding gains are 16.5 and 24.9 points, while NGs gains 20.2 NMI and 21.2 ARI points. On ALOI, ACC increases from the best baseline's 86.4% to 96.9%, accompanied by gains of 3.2 NMI, 10.7 ARI, and 9.3 purity points. CiteSeer achieves the best ACC, NMI, and ARI, while ProteinFold achieves the best ACC, NMI, and purity. The handwritten-digit configurations also consistently achieve the highest mean scores across all four metrics.

Advantages of GBASC. The joint gains in ACC, NMI, ARI, and purity support improved recovery of class-consistent partitions across the evaluated configurations. GBASC also has reported standard deviations of at most 1.0 percentage point in 43 of the 48 entries, indicating small variation across runs for most reported results. Together with the mechanism experiments in Appendix G, these comparisons support the practical value of coupling local refinement with global assignment: GBAS forms better-separated local units, and GBAC calibrates their connections and selects the final grouping. Structural feedback then uses this partition to guide representation learning. This combination provides complementary control of within-unit mixing and between-unit assignment errors in the evaluated multi-view clustering tasks.