Consensus Clustering: A Survey and Empirical Comparison of Consensus Generation Strategies

Ramkumar Selvarathinam
Capital One LLC
Richmond, VA, USA

S. Balamurugan
Hexaware Technologies
Tamil Nadu, India

ABSTRACT
Consensus clustering combines multiple candidate partitions to obtain a final clustering that is intended to be more stable and robust than an individual clustering solution. This paper presents a focused survey of major consensus-generation strategies and complements the survey with an empirical study on high-dimensional text data. The reviewed approaches include median-partition methods, co-association and voting-based methods, graph-based consensus, hierarchical consensus, adaptive and weighted ensembles, semi-supervised consensus, and scalable variants. For the empirical study, a four-class subset of the 20 Newsgroups corpus containing approximately 2,300 documents was represented using TF-IDF and reduced to 200 latent dimensions using Truncated SVD. Thirty K-Means partitions with different random seeds were generated. Three outcomes were evaluated with Adjusted Rand Index (ARI): individual K-Means partitions, a representative median/central partition, and a co-association consensus followed by average-linkage hierarchical clustering. Individual K-Means achieved a mean ARI of 0.206 ± 0.168 and a best-run ARI of 0.556. The representative-partition consensus achieved 0.184, whereas the co-association consensus achieved 0.579. The results show that, in this experiment, aggregating pairwise co-clustering evidence produced a higher-ARI final partition than selecting a single representative partition and slightly exceeded the best individual K-Means run. The study also discusses the trade-offs of consensus strategies and identifies directions for scalable, weighted, adaptive, and representation-aware clustering ensembles.

Keywords
Consensus clustering; clustering ensemble; co-association matrix; K-Means; median partition; text clustering; Adjusted Rand Index; 20 Newsgroups

1. INTRODUCTION
Clustering is a fundamental unsupervised-learning task that organizes observations into groups according to similarity. Because clustering lacks a single externally defined objective solution, results can vary with initialization, algorithm choice, parameter settings, sampling, and feature representation. This variability is particularly visible in high-dimensional data, where distance relationships can become unstable and individual partitions may capture different aspects of the data.

Consensus clustering, also called clustering ensemble, addresses this problem by combining multiple base clusterings into a single consensus partition. The central idea is not distributed-systems consensus; rather, it is the aggregation of alternative partitions of the same observations. Base partitions may be generated by varying random seeds, samples, features, parameters, or clustering algorithms. A consensus function then integrates the resulting assignments.

The literature contains several families of consensus functions. Some select a representative or median partition, some aggregate pairwise co-occurrence information through a co-association matrix, others align labels and vote, and graph-based approaches represent relationships among observations, clusters, or partitions. More recent work extends these ideas with weighting, optimization, adaptive selection, deep representations, and semi-supervised constraints.

This paper makes two practical contributions. First, it organizes representative consensus-clustering approaches into a compact taxonomy emphasizing how the consensus is formed. Second, it empirically compares repeated K-Means, a representative-partition consensus, and a co-association consensus on a high-dimensional text-clustering task. The experiment is intentionally designed as a reproducible illustration of the stability question that motivates clustering ensembles.

2. BACKGROUND AND TAXONOMY OF CONSENSUS CLUSTERING

2.1 Partition Generation and Consensus Generation
A consensus-clustering pipeline generally contains two stages. In partition generation, an ensemble of candidate clusterings is produced by perturbing the clustering process. Diversity can be introduced through different random initializations, data subsamples, feature subsets, parameter values, or base algorithms. In consensus generation, information from these partitions is combined to produce the final assignment. The effectiveness of the ensemble depends on both the quality and diversity of the base partitions and on the consensus function used to aggregate them [3], [24].

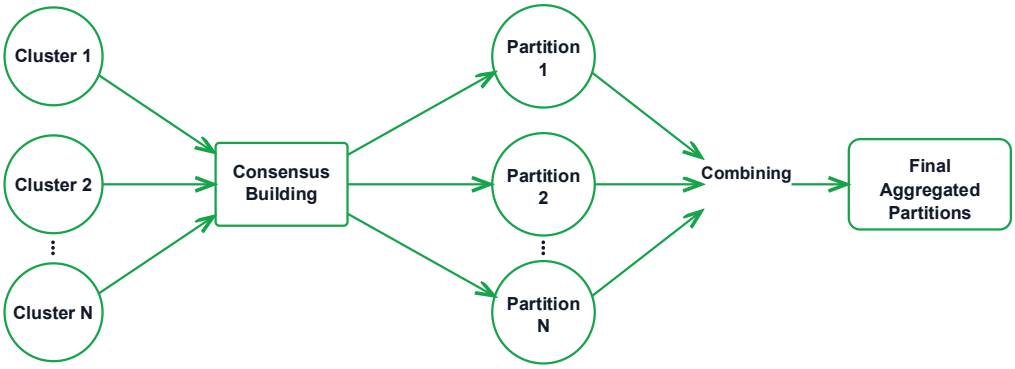

Fig. 1. General consensus-clustering workflow adapted from the authors' original survey draft.

2.2 Median-Partition and Representative-Partition Methods

Median-partition approaches seek a partition that minimizes aggregate disagreement with the ensemble or maximizes overall similarity to the input partitions. A practical representative-partition variant selects the most central existing partition according to pairwise partition similarity. Its main advantage is simplicity: the final answer is one valid base partition and does not require constructing an n × n co-association matrix. Its limitation is equally important: the final result remains bounded by the quality of a single member of the ensemble [6], [8]–[10], [24].

2.3 Co-Association and Evidence-Accumulation Methods

Co-association methods aggregate pairwise evidence. For observations i and j, the co-association value records the fraction of base partitions in which the two observations are assigned to the same cluster. The resulting symmetric matrix can be interpreted as an empirical similarity matrix. A second-stage clustering algorithm, such as agglomerative hierarchical clustering, can then be applied to a dissimilarity matrix obtained as one minus the co-association value. This family avoids cluster-label alignment because it operates on pairwise co-membership rather than raw cluster identifiers.

$$A_{ij} = (1/M) \Sigma_{m=1}^{M} I(c_m(i) = c_m(j))$$

where M is the number of base partitions and I(·) is the indicator function.

Recent approaches further refine the co-association matrix by incorporating structural information and explicitly modeling noise [3]–[5], [22], [24], [25].

2.4 Voting and Relabeling Methods

Voting-based consensus first addresses the label-permutation problem: cluster identifier 1 in one partition does not necessarily correspond to cluster identifier 1 in another. After aligning cluster labels to a reference or common representation, assignments can be combined by majority or weighted voting. Voting is intuitive but can be sensitive to the reference alignment and to low-quality base partitions [13], [24].

2.5 Graph-Based Consensus

Graph-based consensus represents observations and/or clusters as graph structures and obtains the consensus through graph partitioning. Hypergraph and bipartite formulations are common examples. These approaches can encode relationships beyond simple majority agreement, but their computational and parameter requirements can be greater than straightforward evidence accumulation [3], [6], [21], [24].

2.6 Hierarchical Consensus

Hierarchical clustering can be used either to generate diverse base partitions or as the final consensus mechanism. When applied to a co-association-derived dissimilarity matrix, hierarchical clustering provides a natural way to convert pairwise ensemble evidence into a final partition and can additionally expose multilevel structure through a dendrogram [4], [5], [24].

2.7 Weighted, Adaptive, and Selective Ensembles

Treating all base partitions as equally reliable is often unrealistic. Weighted clustering ensembles assign importance at the partition, cluster, feature, or observation level. Recent methods increasingly combine multiple levels of reliability, dynamically learn weights, or select a subset of diverse and high-quality base partitions. These methods aim to reduce the influence of redundant or poor partitions while preserving useful ensemble diversity [16]–[20], [26].

2.8 Semi-Supervised and Representation-Aware Consensus

Semi-supervised consensus clustering incorporates partial prior knowledge, such as pairwise constraints or labels, into partition generation or consensus formation. Representation-aware approaches use learned embeddings, including deep representations, to improve the quality of base partitions before ensemble aggregation. These directions are increasingly important for sparse, high-dimensional, and complex data [21]-[23].

3. REPRESENTATIVE METHODS AND APPLICATIONS

Applications of consensus clustering span bioinformatics, document clustering, image segmentation, social-network and community analysis, anomaly detection, information retrieval, and other settings where a single clustering run may be unstable

Table 1. Comparison of representative consensus-clustering strategies

Method family	Core idea	Strength	Typical limitation
Median/representative partition	Choose a partition central to the ensemble	Simple; low extra memory	Final answer depends on one partition
Co-association	Aggregate pairwise co-clustering frequencies	Label-invariant; uses all partitions	$O(n^2)$ pairwise matrix
Voting/relabeling	Align labels and vote	Intuitive and efficient	Requires label alignment
Graph-based	Partition a graph/hypergraph built from ensemble evidence	Flexible structural representation	Can add algorithmic complexity
Weighted/adaptive	Weight partitions, clusters, points, or features	Suppresses weak/redundant evidence	Weight estimation may add parameters
Semi-supervised	Use labels or constraints with ensemble evidence	Can exploit domain knowledge	Requires reliable side information

The appropriate consensus mechanism depends on data size, dimensionality, availability of side information, and the expected quality and diversity of the base partitions.

3.1 Critical Comparative Analysis of Consensus Strategies

The consensus-clustering approaches reviewed in this study differ in how they represent ensemble information, resolve disagreement among base partitions, and balance computational efficiency against the quality of the final partition. Although these methods share the objective of combining multiple clustering solutions, their underlying assumptions and aggregation mechanisms introduce distinct advantages and limitations.

Representative-Partition Methods: Representative-partition approaches identify a central partition that exhibits the greatest overall agreement with the remaining ensemble members. Their principal advantage is computational simplicity, as the final clustering can be selected without constructing a document-level co-association matrix. However, the selected solution is restricted to an existing base partition and therefore cannot directly reconstruct cluster assignments from pairwise evidence distributed across multiple ensemble members. Consequently, these approaches may be less effective when no individual partition adequately captures the underlying data structure [6], [8]–[10], [24].

Co-Association and Evidence-Accumulation Methods: Co-association methods aggregate the frequency with which pairs of observations are assigned to the same cluster across multiple partitions. Unlike representative-partition selection, this strategy integrates information from the complete ensemble and is invariant to permutations of cluster labels. This property is particularly useful when base clusterings exhibit substantial variation. Nevertheless, conventional dense co-association matrices require quadratic storage with respect to the number of observations, creating scalability challenges for large datasets. Furthermore, equally weighting all partitions may allow redundant or unreliable clustering solutions to influence the consensus [3]–[5], [22], [24], [25].

Voting-Based Methods: Voting approaches align cluster identifiers across partitions and determine final assignments through majority or weighted voting. These methods provide an intuitive aggregation mechanism and can avoid explicitly constructing a dense pairwise similarity matrix. However, their effectiveness depends on the accuracy of label alignment. Incorrect correspondence between clusters can propagate disagreement into the final solution, particularly when base partitions have substantially different cluster structures [13], [24].

Graph-Based and Hierarchical Methods: Graph-based methods represent relationships among observations, clusters, or partitions through graph or hypergraph structures. Their flexibility enables the incorporation of structural relationships that may not be captured by simple voting. However, graph construction, partitioning, and parameter selection can increase implementation and computational complexity. Hierarchical consensus methods provide an alternative mechanism for deriving a final partition from ensemble-derived similarities and can expose multilevel relationships. Their results nevertheless depend on the underlying dissimilarity representation and linkage strategy [3]–[6], [21], [24].

Weighted, Adaptive, and Selective Ensembles: Weighted approaches address the assumption that all base partitions contribute equally useful information. By assigning different importance to partitions, clusters, observations, or features, these methods attempt to reduce the influence of unreliable or redundant evidence. Adaptive and selective approaches extend this principle by adjusting contributions or retaining subsets of the ensemble. Their main challenge lies in estimating reliable weights or selection criteria without introducing excessive computational overhead or additional parameters [16]–[20], [26].

Semi-Supervised and Representation-Aware Methods: Semi-supervised consensus methods incorporate available constraints or partial labels to guide ensemble formation. Representation-aware approaches instead seek improved clustering structures through more informative feature representations, including learned embeddings. These strategies can address limitations of purely unsupervised partition aggregation, but their effectiveness depends on the quality of the supplied constraints or learned representations [21]–[23].

Synthesis and Research Gaps: The reviewed literature indicates that consensus quality depends on three interacting factors: the quality of the base partitions, the diversity of the ensemble, and the effectiveness of the consensus-generation mechanism. No single approach is universally preferable across all datasets and computational settings. Representative-partition approaches offer a relatively simple baseline, while co-association methods provide a mechanism for integrating pairwise evidence across the ensemble. Weighted and adaptive methods introduce additional flexibility but require reliable mechanisms for estimating contributions. A continuing research challenge is therefore to develop consensus strategies that preserve useful ensemble diversity, suppress unreliable evidence, and remain computationally practical for large, high-dimensional datasets.

The empirical study in Section 4 examines one aspect of this broader problem by comparing representative-partition

selection with co-association-based aggregation under a common K-Means ensemble. This comparison isolates the influence of the consensus-generation mechanism within the reported experimental configuration, rather than establishing general superiority across consensus-clustering families.

4. EMPIRICAL STUDY

4.1 Research Question

The empirical study asks whether aggregating repeated K-Means partitions through a co-association matrix can provide higher clustering agreement with the external labels than (i) an average individual K-Means run and (ii) selecting a representative central partition from the same ensemble

4.2 Dataset

A subset of the scikit-learn 20 Newsgroups text dataset was used. Four categories were selected: comp.graphics, rec.sport.baseball, sci.space, and talk.politics.mideast. The subset contains approximately 2,300 documents. The four known newsgroup categories were used only for external evaluation, not for training the unsupervised clustering models.

4.3 Text Representation and Dimensionality Reduction

Documents were vectorized using TF-IDF with max_df = 0.7, min_df = 5, and max_features = 10,000. English stop words from scikit-learn were removed. Truncated Singular Value Decomposition (Latent Semantic Analysis) reduced the sparse TF-IDF representation to 200 dimensions before clustering.

4.4 Base Partition Generation

Thirty K-Means runs were generated with n_clusters = 4. The number of clusters was held constant to match the four selected newsgroup categories, while a different random seed was used for each run. This design isolates initialization-driven variation and provides an ensemble of alternative partitions over the same feature representation.

4.5 Consensus Strategy A: Co-Association Matrix

For every pair of documents, the experiment counted how often the pair appeared in the same K-Means cluster across the 30 runs. Counts were divided by 30 to form a normalized co-association matrix. The matrix was converted to dissimilarity, and average-linkage hierarchical clustering was applied. The final partition was cut at four clusters.

4.6 Consensus Strategy B: Representative Central Partition

The second strategy retained the 30 K-Means label vectors, computed pairwise similarity/distance among partitions, and selected the central partition with the smallest aggregate distance to the other partitions. This approach represents a practical median/representative-partition baseline because the final consensus is one of the original base partitions

4.7 Evaluation Metric

Adjusted Rand Index (ARI) was used to compare predicted cluster assignments with the four ground-truth newsgroup categories. ARI corrects the Rand Index for chance agreement and is invariant to permutations of cluster labels. Higher values indicate greater agreement with the external labels.

5. RESULTS

Table 2. ARI Performance of Baseline and Consensus methods

Method	ARI	Interpretation
Single K-Means (30 runs)	0.206 ± 0.168	Mean ± standard deviation
Best individual K-Means run	0.556	Best observed base partition
Representative/median consensus	0.184	Central base partition
Co-association consensus	0.579	Best observed result

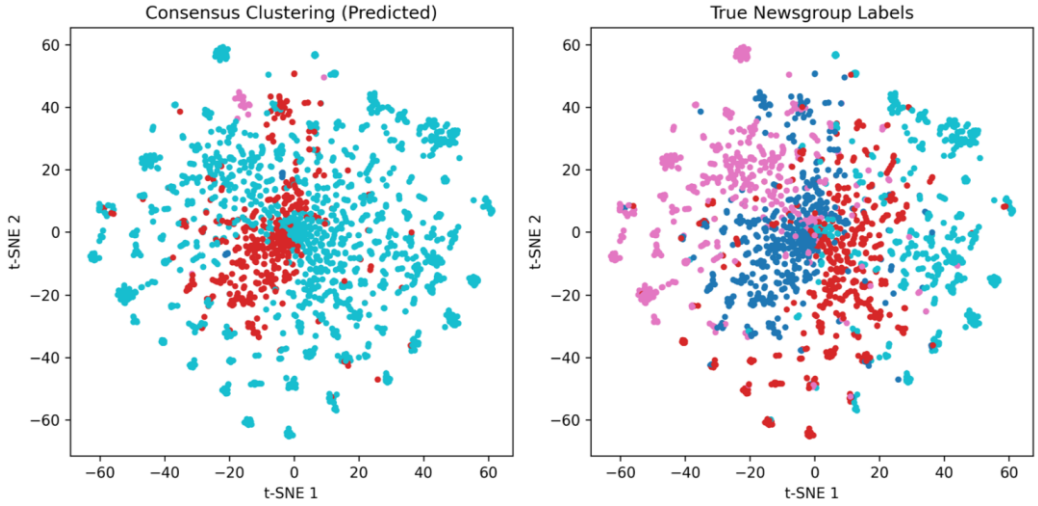

Fig. 2. Co-association consensus clustering (left) compared with true newsgroup labels (right), visualized in two dimensions.

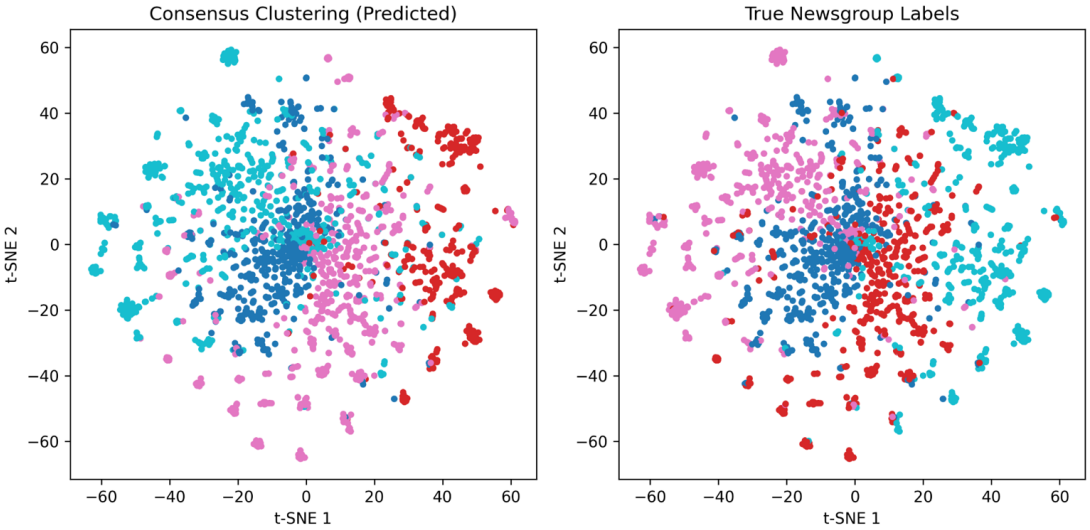

Fig. 3. Representative/median-partition clustering (left) compared with true newsgroup labels (right), visualized in two dimensions

5.1 Performance Comparison of Consensus Strategies

Table 2 summarizes the clustering performance obtained from the 30 K-Means partitions and the two consensus-generation strategies. The mean Adjusted Rand Index (ARI) of the individual K-Means runs was 0.206, with a standard deviation of 0.168. The relatively large standard deviation compared with the mean indicates substantial sensitivity to random initialization within the selected dataset and experimental configuration.

The best individual K-Means partition achieved an ARI of 0.556, demonstrating that certain initializations produced considerably stronger agreement with the external class labels than the average run. However, selecting a favorable individual partition based on its observed ARI is not equivalent to selecting a consensus solution without access to ground-truth labels.

The representative-partition consensus achieved an ARI of 0.184, which was 0.022 points below the mean individual K-Means result. This outcome illustrates an important limitation of central-partition selection: a partition that agrees strongly with other ensemble members does not necessarily provide the strongest correspondence with the underlying external categories. The representative solution is selected according to agreement within the ensemble rather than agreement with ground-truth labels.

In contrast, the co-association consensus achieved an ARI of 0.579, representing the highest agreement with the external class labels among the evaluated methods. The detailed performance differences between the baseline and consensus strategies are presented in Section 5.2.

5.2 Quantitative Analysis of Performance Differences

The absolute performance differences provide additional insight into the effectiveness of the consensus mechanisms.

Table 3. Absolute ARI Difference Among Clustering Approaches

Comparison	Absolute ARI Difference
Co-association vs. mean K-Means	+0.373
Co-association vs. best K-Means	+0.023
Co-association vs. representative partition	+0.395
Representative partition vs. mean K-Means	−0.022

The improvement of the co-association consensus over the mean K-Means result corresponds to approximately 181% of the baseline mean ARI. Its improvement over the best individual run is approximately 4.1% relative to that run's ARI. These percentages describe differences on the ARI scale and should not be interpreted as proportional improvements in clustering accuracy.

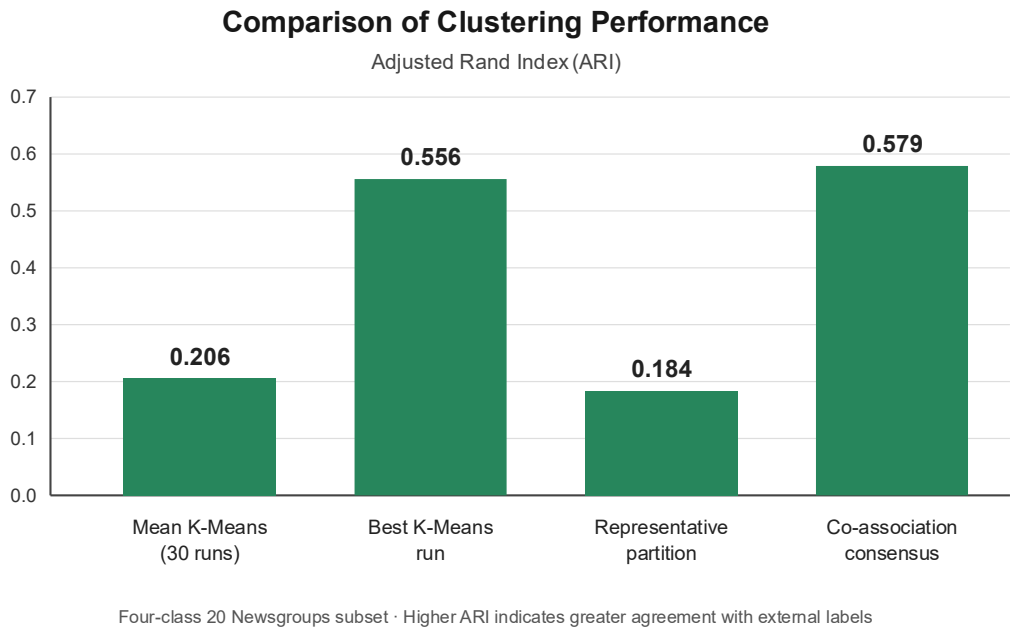

Fig. 4. Adjusted Rand Index comparison of individual K-Means performance and two consensus-generation strategies on the four-class 20 Newsgroups subset.

5.3 Interpretation of Ensemble Aggregation

The observed performance differences can be interpreted in relation to the information retained by each consensus mechanism.

Representative-partition selection identifies one existing partition that is central to the ensemble. Although this approach summarizes agreement among partitions, it cannot directly construct a new assignment by combining pairwise relationships from different partitions. Consequently, the selected partition remains subject to the limitations of its original clustering configuration.

Co-association consensus uses a different representation of ensemble evidence. Rather than selecting one complete partition, it records how frequently pairs of documents are grouped together across the 30 runs. Average-linkage hierarchical clustering subsequently converts this pairwise information into a final four-cluster partition.

One plausible explanation for the observed improvement is that frequently recurring document relationships were reinforced during evidence accumulation, while relationships appearing only in a small number of partitions contributed less strongly. This interpretation is consistent with the design of the co-association method, although the current experiment does not independently measure the contribution of individual document pairs or ensemble diversity.

The result should therefore be understood as evidence supporting co-association aggregation within the specific experimental setting, rather than proof that co-association methods consistently outperform representative-partition methods.

5.4 Experimental Scope and Statistical Interpretation

The reported results correspond to one ensemble configuration involving 30 K-Means partitions generated from a four-class subset of the 20 Newsgroups dataset. Although the observed ARI differences provide useful evidence regarding the behavior of the evaluated consensus strategies, they should be interpreted descriptively because no statistical significance testing was performed across independently generated ensembles.

The findings therefore illustrate differences between consensus-generation mechanisms within the reported experimental setting. Broader generalizability, metric selection, and computational scalability are discussed as threats to validity in Section 7.

6. DISCUSSION

6.1 Implications for Consensus Strategy Selection

The experimental findings highlight the importance of selecting a consensus-generation strategy according to the characteristics of the clustering task. Although representative-partition and co-association methods operate on the same collection of base partitions, they differ in their ability to utilize ensemble information and in their computational requirements.

For applications where computational simplicity and memory efficiency are primary considerations, representative-partition selection provides a practical alternative because it avoids constructing a dense document-level similarity matrix. However, its effectiveness depends on the availability of a sufficiently informative base partition.

Co-association methods offer a different trade-off by utilizing pairwise relationships across the ensemble. The observed results suggest that this approach can be advantageous when individual clustering solutions exhibit substantial variability. Nevertheless, its quadratic storage requirements may limit applicability to large datasets.

These considerations indicate that consensus strategy selection should account for ensemble quality, computational resources, dataset size, and the intended application. The findings motivate further investigation into scalable and adaptive

aggregation strategies rather than suggesting that a single consensus mechanism is universally optimal.

6.2 Interpretation of the Visualizations

The two-dimensional visualizations support the quantitative comparison but should be treated as illustrative rather than as an independent accuracy measure. The co-association visualization shows four visually distinguishable regions with a stronger correspondence to the ground-truth category layout than the representative-partition visualization. Because dimensionality-reduction visualizations can distort distances and neighborhoods, ARI remains the primary evaluation metric.

6.3 Relation to Recent Ensemble-Clustering Research

The experiment uses an intentionally simple, unweighted co-association matrix. Contemporary ensemble-clustering research shows several ways this baseline can be extended: weighting partitions or clusters according to reliability, selecting a high-quality diverse subset of base partitions, optimizing the co-association matrix, and learning representations that improve the base ensemble [16]–[23].

These developments provide a broader research context for the observed differences between consensus-generation mechanisms. They also demonstrate how contemporary methods extend conventional evidence aggregation through reliability estimation, ensemble selection, and representation learning.

7. LIMITATIONS AND THREATS TO VALIDITY

The empirical study has several limitations. First, it uses one dataset and four selected classes, so the numerical results should not be generalized to all clustering tasks. Second, all base partitions are generated by the same K-Means algorithm on the same reduced representation; algorithmic diversity was not evaluated. Third, the number of clusters is fixed at four using knowledge of the evaluation dataset. Fourth, only ARI is reported; complementary metrics such as normalized mutual information and internal validity measures could provide additional perspectives. Fifth, the co-association matrix requires quadratic storage in the number of observations, which can become a scalability constraint. Finally, the comparison is descriptive and does not include statistical significance testing over multiple independently generated ensembles.

8. FUTURE RESEARCH DIRECTIONS

Several extensions follow naturally from this study. An ensemble-size sensitivity experiment can evaluate whether consensus quality stabilizes as the number of base partitions increases. A broader benchmark should include multiple text and non-text datasets, additional external metrics, and runtime/memory measurements. Base-partition diversity can be increased by varying algorithms, feature subsets, SVD dimensions, and cluster counts. Weighted or selective co-association can be used to reduce the effect of weak or redundant partitions. For large datasets, sparse or approximate pairwise representations can address the quadratic memory cost. Finally, deep or transformer-based embeddings could be combined with consensus generation to separate representation quality from consensus quality. Future evaluations should also consider diverse base-clustering generators, weighted co-association strategies, multiple external validity measures, and statistical significance testing [18], [20], [24], [26]–[28].

9. CONCLUSION

This paper presented a focused survey and empirical comparison of consensus-clustering strategies, examining how different mechanisms aggregate multiple base partitions into a final clustering solution. The survey considered representative-partition, co-association, voting-based, graph-based, hierarchical, weighted, adaptive, and semi-supervised approaches, highlighting their underlying principles, advantages, limitations, and computational trade-offs.

The empirical study evaluated repeated K-Means clustering and two consensus-generation strategies using a four-class subset of the 20 Newsgroups dataset. Across 30 K-Means runs, the mean Adjusted Rand Index (ARI) was 0.206 ± 0.168, with the best individual partition achieving 0.556. The representative-partition consensus obtained an ARI of 0.184, whereas the co-association consensus achieved 0.579, producing the highest agreement with the external class labels among the evaluated approaches.

These findings demonstrate that consensus-generation mechanisms can produce substantially different clustering outcomes even when operating on identical base partitions. In the reported experiment, aggregating pairwise co-clustering evidence was more effective than selecting a central representative partition. The results also illustrate that agreement among ensemble members does not necessarily correspond to agreement with external class labels.

From a practical perspective, the study highlights the importance of considering both ensemble diversity and the mechanism used to combine clustering evidence. Representative-partition methods provide a comparatively simple aggregation strategy, whereas co-association methods can exploit relationships distributed across multiple partitions at the cost of increased pairwise storage requirements.

Future research should extend this evaluation across diverse datasets, ensemble configurations, and complementary validation metrics. Particular attention should be given to scalable and adaptive consensus-generation mechanisms that improve reliability while controlling computational cost.

Overall, the study contributes a comparative synthesis of major consensus-generation strategies and an empirical illustration of their differing outcomes. It emphasizes that effective clustering ensembles depend not only on the quality of their constituent partitions but also on how evidence is represented, aggregated, and transformed into a final clustering solution.

10. ACKNOWLEDGMENTS

During the preparation of this manuscript, the authors used OpenAI ChatGPT (GPT-5 series) for language refinement and manuscript organization. The authors reviewed and edited the generated output and take full responsibility for the content of this publication.

11. REFERENCES

[1] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: A review,” ACM Computing Surveys, vol. 31, no. 3, pp. 264–323, 1999, doi: 10.1145/331499.331504.

[2] S. Dudoit and J. Fridlyand, “Bagging to improve the accuracy of a clustering procedure,” Bioinformatics, vol. 19, no. 9, pp. 1090–1099, 2003, doi: 10.1093/bioinformatics/btg038.

[3] A. Strehl and J. Ghosh, “Cluster ensembles—A knowledge reuse framework for combining multiple partitions,” Journal of Machine Learning Research, vol. 3,

pp. 583–617, 2002.

[4] A. L. N. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 835–850, 2005, doi: 10.1109/TPAMI.2005.113.

[5] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data," Machine Learning, vol. 52, no. 1–2, pp. 91–118, 2003, doi: 10.1023/A:1023949509487.

[6] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," ACM Transactions on Knowledge Discovery from Data, vol. 1, no. 1, 2007, doi: 10.1145/1217299.1217303.

[7] V. Filkov and S. Skiena, "Integrating microarray data by consensus clustering," International Journal on Artificial Intelligence Tools, vol. 13, no. 4, pp. 863–880, 2004, doi: 10.1142/S0218213004001867.

[8] M. Bertolacci and A. Wirth, "Are approximation algorithms for consensus clustering worthwhile?" in Proc. 2007 SIAM International Conference on Data Mining, pp. 437–442, 2007, doi: 10.1137/1.9781611972771.41.

[9] N. Ailon, M. Charikar, and A. Newman, "Aggregating inconsistent information: Ranking and clustering," Journal of the ACM, vol. 55, no. 5, Art. no. 23, 2008, doi: 10.1145/1411509.1411513.

[10] B. Mirkin, Mathematical Classification and Clustering: From How to What and Why. Springer, 1998.

[11] C. Bouveyron and C. Brunet-Saumard, "Model-based clustering of high-dimensional data: A review," Computational Statistics & Data Analysis, vol. 71, pp. 52–78, 2014, doi: 10.1016/j.csda.2012.12.008.

[12] M. S. Nathiya and S. C. Punitha, "A survey on consensus clustering," International Journal of Computer Science and Information Security, vol. 7, no. 3, 2010.

[13] E. Dimitriadou, A. Weingessel, and K. Hornik, "Voting-merging: An ensemble method for clustering," in Artificial Neural Networks—ICANN 2001, Lecture Notes in Computer Science, vol. 2130, pp. 217–224. Springer, 2001, doi: 10.1007/3-540-44668-0_31.

[14] A. R. Brannon et al., "Molecular stratification of clear cell renal cell carcinoma by consensus clustering reveals distinct subtypes and survival patterns," Genes & Cancer, vol. 1, no. 2, pp. 152–163, 2010, doi: 10.1177/1947601909359929.

[15] R. Kruse, C. Döring, and M.-J. Lesot, Fundamentals of Fuzzy Clustering and Its Applications. Wiley, 2007.

[16] M. Zhang, "Weighted clustering ensemble: A review," Pattern Recognition, vol. 124, Art. no. 108428, 2022, doi: 10.1016/j.patcog.2021.108428.

[17] D. Aktaş, B. Lokman, T. İnkaya, and G. Dejaegere, "Cluster ensemble selection and consensus clustering: A multi-objective optimization approach," European Journal of Operational Research, vol. 314, no. 3, pp. 1065–1077, 2024, doi: 10.1016/j.ejor.2023.10.029.

[18] F. Xie, F. Nie, W. Yu, and X. Li, "Parameter-free ensemble clustering with dynamic weighting mechanism," Pattern Recognition, vol. 151, Art. no. 110389, 2024, doi: 10.1016/j.patcog.2024.110389.

[19] Q. Gu et al., "An improved weighted ensemble clustering based on two-tier uncertainty measurement," Expert Systems with Applications, vol. 238, Art. no. 121672, 2024, doi: 10.1016/j.eswa.2023.121672.

[20] N. Li et al., "A point-cluster-partition architecture for weighted clustering ensemble," Neural Processing Letters, vol. 56, Art. no. 183, 2024, doi: 10.1007/s11063-024-11618-9.

[21] Y. Shan et al., "Dual-level clustering ensemble algorithm with three consensus strategies," Scientific Reports, vol. 13, Art. no. 22617, 2023, doi: 10.1038/s41598-023-49947-9.

[22] Y. Shi, Z. Yu, C. L. P. Chen, and H. Zeng, "Consensus clustering with co-association matrix optimization," IEEE Transactions on Neural Networks and Learning Systems, 2024, doi: 10.1109/TNNLS.2022.3201975.

[23] M.-S. Chen, J.-Q. Lin, C.-D. Wang, D. Huang, and J.-H. Lai, "Contrastive ensemble clustering," IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 8, pp. 14678–14690, 2025, doi: 10.1109/TNNLS.2025.3531903.

[24] C. X. Gao, S. Wang, Y. Zhu, et al., "Ensemble clustering: A practical tutorial," Behavior Research Methods, vol. 58, Art. no. 286, 2026, doi: 10.3758/s13428-026-03158-y.

[25] J. Qu, Q. Dai, Z. Bian, J. Zhou, and Z. Jiang, "Ensemble clustering method via robust consensus learning," Electronics, vol. 14, no. 23, Art. no. 4764, 2025, doi: 10.3390/electronics14234764.

[26] J. Qu, Q. Dai, Z. Bian, and S. Gu, "Parameter-free ensemble clustering framework with three-level dynamic weighting mechanism," Neurocomputing, vol. 694, Art. no. 133889, 2026, doi: 10.1016/j.neucom.2026.133889.

[27] T.-A. Vu-Le, M. Park, I. Chen, et al., "Improving community detection with CVC, a new cluster ensemble method," Applied Network Science, vol. 11, Art. no. 53, 2026, doi: 10.1007/s41109-026-00809-z.

[28] Y. Ma and Z. Li, "Clustering ensemble method integrating Gaussian mixture model and three-way decision (GMM-3WD-CE)," Scientific Reports, vol. 16, Art. no. 11740, 2026, doi: 10.1038/s41598-026-47453-2.