

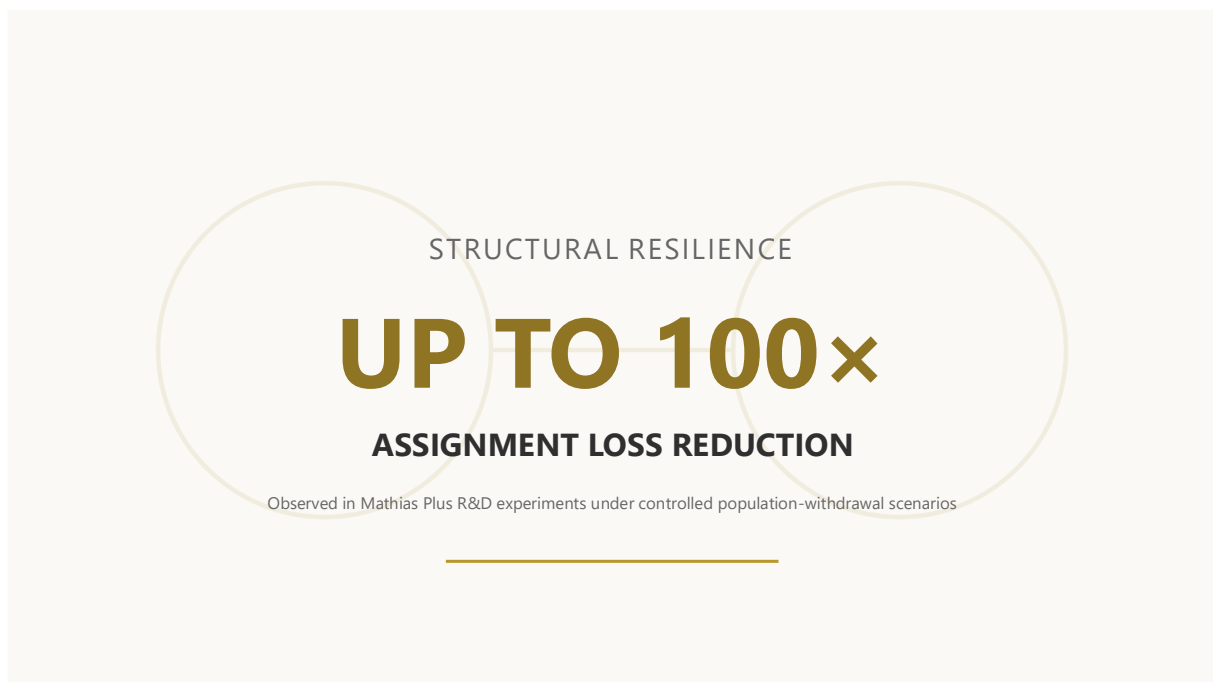
MathIAS+® Recursive Continuity. A New Paradigm for the Structural Resilience of Segmentation Systems



Position Paper.

Author: Hippolyte LAZARD-HOLLY, President and Founder of Mathias Plus SAS

Date: October 9, 2026



Executive Summary

Research on segmentation has historically focused on the quality of the resulting partitions. Considerable progress has been made. Modern algorithms can now automatically discover complex structures within very large populations.

However, organizations do not use segmentations as static objects. They use them to support decisions in environments that evolve continuously.

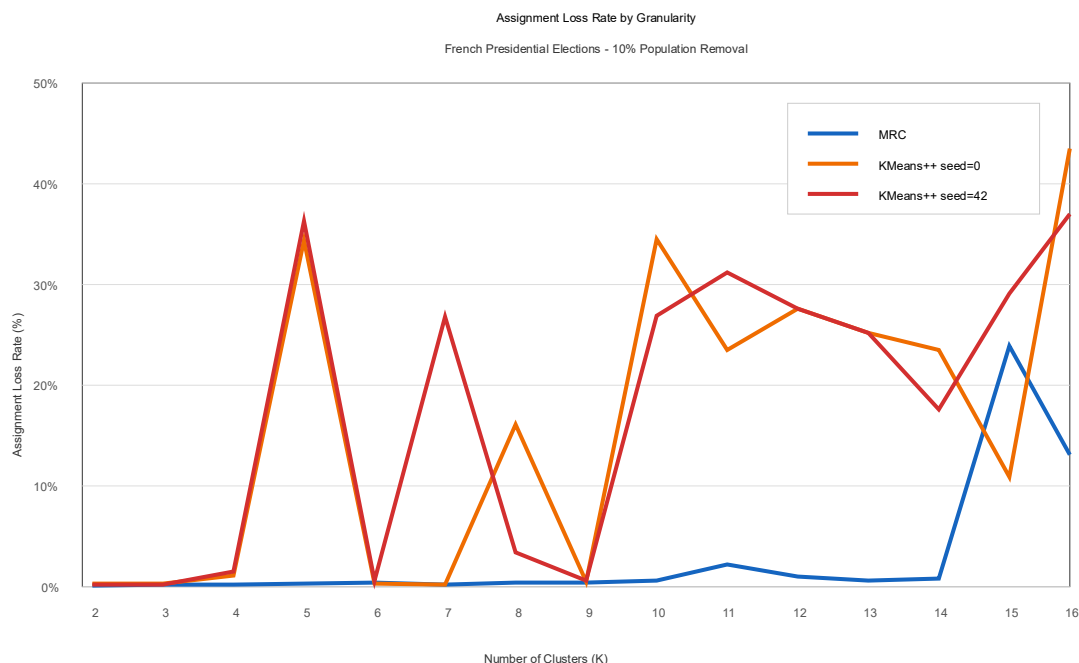
This raises a question that remains insufficiently explored:

What happens to a segmentation when part of the population disappears, changes, or evolves?

When a segmentation is rebuilt, assignments may change substantially. These changes create governance costs, reduce the stability of analyses over time, and complicate the operational use of results.

We propose introducing a new dimension for evaluating segmentation systems: **Structural Resilience**.

This research has led to the **MathIAS+ Recursive Continuity (MRC)** paradigm, a qualification protocol based on Assignment Survival, and its industrialization within **MathIAS+ Structural Intelligence Studio**¹.



¹ SaaS platform available in [Standard Edition](#) or [Enterprise Edition](#) through Microsoft Marketplace.

1. The Paradox of Modern Segmentation

Segmentation science has addressed many of the historical challenges associated with partition quality.

Today, we benefit from:

- high-performance algorithms;
- industrialized tools;
- massive computing capabilities;
- an extensive scientific literature.

Yet a paradox remains.

Most metrics evaluate the instantaneous quality of a segmentation.

Very few evaluate the ability of that segmentation to preserve its structure when its environment changes.

This distinction is fundamental.

A segmentation may be excellent today while becoming difficult to use tomorrow.

2. A Widely Underestimated Operational Pain

Assignment losses are not merely a mathematical curiosity.

They have real consequences.

When a population evolves:

- some observations move to another cluster;
- some groups disappear;
- some boundaries shift.

These phenomena make it more difficult to perform:

- longitudinal comparisons;
- historical analyses;
- decision justification;
- governance exercises.

The issue therefore no longer lies solely in partition quality.

It lies in the partition's ability to survive.

We call this property:

Structural Resilience.

3. Research Without Compromise

A simple response would be to prioritize stability.

However, a perfectly stable segmentation may still be poor.

Stability alone is therefore of little value.

Our objective was different.

We sought to improve segmentation continuity without weakening the historical requirements of clustering.

In other words:

- without sacrificing cluster homogeneity;
- without sacrificing partition quality;
- without sacrificing scalability;
- without sacrificing analytical relevance.

Resilience had to be added.

It could not be purchased at the cost of weakening existing criteria.

4. Why K-Means++ Is the Natural Benchmark

Any research on resilience must begin by defining a credible reference.

Among the most widely used segmentation methods, we prioritized approaches based on minimizing inertia.

This choice is based on a simple property.

Inertia directly measures the geometric homogeneity of clusters. Unlike many specialized criteria, it assumes neither a particular population shape nor a specific data distribution. It therefore provides a general quality indicator applicable across a wide range of contexts.

Within this framework, K-Means++ naturally emerges as the primary benchmark. Its broad industrial adoption, operational robustness, and explicit objective of minimizing inertia make it a particularly relevant reference.

However, using K-Means++ immediately raises a governance question. The algorithm relies on random initialization, which may produce different segmentations on the same dataset. This dependence on initial conditions was historically one of the first challenges addressed by our research program.

For this reason, the qualification protocols presented in this document do not rely on a single K-Means++ execution. Two distinct initialization configurations are systematically evaluated using different random seeds, `random_state=0` and `random_state=42`.

The objective is not only to measure the performance of the resulting segmentations, but also to observe their sensitivity to the underlying stochastic mechanisms.

This methodological precaution explicitly incorporates the governance challenge of unsupervised machine learning into the evaluation of results. A segmentation intended to support operational decisions must be assessed not only according to its analytical quality, but also according to its stability and reproducibility.

A natural alternative would have been to use a deterministic algorithm such as Ward as the primary benchmark.

This option was examined during our research. In experiments where its execution remains practical, Ward does not provide a level of quality comparable to that obtained with K-Means++. This conflicts with the non-compromise requirement adopted in this work.

More importantly, its computational and memory constraints significantly limit its use as a general benchmark for very large populations.

We therefore selected K-Means++ as the primary benchmark for a simple reason: it currently represents one of the most credible industrial references when the objective is to maximize cluster homogeneity at scale.

The research question then becomes:

Can the Structural Resilience of segmentation systems be significantly improved without abandoning the qualities that made K-Means++ successful?

The ambition of MathIAS+ Recursive Continuity is precisely to provide a positive answer to this question.

5. The Invention of Recursive Continuity

Conventional approaches generally construct segmentations independently.

Each new partition restarts the process. The previous structure disappears. Relationships between successive segmentations become difficult to interpret, and analytical lineage is largely lost.

MathIAS+ Recursive Continuity follows a different logic.

The first consequence of this approach is the pursuit of deterministic behavior. At a fixed granularity, the same dataset must produce the same segmentation. Any observed variation must originate in the data itself, not in an algorithmic random mechanism.

This requirement naturally leads to a second property: continuity with respect to the data.

When the population changes only slightly, the segmentation should also change only slightly. The transformations observed should remain proportionate to the perturbations actually introduced into the data.

Within this framework, a segmentation is no longer the isolated outcome of a one-off computation. It becomes the product of a sequence of deterministic and continuous operations whose effects can be tracked, explained, and analyzed.

Recursive continuity across levels of granularity then becomes a natural consequence of this construction.

Instead of rebuilding each partition independently, each level inherits the structure of the previous level. Granularity evolves through progressive transformation rather than through a complete restart of the process.

This change may appear modest.

Its consequences are profound.

The object of analysis is no longer only a partition.

It becomes a structural trajectory.

Segmentation ceases to be a collection of independent results and becomes a coherent system whose successive states remain connected.

Structure evolves.

It does not restart.

6. Measuring What Survives

A new property requires new measures.

The question is no longer:

Are the two segmentations similar?

The question becomes:

How many assignments actually survived?

To answer this question, we use the Hungarian trace.

This metric has an essential characteristic.

It directly models the optimal overlap between two segmentations.

It therefore provides a natural approximation of effective assignment survival.

The Hungarian trace consequently becomes the core of the qualification protocol.

We define **Assignment Survival** as the proportion of assignments preserved between two comparable segmentations.

7. A Global Multi-K Resilience Indicator

Resilience should not be evaluated at a single level of granularity.

A segmentation may be robust for one value of K and fragile for another.

We therefore systematically evaluate the following levels:

$$K = 2, \dots, 16$$

For each level of granularity:

- the segmentation is computed;
- Assignment Survival is measured;
- the results are consolidated.

The average obtained across all levels of granularity constitutes the protocol's primary key indicator.

It provides a synthetic measure of the system's overall behavior.

This approach substantially reduces the effects associated with selecting a particular value of K .

8. Case Study: French Presidential Elections

French presidential elections provide a particularly suitable experimental setting for studying the Structural Resilience of segmentation systems.

The corpus is built from first-round results from the 2017 and 2022 French presidential elections.

Each polling station is described by the percentages obtained by the different candidates, producing a space of 23 homogeneous quantitative variables expressed as percentages of valid votes cast.

This choice offers several methodological advantages.

First, all variables use a common unit of measurement.

The data are therefore used without prior normalization. Observed distances retain their direct electoral meaning, and the resulting analytical structures remain immediately interpretable.

Second, the study is conducted in a fully unsupervised setting.

No political, geographical, or sociological prior is introduced when constructing the segmentations. The observed groups emerge exclusively from regularities present in the data.

Third, despite the absence of initial assumptions, the results remain intuitively understandable to a human observer.

The resulting segmentations can be compared with known political realities, facilitating qualitative analysis of the structures detected and of their evolution.

Finally, presidential elections form a complex system combining both continuity and transformation.

From one election to another, a substantial share of electoral behavior persists while new dynamics emerge. This makes the corpus a particularly relevant laboratory for studying how analytical structures withstand population change.

The electoral corpus therefore possesses a rare property in segmentation: it is algorithmically unsupervised while remaining readily interpretable by humans.

It allows the structural properties of algorithms to be evaluated without relying on a predefined ground truth.

The purpose of this study is not to validate any particular political hypothesis.

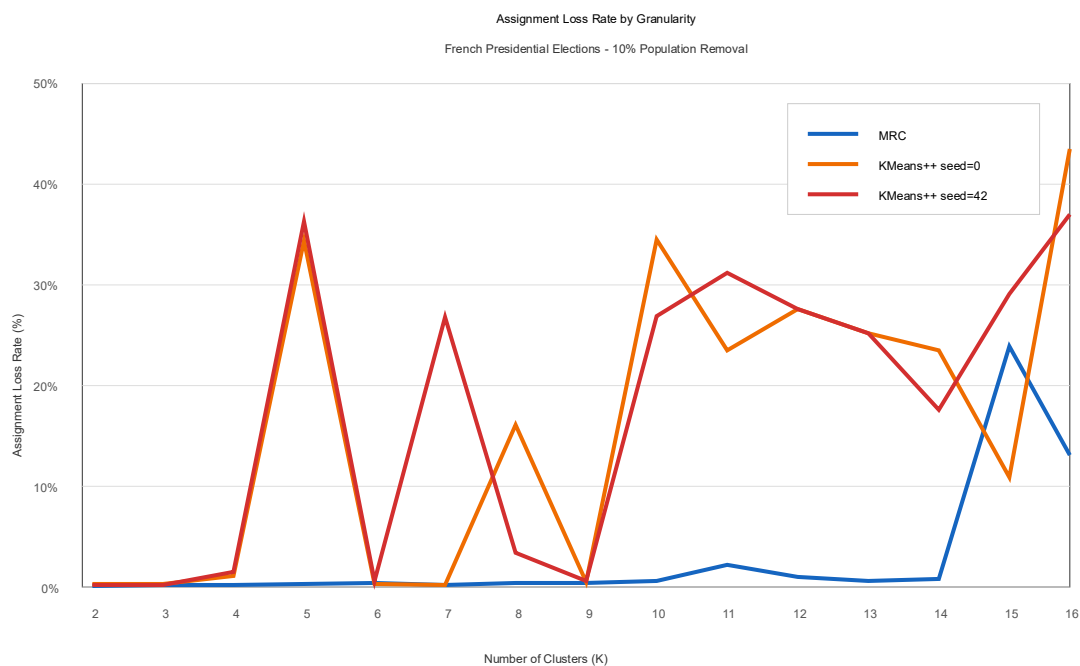
It is to observe how different segmentation approaches behave when confronted with a real, multidimensional, interpretable system subject to measurable transformation over time.

To qualify resilience, 10% of polling stations are withdrawn at random.

Segmentations are computed on the complete population and then on the reduced population. Assignment Survival is subsequently measured using the Hungarian trace for each granularity from $K = 2$ to $K = 16$.

The results are finally consolidated into an average multi-K resilience indicator.

The observed results reveal significant differences between the approaches studied.



MathIAS+ Recursive Continuity achieves an average trace of **97.07%**, corresponding to an average assignment loss rate of **2.93%** across all granularities studied.

The two K-Means++ configurations produce average assignment loss rates of **16.09%** and **17.60%**, respectively.

Their average, used as the consolidated benchmark value in the multi-corpus summary, is **16.85%**.

The consolidated assignment-loss reduction factor therefore reaches **5.75×** in favor of MathIAS+ Recursive Continuity.

From $K = 2$ through $K = 10$, MRC systematically preserves more than 99% of assignments.

Disruptions nevertheless appear at $K = 15$ and $K = 16$, confirming the value of qualification across the complete multi-K trajectory rather than at one arbitrarily selected level of granularity.

Beyond the numerical values, these results illustrate a fundamental difference in behavior.

When part of the population disappears, MathIAS+ Recursive Continuity tends to preserve the existing analytical structure and evolve the segmentations progressively.

Conventional approaches based on independent reconstruction may instead produce substantially larger assignment reorganizations.

The experiment therefore reveals not only a measurable improvement in Assignment Survival, but also the emergence of a broader property: the ability of a segmentation to preserve its structural coherence when confronted with change.

This property is precisely what we call **Structural Resilience**.

9. Empirical Results

Experiments conducted on four corpora with different characteristics, dimensionalities, and sizes lead to a recurring observation.

Under the conditions of the protocol studied, MathIAS+ Recursive Continuity significantly reduces assignment loss compared with the K-Means++ benchmark, while preserving the expected properties of homogeneity and analytical quality.

The protocol is based on the random withdrawal of 10% of the population.

For each corpus, Assignment Survival is measured using the Hungarian trace across all granularities from $K = 2$ to $K = 16$.

The average assignment loss rate is then calculated across these fifteen levels of granularity.

To account for the dependence of K-Means++ on its initialization, two configurations are systematically evaluated using `random_state=0` and `random_state=42`.

The consolidated benchmark value corresponds to the average of the results obtained with these two configurations.

Summary of Results

Corpus	Population	MRC Loss	K-Means++ Loss	Loss Reduction Factor
Digits	1,797	5.16%	25.78%	5.00×
French presidential elections	68,236	2.93%	16.85%	5.75×
Covertypes	581,012	0.20%	20.39%	101.4×
Higgs21	11,000,000	2.97%	36.57%	12.31×

The four corpora cover very different contexts:

- Digits is a small academic benchmark;
- the French electoral corpus represents a real, multidimensional system that remains interpretable by humans;
- Covertypes is a conventional machine learning benchmark containing more than half a million observations;
- Higgs21 extends the evaluation to 11 million observations and confirms the operational scalability of MRC.

These results do not establish a general relationship between corpus size and the resulting loss reduction factor.

They do, however, show that the observed improvement is not confined to one particular domain or scale.

It appears across populations ranging from 1,797 to 11 million observations.

The results therefore suggest that Structural Resilience is a property that is:

- measurable;
- reproducible within the defined experimental framework;
- observable across varied corpora;
- usable within an operational context.

They reinforce the central hypothesis of this work: assignment continuity is not merely a side effect of clustering.

It constitutes an analytical dimension in its own right that can be measured, qualified, and industrialized.

10. From Segmentation to Structural Resilience

For several decades, segmentation has primarily been evaluated through partition quality.

We believe that a new step is now possible.

Quality remains necessary.

It is no longer sufficient.

Organizations also need to understand:

- what changes;
- what survives;
- what remains usable over time.

Assignment continuity therefore becomes a new dimension of analysis.

Structural Resilience becomes a new property of segmentation systems.

The research question consequently evolves:

The objective is no longer only to build good clusters.

It is to build analytical structures capable of surviving change.

11. Industrialization

The results presented in this position paper have not remained theoretical.

They have led to the industrialization of artifacts dedicated to:

- Assignment Survival;
- assignment loss;
- analytical lineages;
- Structural Resilience;
- multi-K analysis.

The value of a new scientific concept can only be fully demonstrated if it has an operational reality.

An analytical property that could not be calculated, observed, or used on real data would retain limited value for organizations.

This is why the work presented in this document was extended through a complete industrialization program.

This program led to the production deployment of [MathIAS+ Structural Intelligence Studio](#), a commercial self-service SaaS platform that enables the analyses presented in this document to be performed on real data.

Organizations can directly apply the qualification protocols to their own populations, measure Assignment Survival, analyze losses associated with changes in data, explore multi-K analytical lineages, and evaluate the Structural Resilience of their segmentation systems.

This industrialization serves two purposes.

First, it transforms a research result into a directly usable operational capability.

Second, it provides empirical validation of the concept itself.

Structural Resilience is no longer presented as an abstract theoretical property.

It becomes an indicator that can be calculated, observed, and reproduced on real data.

The existence of an operational platform capable of producing these analyses at scale is therefore an important part of the demonstration.

It shows that Recursive Continuity can be studied as a scientific object, implemented within an industrial environment, and used to inform real decisions.

In this sense, **MathIAS+ Structural Intelligence Studio**² is not merely a software product.

It is the operational embodiment of the paradigm presented in this document.

An analytical paradigm acquires industrial relevance only when organizations can measure it on their own data. The production deployment of MathIAS+ Structural Intelligence Studio addresses precisely this objective

12. Scope and Conditions of Interpretation

This document reflects the state of research and development carried out by Mathias Plus as of the publication date.

The concepts, methods, protocols, and results described in this position paper are intended to contribute to scientific and industrial discussions on the structural resilience of segmentation systems. They should not be interpreted as universal evidence applicable to all contexts, populations, datasets, or segmentation algorithms.

The experimental results presented in this document are based on the protocols described herein, including random population-withdrawal scenarios involving 10% of the population and multi-granularity evaluations ranging from $K = 2$ to $K = 16$. The reported values should therefore be interpreted within this experimental framework.

The performance observed on the studied corpora should not be considered predictive of results that may be obtained on other datasets. Effective measurement of structural resilience requires a dedicated evaluation performed on the populations under consideration.

Mathias Plus encourages empirical analysis on real-world data and considers that the validation of any segmentation system should be conducted under the operational conditions specific to each organization.

² SaaS platform available in [Standard Edition](#) or [Enterprise Edition](#) through Microsoft Marketplace.

The information provided in this document is for informational purposes only. It does not constitute a guarantee of performance, nor a business, financial, regulatory, or operational recommendation.

The methods, concepts, metrics, software components, and analytical artifacts referenced in this document may evolve as part of the ongoing research, development, and continuous improvement activities conducted by Mathias Plus

APPENDIX A – MATHEMATICAL DEFINITIONS

A.1 Notation

Let:

- X be a complete population of size N
- $X' \subset X$ be a population obtained after removing a fraction of the original population
- $S_K(X)$ be the segmentation of X into K clusters
- $S_K(X')$ be the segmentation of X' into K clusters

Let N' denote the number of observations shared by both populations.

Assignment-survival calculations are performed exclusively on observations present in both X and X' .

A.2 Contingency Matrix

For a given granularity K , we construct the contingency matrix:

$$C = (c_{ij})$$

where:

$$c_{ij} = |S_i(X) \cap S_j(X')|$$

Each coefficient represents the number of observations assigned simultaneously:

- to cluster i in the full population;
- to cluster j in the perturbed population.

A.3 Hungarian Trace

Since cluster identifiers are arbitrary, labels must be realigned before comparison.

We therefore search for the permutation:

$$\pi \in S_K$$

that maximizes the correspondence sum:

$$H(K) = \max_{\pi \in S_K} \sum_{i=1}^K c_{i, \pi(i)}$$

This optimization is performed using the Hungarian algorithm.

We designate:

$$H(K)$$

as the optimal Hungarian Trace.

A.4 Assignment Survival

Assignment Survival for granularity K is defined as:

$$AS(K) = \frac{H(K)}{N'}$$

with:

$$0 \leq AS(K) \leq 1$$

This measure represents the proportion of assignments preserved after population perturbation

A.5 Assignment Loss Rate

Assignment Loss Rate is defined as the complement of Assignment Survival:

$$Loss(K) = 1 - AS(K)$$

This metric directly represents the proportion of observations whose analytical assignment has changed.

For example:

$$AS(K) = 0.97$$

implies:

$$Loss(K) = 0.03$$

or a 3% assignment-loss rate

A.6 Multi-K Resilience

A segmentation may exhibit different behaviors depending on the granularity considered.

To avoid drawing conclusions from a single segmentation level, experiments are conducted across the full granularity range:

$$K \in \{2, 3, \dots, 16\}$$

The global Assignment Survival indicator is defined as:

$$\bar{AS} = \frac{1}{15} \sum_{K=2}^{16} AS(K)$$

where:

- $AS(K)$ = Assignment Survival at granularity K
- $\bar{AS}$ = average Assignment Survival across all evaluated granularities

Equivalently, the global Assignment Loss Rate is defined as:

$$\bar{Loss} = \frac{1}{15} \sum_{K=2}^{16} Loss(K)$$

where:

- $Loss(K)$ = Assignment Loss Rate at granularity K
- $\bar{Loss}$ = average Assignment Loss Rate across all evaluated granularities

This multi-K approach reduces dependence on any single segmentation level and provides a more robust estimate of structural resilience.

A.7 Loss Reduction Factor

To compare two segmentation approaches, A and B , we define the **Loss Reduction Factor (LRF)** as:

$$LRF = \frac{\bar{Loss}_A}{\bar{Loss}_B}$$

where:

- $\bar{Loss}_A$ = average Assignment Loss Rate of approach A
- $\bar{Loss}_B$ = average Assignment Loss Rate of approach B

In this document, the benchmark comparison is:

$$LRF = \frac{\bar{Loss}_{KMeans++}}{\bar{Loss}_{MRC}}$$

A value of:

$$LRF = 5$$

indicates that indicates that assignment loss is reduced by a factor of five with **MathIAS+ Recursive Continuity (MRC)** relative to the K-Means++ benchmark.

A.8 Recursive Continuity

Let:

$$P_K$$

denote the partition produced at granularity K .

Traditional clustering approaches generally construct each partition independently:

$$P_2, P_3, \dots, P_K$$

without any explicit continuity constraint between successive segmentation levels.

By contrast, **MathIAS+ Recursive Continuity (MRC)** relies on a structural dependency relationship between consecutive granularities:

$$P_{K+1} = F(P_K)$$

where:

$$F$$

represents a deterministic transformation.

As a consequence, the partition produced at granularity $K + 1$ inherits the structure generated at granularity K , enabling the construction of a coherent analytical lineage across the entire multi-K trajectory.

Rather than being rebuilt independently at each segmentation level, the structure evolves through deterministic transformations, preserving analytical continuity across the entire multi-K trajectory.