Extended Abstract Track

SRTD: A Symmetric Divergence for Interpretable Comparison of Representation Topology

Editors: List of editors' names

Abstract

Representation Topology Divergence (RTD) has emerged as a powerful tool for analyzing topological differences in point clouds (Barannikov et al., 2021), especially within neural network representations. However, unlike symmetric distance metrics, our two-directional divergence calculations often yield vastly different values. While we currently average these two quantities to enforce symmetry, this approach lacks a clear theoretical justification and interpretability. Its variant, mentioned by Trofimov et al. (2023), has been rarely discussed or explored. Furthermore, unlike CKA (Kornblith et al., 2019), the full potential of RTD has not been thoroughly investigated and applied across various domains of machine learning, particularly in Large Language Models (LLMs). In this paper, we reveal the complementary nature of RTD and its symmetric version. We introduce a more faithful and comprehensive Symmetric Representation Topology Divergence (SRTD), enriching the interpretability of the RTD framework. We explore a series of mathematical properties for SRTD and its lightweight variant inspired by Tulchinskii et al. (2025), SRTD-lite. Through experiments on both synthetic and real-world data, we demonstrate that SRTD and SRTD-lite outperform their one-sided divergence counterparts in terms of computational efficiency and accuracy. Additionally, by applying SRTD to compare the representation spaces of various LLMs, we showcase its strong capability in distinguishing models from different origins.

Keywords: Topological Data Analysis, Representation Learning, Neural Network Analysis, Large Language Models Fingerprinting

1. Introduction

Understanding and comparing the internal representations of neural networks is a central topic in the field of deep learning. As models become increasingly complex, developing tools that can provide insight into their internal mechanisms has become crucial. In recent years, researchers have proposed various methods for measuring representation similarity, ranging from linear decodability probes to kernel-based techniques like Centered Kernel Alignment (CKA) (Kornblith et al., 2019), which haveve become a popular tool for comparing the geometric structures of representation spaces between different network layers or models.

Among these methods, Topological Data Analysis (TDA) offers a unique perspective that goes beyond traditional geometric metrics to focus on the intrinsic 'shape' and connectivity of data.In this context, Representation Topology Divergence (RTD) was introduced as a powerful tool specifically designed to quantify topological differences in point cloud data (Barannikov et al., 2021), such as neural network representations. By comparing the topological features of two representation spaces and their union, RTD effectively measures the structural differences between them. Since computing persistent homology is very time-consuming, it is difficult to use on large-scale datasets.Then, Tulchinskii et al. (2025) proposed RTD-lite as a computationally efficient optimization method that focuses

solely on 0-dimensional homology (clustering and connectivity information), addressing this computational bottleneck.

However, RTD also faces several challenges. The divergence between two point clouds is typically calculated as the average of two directional divergences[1]. The two values being averaged often show significant discrepanciesTable 5(*c*)[2], a phenomenon that no current work has explained, and the same issue is present in RTD-lite.Trofimov et al. (2023) packaged RTD as a differentiable function for gradient optimization. While they also mentioned another variant, which we term Max-RTD, and noted its similar properties, they did not conduct an in-depth investigation. Instead, they merely added it to the loss function to enrich gradient information, leaving this area of research unexplored. Furthermore, RTD has not yet demonstrated its power on AI frontiers such as Large Language Models; previous applications focused on Variational Autoencoders, and its potential has not been fully realized, unlike CKA(Kornblith et al., 2019). To further enhance the theoretical framework of RTD and promote its application in LLMs, we propose **SRTD** and **SRTD-lite**. Our improved version is the first in the RTD series (including the lite series) to have symmetry, interpretability, and the lowest computational overhead, which facilitates more trustworthy representation analysis in LLMs. Our **main contributions** are as follows:

We provide a theoretical explanation for the asymmetry of RTD by formally defining Max-RTD and our proposed SRTD, along with their lightweight variants, and prove their mathematical properties. We then experimentally demonstrate that SRTD is more effective than one-sided divergences when used as a penalty term or for model comparison. Finally, we showcase SRTD's potential by applying it to analyze the representations of LLMs.

2. Preliminary

We consider two point clouds, P and P', of the same size with a one-to-one correspondence. Their pairwise distance matrices are w and $\tilde{w}$, respectively. We define the following Vietoris-Rips complexes: $A = R_\alpha(\mathcal{G}^w)$, $B = R_\alpha(\mathcal{G}^{\tilde{w}})$, $A \cup B = R_\alpha(\mathcal{G}^{\min(w,\tilde{w})})$, and $A \cap B = R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})$. We first revisit the idea behind RTD. RTD computes the persistent homology of the complex represented by the distance matrix m_{min}2[3], and defines the sum of its barcode lengths as $RTD(w, \tilde{w})$, with the other direction defined symmetrically. RTD focuses on the difference in topological features between A, B and their union $A \cup B$. Trofimov et al. (2023) mentioned integrating an RTD variant as an optimization term into the original RTD loss to enrich gradient information. This variant focuses on the relationship between A, B and their intersection $A \cap B$, and exhibits behavior similar to RTD. Here, we provide a formal definition for this variant Definition 6.

3. Symmetric Representation Topology Divergence (SRTD)

In practice, we observe a complementary phenomenon between RTD and Max-RTD 5(*c*). When $RTD(w, \tilde{w}) > RTD(\tilde{w}, w)$, we consistently find that $Max\text{-}RTD(w, \tilde{w}) < Max\text{-}RTD(\tilde{w}, w)$. This suggests that the topological structural differences between $A \cup B$ and $A \cap B$ seem to

1. For example, $RTD(P, P') = \frac{RTD(w,\tilde{w})+RTD(\tilde{w},w)}{2}$. The same applies to Max-RTD and SRTD.
2. The **Min** column in the table shows the difference between the divergences in the two directions.
3. The three important matrices, m_{min}, m_{max}, and m_{sym}, are defined in Appendix A

be the core reason for the asymmetry in RTD. Therefore, we propose to directly measure this difference as the Symmetric Representation Topology Divergence (SRTD) of P and P'.

Definition 1 (SRTD) *For two point clouds P and P' with a one-to-one correspondence, the distance matrix of their auxiliary graph $\hat{\mathcal{G}}'_{sym}$ is $m_{sym}1$. The sum of the lengths of its persistent homology barcodes is defined as $SRTD(P, P')$. Its chain complex is homotopy equivalent to the mapping cone of the inclusion map $f' : C_*(A \cap B) \to C_*(A \cup B)$.*

Tulchinskii et al. (2025) proposed RTD-Lite, pointing out that $\frac{mst(w)+mst(\tilde{w})-2mst(\min(w,\tilde{w}))}{2}$ is equivalent to the information related only to clustering (i.e., 0-dimensional homology) in $\frac{RTD_1(w,\tilde{w})+RTD_1(\tilde{w},w)}{2}$, where $mst(w)$ is the length of the minimum spanning tree for the distance matrix w. Building upon RTD-lite, we can similarly define its 'max' variant, Max-RTD-Lite, and a symmetric version, SRTD-Lite, which focuses on the clustering differences between $\min(w, \tilde{w})$ and $\max(w, \tilde{w})$.

Definition 2 (SRTD-Lite) *By comparing the minimum spanning trees of $\min(w, \tilde{w})$ and $\max(w, \tilde{w})$ (see Algorithm A.6), we can obtain a series of barcodes, $SRTD\text{-}L\text{-}barcode(w, \tilde{w})$. We define the sum of the lengths of these barcodes as $SRTD\text{-}Lite(w, \tilde{w})$.*

3.1. Mathematical Properties

SRTD, RTD, and Max-RTD satisfy some elegant mathematical properties. The mapping cones corresponding to their auxiliary graphs fit into the following long exact sequence:

$$\cdots \to H_n(R_\alpha(\mathcal{G}^w), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})) \xrightarrow{\gamma_n} H_n(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})}), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})}))$$
$$\xrightarrow{\beta_n} H_n(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})}), R_\alpha(\mathcal{G}^w)) \xrightarrow{\delta_n} H_{n-1}(R_\alpha(\mathcal{G}^w), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})) \xrightarrow{\gamma_{n-1}} \cdots$$

Theorem 3 *For any dimension i, point clouds P, P' and distance matrices $w, \tilde{w}$, the three divergences satisfy the following relationship:*

$$RTD_i(w, \tilde{w}) + Max\text{-}RTD_i(w, \tilde{w}) - SRTD_i(P, P') = \int_0^\infty (\dim(\ker(\gamma_i)) + \dim(\ker(\gamma_{i-1})))d\alpha$$

By swapping the positions of w and $\tilde{w}$ in Theorem 3, we obtain a similar equality. We denote $RTD_i(w, \tilde{w}) + Max\text{-}RTD_i(w, \tilde{w})$ as $minmax(w, \tilde{w})$,and $RTD_i(\tilde{w}, w) + Max\text{-}RTD_i(\tilde{w}, w)$as $minmax(\tilde{w}, w)$. Both are strictly greater than SRTD, but in our experiments, we find this gap to be very small, as shown in the Table 5(*b*).

The introduction of SRTD provides a more mathematically elegant explanation for the RTD family of divergences. The terms $minmax(w, \tilde{w})$ and $minmax(\tilde{w}, w)$ can be seen as mixed divergences calculated from the perspectives of w and $\tilde{w}$, respectively. Their minimal difference stems from their large shared component, $SRTD(w, \tilde{w})$. Their 'private' parts can be understood as topological features that do not exist in $\mathcal{G}^{\max(w,\tilde{w})}$, are born only in $\mathcal{G}^w$ or $\mathcal{G}^{\tilde{w}}$, and die in $\mathcal{G}^{\min(w,\tilde{w})}$. Consequently, calculating the divergence from different directions does not yield significant discrepancies, and this difference becomes interpretable. In the lite version, this relationship is even more elegant:

Corollary 4 $Max\text{-}RTD\text{-}Lite(w, \tilde{w}) + RTD\text{-}Lite(w, \tilde{w}) = SRTD\text{-}Lite(w, \tilde{w})$

Corollary 5 $Max\text{-}RTD\text{-}Lite(P, P') \geq SRTD\text{-}Lite(P, P') \geq RTD\text{-}Lite(P, P')$

4. Experiments

4.1. Behavioral Similarity on Synthetic Data

Clusters Experiment We conduct an experiment on a 300-point point cloud, using a single-cluster point cloud as a baseline. Our results first confirm the behavioral similarity of the three divergences, the complementarity between RTD and Max-RTD, and that the gap in Theorem 4 is negligible. Furthermore, we find that the RTD-lite divergence exhibits anomalous behavior, with its direction of change being completely opposite to the expected trend. In contrast, both Max-RTD-lite and SRTD, which incorporate information from $max(w, \tilde{w})$, correctly reflect the trend. This finding highlights the inadequacy of calculating divergence based solely on $min(w, \tilde{w})$. A detailed description is provided in Appendix F.

Sensitivity Experiment In our sensitivity experiment, we use 5 point clusters where points move radially outwards from their respective fixed centers. The distance of each point to its center, denoted by α, is varied from 0.1 to 12. When compared to the baseline configuration at $\alpha = 0.1$, we find that as the points move further apart, RTD becomes progressively less sensitive to the changes. However, SRTD and Max-RTD remain effective at detecting these structural variations, see Appendix E for details.

4.2. UMAP Embedding Analysis

We conduct an experiment on the 2D representations from UMAP with varying numbers of neighbors. The results show the behavioral similarity among the three divergences, the complementarity between RTD and Max-RTD, and that the gap in Theorem 4 is negligible. A detailed experimental description is provided in Appendix C

4.3. Gradient optimization for autoencoder

Following RTD-AE and RTD-lite (Trofimov et al., 2023; Tulchinskii et al., 2025), we use a small autoencoder to project COIL-20 and F-MNIST into a 16-dimensional space with SRTD and SRTD-lite. Both SRTD and SRTD-lite achieve the best performance in their respective series. Furthermore, SRTD is nearly twice as fast as RTD. Detailed experimental settings and descriptions can be found in Appendix D.

5. LLM fingerprinting

We propose using SRTD to reveal the homology of large language models, for which we conducted preliminary validation experiments. Zhang et al. (2024) has shown that Centered Kernel Alignment is significantly effective in identifying pruned and fine-tuned versions of various large language models. In our initial experiments, we selected a few models and their instruction-tuned versions(13 models) and extracted their representations on the TrustfulQA dataset(Lin et al., 2021). We found that SRTD has the potential to identify models from different origins, as illustrated in Appendix G.

References

Serguei Barannikov, Ilya Trofimov, Nikita Balabin, and Evgeny Burnaev. Representation topology divergence: A method for comparing neural network representations. *arXiv preprint arXiv:2201.00058*, 2021.

Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4: 667963, 2021.

Sebastian Damrich and Fred A Hamprecht. On umap's true loss function. *Advances in Neural Information Processing Systems*, 34:5798–5809, 2021.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International conference on machine learning*, pages 3519–3529. PMlR, 2019.

Stephanie Lin, Jacob Hilton, and Owain Evans. Truthfulqa: Measuring how models mimic human falsehoods. *arXiv preprint arXiv:2109.07958*, 2021.

Ilya Trofimov, Daniil Cherniavskii, Eduard Tulchinskii, Nikita Balabin, Evgeny Burnaev, and Serguei Barannikov. Learning topology-preserving data representations. *arXiv preprint arXiv:2302.00136*, 2023.

Eduard Tulchinskii, Daria Voronkova, Ilya Trofimov, Evgeny Burnaev, and Serguei Barannikov. Rtd-lite: Scalable topological analysis for comparing weighted graphs in learning tasks. *arXiv preprint arXiv:2503.11910*, 2025.

Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. Understanding how dimension reduction tools work: an empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization. *Journal of Machine Learning Research*, 22 (201):1–73, 2021.

Jie Zhang, Dongrui Liu, Chen Qian, Linfeng Zhang, Yong Liu, Yu Qiao, and Jing Shao. Reef: Representation encoding fingerprints for large language models. *arXiv preprint arXiv:2410.14273*, 2024.

Appendix A. Important Matrix and Algorithm

In the definitions below, for any given matrix M, we denote M^+ as the matrix obtained by replacing its strictly upper triangular part with infinity (∞).

A.1. Symmetric Auxiliary Matrix

This matrix is central to the definition of Symmetric Representation Topology Divergence (SRTD).

$$m_{sym} = \begin{pmatrix} \max(w, \tilde{w}) & (\max(w, \tilde{w})^+)^T & 0 \\ \max(w, \tilde{w})^+ & \min(w, \tilde{w}) & \infty \\ 0 & \infty & 0 \end{pmatrix} \tag{1}$$

A.2. Min Auxiliary Matrix

This matrix is used to compute the standard Representation Topology Divergence (RTD).

$$m_{min} = \begin{pmatrix} w & (w^+)^T & 0 \\ w^+ & \min(w, \tilde{w}) & \infty \\ 0 & \infty & 0 \end{pmatrix} \tag{2}$$

A.3. Max Auxiliary Matrix

This matrix is used in the definition of Max-RTD.

$$m_{max} = \begin{pmatrix} \max(w, \tilde{w}) & (\max(w, \tilde{w})^+)^T & 0 \\ \max(w, \tilde{w})^+ & w & \infty \\ 0 & \infty & 0 \end{pmatrix} \tag{3}$$

A.4. Max-RTD

Definition 6 (Max-RTD) *For two point clouds P and P' with a one-to-one correspondence, the distance matrix of their auxiliary graph $\hat{\mathcal{G}}'_{max}$ is given by m_{max}3. The sum of the lengths of the persistent homology barcodes of $\hat{\mathcal{G}}'_{max}$ is defined as $Max\text{-}RTD(w, \tilde{w})$. Its chain complex is homotopy equivalent to the mapping cone of the inclusion map $f' : C_*(A \cap B) \to C_*(A)$.*

A.5. SRTD algorithm

Algorithm 1: Symmetric Divergence Score Calculation (SRTD)

Input: Pairwise distance matrices $w, \tilde{w}$
Output: Symmetric divergence scores Sym-Divergence$_i$
Normalize $w, \tilde{w}$ by their 0.9 quantiles;
Let $m_{\max}+ \leftarrow \max(w, \tilde{w})$, and replace its upper triangular part with $+\infty$;
Construct the symmetric auxiliary matrix m_{sym}1 Compute barcodes
Sym-Barcode$_i \leftarrow B(vr(m_{sym}), i)$;
Compute divergence Sym-Divergence$_i \leftarrow \sum_{(b,d)\in \text{Sym-Barcode}_i}(d - b)$;

A.6. SRTD_lite Barcode Algorithm

Algorithm 2 Computation of SRTD-Lite Barcode

Input: D_1, D_2 — weight matrices of two models.
Require:
$MST(\cdot)$ — function computing Minimal Spanning Tree, returns list of edges.
$Sort(\cdot)$ — function sorting list of edges by their weights.
Output: Multiset of pairs (intervals constructing $SRTD\text{-}L\text{-}Barcode(D_1, D_2)$)

procedure SRTD-L-Barcode(D_1, D_2)
  $D_1', D_2' \leftarrow D_1, D_2$ divided by their 0.9 quantiles
  $D_{\min} \leftarrow$ element-wise minimum of D_1' and D_2'
  $D_{\max} \leftarrow$ element-wise maximum of D_1' and D_2'
  $E_{\min} \leftarrow Sort(MST(D_{\min}))$
  $E_{\max} \leftarrow Sort(MST(D_{\max}))$
  $BarcodeSet \leftarrow []$
  $SubTree \leftarrow$ empty graph with N vertices
  for each edge $e = (u, v)$ with weight w_{birth} in $E_{\min}$ **do**
    if u and v are not connected in $SubTree$ **then**
      $TemporaryGraph \leftarrow$ copy($SubTree$)
      for each edge $e' = (u', v')$ with weight w_{death} in $E_{\max}$ **do**
        Add e' to $TemporaryGraph$
        if u and v are connected in $TemporaryGraph$ **then**
          Add (w_{birth}, w_{death}) to $BarcodeSet$
          break
        end if
      end for
      Add e to $SubTree$
    end if
  end for
  return $BarcodeSet$
end procedure

Appendix B. Proofs

B.1. statement in definition

We first prove the following lemmas, they are stated in Definition 6 and Definition 1:

Lemma 7 *There exists a specially constructed auxiliary graph $\hat{\mathcal{G}}'_{max}$ such that its chain complex is homotopy equivalent to the mapping cone $Cone(f')$, where $f' : C_*(A \cap B) \to C_*(A)$ is a chain map induced by the inclusion.*

$$R_\alpha(\hat{\mathcal{G}}'_{max}) \sim Cone\left(R_\alpha(\mathcal{G}^{\max(w,\tilde{w})}) \to R_\alpha(\mathcal{G}^w)\right)$$

Lemma 8 *Similarly, there exists a specially constructed auxiliary graph $\hat{\mathcal{G}}'_{sym}$ such that its chain complex is homotopy equivalent to the mapping cone $Cone(f')$, where $f' : C_*(A \cap B) \to$*

$C_*(A \cup B)$ is a chain map induced by the inclusion.

$$R_\alpha(\hat{\mathcal{G}}'_{sym}) \sim Cone\left(R_\alpha(\mathcal{G}^{\max(w,\tilde{w})}) \to R_\alpha(\mathcal{G}^{\min(w,\tilde{w})})\right)$$

Proof

The mapping cone we are interested in is constructed from the direct sum of the following chain complexes:

$$\text{Cone}(f') = C_*(A \cap B)[-1] \oplus C_*(A)$$

Following the construction from the RTD paper, we can propose two auxiliary graph schemes: The vertex set of the auxiliary graph $\hat{\mathcal{G}}'_{max}$ is composed of the original vertices v'_i, mirrored vertices v_i, and a special vertex O. Its distance rules are defined as follows: $d'_{v_iv_j} = \max(w_{ij}, \tilde{w}_{ij})$,$d'_{v'_iv'_j} = w_{ij}$,$d'_{v_iv'_i} = 0$,$d'_{Ov_i} = 0$, $d'_{Ov'_i} = +\infty$,$d'_{v_iv'_j} = \max(w_{ij}, \tilde{w}_{ij})$

The vertex set of the auxiliary graph $\hat{\mathcal{G}}'_{sym}$ is composed of twice the number of original vertices and O. $d'_{v_iv_j} = \max(w_{ij}, \tilde{w}_{ij})$,$d'_{v'_iv'_j} = \min(w_{ij}, \tilde{w}_{ij})$,$d'_{v_iv'_i} = 0$,$d'_{Ov_i} = 0$, $d'_{Ov'_i} = +\infty$,$d'_{v_iv'_j} = \max(w_{ij}, \tilde{w}_{ij})$

For the auxiliary graph $R_\alpha(\hat{\mathcal{G}}'_{max})$, there are three types of simplices:

- $A_{i_1} \dots A_{i_k} A'_{i_k} \dots A'_{i_n}$, where $\max(w_{A_{i_r}A_{i_s}}, \tilde{w}_{A_{i_r}A_{i_s}}) \leq \alpha$ for $r \leq k$, and $w_{A_{i_r}A_{i_s}} \leq \alpha$ for $r, s \geq k$.
- $A_{i_1} \dots A_{i_k} A'_{i_{k+1}} \dots A'_{i_n}$, where $\max(w_{A_{i_r}A_{i_s}}, \tilde{w}_{A_{i_r}A_{i_s}}) \leq \alpha$ for $r \leq k$, and $w_{A_{i_r}A_{i_s}} \leq \alpha$ for $r, s \geq k+1$.
- $OA_{i_1}A_{i_2} \dots A_{i_n}$, where $\max(w_{A_{i_r}A_{i_s}}, \tilde{w}_{A_{i_r}A_{i_s}}) \leq \alpha$.

Forward Map

$$\psi' : \text{Cone}(f') \to R_\alpha(\hat{\mathcal{G}}'_{max})$$

- For $c \in C_*(A \cap B)[-1]$ (of the form $A_{i_1} \dots A_{i_n}[-1]$):

$$\psi'(c) = OA_{i_1} \dots A_{i_n} + \sum_{k=1}^{n} A_{i_1} \dots A_{i_k} A'_{i_k} \dots A'_{i_n}$$

- For $a \in C_*(A)$ (of the form $A_{i_1} \dots A_{i_n}$):

$$\psi'(a) = A'_{i_1} \dots A'_{i_n}$$

Backward Map

$$\tilde{\psi}' : R_\alpha(\hat{\mathcal{G}}'_{max}) \to \text{Cone}(f')$$

- $\tilde{\psi}'(OA_{i_1} \dots A_{i_n}) = A_{i_1} \dots A_{i_n}[-1]$
- $\tilde{\psi}'(A'_{i_1} \dots A'_{i_n}) = A_{i_1} \dots A_{i_n}$
- $\tilde{\psi}'(\Delta) = 0$ (for all other types of simplices Δ)

Homotopy Operator H For the second type of simplex:

$$H : A_{i_1} \dots A_{i_k} A'_{i_{k+1}} \dots A'_{i_n} \to \sum_{l=1}^{k} A_{i_1} \dots A_{i_l} A'_{i_l} \dots A'_{i_n}, 1 \le k \le n$$

For all other simplices:

$$H(\Delta) = 0$$

Therefore, $\tilde{\psi}' \circ \psi' = \mathrm{Id}$ and $\psi' \circ \tilde{\psi}' - \mathrm{Id} = H\partial - \partial H$. This proves Lemma 7, and Lemma 8 can be proven similarly. ■

B.2. proof of Theorem 3

Lets proof Theorem 3. To proof the theorem,we just need to proof the following theorem:

Theorem 9 *For any dimension i, the Betti numbers of the three auxiliary graphs satisfy the following relation:*

$$\beta_i^{\min}(\alpha) + \beta_i^{\max}(\alpha) - \beta_i^{sym}(\alpha) = \dim(ker(\gamma_i)) + \dim(ker(\gamma_{i-1}))$$

Proof We have the following inclusion of simplicial complexes:

$$R_\alpha(\mathcal{G}^{\max(w,\tilde{w})}) \subseteq R_\alpha(\mathcal{G}^{w}) \subseteq R_\alpha(\mathcal{G}^{\min(w,\tilde{w})})$$

This forms a triple of complexes, which gives rise to a standard short exact sequence of their chain complexes:

$$0 \to C_*(R_\alpha(\mathcal{G}^{w}), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})) \to C_*(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})}), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})) \to C_*(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})}), R_\alpha(\mathcal{G}^{w})) \to 0$$

This, in turn, induces the following long exact sequence in homology:

$$\begin{aligned}\dots \to H_n(R_\alpha(\mathcal{G}^{w}), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})) &\to H_n(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})}), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})) \\ &\to H_n(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})}), R_\alpha(\mathcal{G}^{w})) \xrightarrow{\partial_*} H_{n-1}(R_\alpha(\mathcal{G}^{w}), R_\alpha(\mathcal{G}^{\max(w,\tilde{w})})) \to \dots\end{aligned}$$

Since the relative homology groups are isomorphic to the homology groups of the corresponding mapping cones, we have the following long exact sequence for the auxiliary graphs:

$$\dots \to H_i(R_\alpha(\hat{\mathcal{G}}'_{max})) \xrightarrow{\gamma_i} H_i(R_\alpha(\hat{\mathcal{G}}'_{sym})) \xrightarrow{\beta_i} H_i(R_\alpha(\hat{\mathcal{G}}'_{min})) \xrightarrow{\delta_i} H_{i-1}(R_\alpha(\hat{\mathcal{G}}'_{max})) \to \dots$$

where $\gamma_i, \beta_i, \delta_i$ are the homomorphism maps in the sequence. For any segment of an exact sequence of vector spaces $U \xrightarrow{f} V \xrightarrow{g} W$, we have $\mathrm{im}(f) = \ker(g)$. By the rank-nullity theorem, $\dim(V) = \dim(\ker(g)) + \dim(\mathrm{im}(g))$. Substituting $\mathrm{im}(f) = \ker(g)$, we get $\dim(V) = \dim(\mathrm{im}(f)) + \dim(\mathrm{im}(g))$. Therefore, the dimensions of the homology groups of the auxiliary graphs (i.e., the Betti numbers $\beta_i(\alpha)$) can be expressed as:

$$\beta_i^{\max}(\alpha) = \dim(H_i(R_\alpha(\hat{\mathcal{G}}'_{max}))) = \dim(\mathrm{im}(\delta_{i+1})) + \dim(\mathrm{im}(\gamma_i)) \tag{4}$$

$$\beta_i^{\mathrm{sym}}(\alpha) = \dim(H_i(R_\alpha(\hat{\mathcal{G}}'_{sym}))) = \dim(\mathrm{im}(\gamma_i)) + \dim(\mathrm{im}(\beta_i)) \tag{5}$$

$$\beta_i^{\min}(\alpha) = \dim(H_i(R_\alpha(\hat{\mathcal{G}}'_{min}))) = \dim(\mathrm{im}(\beta_i)) + \dim(\mathrm{im}(\delta_i)) \tag{6}$$

By substituting equations (4), (5), and (6), we obtain:

$$\begin{aligned}
&\beta_i^{\min}(\alpha) + \beta_i^{\max}(\alpha) - \beta_i^{\text{sym}}(\alpha) \\
&= \big(\dim(\text{im}(\beta_i)) + \dim(\text{im}(\delta_i))\big) \\
&\quad + \big(\dim(\text{im}(\delta_{i+1})) + \dim(\text{im}(\gamma_i))\big) \\
&\quad - \big(\dim(\text{im}(\gamma_i)) + \dim(\text{im}(\beta_i))\big) \\
&= \dim(\text{im}(\delta_{i+1})) + \dim(\text{im}(\delta_i)) \\
&= \dim(\ker(\gamma_i)) + \dim(\ker(\gamma_{i-1}))
\end{aligned}$$

By integrating both sides of Theorem 9 with respect to filtration radius α, we obtain its conclusion. This completes the proof of Theorem 9 and Theorem 3. ■

B.3. proof of corollary

proof of Corollary 4 From definition, we have

$$RTD\text{-}lite(P, P') = \frac{(mst(\mathcal{G}^w) - mst(\mathcal{G}^{\min(w,\tilde{w})})) + (mst(\mathcal{G}^{\tilde{w}}) - mst(\mathcal{G}^{\min(w,\tilde{w})}))}{2}$$

$$Max\text{-}RTD\text{-}lite(P, P') = \frac{(mst(\mathcal{G}^{\max(w,\tilde{w})}) - mst(\mathcal{G}^w)) + (mst(\mathcal{G}^{\max(w,\tilde{w})}) - mst(\mathcal{G}^{\tilde{w}}))}{2}$$

$$SRTD\text{-}lite(P, P') = mst(\mathcal{G}^{\max(w,\tilde{w})}) - mst(\mathcal{G}^{\min(w,\tilde{w})})$$

Summing the three equations above completes the proof.

proof of Corollary 5 This corollary holds if and only if the following expression is true, where A and B are two non-negative, symmetric distance matrices of the same size with zeros on the diagonal.

Proof

$$\text{MST}(\max(A, B)) + \text{MST}(\min(A, B)) \geq \text{MST}(A) + \text{MST}(B). \qquad (\star)$$

Let the graph have n vertices and an edge set E. We can view a weight matrix W as a function that assigns a non-negative weight W_e to each edge $e \in E$. For any non-negative weight matrix W, let $E_{\leq t}(W) := \{e \in E : W_e \leq t\}$ be the set of edges with weight at most t, and let $\kappa_W(t)$ be the number of connected components in the graph $(V, E_{\leq t}(W))$. A standard result from Kruskal's algorithm gives the MST weight as an integral:

$$\text{MST}(W) = \int_0^\infty \big(\kappa_W(t) - 1\big)\, dt. \qquad (7)$$

The element-wise min and max operations on weight matrices correspond to the union and intersection of their threshold edge sets:

$$\begin{aligned}
E_{\leq t}(\max(A, B)) &= E_{\leq t}(A) \cap E_{\leq t}(B), \\
E_{\leq t}(\min(A, B)) &= E_{\leq t}(A) \cup E_{\leq t}(B).
\end{aligned} \qquad (8)$$

Let $\kappa(S)$ be the number of connected components of the graph induced by an edge set $S \subseteq E$. A fundamental result in graph theory and matroid theory is that the rank function $r(S) = n - \kappa(S)$ is submodular. Consequently, $\kappa(S)$ is supermodular:

$$\kappa(X \cap Y) + \kappa(X \cup Y) \geq \kappa(X) + \kappa(Y), \quad \forall X, Y \subseteq E. \tag{9}$$

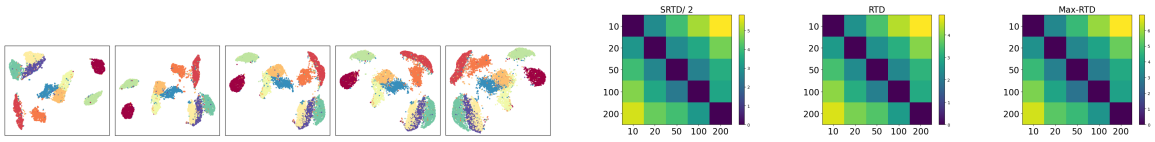
Substituting (8) into (9) with $X = E_{\leq t}(A)$ and $Y = E_{\leq t}(B)$, we get for every $t \geq 0$:

$$\kappa_{\max(A,B)}(t) + \kappa_{\min(A,B)}(t) \geq \kappa_A(t) + \kappa_B(t).$$

Integrating over $t \in [0, \infty)$, and applying the formula (7) yields the desired inequality ($\star$). ■

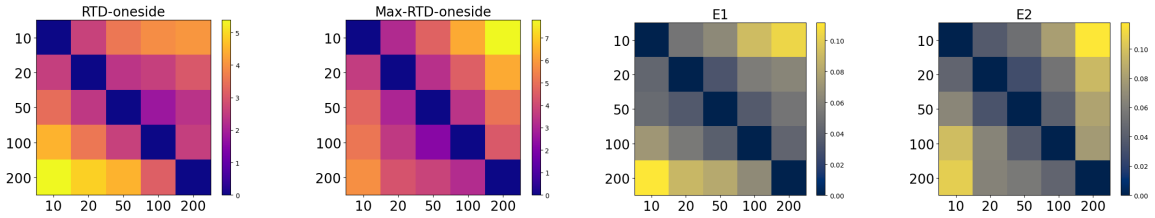
Appendix C. UMAP Experiment

UMAP is a state-of-the-art dimensionality reduction technique for visualization(Damrich and Hamprecht, 2021), excelling at preserving both the local and global structure of the data. We select a range for the `n_neighbors` parameter: (10, 20, 50, 100, and 200). We then compute the pairwise RTD, Max-RTD, and SRTD between their 2D representations. From Figure 1(b), we can observe the similarity among the three divergences. From Figure 2(a) shows $RTD(w, \tilde{w}) - RTD(\tilde{w}, w)$ (left column) and $Max - RTD(w, \tilde{w}) - Max - RTD(\tilde{w}, w)$ (right column), reflecting their asymmetry and complementarity.From Figure 2(b), we observe the minimal difference between $minmax(w, \tilde{w})$, $minimax(\tilde{w}, w)$, and $SRTD$,Definition of E1 is the same as above.



(a) UMAP(10,20,50,100,200) (b) Divergence scores

Figure 1: Similarity of 3 divergence measures.



(a) Symmetry and Complementarity (b) Minimal Difference Between SRTD and Minmax Divergence

Figure 2: Sensitivity analysis of divergence measures.

Appendix D. Experiment on Autoencoder and Experimental Setup

D.1. Experiment on autoencoder

Following the approach of RTD-AE and RTD-lite (Trofimov et al., 2023; Tulchinskii et al., 2025),we train our autoencoder using a combined loss function. This objective includes a standard reconstruction loss alongside our proposed SRTD (or SRTD_lite) divergence, which is computed between the high-dimensional input data and its low-dimensional latent representation. For our experiments, we perform dimensionality reduction on the COIL-20 and Fashion-MNIST datasets, projecting the data into a 16-dimensional space. To evaluate the quality of the reduction, we compare the original and latent representations using the following metrics: (1) linear correlation of pairwise distances, (2) the Wasserstein distance of the H_0 persistent homology barcodes (Chazal and Michel, 2021), (3) triplet distance ranking accuracy (Wang et al., 2021), (4) RTD (Barannikov et al., 2021) (5) SRTD. The results of RTD series are summarized in Table 1 and Table 2,. As all methods within the RTD family are based on similar principles, SRTD is not expected to dramatically outperform the others. Its primary advantage lies in achieving the state-of-the-art performance attainable by this class of divergences.

Table 1: Dimensionality Reduction Quality Metrics(COIL-20).

Method	**Dist Corr**	**Triplet Acc**	**H0 Wass**	**RTD**	**SRTD**
RTD	0.942	0.893 ± 0.01	40.1±0.0	1.28 ± 0.4	1.29 ± 0.4
Max-RTD	0.924	0.879 ± 0.01	32.3±0.0	1.17 ± 0.3	1.17 ± 0.3
SRTD	0.948	0.899 ± 0.01	36.7±0.0	1.21 ± 0.4	1.21 ± 0.4

Table 2: Dimensionality Reduction Quality Metrics(F-mnist).

Method	**Dist Corr**	**Triplet Acc**	**H0 Wass**	**RTD**	**SRTD**
RTD	0.954	0.907 ± 0.00	98.2 ± 4.3	1.28 ± 0.1	1.35 ± 0.2
Max-RTD	0.937	0.895 ± 0.01	94.1 ± 4.1	1.51 ± 0.1	1.55 ± 0.1
SRTD	0.957	0.910 ± 0.01	94.0 ± 2.7	1.29 ± 0.1	1.34 ± 0.2

Table 3and Table 4 illustrate the dimensionality reduction performance of the lite series divergences.

Table 3: Dimensionality Reduction Quality Metrics(COIL-20).

Method	**Dist Corr**	**Triplet Acc**	**H0 Wass**	**RTD**	**SRTD**
RTD_lite	0.904	0.855 ± 0.01	26.0±0.0	0.99 ± 0.3	1.00 ± 0.3
Max-RTD_lite	0.935	0.886 ± 0.01	29.9 ± 0.0	1.03 ± 0.3	1.04 ± 0.3
SRTD_lite	0.930	0.882 ± 0.01	28.2 ± 0.0	1.00 ± 0.2	1.01 ± 0.2

Table 4: Dimensionality Reduction Quality Metrics(F-mnist).

Method	**Dist Corr**	**Triplet Acc**	**H0 Wass**	**RTD**	**SRTD**
RTD_lite	0.937	0.896 ± 0.01	90.2 ± 3.9	1.38 ± 0.1	1.43 ± 0.1
Max-RTD_lite	0.940	0.897 ± 0.00	92.0 ± 3.6	1.47 ± 0.1	1.51 ± 0.2
SRTD_lite	0.941	0.897 ± 0.00	91.4 ± 5.1	1.42 ± 0.1	1.47 ± 0.1

D.2. Experimental Setup

Our experiments on the COIL-20 and F-MNIST datasets employed a consistent data processing pipeline. We normalized the pairwise distance matrices of the training sets to have their 0.9 quantiles equal to 1. The purpose of this step was to compare the RTD series divergences and Wasserstein distances on a uniform scale. Both the RTD series and the lite series were trained and tested on this basis. Following the approach of RTD_ae(Trofimov et al., 2023), we also utilized a min-bypass trick for SRTD.

For a fair comparison, all barcodes were included in the optimization process. **Our experiments were designed to measure whether SRTD could achieve the same level of performance as the RTD family of divergences, all while reducing computational costs. As our follow-up work is focused on using SRTD to study large language model representations, we did not perform a detailed comparison with other dimensionality reduction methods.**

The specific parameters used in our experiments are detailed below:

Table 5: Experimental Parameters

Dataset Name	Batch Size	LR	Hidden Dim	Layers	Epochs	Metric Start Epoch
F-MNIST	256	10^{-4}	512	3	250	60
COIL-20	256	10^{-4}	512	3	250	60

Table 6: Dataset Characteristics

Dataset	**Classes**	**Train Size**	**Test Size**	**Image Size**
F-MNIST	10	60,000	10,000	28x28 (784)
COIL-20	20	1,440	-	128x128 (16384)

Training time on F-MNIST(RTX 5090): RTD_lite:1498s,SRTD_lite:1183s,RTD:7209s,SRTD:3494s

Appendix E. Sensitive Experiment

The max-divergence offers more than just a uniform vertical shift of the min-divergence. As $max(w, \tilde{w})$ enhances the overall separation of the point cloud, it is more sensitive to changes in the data structure compared to $min(w, \tilde{w})$. To demonstrate this, we designed a sensitivity experiment. We select 5 fixed centers, with 50 points positioned on the circumference of a circle of radius α around each center. We maintain a constant direction for each point

relative to its respective center(baseline $\alpha = 0.1$), only increasing their distance to the center to generate a series of point clouds, as shown in Figure 3. As α gradually increases, the standard RTD becomes less effective at capturing the changes. However, $Max\text{-}RTD$ and $SRTD$, which incorporate information from $max(w, \tilde{w})$, continue to detect these changes effectively.

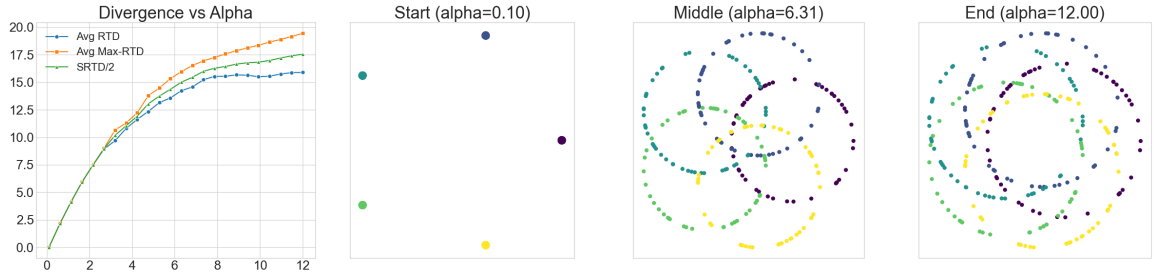


Figure 3: sensitivity experiment

Appendix F. Clusters Experiment

We generated point clouds with varying numbers of clusters. The initial point cloud consists of 300 points, which are subsequently arranged into 2, 3, 4, ..., up to 12 clusters. We compare the initial point cloud (w) as a baseline against the subsequent point clouds ($\tilde{w}$). Let $E_1 = \frac{\text{minmax}(w,\tilde{w}) - \text{SRTD}(w,\tilde{w})}{2}$, and E_2 is defined similarly. 'Percentage' represents the ratio $\frac{\text{minmax}(w,\tilde{w}) - \text{SRTD}(w,\tilde{w})}{\text{SRTD}(w,\tilde{w})}$ expressed as a percentage. As shown in Figure 5(a), the behaviors of the three divergence metrics are highly similar. Table 5(b) indicates that the right-hand side of theorem 3 actually constitutes a very small proportion. Furthermore, Table 5(c) reveals that RTD and $Max\text{-}RTD$ exhibit a high degree of asymmetry and complementarity.

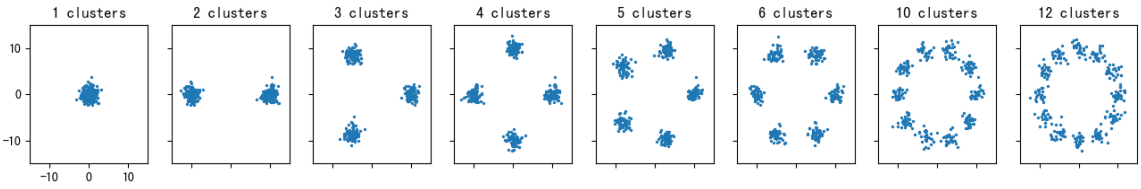
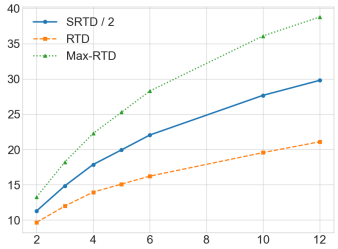


Figure 4: Divergence scores on synthetic cluster datasets.

Here Figure 6 are the performances of the lite series divergences and CKA on point clouds of clusters, with a single-cluster point cloud as the baseline. It can be seen that RTD_lite shows an anomaly here; while the RTD series, along with max_rtd_lite and srtd_lite, are able to recognize that similarity gradually decreases as the number of point clouds increases, RTD_lite exhibits a completely different trend. CKA fails to identify any pattern in this task. This once again reflects the bias and anomalies that can arise when analyzing differences solely by using $\min(w, \tilde{w})$. By introducing $\min(w, \tilde{w})$, both max_rtd_lite and srtd_lite manage to avoid this anomaly and identify the correct pattern.



(a) Divergence

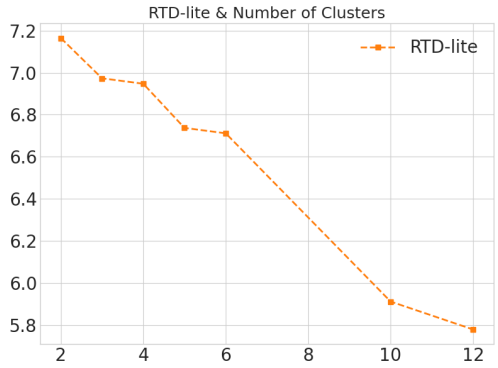
Clust.	E_1	E_2	Percent
2	0.357	0.000	3.16(0.00)
3	0.493	0.013	3.32(0.09)
4	0.441	0.061	2.47(0.34)
5	0.451	0.039	2.26(0.20)
6	0.347	0.060	1.57(0.27)
10	0.263	0.043	0.95(0.15)
12	0.226	0.046	0.76(0.15)

(b) Diff

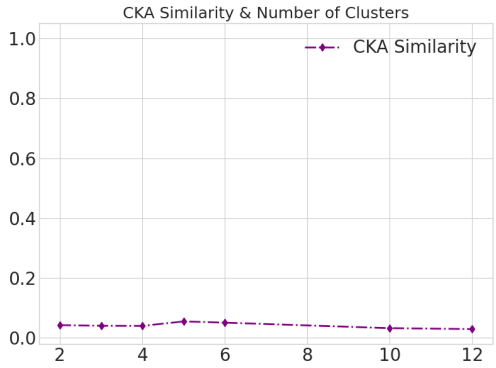
Min	Max
13.0976	-12.3839
11.2554	-10.2954
10.8131	-10.0535
10.3320	-9.5084
9.4315	-8.8572
8.3074	-7.8674
7.6888	-7.3296

(c) Asymmetry

Figure 5: Comprehensive analysis of synthetic cluster datasets. In table (b), $E_1 = \frac{\text{minmax}(w,\tilde{w}) - \text{SRTD}(w,\tilde{w})}{2}$, E_2 similarly

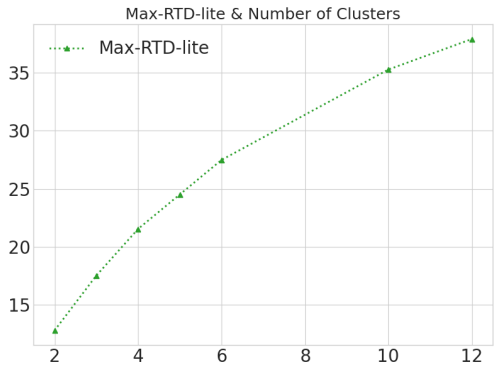


(a) RTD-lite



(b) CKA Similarity

(c) SRTD-lite



(d) Max-RTD-lite

Figure 6: scores on synthetic cluster datasets.

Here are the barcodes of the RTD series divergences from the clusters experimentFigure 7. It can be observed that the shape of the SRTD barcode (line 1) resembles a combination of the $RTD(w, \tilde{w})$ and $Max\text{-}RTD(w, \tilde{w})$ barcode(line 2 and 4), or the $RTD(\tilde{w}, w)$ and $Max\text{-}RTD(\tilde{w}, w)$ barcodes(line 3 and 5). The SRTD barcode is also richer in quantity, which provides evidence that SRTD can simultaneously extract the common features of both divergences in a single analysis.

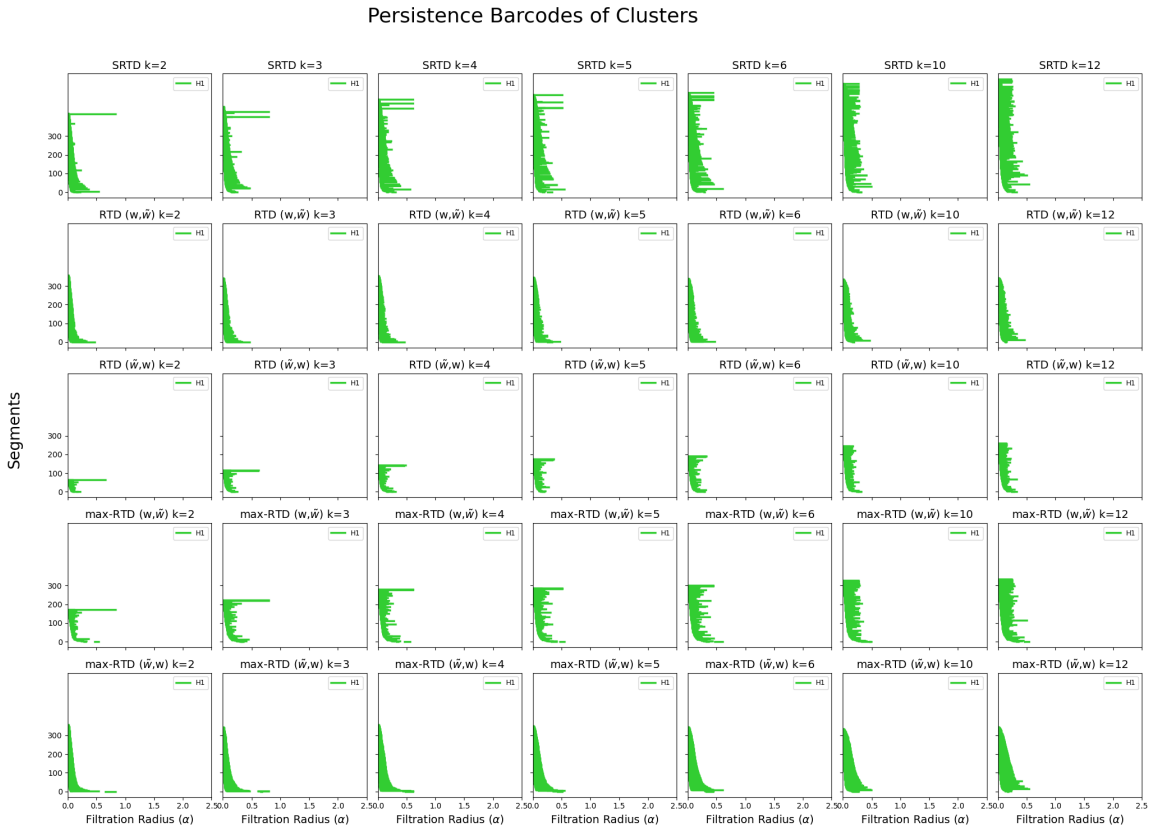
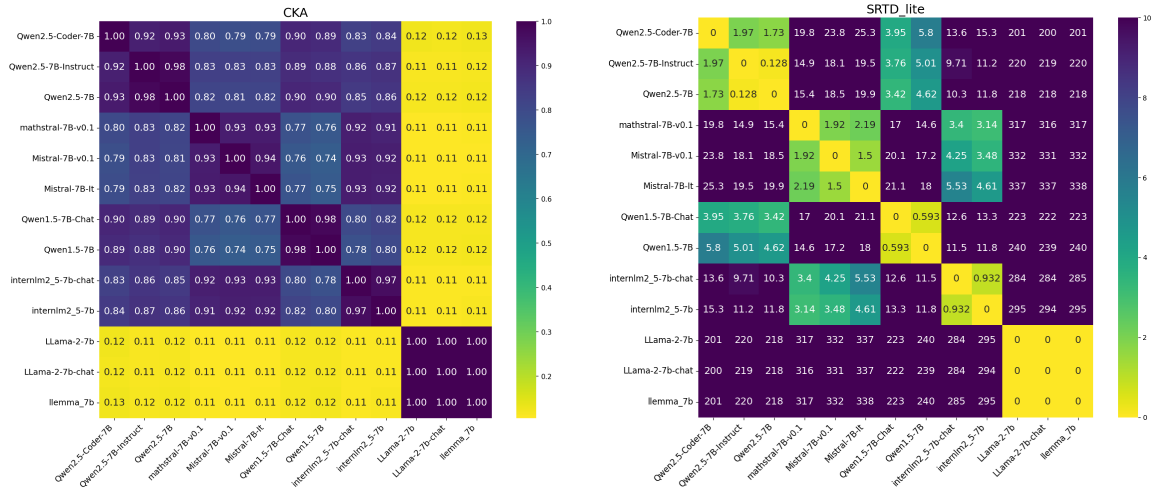


Figure 7: RTD series barrcodes

Appendix G. LLM Fingerprinting

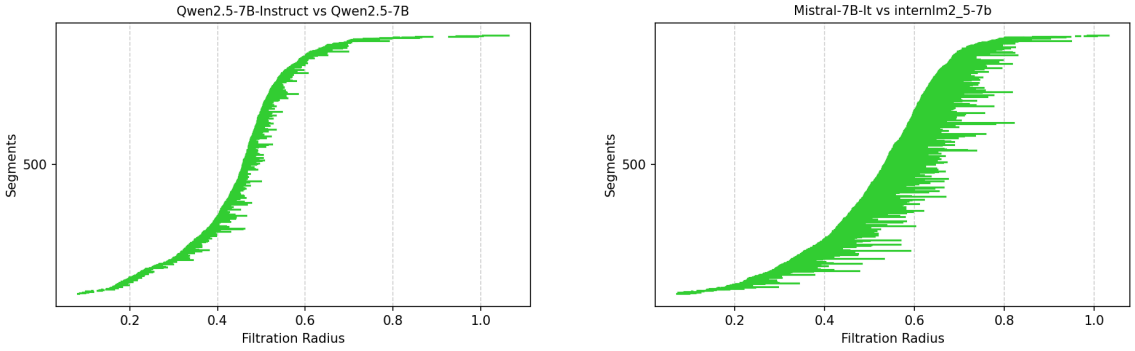
We selected the following models for our experiment: `Qwen2.5-Coder-7B`, `Qwen2.5-7B-Instruct`, `Qwen2.5-7B`, `mathstral-7B-v0.1`, `Mistral-7B-v0.1`, `Mistral-7B-It`, `Qwen1.5-7B-Chat`, `Qwen1.5-7B`, `internlm2_5-7b-chat`, `internlm2_5-7b`, `LLama-2-7b`, `LLama-2-7b-chat`, and `llemma_7b`. To analyze them, we randomly sampled 1000 question-answer pairs from the TrustfulQA dataset and fed them into each model to extract their sixth-layer representations. We chose these high-level representations because they tend to remain relatively stable throughout the training process. Subsequently, we employed REEF and SRTD_lite to perform a similarity comparison on these representations.We filtered out barcodes with a length less than 0.04 and calculated the sum of squares of the remaining barcodes. This soft and hard filtering approach penalizes longer barcodes more. Our findings indicate that

REEF often assigns high similarity scores to the vast majority of high-level representations, demonstrating poor reliability. In contrast, SRTD_lite can very clearly distinguish between models from the same family and those from different families, as illustrated in Figure 8. Figure Figure 9 shows the srtd_lite barcode for the homologous and non-homologous models.



(*a*) CKA (*b*) SRTD

Figure 8: LLM fingerprinting



(*a*) Qwen2.5-7B and Qwen2.5-7B-it (*b*) Mistral-7B-it and internlm2.5-7b

Figure 9: Barcodes example