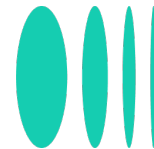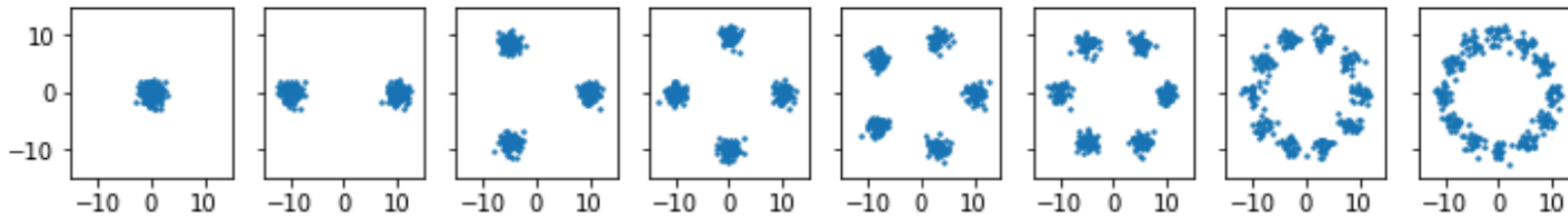# Representation Topology Divergence: a Method for Comparing Data Representations

Barannikov, S., Trofimov, I., Balabin, N., & Burnaev, E.
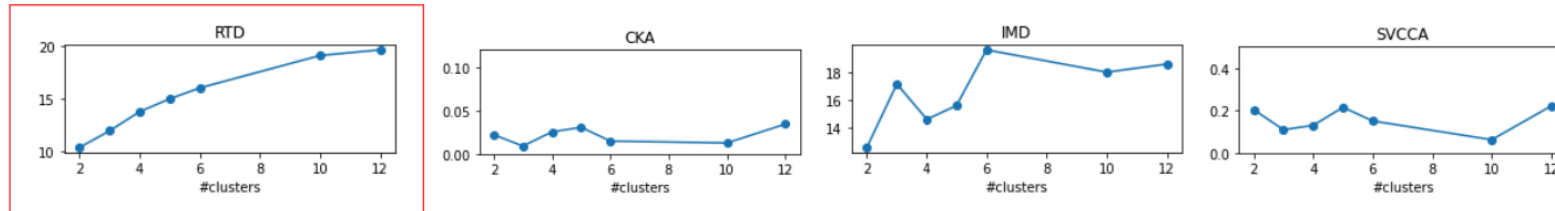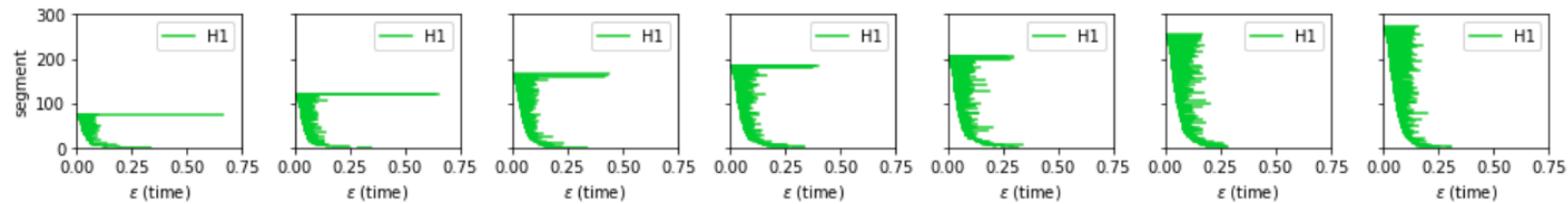
(a) Point clouds used in "clusters" experiment.

(b) Representations' comparison measures. Ideally, the measure should change monotonically with the increase of topological discrepancy.

(c) R-Cross-Barcode$(P, \tilde{P})$ for the "clusters" experiments. $\tilde{P}$ - is the point cloud having one cluster, P - 2, 3, 4, 5, 6, 10, 12 clusters.

Skoltech

A Method for Comparing Data Representations

- We introduce *R-Cross-Barcode*$(P, \tilde{P})$ - a Topological Data Analysis (TDA) tool, that measures the differences in the multi-scale topology of two point clouds $P, \tilde{P}$, with a point-to-point correspondence between clouds;
- Based on the *R-Cross-Barcode*$(P, \tilde{P})$, we define the *Representation Topology Divergence (RTD)*, the quantity measuring the multi-scale topological dissimilarity between two representations;
- RTD agrees with an intuitive notion of neural network representations similarity.

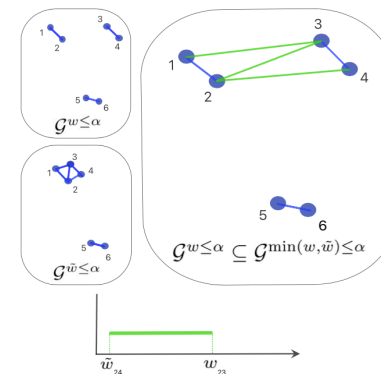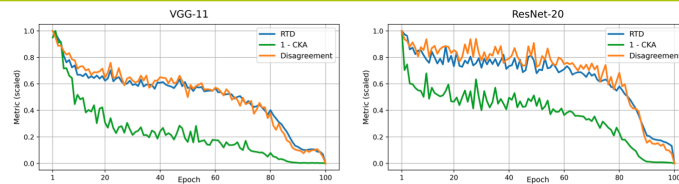# A Method for Comparing Data Representations

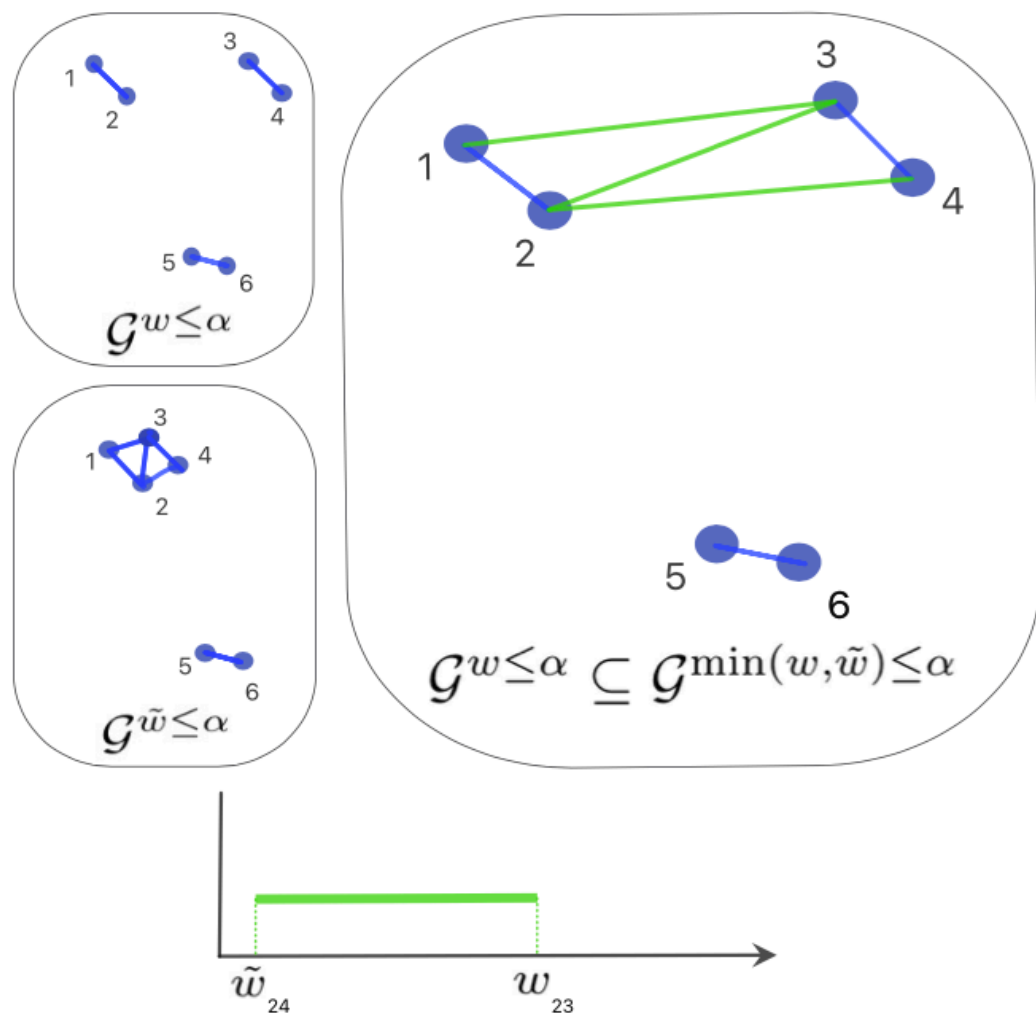

(a) Point clouds used in "clusters" experiment.

(b) Representations' comparison measures. Ideally, the measure should change monotonically with the increase of topological discrepancy.
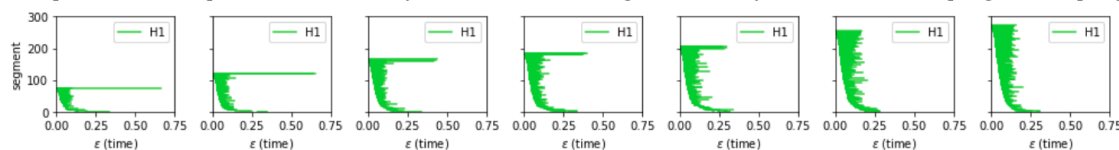
(c) R-Cross-Barcode($P, \tilde{P}$) for the "clusters" experiments. $\tilde{P}$ - is the point cloud having one cluster, P - 2, 3, 4, 5, 6, 10, 12 clusters.
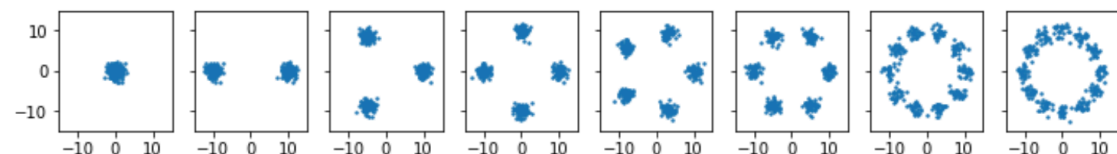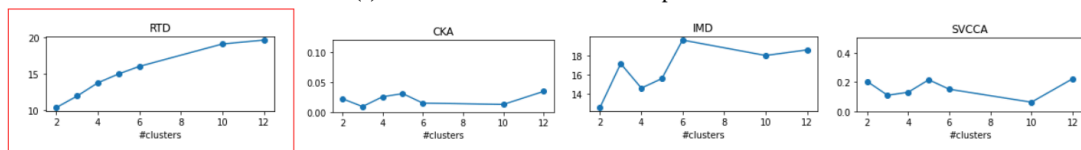
- In contrast to most existing approaches, RTD is sensitive to differences in topological structures (clusters, voids, cavities, tunnels, etc.) of the representations;
- RTD enjoys a very good correlation with the disagreement of models predictions.
- We apply RTD to compare representations in computer vision and NLP domains and various problems: training dynamics analysis, data distribution shift, transfer learning, ensemble learning, and disentanglement. Experiments show that RTD often outperforms CKA, IMD, and SVCCA.

Skoltech

Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathscr{P}}, M_{\tilde{\mathscr{P}}}$

- Let $\mathscr{P}, \tilde{\mathscr{P}}$ be two representations giving two embeddings of the same data $\mathscr{V}$. The two embeddings belong in general to different ambient spaces and we have the natural point-to-point correspondence between $\mathscr{P}$ and $\tilde{\mathscr{P}}$.

Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathscr{P}}, M_{\tilde{\mathscr{P}}}$

- Let $\mathscr{P}, \tilde{\mathscr{P}}$ be two representations giving two embeddings of the same data $\mathscr{V}$. The two embeddings belong in general to different ambient spaces and we have the natural point-to-point correspondence between $\mathscr{P}$ and $\tilde{\mathscr{P}}$.
- Given a sample of data V, the two representations $P = \mathscr{P}(V), \tilde{P} = \tilde{\mathscr{P}}(V)$ define two weighted graphs $\mathscr{G}^w, \mathscr{G}^{\tilde{w}}$ with the same vertex set V, and with edge weights $w_{AB} = \mathrm{dist}(P(A), P(B)), \tilde{w}_{AB} = \mathrm{dist}(\tilde{P}(A), \tilde{P}(B))$.
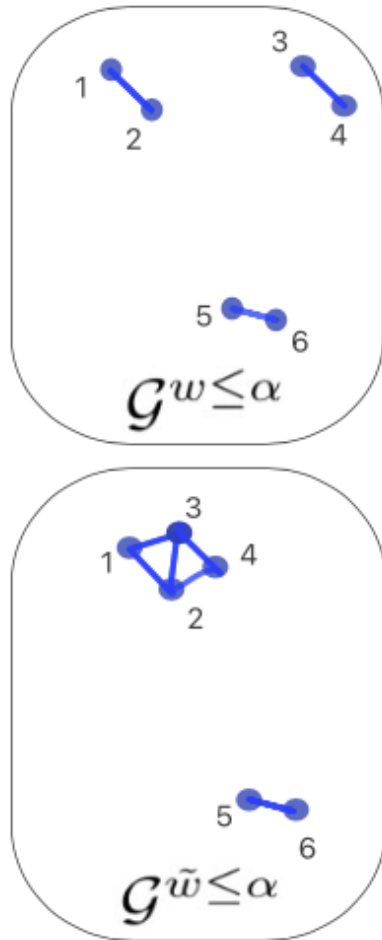
Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathscr{P}}, M_{\tilde{\mathscr{P}}}$

- Let $\mathscr{P}, \tilde{\mathscr{P}}$ be two representations giving two embeddings of the same data $\mathscr{V}$. The two embeddings belong in general to different ambient spaces and we have the natural point-to-point correspondence between $\mathscr{P}$ and $\tilde{\mathscr{P}}$.

- Given a sample of data V, the two representations $P = \mathscr{P}(V), \tilde{P} = \tilde{\mathscr{P}}(V)$ define two weighted graphs $\mathscr{G}^w$, $\mathscr{G}^{\tilde{w}}$ with the same vertex set V, $w_{AB} = \text{dist}(P(A), P(B))$, $\tilde{w}_{AB} = \text{dist}(\tilde{P}(A), \tilde{P}(B))$.

- Simplicial approximation $M_{\mathscr{P}}^{\alpha}$ to the manifold $M_{\mathscr{P}}$ at threshold $\alpha > 0$ consists of simplexes with vertices from $P(V)$ whose edges in $\mathscr{G}^w$ have weights not exceeding α.
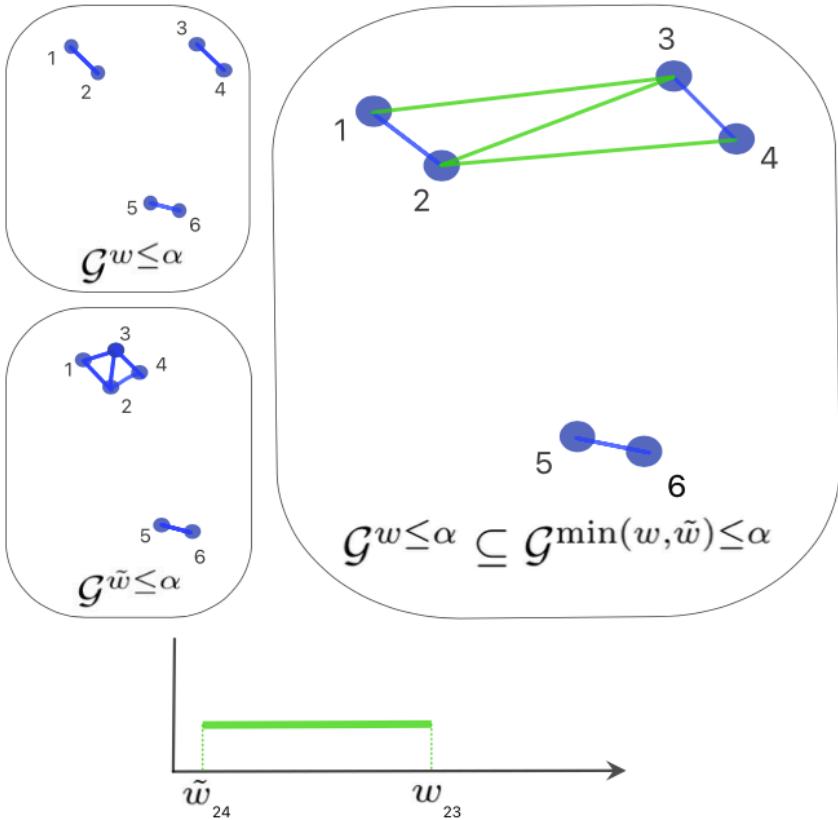
Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathcal{P}}, M_{\tilde{\mathcal{P}}}$

$$\mathcal{G}^{w \leq \alpha}$$

$$\mathcal{G}^{\tilde{w} \leq \alpha}$$

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

$\tilde{w}_{24}$ w_{23}

- Simplicial approximation $M_{\mathcal{P}}^{\alpha}$ to the manifold $M_{\mathcal{P}}$ at threshold $\alpha > 0$ consists of simplexes with vertices from $P(V)$ whose edges in $\mathcal{G}^w$ have weights not exceeding α.
- Simplices in $M_{\tilde{\mathcal{P}}}^{\alpha}$ which are absent in $M_{\mathcal{P}}^{\alpha}$ form topological features (paths, 2-membranes,…, k-membranes).

Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathscr{P}}, M_{\tilde{\mathscr{P}}}$

- Simplicial approximation $M_{\mathscr{P}}^{\alpha}$ to the manifold $M_{\mathscr{P}}$ at threshold $\alpha > 0$ consists of simplexes with vertices from $P(V)$ whose edges in $\mathscr{G}^w$ have weights not exceeding α.
- Simplices in $M_{\tilde{\mathscr{P}}}^{\alpha}$ which are absent in $M_{\mathscr{P}}^{\alpha}$ form topological features (paths, 2-membranes,…, k-membranes).
- We track the appearence and disappearence of such features with increase of the threshold, across all thresholds $\alpha > 0$.

Skoltec

9

Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathscr{P}}, M_{\tilde{\mathscr{P}}}$

- Simplicial approximation $M_{\mathscr{P}}^{\alpha}$ to the manifold $M_{\mathscr{P}}$ at threshold $\alpha > 0$ consists of simplexes with vertices from $P(V)$ whose edges in $\mathscr{G}^w$ have weights not exceeding α.
- Simplices in $M_{\tilde{\mathscr{P}}}^{\alpha}$ which are absent in $M_{\mathscr{P}}^{\alpha}$ form topological features (paths, 2-membranes,…, k-membranes).
- We track the appearence and disappearence of such features with increase of the threshold, across all thresholds $\alpha > 0$.
- The longer the lifespan of such topological feature across the change of threshold the bigger the described by this feature discrepancy between the two manifolds
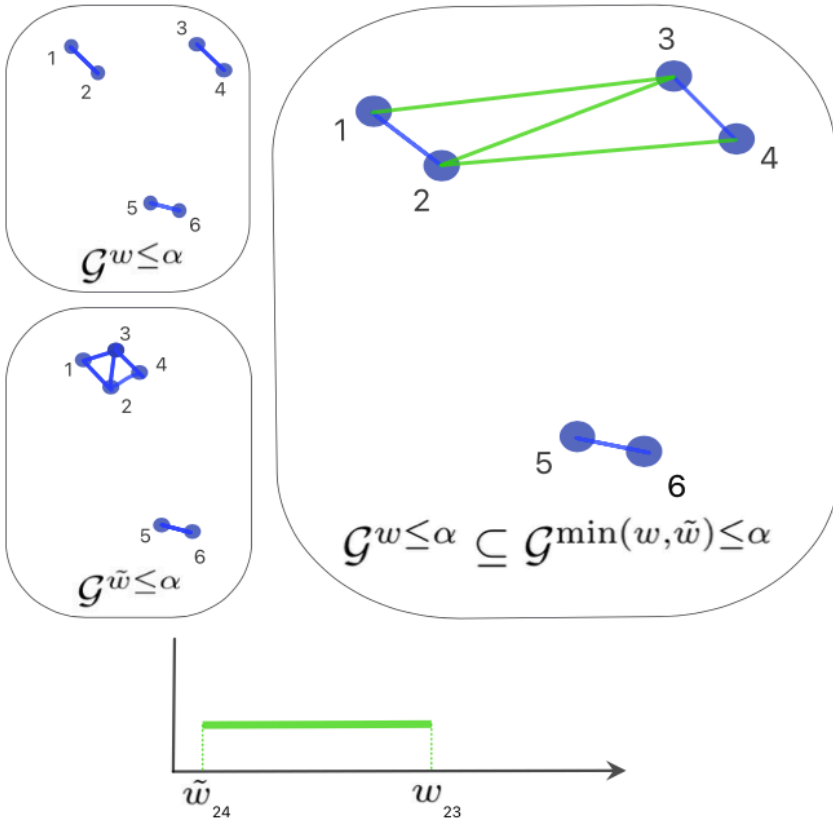
Skoltec

10

Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathcal{P}}, M_{\tilde{\mathcal{P}}}$

- Simplicial approximation $M_{\mathcal{P}}^{\alpha}$ to the manifold $M_{\mathcal{P}}$ at threshold $\alpha > 0$ consists of simplexes with vertices from $P(V)$ whose edges in $\mathcal{G}^w$ have weights not exceeding α.

- Simplices in $M_{\tilde{\mathcal{P}}}^{\alpha}$ which are absent in $M_{\mathcal{P}}^{\alpha}$ form topological features (paths, 2-membranes,…, k-membranes).

- We track the appearence and disappearence of such features with increase of the threshold, across all thresholds $\alpha > 0$.

- The longer the lifespan of such topological feature across the change of threshold the bigger the described by this feature discrepancy between the two manifolds.

- The *R-Cross-Barcode$_i$*$(P, \tilde{P})$ is the set of intervals recording the « appearences » and « disappearences» thresholds of such *i*-dimensional topological features. To calculate it we introduce the auxiliary graph $\hat{\mathcal{G}}^{w,\tilde{w}}$.

Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathcal{P}}, M_{\tilde{\mathcal{P}}}$
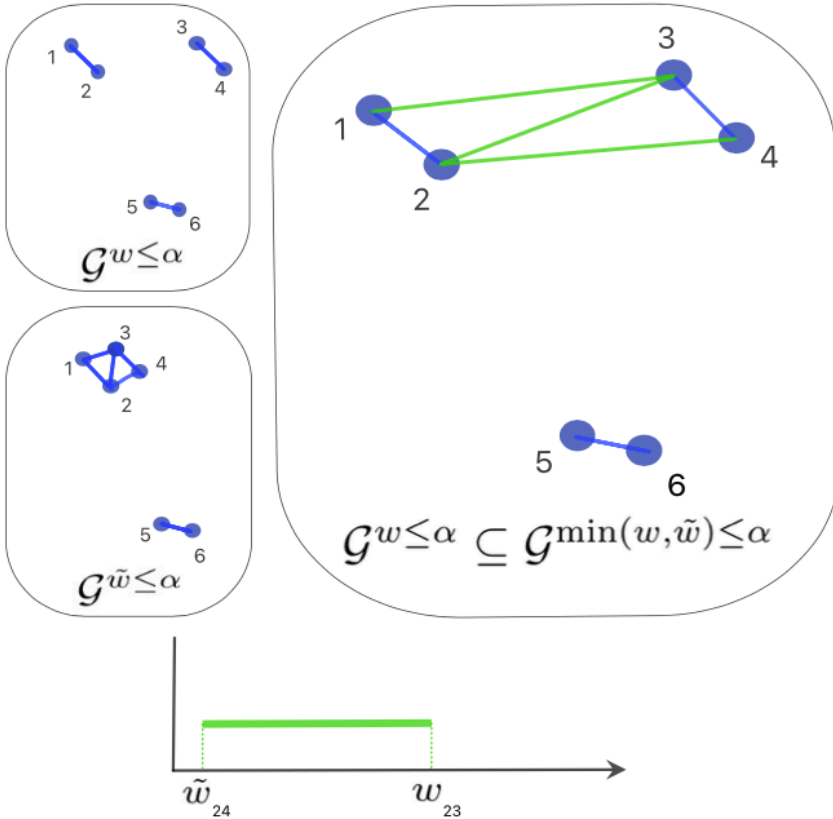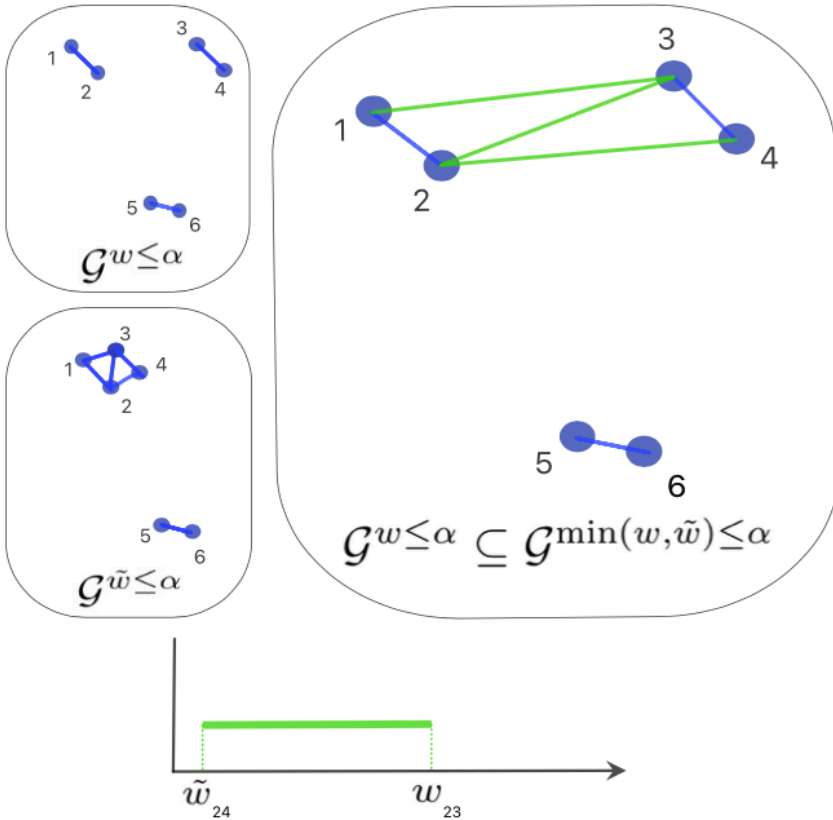

Concretely, to compare the multi-scale topology of the two weighted graphs $\mathcal{G}^w$ and $\mathcal{G}^{\tilde{w}}$ we introduce the weighted graph $\hat{\mathcal{G}}^{w,\tilde{w}}$ with doubled set of vertices and with the edge weights defined as follows. For convenience, fix a numbering of vertices $\mathrm{Vert}(\mathcal{G}) = \{A_1, \ldots, A_N\}$. For each vertex $A \in \mathrm{Vert}(\mathcal{G})$ we add the extra vertex A' together with A to $\hat{\mathcal{G}}$, plus the unique additional vertex O, and define the distance-like edge weights in $\hat{\mathcal{G}}^{w,\tilde{w}}$ as:

$$d_{A_i'A_j'} = \min(w_{A_iA_j}, \tilde{w}_{A_iA_j}), \; d_{A_iA_j'} = d_{A_iA_j} = w_{A_iA_j},$$

$$d_{A_iA_i'} = d_{OA_i} = 0, \; d_{A_jA_i'} = d_{OA_i'} = +\infty \qquad (1)$$

where $i < j$ and $O \in \mathrm{Vert}(\hat{\mathcal{G}}^{w,\tilde{w}})$ is the additional vertex.

Comparing Data Representations

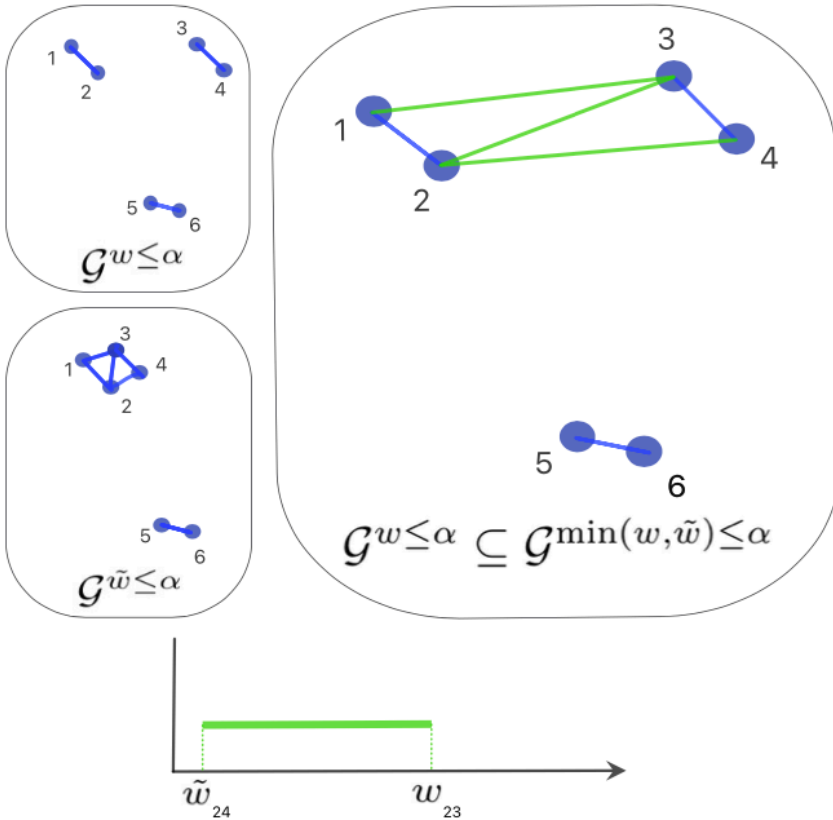IDEA: compare (simplicial approximations to) manifolds $M_{\mathcal{P}}, M_{\tilde{\mathcal{P}}}$

Recall that the Vietoris-Rips complex of a graph $\mathcal{G}$ equipped with edge weights' matrix m is the collection of $k-$simplexes, $k \geq 0$, which are $(k+1)-$element subsets of the set of vertices of $\mathcal{G}$, with the filtration threshold of a simplex defined by the maximal weight on the edges:

$$R_\alpha(\mathcal{G}^m) = \left\{ \{A_{i_0}, \ldots, A_{i_k}\}, A_i \in \text{Vert}(\mathcal{G}) | m_{A_i A_j} \leq \alpha \right\}$$

Our simplicial approximation to the manifold $M_{\mathcal{P}}$ at threshold α is the union of all simplexes from the simplicial complex $R_\alpha(\mathcal{G}^w)$, and similarly the approximation to $M_{\tilde{\mathcal{P}}}$ is the union of all simplexes from $R_\alpha(\mathcal{G}^{\tilde{w}})$.

Comparing Data Representations

IDEA: compare (simplicial approximations to) manifolds $M_{\mathcal{P}}, M_{\tilde{\mathcal{P}}}$

Recall that the Vietoris-Rips complex of a graph $\mathcal{G}$ equipped with edge weights' matrix m is the collection of $k-$simplexes, $k \geq 0$, which are $(k+1)-$element subsets of the set of vertices of $\mathcal{G}$, with the filtration threshold of a simplex defined by the maximal weight on the edges:

$$R_\alpha(\mathcal{G}^m) = \left\{ \{A_{i_0}, \ldots, A_{i_k}\}, A_i \in \text{Vert}(\mathcal{G}) | m_{A_i A_j} \leq \alpha \right\}$$

Our simplicial approximation to the manifold $M_{\mathcal{P}}$ at threshold α is the union of all simplexes from the simplicial complex $R_\alpha(\mathcal{G}^w)$, and similarly the approximation to $M_{\tilde{\mathcal{P}}}$ is the union of all simplexes from $R_\alpha(\mathcal{G}^{\tilde{w}})$.

Definition. The *R-Cross-Barcode$_i$($P, \tilde{P}$)* is the set of intervals recording the « appearances » and « disappearances» thresholds of *i*-dimensional topological features in the filtered simplicial complex $R_\alpha(\hat{\mathcal{G}}^{w,\tilde{w}})$.

Algorithm for $R\text{-}Cross\text{-}Barcode_i(P, \tilde{P})$

Algorithm 1 R-Cross-Barcode$_i(P, \tilde{P})$

Input: $w, \tilde{w}$: matrices of pairwise distances within point clouds $P, \tilde{P}$

Require: $\mathrm{vr}(m)$: function computing filtered complex from pairwise distances matrix m

Require: $\mathrm{B}(C, i)$: function computing persistence intervals of filtered complex C in dimension i

$w, \tilde{w} \leftarrow w, \tilde{w}$ divided by their 0.9 quantiles

$$m \leftarrow \begin{pmatrix} w & (w_+)^\mathsf{T} & 0 \\ w_+ & \min(w, \tilde{w}) & +\infty \\ 0 & +\infty & 0 \end{pmatrix}$$

$R\text{-}Cross\text{-}Barcode_i \leftarrow \mathrm{B}(\mathrm{vr}(m), i)$

Return: intervals list $R\text{-}Cross\text{-}Barcode_i(P, \tilde{P})$ representing "births" and "deaths" of topological discrepancies between P and $\tilde{P}$.

Definition. The $R\text{-}Cross\text{-}Barcode_i(P, \tilde{P})$ is the set of intervals recording the « appearances » and « disappearances» thresholds of i-dimensional topological features in the filtered simplicial complex $R_\alpha(\hat{\mathscr{G}}^{w, \tilde{w}})$.

Skoltech

Basic Properties of *R-Cross-Barcode*$_i(P, \tilde{P})$
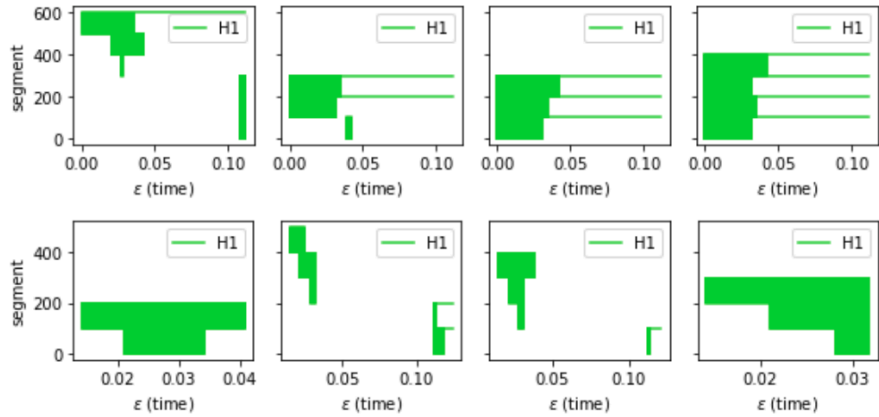
- if $P(A) = \tilde{P}(A)$ for any object $A \in V$, then R-Cross-Barcode$_*(P, \tilde{P}) = \varnothing$;
- if all distances within $\tilde{P}(V)$ are zero i.e. all objects are represented by the same point in $\tilde{P}$, then for all $k \geq 0$: R-Cross-Barcode$_{k+1}(P, \tilde{P}) = $ Barcode$_k(P)$ the standard barcode of the point cloud P;
- for any value of threshold α, the following sequence of natural linear maps of homology groups

$$\xrightarrow{r_{3i+3}} H_i(R_\alpha(\mathcal{G}^w)) \xrightarrow{r_{3i+2}} H_i(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})})) \xrightarrow{r_{3i+1}}$$

$$\xrightarrow{r_{3i+1}} H_i(R_\alpha(\hat{\mathcal{G}}^{w,\tilde{w}})) \xrightarrow{r_{3i}} H_{i-1}(R_\alpha(\mathcal{G}^w)) \xrightarrow{r_{3i-1}}$$

$$\xrightarrow{r_{3i-1}} \ldots \xrightarrow{r_1} H_0(R_\alpha(\mathcal{G}^{\min(w,\tilde{w})})) \xrightarrow{r_0} 0 \quad (2)$$

is exact, i.e. for any j the kernel of the map r_j is the image of the map r_{j+1}.

Representation Topology Divergence: the Algorithm

Algorithm 2 $RTD(\mathcal{P}, \tilde{\mathcal{P}})$, see section 2.4 for details, suggested default values: $b = 500, n = 10$

Input: $\mathcal{P} \in \mathbb{R}^{|\mathcal{V}| \times D}, \tilde{\mathcal{P}} \in \mathbb{R}^{|\mathcal{V}| \times \tilde{D}}$: data representations

   for $j = 1$ **to** n **do**

      $V_j \leftarrow$ random choice $(\mathcal{V}, b)$

      $P_j, \tilde{P}_j \leftarrow \mathcal{P}(V_j), \tilde{\mathcal{P}}(V_j)$

      $\mathcal{B}_j \leftarrow R\text{-}Cross\text{-}Barcode_1(P_j, \tilde{P}_j)$ intervals' list calculated by Algorithm 1

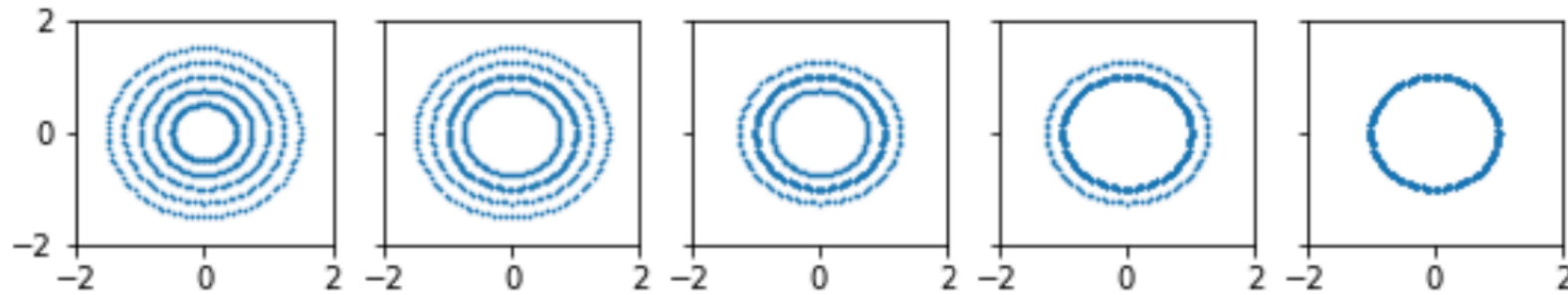 $rtd_j \leftarrow$ sum of lengths of all intervals in $\mathcal{B}_j$

   end for

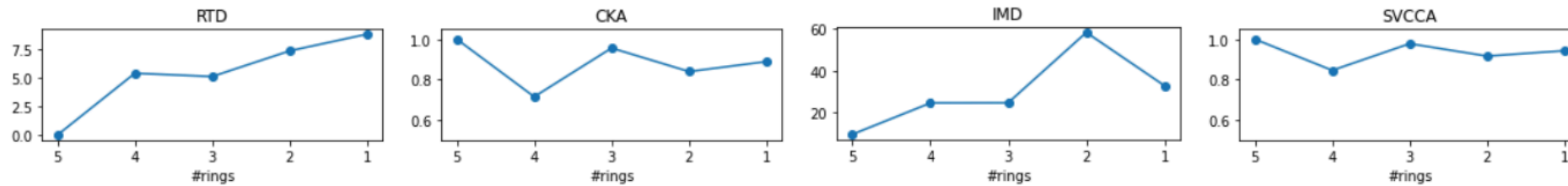 $RTD_1(\mathcal{P}, \tilde{\mathcal{P}}) \leftarrow \text{mean}(rtd)$

Return: number $\mathbf{RTD}_1(\mathcal{P}, \tilde{\mathcal{P}})$ representing discrepancy between the representations $\mathcal{P}, \tilde{\mathcal{P}}$
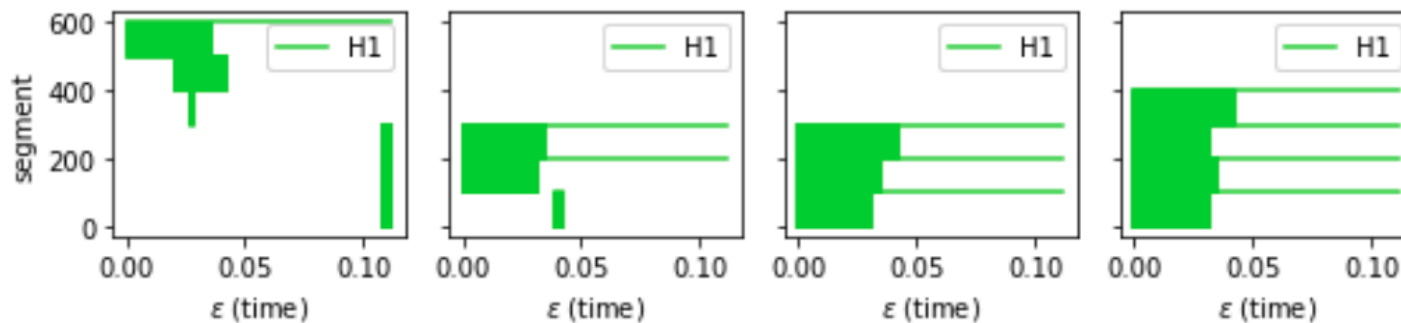
Experiments. 'Five rings vs 5,4,3,2,1 rings

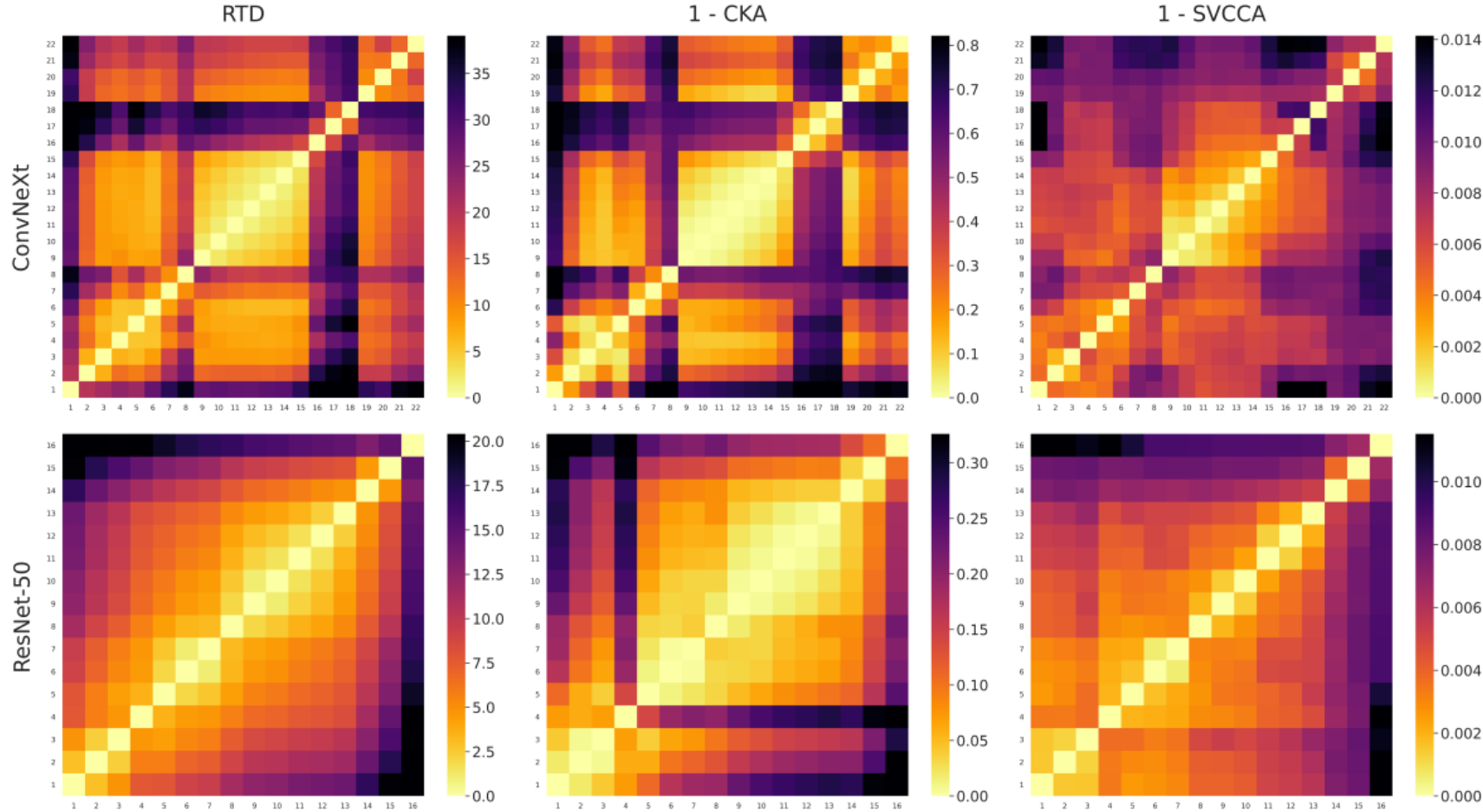(a) Point clouds used in "rings" experiment.

(b) Representations' comparison measures. Ideally, the measure should change monotonically with the increase of topological discrepancy.

Figure 12: RTD perfectly detects changes in topology, while rival measures fail. Five rings are compared with 5,4,3,2,1 rings.

18

Experiments. Internal similarity of Neural Network layers

The differences of representations between the layers within networks trained on the ImageNet-1k dataset.
The columns correspond to the metrics, and the rows – to the architectures (ConvNeXt, ResNet-50). We observe that RTD catches architecture's block structure better than CKA, SVCCA. The ResNet-50 architecture has sequence of blocks in form [3, 4, 6, 3] and it can be seen that RTD highlights it with sub-squares of corresponding sizes.

Conclusions

1. In this paper, we have proposed a topologically-inspired approach to compare neural network representations.The most widely used methods for this problem are statistical: Canonical Correlation Analysis (CCA) and Centered Kernel Alignment (CKA). But the problem itself is a geometric one: the comparison of two neural representations of the same objects is de-facto the comparison of two points clouds from different spaces having point-to-point correspondence between clouds. The natural way is to compare their geometrical and topological features with due account of their localization that is exactly what was done by the R-Cross-Barcode and RTD.

2. We demonstrated that RTD agrees with the natural assessment of representations similarity.

3. We used the RTD to gain insights into neural network representations in computer vision and NLP domains for various problems: training dynamics analysis, data distribution shift, transfer learning, ensemble learning, and disentanglement assessment.

4. RTD correlates strikingly well with t*he disagreement of models' predictions*;

5. Finally, R-Cross-Barcode and RTD are general tools that are not limited only to the comparison of representations. They could be applied to other problems involving comparison of two point clouds with point-to-point correspondence, for example, in 3D computer vision.

Skoltech

Thank you for your attention!

Skoltech