
Characterizing and Measuring the Similarity of Neural Networks with Persistent Homology

Supplementary Material

Anonymous Author(s)

Affiliation

Address

email

1 Appendix I

2 This Appendix contains mathematical definitions.

3 **Definition 1 (simplex)** *A k -simplex is a k -dimensional polytope which is the convex hull of its $k + 1$*
4 *vertices. i.e. the set of all convex combinations $\lambda_0 v_0 + \lambda_1 v_1 + \dots + \lambda_k v_k$ where $\lambda_0 + \lambda_1 + \dots + \lambda_k = 1$*
5 *and $0 \leq \lambda_j \leq 1 \ \forall j \in \{0, 1, \dots, k\}$.*

6 Some examples of simplices are:

- 7 • 0-simplex is a point.
- 8 • 1-simplex is a line segment.
- 9 • 2-simplex is a triangle.
- 10 • 3-simplex is a tetrahedron.

11 **Definition 2 (simplicial complex)** *A simplicial complex \mathcal{K} is a set of simplices that satisfies the*
12 *following conditions:*

- 13 1. *Every subset (or face) of a simplex in \mathcal{K} also belongs to \mathcal{K} .*
- 14 2. *For any two simplices σ_1 and σ_2 in \mathcal{K} , if $\sigma_1 \cap \sigma_2 \neq \emptyset$, then $\sigma_1 \cap \sigma_2$ is a common subset, or*
15 *face, of both σ_1 and σ_2 .*

16 **Definition 3 (directed flag complex)** *Let $G = (V, E)$ be a directed graph. The directed flag complex*
17 *$FC(G)$ is defined to be the ordered simplicial complex whose k -simplices are all ordered $(k + 1)$ -*
18 *cliques, i.e., $(k + 1)$ -tuples $\sigma = (v_0, v_1, \dots, v_k)$, such that $v_i \in V \ \forall i$, and $(v_i, v_j) \in E$ for $i < j$.*

19 We define the boundary, ∂ , as a function that maps i -simplex to the sum of its $(i-1)$ -dimensional faces.
20 Formally speaking, for an i -simplex $\sigma = [v_0, \dots, v_i]$, its *boundary* (∂) is:

$$\partial_i \sigma = \sum_{j=0}^i [v_0, \dots, \hat{v}_j, \dots, v_i] \quad (1)$$

21 where the hat indicates the v_j is omitted.

22 We can expand this definition to i -chains. For an i -chain $c = c_i \sigma_i$, $\partial_i(c) = \sum_i c_i \partial_i \sigma_i$.

23 We can now distinguish two special types of chains using the boundary map that will be useful to
24 define homology:

- 25 • The first one is an i -cycle, which is defined as an i -chain with empty boundary. In other
26 words, an i -chain c is an i -cycle if and only if $\partial_i(c) = 0$, i.e. $c \in \text{Ker}(\partial_i)$.
- 27 • An i -chain c is i -boundary if there exists an $(i + 1)$ -chain d such that $c = \partial_{i+1}(d)$, i.e.
28 $c \in \text{Im}(\partial_{i+1})$.

29 **Definition 4 (graph)** A graph G is a pair (V, E) , where V is a finite set referred to as the vertices or
30 nodes of G , and E is a subset of the set of unordered pairs $e = \{u, v\}$ of distinct points in V , which
31 we call the edges of G . Geometrically the pair $\{u, v\}$ indicates that the vertices u and v are adjacent
32 in G . A directed graph, or a digraph, is similarly a pair (V, E) of vertices V and edges E , except the
33 edges are ordered pairs of distinct vertices, i.e., the pair (u, v) indicates that there is an edge from u to
34 v in G . In a digraph, we allow reciprocal edges, i.e., both (u, v) and (v, u) may be edges in G , but we
35 exclude loops, i.e., edges of the form (v, v) .

36 **Definition 5 (homology group)** Given these two special subspaces, i -cycles $Z_i(K)$ and i -boundaries
37 $B_i(K)$ of $C_i(K)$, we now take the quotient space of $B_i(K)$ as a subset of $Z_i(K)$. In this quotient space,
38 there are only the i -cycles that do not bound an $(i + 1)$ -complex, or i -voids of K . This quotient space
39 is called i -th homology group of the simplicial complex K :

$$H_i(K) = \frac{Z_i(K)}{B_i(K)} = \frac{\text{Ker}(\partial_i)}{\text{Im}(\partial_{i+1})} \quad (2)$$

40 where Ker and Im are the function kernel and image respectively.

41 The dimension of i -th homology is called the i -th Betti number of K , $\beta_i(K)$, where:

$$\beta_i(K) = \dim(\text{Ker}(\partial_i)) - \dim(\text{Im}(\partial_{i+1})) \quad (3)$$

42 **Definition 6 (Wasserstein distance)** The p -Wasserstein distance between two PDs D_1 and D_2 is the
43 infimum over all bijections: $\gamma: D_1 \rightarrow D_2$ of:

$$d_W(D_1, D_2) = \left(\sum_{x \in D_1} \|x - \gamma(x)\|_\infty^p \right)^{1/p} \quad (4)$$

44 where $\| - \|_\infty$ is defined for $(x, y) \in \mathbb{R}^2$ by $\max\{|x|, |y|\}$. The limit $p \rightarrow \infty$ defines the Bottleneck
45 distance. More explicitly, it is the infimum over the same set of bijections of the value

$$d_B(D_1, D_2) = \sup_{x \in D_1} \|x - \gamma(x)\|_\infty. \quad (5)$$

46 **Definition 7 (Persistence landscape)** Given a collection of intervals $\{(b_i, d_i)\}_{i \in I}$ that compose a
47 PD, its persistence landscape is the set of functions $\lambda_k: \mathbb{R} \rightarrow \mathbb{R}$ defined by letting $\lambda_k(t)$ be the k -th
48 largest value of the set $\{\Lambda_i(t)\}_{i \in I}$ where:

$$\Lambda_i(t) = [\min\{t - b_i, d_i - t\}]_+ \quad (6)$$

49 and $c_+ := \max(c, 0)$. The function λ_k is referred to as the k -layer of the persistence landscape.

50 Now we define a vectorization of the set of real-valued function that compose PDs on $\mathbb{N} \times \mathbb{R}$. For
51 any $p = 1, \dots, \infty$ we can restrict attention to PDs D whose associated persistence landscape λ is
52 p -integrable, that is to say,

$$\|\lambda\|_p = \left(\sum_{i \in \mathbb{N}} \|\lambda_i\|_p^p \right)^{1/p} \quad (7)$$

53 is finite. In this case, we refer to Equation (7) as the p -landscape norm of D . For $p = 2$, we define the
54 value of the landscape kernel or similarity of two vectorized PDs D and E as

$$\langle \lambda, \mu \rangle = \left(\sum_{i \in \mathbb{N}} \int_{\mathbb{R}} |\lambda_i(x) - \mu_i(x)|^2 dx \right)^{1/2} \quad (8)$$

55 where λ and μ are their associated persistence landscapes.

56 λ_k is geometrically described as follows. For each $i \in I$, we draw an isosceles triangle with base the
 57 interval (b_i, d_i) on the horizontal t -axis, and sides with slope 1 and -1 . This subdivides the plane
 58 into a number of polygonal regions that we label by the number of triangles contained on it. If P_k is
 59 the union of the polygonal regions with values at least k , then the graph of λ_k is the upper contour of
 60 P_k , with $\lambda_k(a) = 0$ if the vertical line $t = a$ does not intersect P_k .

61 **Definition 8 (Weighted silhouette)** Let $D = \{(b_i, d_i)\}_{i \in I}$ be a PD and $w = \{w_i\}_{i \in I}$ a set of positive
 62 real numbers. The silhouette of D weighted by w is the function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ defined by:

$$\phi(t) = \frac{\sum_{i \in I} w_i \Lambda_i(t)}{\sum_{i \in I} w_i}, \quad (9)$$

63 where

$$\Lambda_i(t) = [\min\{t - b_i, d_i - t\}]_+ \quad (10)$$

64 and $c_+ := \max(c, 0)$. When $w_i = |d_i - b_i|^p$ for $0 < p \leq \infty$ we refer to ϕ as the p -power-weighted
 65 silhouette of D . It defines a vectorization of the set of PDs on the vector space of continuous
 66 real-valued functions on \mathbb{R} .

67 **Definition 9 (Heat vectorizations)** Considering PD as the support of Dirac deltas, one can con-
 68 struct, for any $t > 0$, two vectorizations of the set of PDs to the set of continuous real-valued function
 69 on the first quadrant $\mathbb{R}_{>0}^2$. The heat vectorization is constructed for every PD D by solving the heat
 70 equation:

$$\begin{aligned} \Delta_x(u) &= \partial_t u && \text{on } \Omega \times \mathbb{R}_{>0} \\ u &= 0 && \text{on } \{x_1 = x_2\} \times \mathbb{R}_{\geq 0} \\ u &= \sum_{p \in D} \delta_p && \text{on } \Omega \times 0 \end{aligned} \quad (11)$$

71 where $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \leq x_2\}$, then solving the same equation after precomposing the data of
 72 Equation (11) with the change of coordinates $(x_1, x_2) \mapsto (x_2, x_1)$, and defining the image of D to be
 73 the difference between these two solutions at the chosen time t .

74 We recall that the solution to the heat equation with initial condition given by a Dirac delta supported
 75 at $p \in \mathbb{R}^2$ is:

$$\frac{1}{4\pi t} \exp\left(-\frac{\|p - x\|^2}{4t}\right) \quad (12)$$

76 To highlight the connection with normally distributed random variables, it is customary to use the
 77 change of variable $\sigma = \sqrt{2t}$.

78 For a complete reference on vectorized persistence summaries and PH approximated metrics, see
 79 Tauzin et al. [2], Berry et al. [1] and Giotto-TDA package documentation appendix¹.

80 Figure 1 shows a neural network filtration example. Note that most of the edges have been omitted
 81 for clarity and, for the same reason, ε evolution has been discretized.

¹<https://giotto-ai.github.io/gtda-docs/0.3.1/theory/glossary.html#persistence-landscape>

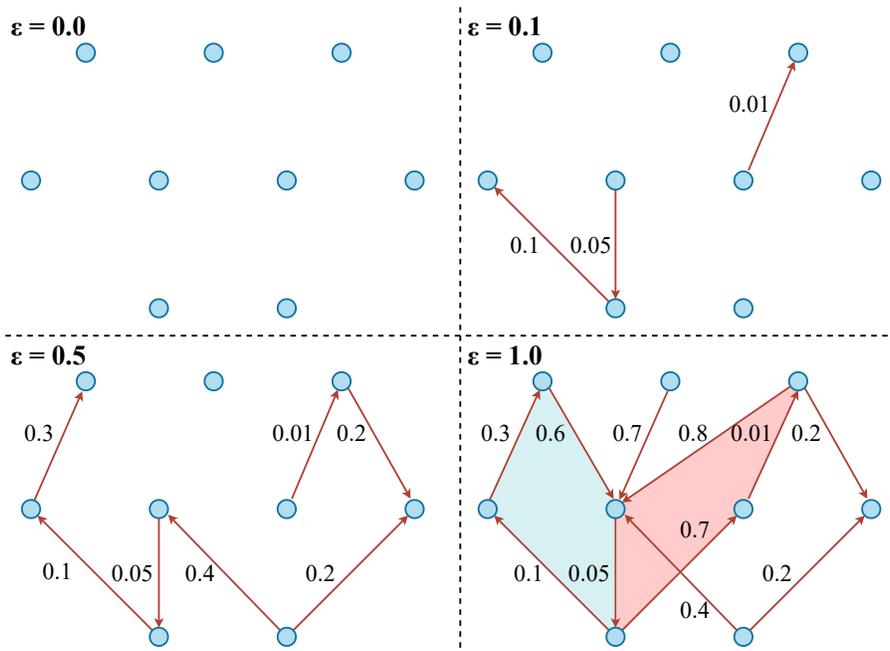


Figure 1: MLP Simplicial complex filtration example.

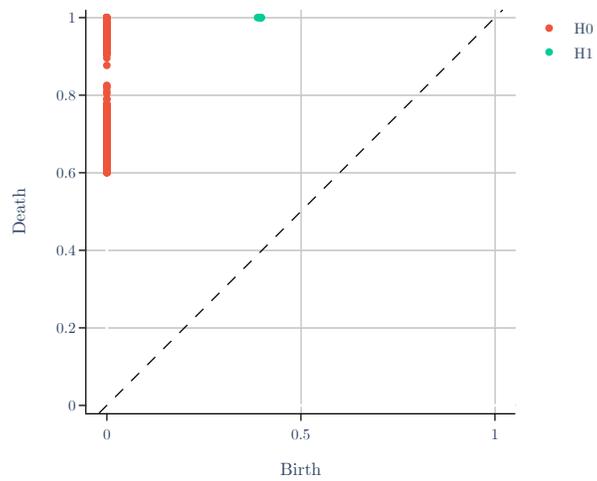
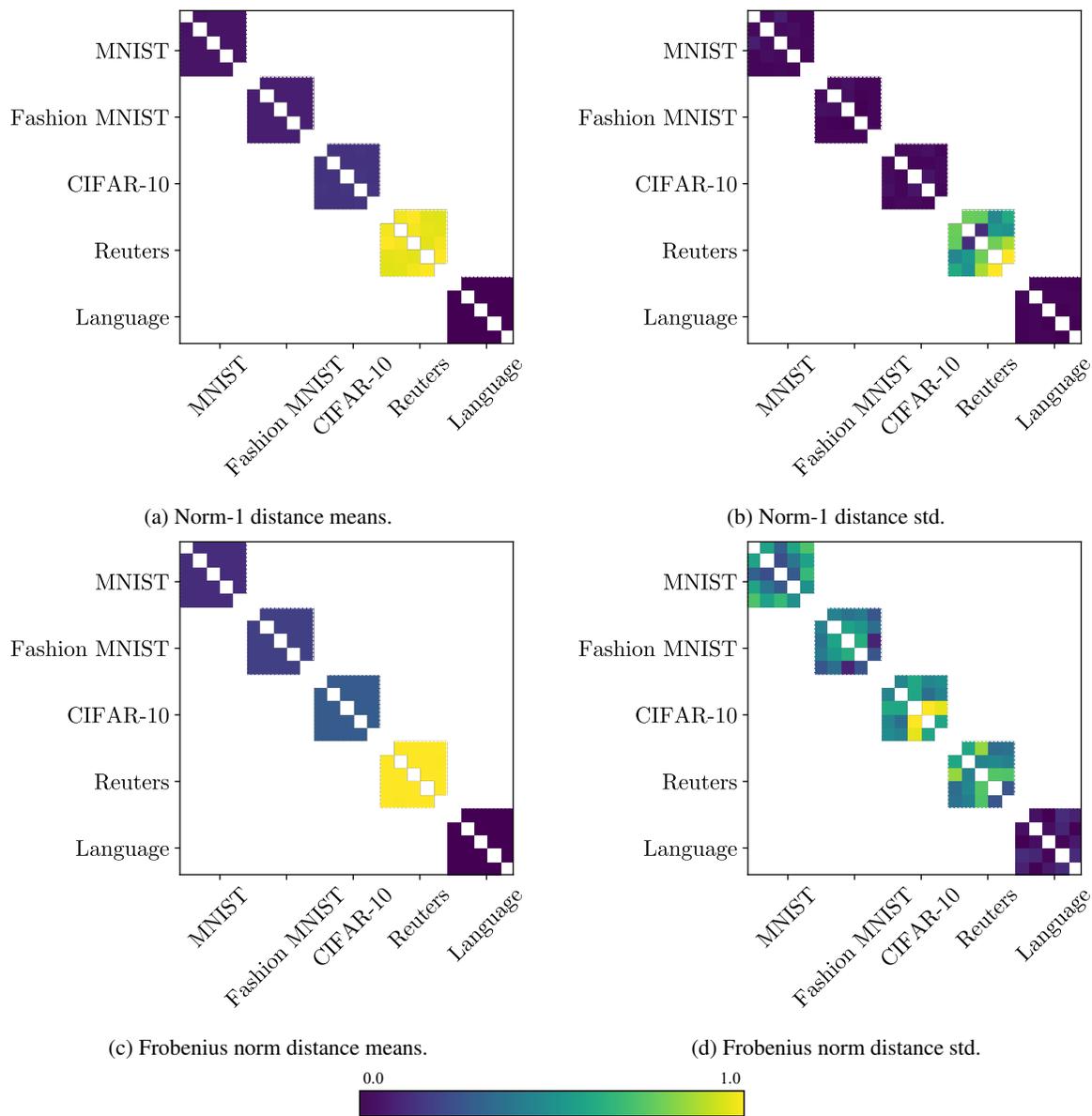


Figure 2: A Persistence Homology diagram.

Figure 3: Norm distances among control experiments. 5 runs \times 5 randomizations.

Norm	Minimum	Maximum	Mean	Standard deviation
1-Norm	0.6683	4.9159	1.9733	1.5693
Frobenius	0.0670	0.9886	0.4514	0.3074

Table 1: Normalized difference comparison of self-norm against the maximum mean distance of the experiment.

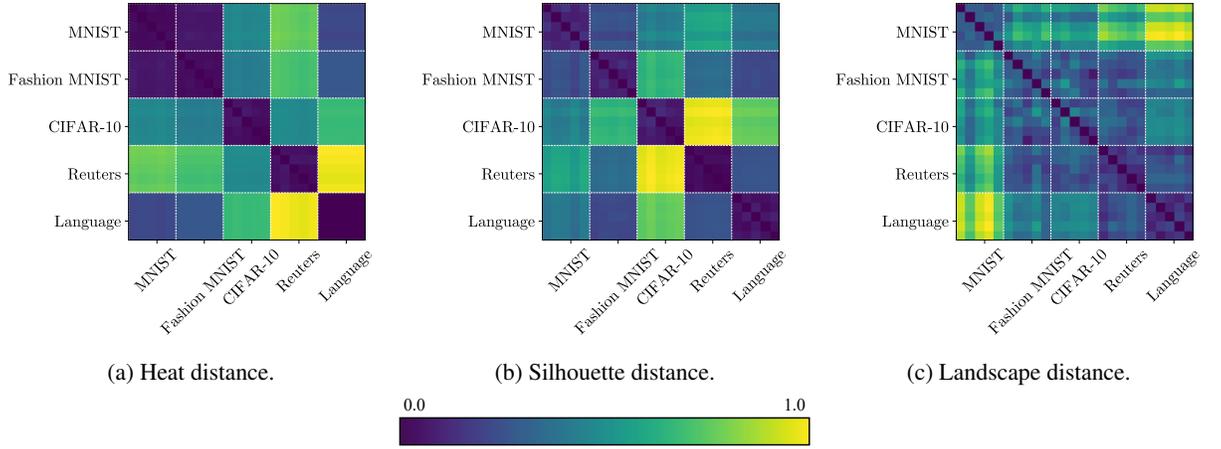


Figure 4: Topological distance means of control experiments.

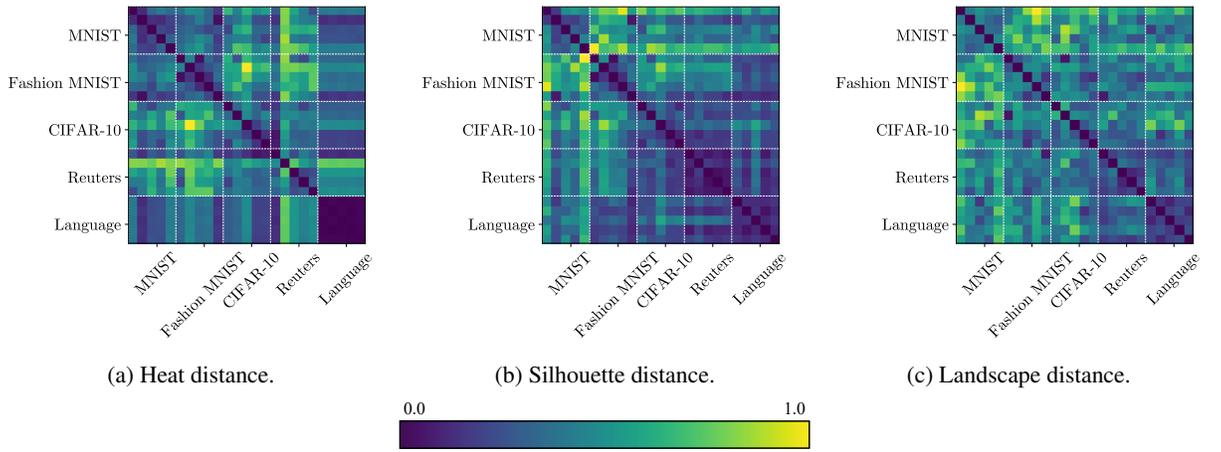


Figure 5: Topological distance standard deviations of control experiments.

83 **Appendix III**

84 This Appendix contains statistics about experiment groups by dataset.

Experiment group	Experiment	Value	Index
1	Layer size	128	1
		256	2
		512	3
		1024	4
2	Number of layers	2	5
		4	6
		6	7
		8	8
		10	9
3	Input order	NA	10-14
4	Number of labels	2	15
		4 (M, FM, C), 6 (R), 3 (L)	16
		6 (M, FM, C), 12 (R), 4 (L)	17
		8 (M, FM, C), 23 (R), 6 (L)	18
		10 (M, FM, C), 46 (R), 7 (L)	19

Table 2: Indices of the experiments of the distance matrices. M is for MNIST, FM for Fashion MNIST, C for CIFAR-10, R for Reuters and L for Language Identification.

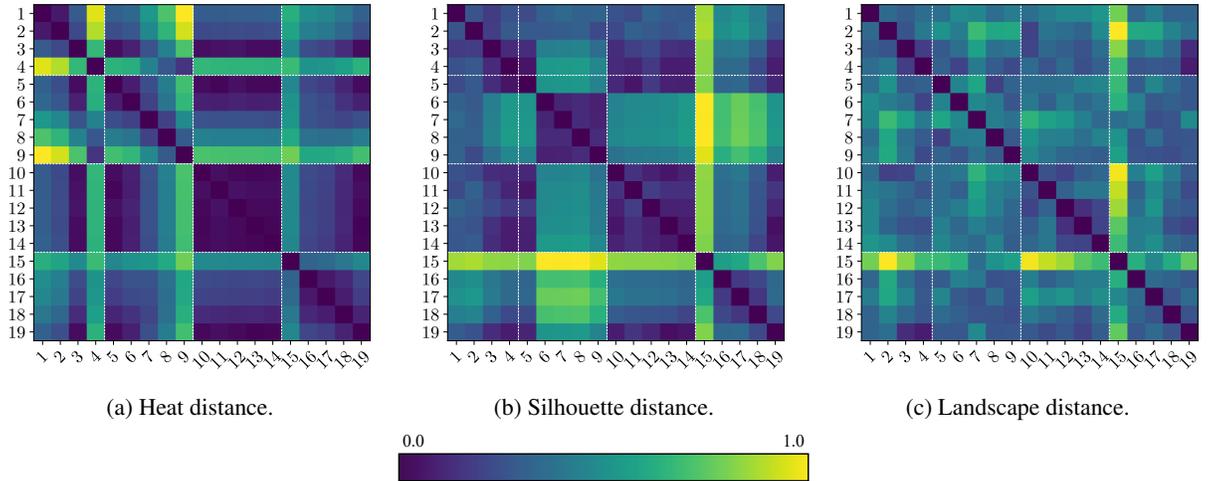


Figure 6: MNIST neural networks' Persistent Homology distance matrices means.

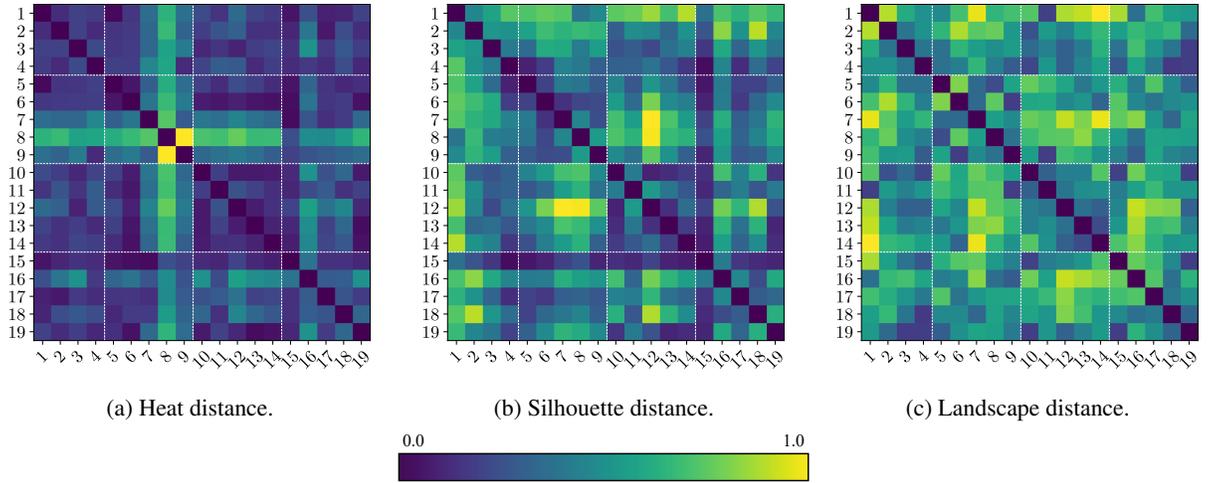


Figure 7: MNIST neural networks' Persistent Homology distance matrices standard deviations.

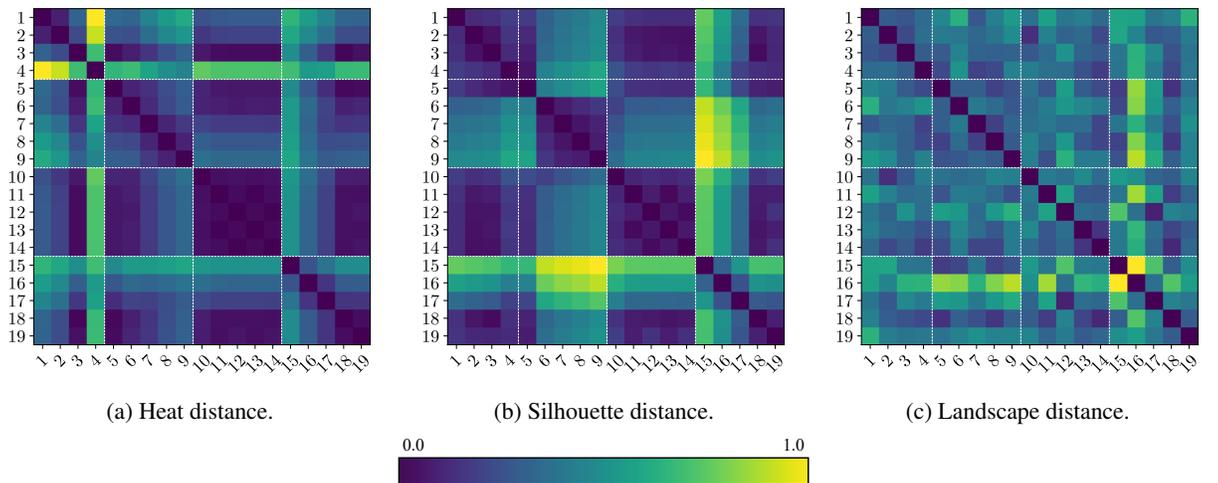


Figure 8: Fashion MNIST neural networks' Persistent Homology distance matrices means.

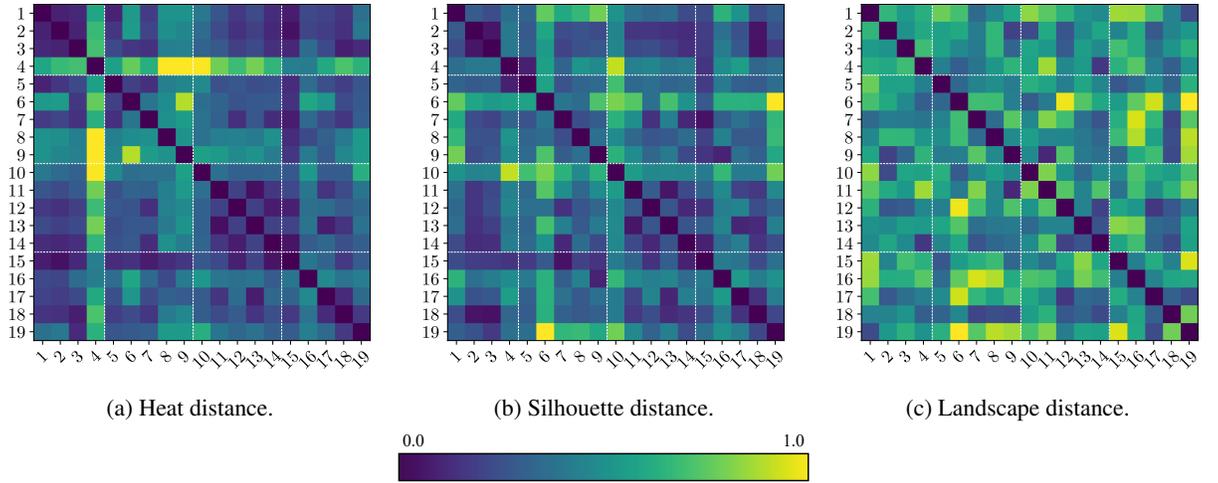


Figure 9: Fashion MNIST neural networks' Persistent Homology distance matrices standard deviations.

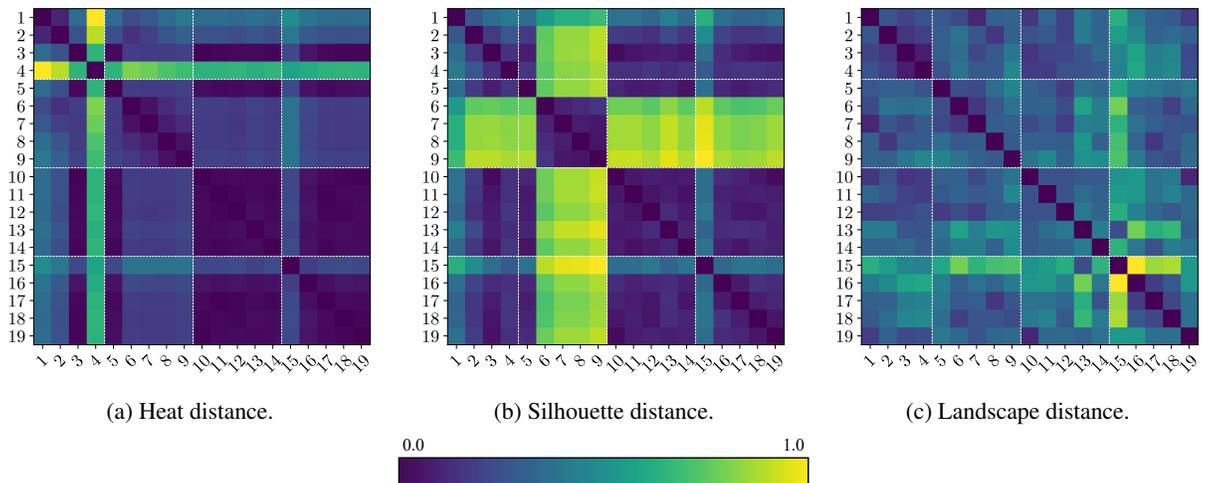


Figure 10: CIFAR-10 neural networks' Persistent Homology distance matrices means.

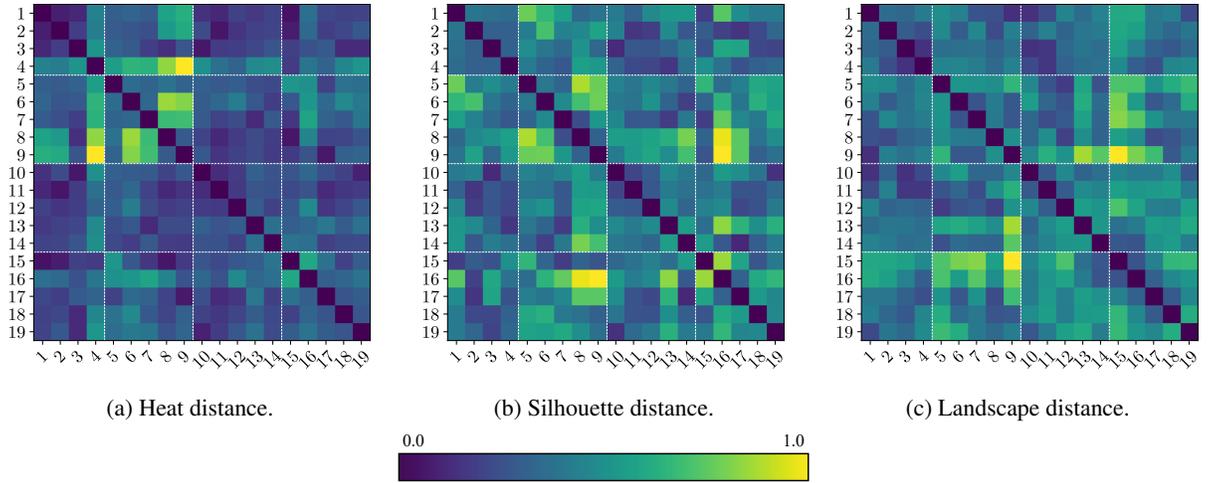


Figure 11: CIFAR-10 neural networks' Persistent Homology distance matrices standard deviations.

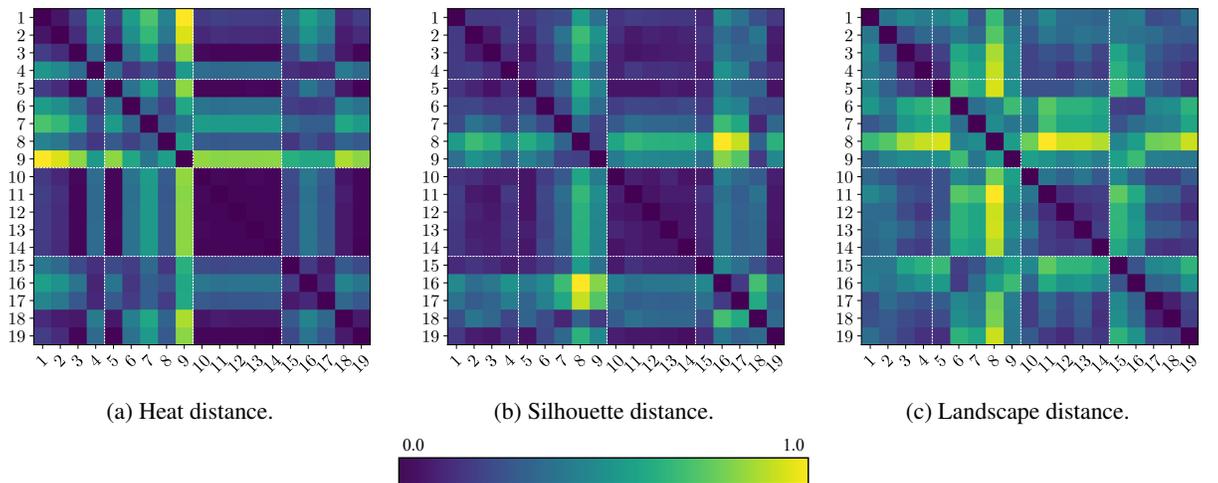


Figure 12: Language Identification neural networks' Persistent Homology distance matrices means.

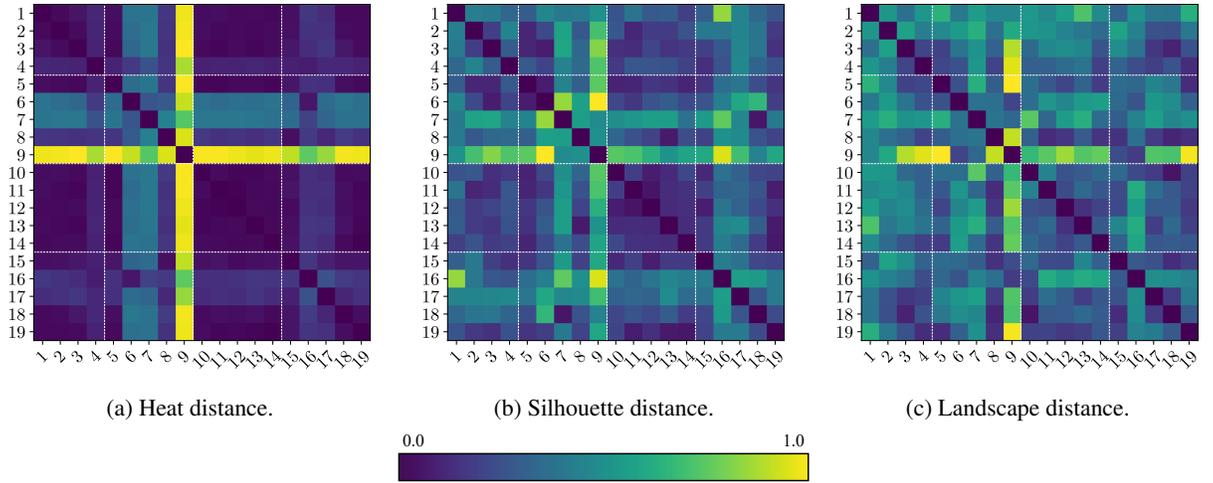


Figure 13: Language Identification neural networks' Persistent Homology distance matrices standard deviations.

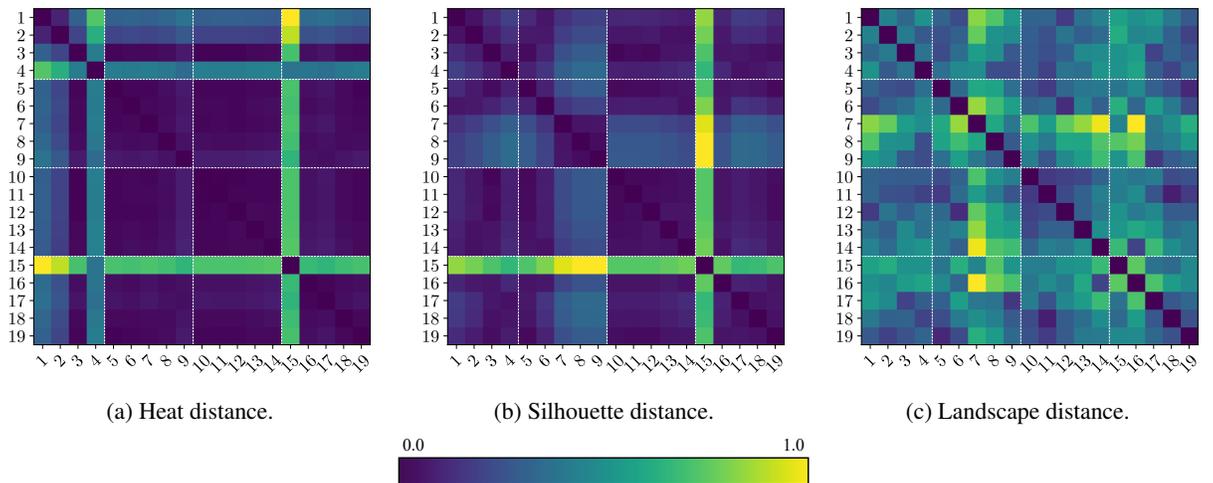


Figure 14: Reuters neural networks' Persistent Homology distance matrices means.

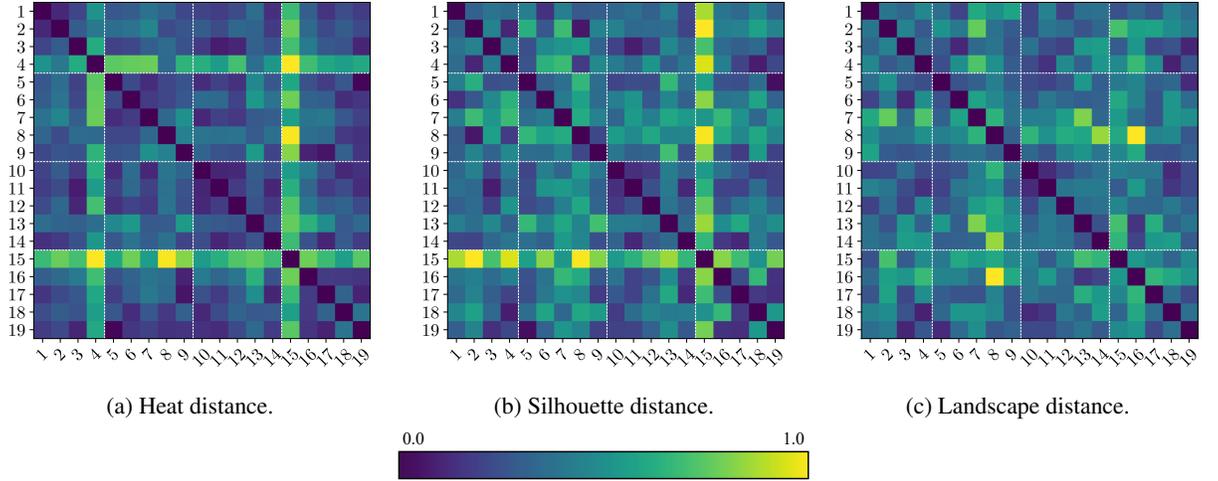


Figure 15: Reuters neural networks' Persistent Homology distance matrices standard deviations.

Discretization	Experiment	Mean	Standard deviation
Heat	Layer size	0.5128	0.3391
Heat	Number layers	0.3633	0.1933
Heat	Input order	0.0291	0.0100
Heat	Number labels	0.2279	0.1352
Landscape	Layer size	0.3077	0.0727
Landscape	Number layers	0.3776	0.1096
Landscape	Input order	0.2719	0.0822
Landscape	Number labels	0.4266	0.1705
Silhouette	Layer size	0.2110	0.0552
Silhouette	Number layers	0.2681	0.1953
Silhouette	Input order	0.1115	0.0364
Silhouette	Number labels	0.4126	0.2270

Table 3: MNIST statistics of experiment groups.

Discretization	Experiment	Mean	Standard deviation
Heat	Layer size	0.5380	0.3487
Heat	Number layers	0.1774	0.0739
Heat	Input order	0.0308	0.0132
Heat	Number labels	0.2679	0.1427
Landscape	Layer size	0.2976	0.0482
Landscape	Number layers	0.2886	0.0631
Landscape	Input order	0.3583	0.1114
Landscape	Number labels	0.5301	0.2149
Silhouette	Layer size	0.1265	0.0409
Silhouette	Number layers	0.2592	0.2024
Silhouette	Input order	0.0824	0.0353
Silhouette	Number labels	0.4150	0.1940

Table 4: Fashion MNIST statistics of experiment groups.

Discretization	Experiment	Mean	Standard deviation
Heat	Layer size	0.5414	0.3319
Heat	Number layers	0.1314	0.0529
Heat	Input order	0.0243	0.0068
Heat	Number labels	0.1084	0.0870
Landscape	Layer size	0.1907	0.0630
Landscape	Number layers	0.2967	0.0789
Landscape	Input order	0.3119	0.0584
Landscape	Number labels	0.5193	0.2803
Silhouette	Layer size	0.2705	0.0964
Silhouette	Number layers	0.3885	0.3626
Silhouette	Input order	0.0769	0.0204
Silhouette	Number labels	0.2049	0.1341

Table 5: CIFAR-10 statistics of experiment groups.

Discretization	Experiment	Mean	Standard deviation
Heat	Layer size	0.2856	0.1771
Heat	Number layers	0.4396	0.1811
Heat	Input order	0.0159	0.0040
Heat	Number labels	0.2374	0.1243
Landscape	Layer size	0.3246	0.1254
Landscape	Number layers	0.5643	0.1643
Landscape	Input order	0.2320	0.0982
Landscape	Number labels	0.3519	0.1671
Silhouette	Layer size	0.1501	0.0384
Silhouette	Number layers	0.3374	0.1420
Silhouette	Input order	0.0699	0.0159
Silhouette	Number labels	0.3737	0.1719

Table 6: Language Identification statistics of experiment groups.

Discretization	Experiment	Mean	Standard deviation
Heat	Layer size	0.4004	0.2220
Heat	Number layers	0.0412	0.0211
Heat	Input order	0.0166	0.0051
Heat	Number labels	0.2927	0.3185
Landscape	Layer size	0.3950	0.0733
Landscape	Number layers	0.5265	0.1678
Landscape	Input order	0.3060	0.1028
Landscape	Number labels	0.4534	0.1651
Silhouette	Layer size	0.1104	0.0470
Silhouette	Number layers	0.1596	0.0922
Silhouette	Input order	0.0387	0.0112
Silhouette	Number labels	0.3206	0.3158

Table 7: Reuters statistics of experiment groups.

85 **References**

- 86 [1] E. Berry, Y.-C. Chen, J. Cisewski-Kehe, and B. T. Fasy. Functional summaries of persistence
87 diagrams. *Journal of Applied and Computational Topology*, 4:211–262, 2020.
- 88 [2] G. Tauzin, U. Lupo, L. Tunstall, J. B. Pérez, M. Caorsi, A. Medina-Mardones, A. Dassatti, and
89 K. Hess. giotto-tda: A topological data analysis toolkit for machine learning and data exploration,
90 2020.