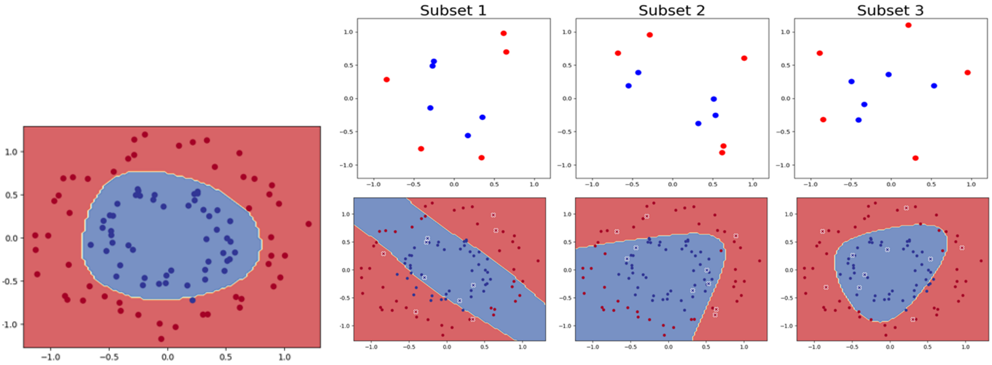
Comparison of finite metric spaces via persistence matching diagrams

Álvaro Torras-Casas
Inserm, INRAE, Centre for Research in Epidemiology and Statistics (CRESS), Université Paris Cité and Université Sorbonne Paris Nord

Thursday 6th March 2025
CUNEF Universidad, Madrid.
joint work with CIMAgroup (Prof. Rocío González-Díaz), Universidad de Sevilla

(Topological) Data quality

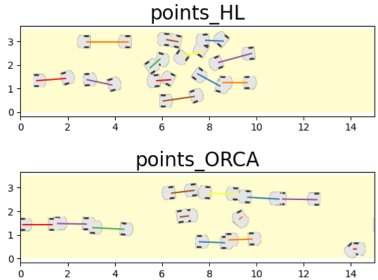
- **Aim:** replace datasets by smaller subsets that capture the "same information"
- **Question:** can topology help to answer this question?
- **Example:** two classes and a small feed forward neural network.



- **Difficult question:** we focused on dimension 0.

Fleet behavior: Navground

- Characterise autonomous wheelchair simulations via Navground (J. Guzzi, SUPSI)
- **Ultimate aim:** detect order vs chaos and predict likelihood of collisions or deadlocks.
- **Motivation:** spontaneous formation of groups in macroscopic behavior.

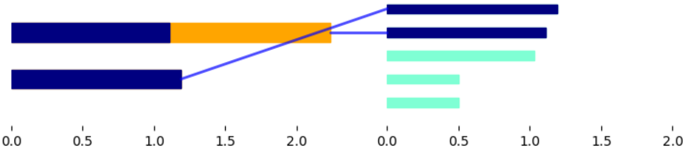


HORIZON-CL4-HUMAN-01 grant agreement no.101070028

Funded by the European Union

Idea: relate zero-dimensional persistent homology barcodes

- Consider a pair of metric spaces $(\mathbb{Y}, d^{\mathbb{Y}})$ and $(\mathbb{X}, d^{\mathbb{X}})$
- Connected components can be understood by $\mathrm{PH}_0(\mathbb{Y})$ and $\mathrm{PH}_0(\mathbb{X})$ as well as their respective $\mathbf{B}(\mathbb{Y})$ and $\mathbf{B}(\mathbb{X})$ barcodes.
- Question: can we relate $\mathbf{B}(\mathbb{Y})$ and $\mathbf{B}(\mathbb{X})$?



- Is this well defined? and stable? so what?

Outline

1. Motivation: Data quality and fleet behavior
2. **Merge trees and zero-dimensional persistent homology**
3. The induced matching diagram (for subsets of metric spaces)
4. The induced block function and the matching diagram
5. Stability of $D(f)$
6. Example I: exploring topological data quality
7. Example II: Navground Analysis

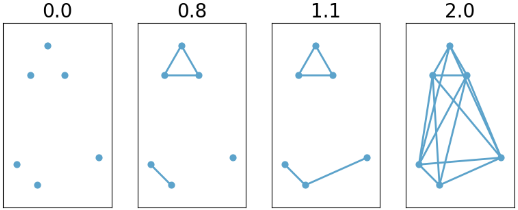
Vietoris-Rips filtration (1-skeleton)

- Consider a finite metric space $(\mathbb{X}, d^{\mathbb{X}})$.
- Let $\mathrm{VR}^1(\mathbb{X})$ be the one-skeleton of the **Vietoris-Rips filtration**, i.e. $\mathrm{VR}^1(\mathbb{X})$ is a family of graphs

$$\mathrm{VR}^1(\mathbb{X}) = \{\mathrm{VR}^1_r(\mathbb{X})\}_{r \in [0,\infty)}$$

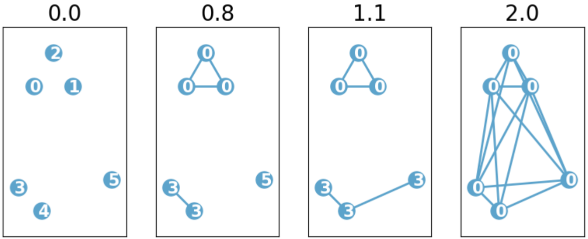
  where there are inclusions $\mathrm{VR}^1_r(\mathbb{X}) \subseteq \mathrm{VR}^1_s(\mathbb{X})$ for all $r \leq s$.
- Given $r \geq 0$, the graph $\mathrm{VR}^1_r(\mathbb{X})$ has
  - **vertices** points from $\mathbb{X}$
  - **edges** $[x, y]$ for $x, y \in \mathbb{X}$ such that $d^{\mathbb{X}}(x, y) \leq r$.

0.00	0.50	0.50	1.32	1.60	1.56
0.50	0.00	0.50	1.48	1.65	1.30
0.50	0.50	0.00	1.79	2.04	1.80
1.32	1.48	1.79	0.00	0.42	1.20
1.60	1.65	2.04	0.42	0.00	0.98
1.56	1.30	1.80	1.20	0.98	0.00



Evolution of connected components

- Let $(\mathbb{X}, d^{\mathbb{X}})$ be a finite metric space.
- Suppose that $\mathbb{X} = \{x_0, \cdots, x_n\}$ for some $n \in \mathbb{Z}_{>0}$.
- Given $r \geq 0$, we consider an equivalence relation $\sim_r$ on $\mathbb{X}$ given by $x \sim_r y$ if and only if both x and y are path connected in $\mathrm{VR}^1_r(\mathbb{X})$.
- We can represent equivalence classes of $\sim_r$ by using the minimum index on connected components:



- We define $\pi_0(\mathrm{VR}^1_r(\mathbb{X})) = \mathbb{X}/\sim_r$ and denote by $[x_i]_r$ the coset of $x_i \in \mathbb{X}$.

0-dimensional persistent homology

- Let $\mathrm{H}_0(\mathrm{VR}^1_r(\mathbb{X})) = \mathbb{Z}_2[\pi_0(\mathrm{VR}^1_r(\mathbb{X}))] = \langle [x_i]_r \mid x_i \in \mathbb{X} \rangle_{\mathbb{Z}_2}$.
- We define the 0-dimensional *persistent homology* $\mathrm{PH}_0(X)$ to be
  - the collection of $\mathbb{Z}_2$-vector spaces $\mathrm{PH}_0(X)_r := \mathrm{H}_0(\mathrm{VR}_r(X))$ for all $r \in [0, \infty)$ together with
  - the *structure maps* $\rho_{rs} \colon \mathrm{PH}_0(X)_r \to \mathrm{PH}_0(X)_s$ for all $r \leq s$ that are induced by the inclusions $\mathrm{VR}_r(X) \subseteq \mathrm{VR}_s(X)$.

Example

Let $\mathrm{PH}_{0.8}(\mathbb{X}) = \langle [x_0], [x_3], [x_5] \rangle_{\mathbb{Z}_2}$ and $\mathrm{PH}_{1.1}(\mathbb{X}) = \langle [x_0], [x_3] \rangle_{\mathbb{Z}_2}$, together with the structure map $\rho_{0.8,1.1}$ is defined from assignments

$$[x_0] \mapsto [x_0] \quad [x_3] \mapsto [x_3] \quad [x_5] \mapsto [x_3]\,.$$

Notice: component $[x_5]$ has merged to $[x_3]$ at 1.1

Barcode of 0-dimensional persistent homology

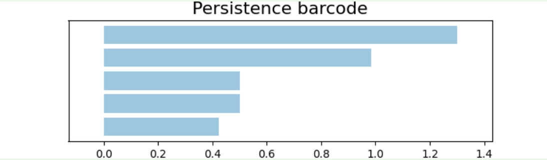
- In our case, all classes from $\mathrm{PH}_0(\mathbb{X})$ are **born** at 0 where $\pi_0(\mathrm{VR}_0^1(\mathbb{X})) = \{[x_0], \ldots, [x_n]\}$.
- We say that $[x_j] \in \pi_0(\mathrm{VR}_0^1(\mathbb{X}))$ **dies** at $b > 0$ if
  1) $\rho_{0r}([x_i]) = [x_i]$ for all $0 \leq r < b$, and
  2) $\rho_{0b}([x_i]) = [x_j]$ for some $j < i$.
- **Persistence barcode:** is a multiset $\mathbf{B}(\mathbb{X}) = (S^{\mathbb{X}}, \mu^{\mathbb{X}})$ where $S^{\mathbb{X}} \subset \mathbb{R}^+$ and $\mu^{\mathbb{X}} \colon S^{\mathbb{X}} \to \mathbb{Z}$ a multiplicity function such that,
- $\mu^{\mathbb{X}}(b) = \#\{[x_i] \in \pi_0(\mathrm{VR}_0^1(\mathbb{X})) \mid [x_i] \text{ dies at } b\}$
- Given a multiset (S, μ), its representation is a set

$$\mathrm{Rep}(S, \mu) = \{(i, x) \in \mathbb{Z} \times S \mid x \in S \text{ and } 1 \leq i \leq \mu(x)\}.$$

Example of 0-dimensional persistent homology

Example

- $\mathbf{B}(\mathbb{X}) = (S^{\mathbb{X}}, \mu^{\mathbb{X}})$ with $S^{\mathbb{X}} = \{0.42, 0.5, 0.98, 1.3\}$ and
- $\mu^{\mathbb{X}} \colon S^{\mathbb{X}} \to \mathbb{Z}$ equal to 1 everywhere,
- except $\mu^{\mathbb{X}}(0.5) = 2$.
- $\mathrm{Rep}\mathbf{B}(\mathbb{X}) = \{(1, 0.42), (1, 0.5), (2, 0.5), (1, 0.98), (1, 1.3)\} \subset \mathbb{Z} \times S^{\mathbb{X}}$



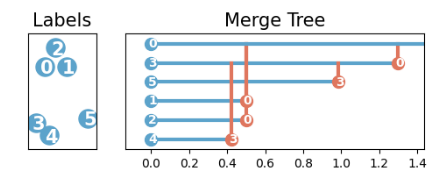
Triplet Merge Trees

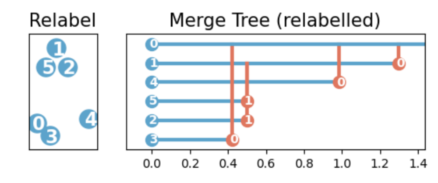
- Given $x_i, x_j \in \mathbb{X}$ with $i < j$ and some $r \geq 0$, we write a **triplet** (x_j, r, x_i) to indicate that component x_j is alive up to value r, where it merges with x_i. That is, $[x_j] \in \mathrm{PH}_0(\mathbb{X})_0 = \mathbb{Z}_2[\mathbb{X}]$ is such that $\rho_{0s}[x_j] = [x_j]$ for all $s < r$ and $\rho_{0r}[x_j] = [x_i]$.
- We denote by $\mathrm{TMT}(\mathbb{X})$ the set of triples from $\mathrm{VR}(\mathbb{X})$.

Example

Consider $\mathbb{X}$ and a relabelling $\widetilde{\mathbb{X}}$:

- $\mathrm{TMT}(\mathbb{X}) = \{(x_4, 0.42, x_3), (x_2, 0.5, x_0), (x_1, 0.5, x_0), (x_5, 0.98, x_3), (x_3, 1.3, x_0)\}$
- $\mathrm{TMT}(\widetilde{\mathbb{X}}) = \{(x_3, 0.42, x_0), (x_2, 0.5, x_1), (x_5, 0.5, x_1), (x_4, 0.98, x_0), (x_1, 1.3, x_0)\}$





Barcode decomposition of $\mathrm{PH}_0(\mathbb{X})$

- We define the **interval module** κ_b, for $b > 0$ or $b = \infty$, as
  - $\kappa_{br} = \mathbb{Z}_2$ for all $0 \leq r < b$ and is zero otherwise, and
  - the structure maps are the identities $\mathbb{Z}_2 \to \mathbb{Z}_2$ whenever possible.
- Since $\mathbb{X}$ is finite, $\mathrm{PH}_0(\mathbb{X})_r$ is **tame**, in particular, it decomposes as

$$\mathrm{PH}_0(\mathbb{X}) \simeq \left(\bigoplus_{(i,b) \in \mathrm{Rep}\mathbf{B}(\mathbb{X})} \kappa_b \right) \oplus \kappa_\infty$$

$$\simeq \left(\bigoplus_{(x_j, b_j, x_i) \in \mathrm{TMT}(\mathbb{X})} \kappa_{b_j} \right) \oplus \kappa_\infty$$

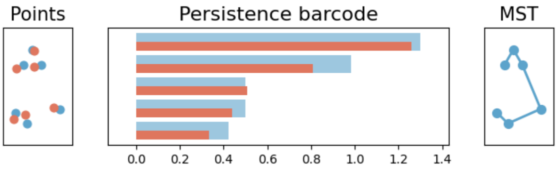
- The intervals on the barcode $\mathbf{B}(\mathbb{X})$ do not depend on the particular labelling.
- **Computation:** of $\mathbf{B}(\mathbb{X})$ and $\mathrm{TMT}(\mathbb{X})$ efficiently computed via minimum spanning tree $\mathrm{MST}(\mathbb{X})$ (e.g. using union-find data from Kruskal's method)

Stability of $\mathrm{PH}_0(\mathbb{X})$

- The barcode is stable $d_B(\mathbf{B}(\mathbb{X}), \mathbf{B}(\mathbb{Y})) \leq d_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$, where

$$d_B(\mathbf{B}(\mathbb{X}), \mathbf{B}(\mathbb{Y})) = \inf\{\varepsilon > 0 \mid \exists\ \varepsilon\text{-matching between } \mathbf{B}(\mathbb{X}) \text{ and } \mathbf{B}(\mathbb{Y})\}$$

- A ε-matching $\mu\colon \mathbf{B}(\mathbb{X}) \nrightarrow \mathbf{B}(\mathbb{Y})$ consists of a bijection $\mu\colon A \to B$ where $A \subset \mathrm{Rep}\mathbf{B}(\mathbb{X})$ and $B \subset \mathrm{Rep}\mathbf{B}(\mathbb{Y})$, such that:
  - $\mu(i, b) = (j, b')$ implies $|b - b'| < \varepsilon$
  - $(i, a) \in \mathrm{Rep}\mathbf{B}(\mathbb{X}) \setminus \mu^{-1}(B)$ implies $|a| < \varepsilon$,
  - $(j, b) \in \mathrm{Rep}\mathbf{B}(\mathbb{Y}) \setminus \mu(A)$ implies $|b| < \varepsilon$.



Outline

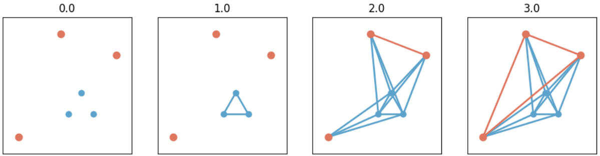
1. Motivation: Data quality and fleet behavior
2. Merge trees and zero-dimensional persistent homology
3. **The induced matching diagram (for subsets of metric spaces)**
4. The induced block function and the matching diagram
5. Stability of $D(f)$
6. Example I: exploring topological data quality
7. Example II: Navground Analysis

Subsets of metric spaces

- Consider a pseudo-metric space $(\mathbb{X}, d^{\mathbb{X}})$ together with a subset $\mathbb{Y} \subseteq \mathbb{X}$.
- There is an inclusion $\mathrm{VR}^1(\mathbb{Y}) \subseteq \mathrm{VR}^1(\mathbb{X})$.

Example

- Let $\mathbb{Y} \subset \mathbb{X} \subset \mathbb{R}^2$ where $\mathbb{Y}$ is indicated by the red points.
- The filtration of the pair $\mathrm{VR}(\mathbb{Y}) \subseteq \mathrm{VR}(\mathbb{X})$ is depicted below.



Persistence morphism induced by $\mathbb{Y} \subset \mathbb{X}$

- $\mathrm{VR}(\mathbb{Y}) \subset \mathrm{VR}(\mathbb{X})$ induces a **persistence morphism** $f \colon \mathrm{PH}_0(\mathbb{Y}) \to \mathrm{PH}_0(\mathbb{X})$, that is
  - a collection of $\mathbb{Z}_2$-linear maps $f_r \colon \mathrm{PH}_0(\mathbb{Y})_r \to \mathrm{PH}_0(\mathbb{X})_r$ for all $r \in \mathbb{R}$ and such that
  - given $r < s$, there is a commutative diagram

$$\begin{array}{ccc} \mathrm{PH}_0(\mathbb{Y})_r & \xrightarrow{f_r} & \mathrm{PH}_0(\mathbb{X})_r \\ \downarrow{\scriptstyle \rho_{rs}} & & \downarrow{\scriptstyle \rho_{rs}} \\ \mathrm{PH}_0(\mathbb{Y})_s & \xrightarrow{f_s} & \mathrm{PH}_0(\mathbb{X})_s \end{array}$$

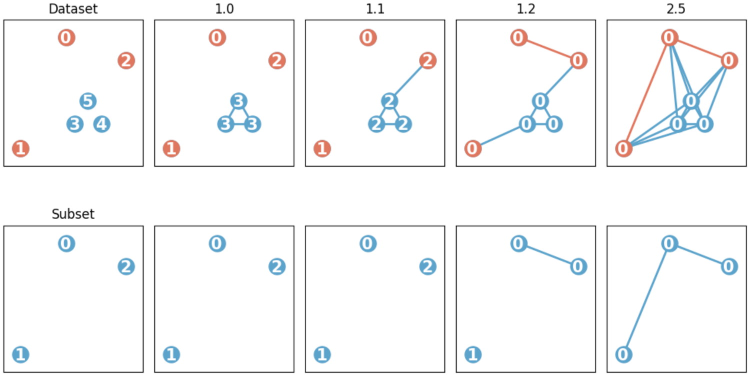
- We can define $\mathrm{im}(f)$, $\ker(f)$ and $\mathrm{coker}(f)$, e.g. $\mathrm{im}(f)_r := \mathrm{im}(f_r)$ for all $r \in \mathbb{R}$.
- Intuitively, $f \colon \mathrm{PH}_0(\mathbb{Y}) \to \mathrm{PH}_0(\mathbb{X})$ relates the (single-linkage) clusters from $\mathrm{VR}^1(\mathbb{Y})$ and $\mathrm{VR}^1(\mathbb{X})$.

Subset labels and connected components

- $\mathbb{Y} = \{x_0, \cdots, x_\ell\}$ and $\mathbb{X} = \{x_0, \cdots, x_n\}$ for $\ell \leq n$.

Example

Depiction of $\pi_0(\mathrm{VR}_r(X))$ and $\pi_0(\mathrm{VR}_r(Z))$ for varying $r \geq 0$.



Decomposition of morphisms in dimension 0

- Using the barcode decompositions, we can write f as

$$f : \left(\bigoplus_{(i,b_i)\in \mathrm{Rep}\mathbf{B}(\mathbb{Y})} \kappa_{b_i} \right) \oplus \kappa_\infty \longrightarrow \left(\bigoplus_{(i,b_i)\in \mathrm{Rep}\mathbf{B}(\mathbb{X})} \kappa_{b_i} \right) \oplus \kappa_\infty .$$

- This does not give information about the relation between $\mathbf{B}(\mathbb{Y})$ and $\mathbf{B}(\mathbb{X})$.
- On the other hand, using a Theorem by Jacquard, Nanda and Tillmann [4], there exists $r_a^b \geq 0$ and $d_a^+, d_b^- \geq 0$ such that

$$f \simeq \left(\bigoplus_{b>0} \bigoplus_{a\geq b} \bigoplus_{i\in[[r_a^b]]} (\kappa_a \to \kappa_b) \right) \oplus \left(\bigoplus_{b>0} \bigoplus_{j\in[[d_b^-]]} (0 \to \kappa_b) \right) \oplus \left(\bigoplus_{a>0} \bigoplus_{j\in[[d_a^+]]} (\kappa_a \to 0) \right) \oplus \left(\kappa_\infty \to \kappa_\infty \right).$$

- Where $[[n]] = \{0, \ldots, n-1\}$ for $n \in \mathbb{Z}$ and
- $\kappa_a \to \kappa_b$ for $a \geq b$ are the natural persistence morphisms.

The matching diagram $D(f)$ (for $\mathbb{Y} \subset \mathbb{X}$)

- Using $\mathbb{Y} \subset \mathbb{X}$, it follows that $d_a^+ = 0$ for all $a \in \mathbb{R}$, and so

$$f \simeq \left(\bigoplus_{b>0} \bigoplus_{a \geq b} \bigoplus_{i \in [[r_a^b]]} (\kappa_a \to \kappa_b) \right) \oplus \left(\bigoplus_{b>0} \bigoplus_{j \in [[d_b^-]]} (0 \to \kappa_b) \right) \oplus (\kappa_\infty \to \kappa_\infty).$$

- We define the **matching diagram** $D(f)$ as a multiset (S^D, μ^D) consisting of
  - a set $S^D = \{(a,b) \mid r_b^a \neq 0\} \cup \{(\infty, b) \mid d_b^- \neq 0\} \subset \mathbb{R}^2$,
  - a multiplicity function $\mu^D \colon S^D \to \mathbb{Z}_{>0}$ given by

$$\mu^D(a,b) = r_b^a, \text{ and } \mu^D(\infty, b) = d_b^-.$$

- There is an "injection" of barcodes $\mathbf{B}(\mathbb{Y}) \hookrightarrow \mathbf{B}(\mathbb{X})$
  - **Well-defined:** $a \geq 0$ fixed: $\sum_{b \leq a} \mu^D(a,b) = \mu^{\mathbb{Y}}(a)$
  - **Injective:** $b \geq 0$ fixed: $\sum_{b \leq a} \mu^D(a,b) \leq \mu^{\mathbb{X}}(b)$

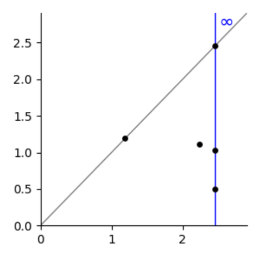
Example of matching diagram

Example

- We decompose $f : \mathrm{PH}_0(\mathbb{Y}) \to \mathrm{PH}_0(\mathbb{X})$ as

$$f \simeq (\kappa_{2.24} \to \kappa_{1.11}) \oplus (\kappa_{1.19} \to \kappa_{1.19}) \oplus (0 \to \kappa_{1.03}) \oplus (0 \to \kappa_{0.5}) \oplus (0 \to \kappa_{0.5})$$

- $S^D = \{(2.24, 1.11), (1.19, 1.19), (0, 1.03), (0, 0.5)\}$
- Except $\mu^D(0, 0.5) = 2$, all multiplicities are equal to 1.



Relation of $D(f)$ with $\mathrm{im}(f)$, $\ker(f)$, $\mathrm{coker}(f)$, $\mathrm{PH}_0(\mathbb{Y})$ and $\mathrm{PH}_0(\mathbb{X})$

- We can rewrite f as follows:

$$f \simeq \Bigg(\bigoplus_{\substack{(a,b)\in S^D \\ a\neq\infty}} \bigoplus_{i\in[[\mu^D(a,b)]]} (\kappa_a \to \kappa_b) \Bigg) \oplus \Bigg(\bigoplus_{(\infty,b)\in S^D} \bigoplus_{i\in[[\mu^D(\infty,b)]]} (0 \to \kappa_b) \Bigg) \oplus (\kappa_\infty \to \kappa_\infty).$$

- $\mathrm{im}(f) \simeq \Bigg(\bigoplus_{\substack{(a,b)\in S^D \\ a\neq\infty}} \bigoplus_{i\in[[\mu^D(a,b)]]} \kappa_b \Bigg) \oplus \kappa_\infty$
- $\ker(f) \simeq \bigoplus_{\substack{(a,b)\in S^D \\ a\neq\infty}} \bigoplus_{i\in[[\mu^D(a,b)]]} \kappa_{[b,a)}$
- $\mathrm{coker}(f) \simeq \bigoplus_{(\infty,b)\in S^D} \bigoplus_{i\in[[\mu^D(\infty,b)]]} \kappa_b$
- $\mathrm{PH}_0(\mathbb{Y}) \simeq \Bigg(\bigoplus_{\substack{(a,b)\in S^D \\ a\neq\infty}} \bigoplus_{i\in[[\mu^D(a,b)]]} \kappa_a \Bigg) \oplus \kappa_\infty$
- $\mathrm{PH}_0(\mathbb{X}) \simeq \Bigg(\bigoplus_{\substack{(a,b)\in S^D \\ a\neq\infty}} \bigoplus_{i\in[[\mu^D((a,b))]]} \kappa_b \Bigg) \oplus \Bigg(\bigoplus_{(\infty,b)\in S^D} \bigoplus_{i\in[[\mu^D(\infty,b)]]} \kappa_b \Bigg) \oplus \kappa_\infty$

Outline

The induced block function $\mathcal{M}_f^0$

- Let $(\mathbb{Y}, d^{\mathbb{Y}})$ and $(\mathbb{X}, d^{\mathbb{X}})$ be a pair of finite metric spaces.
- Let $f_0 \colon \mathbb{Y} \to \mathbb{X}$ be a set map (ignoring metrics).
- We define the induced block function $\mathcal{M}_f^0 \colon \mathbb{R}^2 \to \mathbb{Z}_{\geq 0}$ as, given $a, b \in \mathbb{R}$,

$$\mathcal{M}_f^0(a,b) = \dim \left(\frac{f(\ker_a^+(V)) \cap \ker_b^+(U)}{f(\ker_a^-(V)) \cap \ker_b^+(U) + f(\ker_a^+(V)) \cap \ker_b^-(U)} \right).$$

- Where: $f(\ker_a^{\pm}(\mathbb{Y}))$ and $\ker_b^{\pm}(\mathbb{X})$ are subspaces from $\mathrm{PH}_0(\mathbb{X}) \simeq \mathbb{Z}_2[\mathbb{X}]$ given by
  - $f(\ker_a^+(\mathbb{Y})) = \langle [f_0(x_j)] + [f_0(x_i)] | (x_j, r, x_i) \in \mathrm{TMT}(\mathbb{Y}) \text{ such that } r \leq a \rangle$
  - $f(\ker_a^-(\mathbb{Y})) = \langle [f_0(x_j)] + [f_0(x_i)] | (x_j, r, x_i) \in \mathrm{TMT}(\mathbb{Y}) \text{ such that } r < a \rangle$
  - $\ker_b^+(\mathbb{X}) = \langle [x_j] + [x_i] | (x_j, r, x_i) \in \mathrm{TMT}(\mathbb{X}) \text{ such that } r \leq a \rangle$
  - $\ker_a^-(\mathbb{X}) = \langle [x_j] + [x_i] | (x_j, r, x_i) \in \mathrm{TMT}(\mathbb{X}) \text{ such that } r < a \rangle$
- **NB:** One can check $f(\ker_b^+(\mathbb{Y})) = f(\ker(\rho_{0b}^{\mathbb{Y}}))$, etc, which means that $\mathcal{M}_f^0$ does not depend on the particular indexing.

$\mathcal{M}_f^0$ and $D(f)$

Define the **matching diagram** as a multiset $D(f) = (S^D, \mu^D)$ where

- $S^D \subseteq \overline{\mathbb{R}} \times \mathbb{R}$ with non-zero multiplicity, where:
- $\mu^D((a,b)) = \mathcal{M}_f^0(a,b)$ for all $(a,b) \in \mathbb{R}^2$ and
- $\mu^D((\infty, b)) = \mu^{\mathbb{X}}(b) - \sum_{a \in \mathbb{R}} \mathcal{M}_f^0(a,b)$ for all $b \geq 0$.

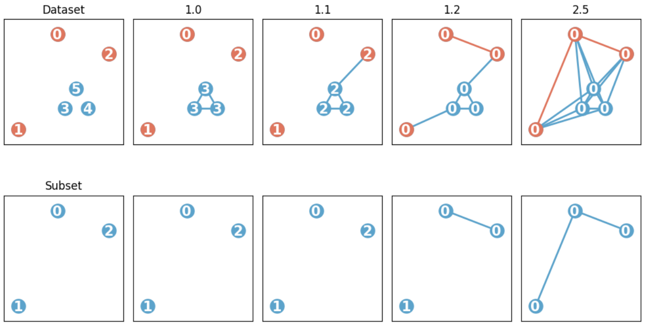
Remarks:

- $\mathcal{M}_f^0$ and $D(f)$ are defined for any pair of metric spaces $(\mathbb{Y}, d^{\mathbb{Y}})$ and $(\mathbb{X}, d^{\mathbb{X}})$ together with a set map $f_0 \colon \mathbb{Y} \to \mathbb{X}$.
- In particular, there might be no persistence morphism.
- If $f_0 \colon \mathbb{Y} \to \mathbb{X}$ is one-Lipschitz, then it induces a persistence morphism $f \colon \mathrm{PH}_0(\mathbb{Y}) \to \mathrm{PH}_0(\mathbb{X})$ and $D(f)$ coincides with the definition using the direct sum decomposition of f.

Example computation of $\mathcal{M}_f^0$ (Part 0)

Example

Consider $\mathbb{Y} \subset \mathbb{X}$ as in the previous section.
One can compute $\mathrm{TMT}(\mathbb{Y}) = \{(x_2, 1.19, x_0), (x_1, 2.24, x_0)\}$ and
$\mathrm{TMT}(\mathbb{X}) = \{(x_5, 0.5, x_3), (x_4, 0.5, x_3), (x_3, 1.03, x_2), (x_2, 1.11, x_1), (x_1, 1.19, x_0)\}$.



Example computation of $\mathcal{M}_f^0$ (Part I)

Example

Recall $\mathrm{TMT}(\mathbb{Y}) = \{(x_2, 1.19, x_0), (x_1, 2.24, x_0)\}$ and $\mathrm{TMT}(\mathbb{X}) = \{(x_5, 0.5, x_3), (x_4, 0.5, x_3), (x_3, 1.03, x_2), (x_2, 1.11, x_1), (x_1, 1.19, x_0)\}$. Then, we have

$$\begin{aligned} f(\ker^-(\mathbb{Y})_{1.19}) &= 0 \\ f(\ker^+(\mathbb{Y})_{1.19}) &= \langle [x_2] + [x_0] \rangle \\ \ker^-(\mathbb{X})_{1.19} &= \langle [x_5] + [x_3], [x_4] + [x_3], [x_3] + [x_2], [x_2] + [x_1] \rangle \\ \ker^+(\mathbb{X})_{1.19} &= \langle [x_5] + [x_3], [x_4] + [x_3], [x_3] + [x_2], [x_2] + [x_1], [x_1] + [x_0] \rangle \end{aligned}$$

So that

$$\mathcal{M}_f^0(1.19, 1.19) = \dim \left(\frac{\langle [x_2] + [x_0] \rangle}{0 + 0} \right) = \dim(\langle [x_2] + [x_0] \rangle) = 1\,.$$

Example computation of $\mathcal{M}_f^0$ (Part II)

Example

Recall $\mathrm{TMT}(\mathbb{Y}) = \{(x_2, 1.19, x_0), (x_1, 2.24, x_0)\}$ and $\mathrm{TMT}(\mathbb{X}) = \{(x_5, 0.5, x_3), (x_4, 0.5, x_3), (x_3, 1.03, x_2), (x_2, 1.11, x_1), (x_1, 1.19, x_0)\}$. Then, we have

$$\begin{aligned}
f(\ker^-(\mathbb{Y})_{2.24}) =& \langle [x_2] + [x_0] \rangle \\
f(\ker^+(\mathbb{Y})_{2.24}) =& \langle [x_2] + [x_0], [x_1] + [x_0] \rangle \\
\ker^-(\mathbb{X})_{1.19} =& \langle [x_5] + [x_3], [x_4] + [x_3], [x_3] + [x_2], [x_2] + [x_1] \rangle \\
\ker^+(\mathbb{X})_{1.19} =& \langle [x_5] + [x_3], [x_4] + [x_3], [x_3] + [x_2], [x_2] + [x_1], [x_1] + [x_0] \rangle
\end{aligned}$$

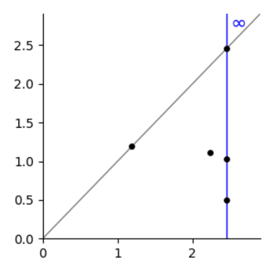
So that

$$\mathcal{M}_f^0(1.19, 2.24) = \dim \left(\frac{\langle [x_2] + [x_0], [x_1] + [x_0] \rangle}{\langle [x_2] + [x_0] \rangle + \langle [x_2] + [x_1] \rangle} \right) = 0\,.$$

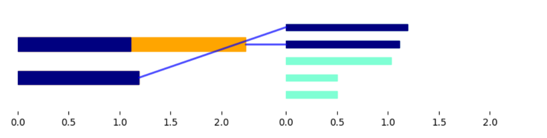
Example computation of $\mathcal{M}_f^0$ (Part III)

Example

All in all, one can check that $\mathcal{M}_f^0(1.19, 1.19) = \mathcal{M}_f^0(2.24, 1.11) = 1$ while $\mathcal{M}_f^0(a, b) = 0$ for any other pair in $\mathbb{R}^2$. Thus, we recover the matching diagram $D(f)$ from the previous section



$$\mathcal{M}_f^0 \colon \mathbf{B}(\mathbb{Y}) \hookrightarrow \mathbf{B}(\mathbb{X})$$



Computation of $\mathcal{M}^0_f$

- **Github:** https://github.com/Cimagroup/tdqual
- Small python script: 282 lines, most for generating plots.
- **Algorithm:**
  - Compute the minimum spanning trees $\mathrm{MST}(\mathbb{Y})$ and $\mathrm{MST}(\mathbb{X})$ of $\mathrm{VR}(\mathbb{Y})$ and $\mathrm{VR}(\mathbb{X})$ respectively.
  - Compute $\mathrm{TMT}(\mathbb{Y})$ and $\mathrm{TMT}(\mathbb{X})$ from $\mathrm{MST}(\mathbb{Y})$ and $\mathrm{MST}(\mathbb{X})$.
  - Compute the matrix F associated to $\mathrm{PH}_0(\mathbb{Y}) \to \mathrm{PH}_0(\mathbb{X})$ using $\mathrm{TMT}(\mathbb{Y})$ and $\mathrm{TMT}(\mathbb{X})$.
  - Perform a Gaussian reduction of F and obtain R

Using R we obtain $\mathcal{M}^0_f$

$$\mathcal{M}^0_f(a,b) = \# \left\{ \begin{array}{l} \text{pivots from } R \text{ in columns labelled by } (i,a) \in \mathrm{Rep}\mathbf{B}(\mathbb{Y}) \\ \text{and rows labelled by } (j,b) \in \mathrm{Rep}\mathbf{B}(\mathbb{X}) \\ \text{for all } t \in [[\mu^{\mathbb{Y}}(a)]] \text{ and } r \in [[\mu^{\mathbb{X}}(b)]] \end{array} \right\}.$$

Runtimes

Size	Prop	Time (s) Dim 100	Time (s) Dim 200	Time (s) Dim 500	Time (s) Dim 1000
1000	0.1	0.2137	0.2251	0.2906	0.4167
	0.2	0.2205	0.2337	0.3025	0.4266
	0.5	0.2614	0.2770	0.3672	0.5181
	0.8	0.3420	0.3647	0.4785	0.6791
5000	0.1	6.8683	7.4391	9.3776	13.2139
	0.2	7.1100	7.6029	9.6264	13.5654
	0.5	8.3384	9.0599	11.3875	15.9382
	0.8	11.0920	11.9642	15.1606	21.0903
10000	0.1	30.9312	33.8928	43.3496	59.5236
	0.2	31.7921	34.5644	44.4252	61.1580
	0.5	37.6867	41.3084	52.5283	72.4249
	0.8	50.1831	54.3484	69.9975	96.1254

Outline

Stability of $D(f)$: (Recap) Gromov-Hausdorff distance

- Let (M, d^M) be a finite metric space.
- Given a subset $A \subseteq M$ and a point $x \in M$ we define the distance

$$d^M(x, M) = \min\{d^M(x, m) \text{ for } m \in M\}\,.$$

- Given two subsets $A, B \subseteq M$, we define the **Hausdorff distance**

$$d_H^M(A, B) = \max\{\max\{d^M(a, B) | a \in A\}, \max\{d^M(b, A) | b \in B\}\}$$

- Given two metric spaces (Z, d^Z) and $(Z', d^{Z'})$, we consider isometric embeddings $\gamma_Z : Z \hookrightarrow M$ and $\gamma_{Z'} : Z' \hookrightarrow M$ for some (M, d^M)
- We define the **Gromov-Hausdorff distance**

$$d_{GH}(Z, Z') = \inf_{M, \gamma_Z, \gamma_{Z'}} \left\{d_H^M(\gamma_Z(Z), \gamma_{Z'}(Z'))\right\}.$$

Stability of $D(f)$

Given (X, Z) and (X', Z') such that $X \subseteq Z$ and $X' \subseteq Z'$, we define the Gromov-Hausdorff distance $d_{GH}((X, Z), (X', Z'))$ as

$$\inf_{M,\gamma_Z,\gamma_{Z'}} \left\{ \max \left\{ d_H^M(\gamma_Z(X), \gamma_{Z'}(X')), d_H^M(\gamma_Z(Z), \gamma_{Z'}(Z')) \right\} \right\}.$$

Theorem

Suppose $d_{GH}((X, Z), (X', Z')) < \varepsilon$. There exists a partial matching $\sigma^{D(f)} : \mathit{Rep}D(f) \nrightarrow \mathit{Rep}D(f')$ such that

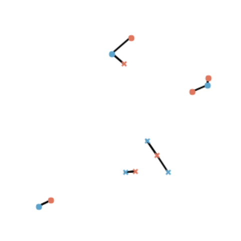
- *for $(i, (a, b)) \in \mathit{coim}(\sigma^{D(f)})$, writing $\sigma^{D(f)}(i, (a, b)) = (j, (a', b'))$, we have that (either $a = a' = \infty$ or $|a - a'| < 2\varepsilon$) and $|b - b'| < 2\varepsilon$.*
- *for $(i, (a, b)) \in \mathit{Rep}D(f) \setminus \mathit{coim}(\sigma^{D(f)})$ then (either $a = \infty$ or $a < 2\varepsilon$) and $b < 2\varepsilon$.*
- *for $(j, (a', b')) \in \mathit{Rep}D(f') \setminus \mathit{im}(\sigma^{D(f)})$ then (either $a' = \infty$ or $a' < 2\varepsilon$) and $b' < 2\varepsilon$.*

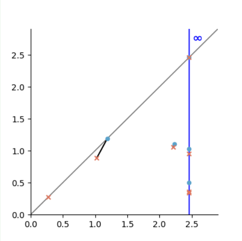
Stability example

Example

Suppose that $\varepsilon \sim 0.27$. We consider two point clouds and check stability.

a	b	a'	b'	$\lvert a - a' \rvert$	$\lvert b - b' \rvert$
1.19	1.19	1.03	0.89	0.16	**0.30**
2.24	1.11	2.21	1.06	0.02	0.05
–	–	0.27	0.27	–	–
inf	1.03	inf	0.95	nan	0.07
inf	0.50	inf	0.36	nan	0.14
inf	0.50	inf	0.34	nan	0.16



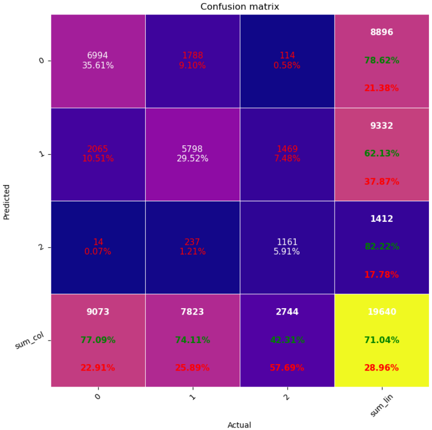


Outline

1. Motivation: Data quality and fleet behavior
2. Merge trees and zero-dimensional persistent homology
3. The induced matching diagram (for subsets of metric spaces)
4. The induced block function and the matching diagram
5. Stability of $D(f)$
6. **Example I: exploring topological data quality**
7. Example II: Navground Analysis

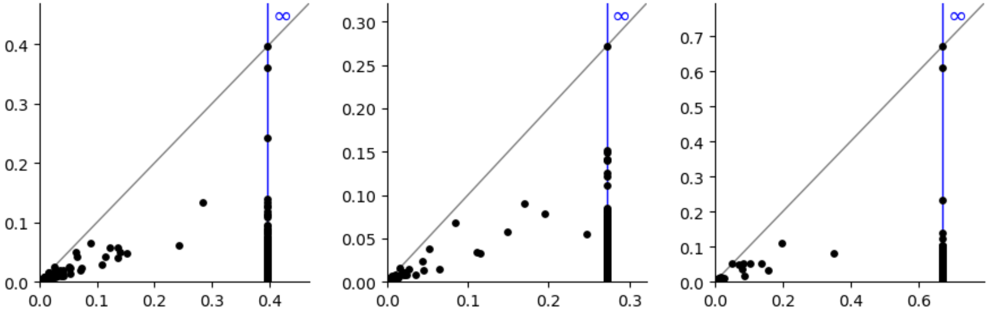
Example: Housing dataset

- $\mathcal{H}$ contains 20648 real-valued 8-dimensional samples.
- Divided into three classes.
- We took a random sample $X^{\mathcal{H}}$ of size 1000 as training set.
- Trained a MLP and obtained the following confusion matrix:



Example: Housing Dataset (Continued)

- We computed the matching diagram for the three classes separately.



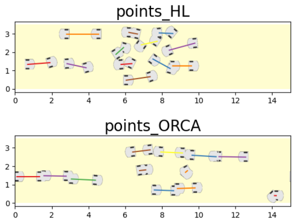
- The third class has a very large interval in $\mathrm{coker}(f)$ compared to the other two.
- Also, the larger interval in $\ker(f)$ is in the third class as well.

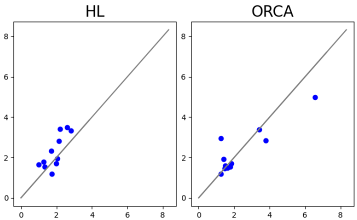
Outline

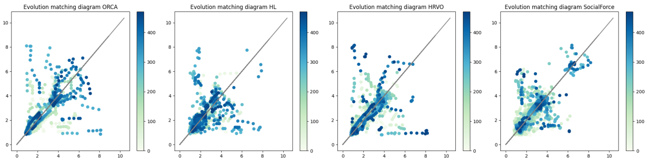
Fleet analysis

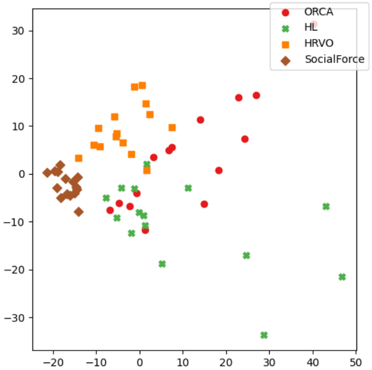
- We consider the output from Navground simulations.
- This consists of point clouds $\{\mathbb{X}^t\}_{t=0}^N$ where
- $\mathbb{X}^t \subset \mathbb{R}^6$, given by positions, angle, speed components and angular speed.
- Given two timesteps $T_1 < T_2$ let the Euclidean distances of $\mathbb{X}^{T_1}$ and $\mathbb{X}^{T_2}$ be given by matrices D^{T_1} and D^{T_2} respectively.
- Agent movements leads to an isomorphism $\mathbb{X}^{T_1} \to \mathbb{X}^{T_2}$
- We analyse the evolution of $\{\mathbb{X}^t\}_{t=0}^N$ by keeping the difference $T_2 - T_1$ constant and changing T_1.
- We tried this approach in three scenarios: Corridor, Cross and CrossTorus.
- We compare four behaviors: ORCA, HL, HRVO and Social Force.

Example: Corridor using persistence images

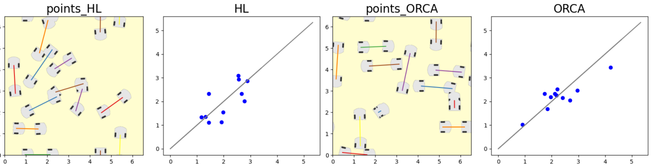


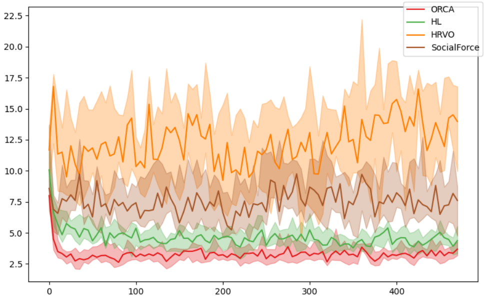


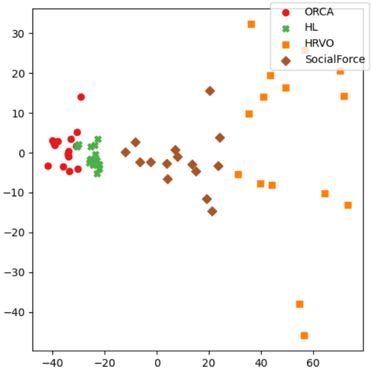




Example: CrossTorus using absolute sum of differences in $D(f)$







Conclusions and future work

- $D(f)$ is versatile; can be applied in several examples.
- $D(f)$ can be computed using minimum spanning trees and a simple Gausian reduction.
- Stability also works for set injections $\mathbb{Y} \hookrightarrow \mathbb{X}$, can it be generalised?
- What about metrics that do not satisfy the triangle inequality? e.g. when using Dynamic Time Warping?
- **Ongoing work:** extend matching diagrams to higher dimensions or when points are allowed to be born later than 0.
- It would be good to see/work on more applications of $D(f)$ in the future!
- What about other clustering techniques?

References

- Á. Torras-Casas, E. Paluzo-Hidalgo, R. González-Díaz. *Topological data quality via 0-dimensional persistence matching*, arXiv:2306.02411v2.
- Á. Torras-Casas, R. González-Díaz. *Properties and Stability of Persistence Matching Diagrams*, arXiv:2409.14954.
- D. Smirnov, D. Morozov, *Triplet Merge Trees* Topological Methods in Data Analysis and Visualization V, 2020
- E. Jacquard and V. Nanda and U. Tillmann *The space of barcode bases for persistence modules*, Journal of Applied and Computational Topology, Volume 7, 2023.
- R. González-Díaz, M. Soriano-Trigueros, Á. Torras-Casas, *Partial Matchings induced by Morphisms between Persistence Morphisms*, Computational Geometry, Volume 112, 2023.

Thank You

Email: alvaro.torras-casas@inserm.fr

Website: https://alvaro-torras-casas.org/