Density-based persistent homology

Ximena Fernández*
joint work with E. Borghini, P. Groisman and G. Mindlin

2nd Workshop on Topological Methods in Data Analysis
6th October 2021

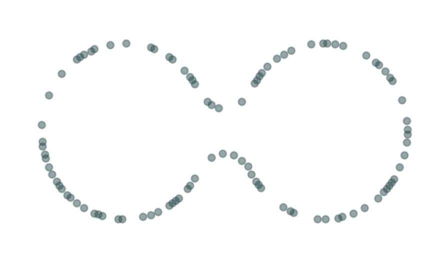
*EPSRC Centre for Topological Data Analysis

Durham University

The problem

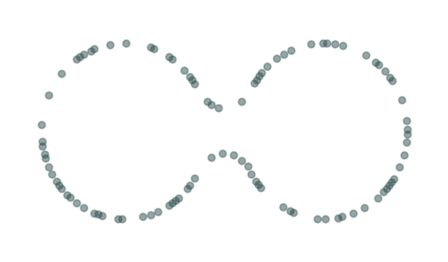
Homology inference

$\mathbb{X}_n = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}^D$ a finite sample.

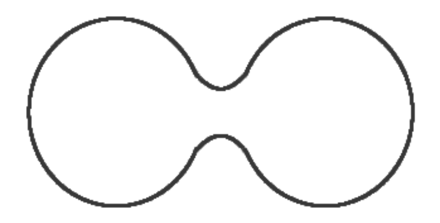


Homology inference

$\mathbb{X}_n = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{R}^D$ a finite sample.

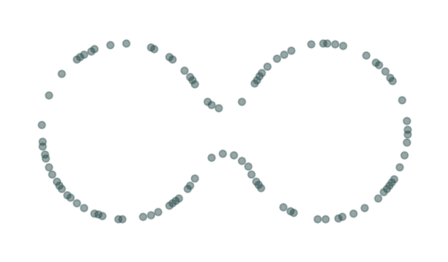


$\mathbb{X}_n \subseteq \mathcal{M}$ a d-dimensional Riemannian manifold.

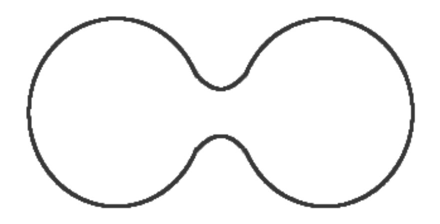


Homology inference

$\mathbb{X}_n = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}^D$ a finite sample.



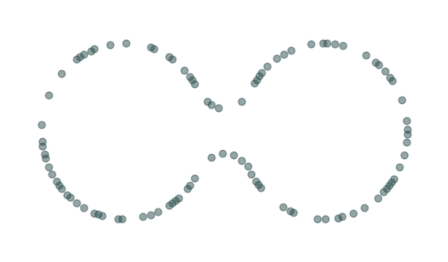
$\mathbb{X}_n \subseteq \mathcal{M}$ a d-dimensional Riemannian manifold.



Q: How to infer the homology of $\mathcal{M}$ from the sample $\mathbb{X}_n$?

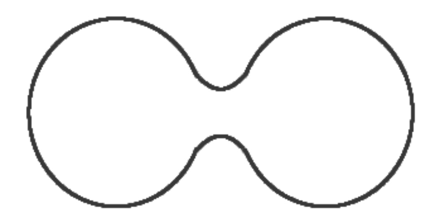
Homology inference

$\mathbb{X}_n = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}^D$ a finite sample.



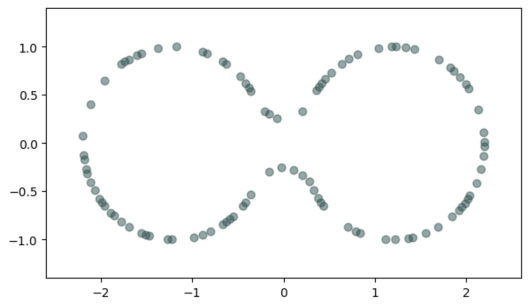
A: Compute persistent homology of $\mathbb{X}_n$.

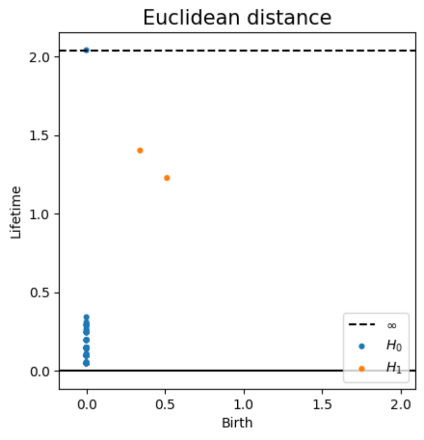
$\mathbb{X}_n \subseteq \mathcal{M}$ a d-dimensional Riemannian manifold.



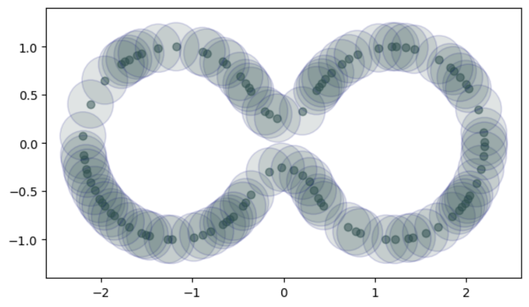
Q: How to infer the homology of $\mathcal{M}$ from the sample $\mathbb{X}_n$?

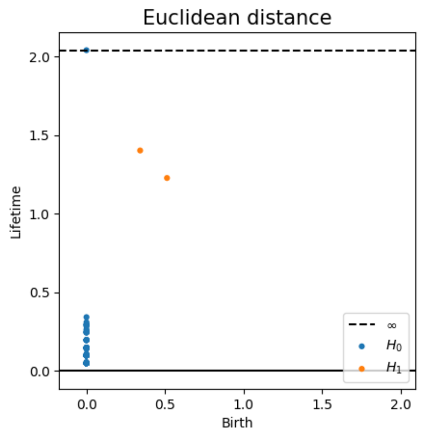
Ambient persistent homology





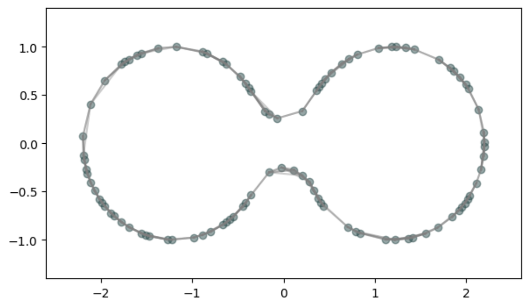
Ambient persistent homology

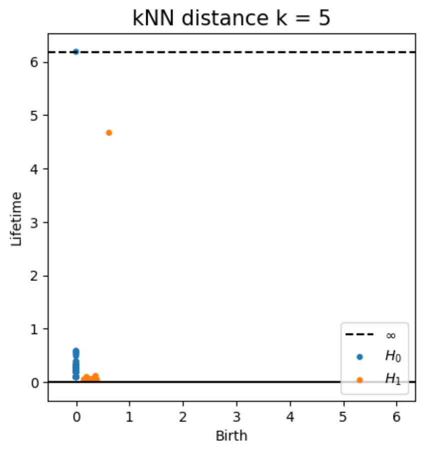




- $\mathrm{Rips}_\epsilon(\mathcal{M}, d_E) \simeq \mathcal{M}$ for $\epsilon < 2\mathrm{rch}(\mathcal{M})$

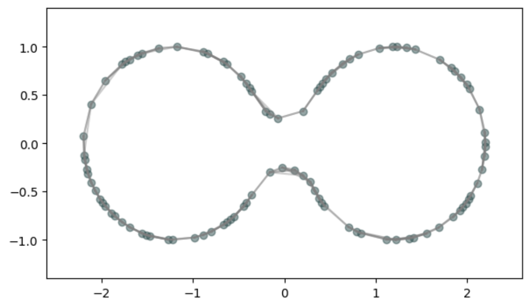
Intrinsic persistent homology

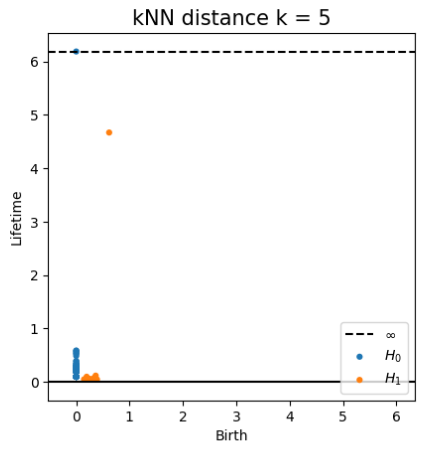




$$d_{kNN}(x, y) = \inf_{\gamma : x \to y} \sum |x_i - x_{i+1}| \sim d_{\mathcal{M}}(x, y)$$

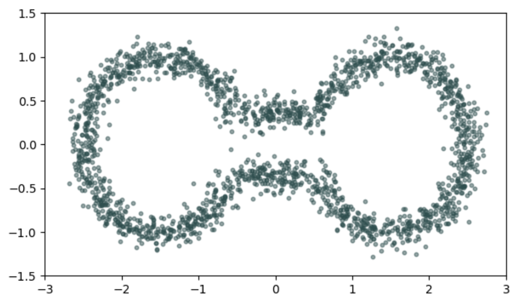
Intrinsic persistent homology

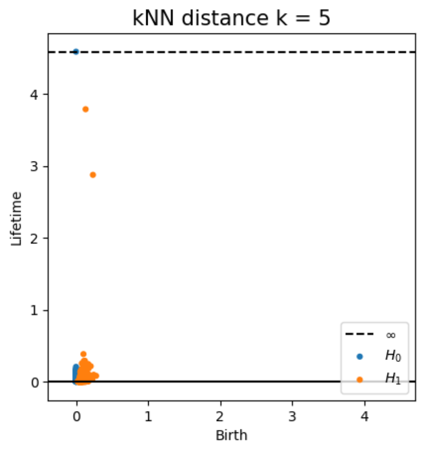




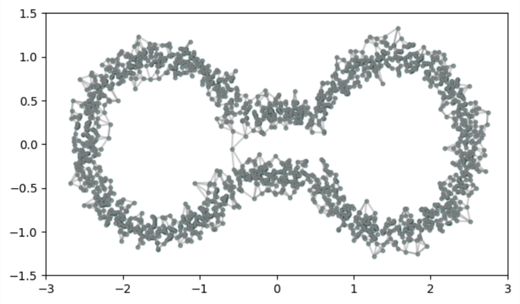
- $\mathrm{Rips}_\epsilon(\mathcal{M}, d_{\mathcal{M}}) \simeq \mathcal{M}$ for $\epsilon < \mathrm{conv}(\mathcal{M}, d_{\mathcal{M}})$

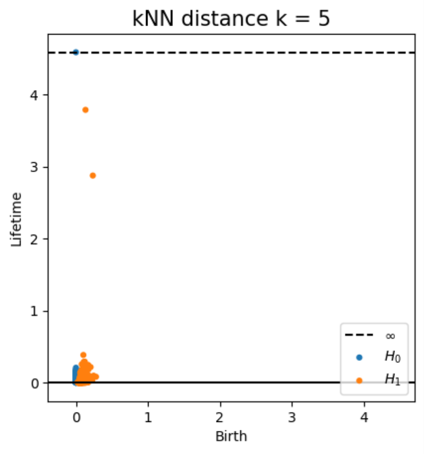
The problem of noise



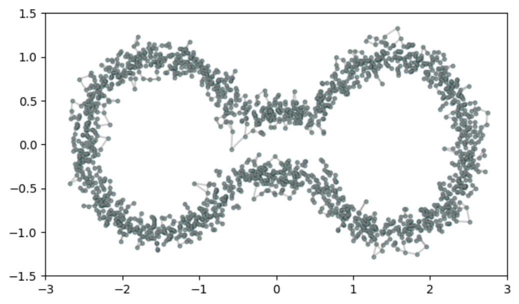


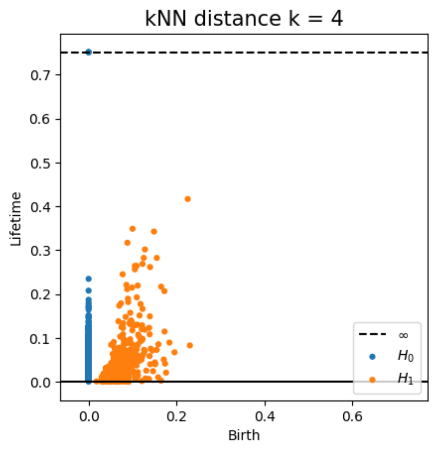
The problem of noise





The problem of noise





Density-based manifold learning

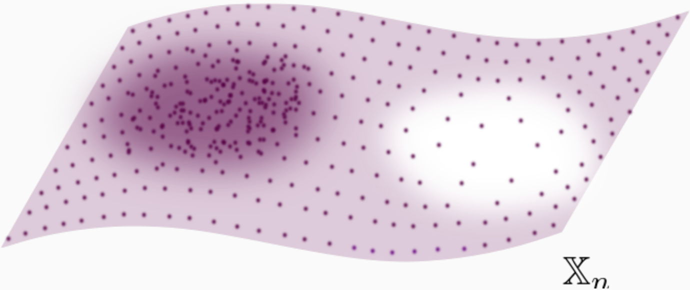
Fermat distance

Let $\mathbb{X}_n \subseteq \mathbb{R}^D$ a sample of points.

For $p > 1$, the **Fermat distance** between $x, y \in \mathbb{R}^D$ is defined by

$$d_{\mathbb{X}_n,p}(x,y) = \inf_{\gamma} \sum_{i=0}^{r} |x_{i+1} - x_i|^p$$

over all paths $\gamma = (x_0, \ldots, x_{r+1})$ of finite length with $x_0 = x$, $x_{r+1} = y$ and $\{x_1, x_2, \ldots, x_r\} \subseteq \mathbb{X}_n$.



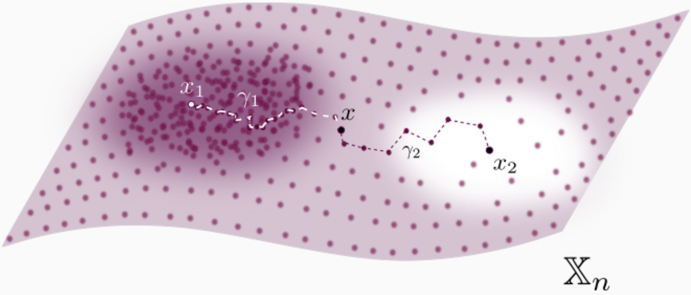
Fermat distance

Let $\mathbb{X}_n \subseteq \mathbb{R}^D$ a sample of points.

For $p > 1$, the **Fermat distance** between $x, y \in \mathbb{R}^D$ is defined by

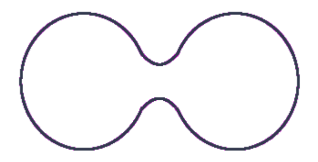
$$d_{\mathbb{X}_n,p}(x,y) = \inf_{\gamma} \sum_{i=0}^{r} |x_{i+1} - x_i|^p$$

over all paths $\gamma = (x_0, \ldots, x_{r+1})$ of finite length with $x_0 = x$, $x_{r+1} = y$ and $\{x_1, x_2, \ldots, x_r\} \subseteq \mathbb{X}_n$.



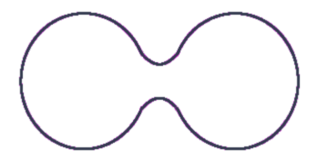
Example

Manifold

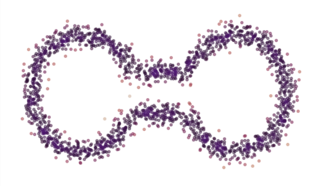


Example

Manifold

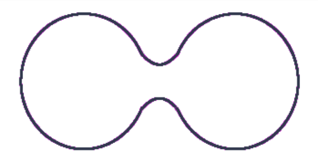


Sample

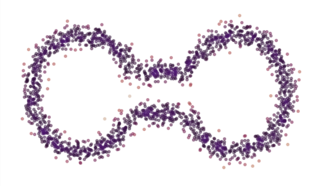


Example

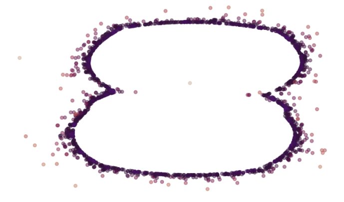
Manifold



Sample

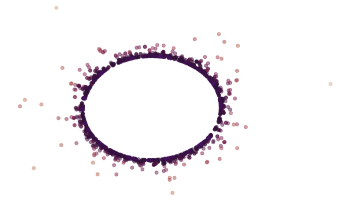


Fermat p = 2.0

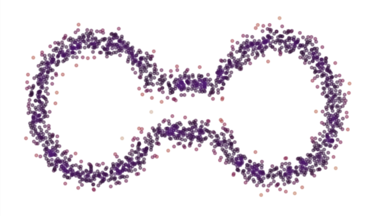


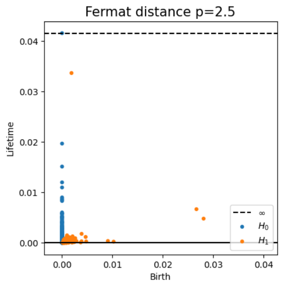
Fermat p = 2.5

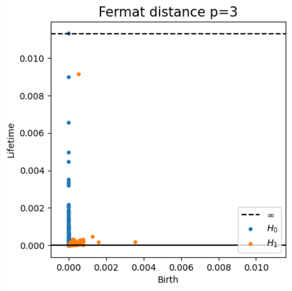
Fermat p = 3.0

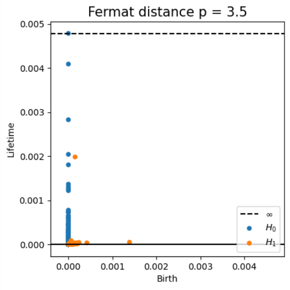


Example




Fermat distance p=2.5
Lifetime
0.04
0.03
0.02
0.01
0.00
0.00 0.01 0.02 0.03 0.04
Birth
∞
H_0
H_1


Fermat distance p=3
Lifetime
0.010
0.008
0.006
0.004
0.002
0.000
0.000 0.002 0.004 0.006 0.008 0.010
Birth
∞
H_0
H_1


Fermat distance p = 3.5
Lifetime
0.005
0.004
0.003
0.002
0.001
0.000
0.000 0.001 0.002 0.003 0.004
Birth
∞
H_0
H_1

Properties of Fermat distance

$\mathbb{X} \subseteq \mathbb{R}^D$ sample of points.

$$d_{\mathbb{X}_n,p}(x,y) = \inf_{\gamma:x\to y} \sum_{i=0}^{r} |x_{i+1} - x_i|^p$$

Properties of Fermat distance

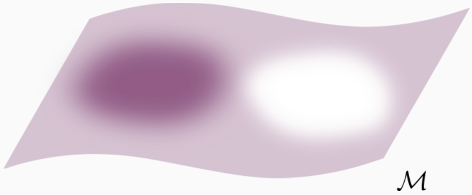
$\mathbb{X} \subseteq \mathbb{R}^D$ sample of points.

$$d_{\mathbb{X}_n,p}(x,y) = \inf_{\gamma:x\to y} \sum_{i=0}^{r} |x_{i+1} - x_i|^p$$

$\mathcal{M} \subseteq \mathbb{R}^D$ a d-dimensional manifold.

$f : \mathcal{M} \to \mathbb{R}$ a density function.

$$d_{\mathcal{M},f,p}(x,y) = \inf_{\Gamma:x\to y} \int_\Gamma \frac{1}{f^{(p-1)/d}}$$



Properties of Fermat distance

$\mathbb{X} \subseteq \mathbb{R}^D$ sample of points.

$$d_{\mathbb{X}_n,p}(x,y) = \inf_{\gamma : x \to y} \sum_{i=0}^{r} |x_{i+1} - x_i|^p$$

$\mathbb{X}_n$

$\mathcal{M} \subseteq \mathbb{R}^D$ a d-dimensional manifold.
$f : \mathcal{M} \to \mathbb{R}$ a density function.

$$d_{\mathcal{M},f,p}(x,y) = \inf_{\Gamma : x \to y} \int_{\Gamma} \frac{1}{f^{(p-1)/d}}$$

$\mathcal{M}$

- **Convergence of metric spaces:**

$$\left(\mathbb{X}_n, C(n,p,d)\, d_{\mathbb{X}_n,p}\right) \xrightarrow[n \to \infty]{GH} \left(\mathcal{M}, d_{\mathcal{M},f,p}\right)$$

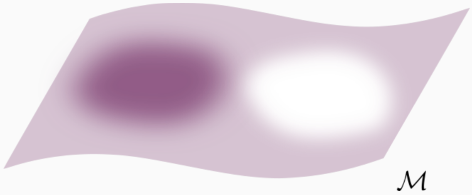
Properties of Fermat distance

$\mathbb{X} \subseteq \mathbb{R}^D$ sample of points.

$$d_{\mathbb{X}_n,p}(x,y) = \inf_{\gamma:x\to y} \sum_{i=0}^{r} |x_{i+1} - x_i|^p$$

$\mathcal{M} \subseteq \mathbb{R}^D$ a d-dimensional manifold.
$f : \mathcal{M} \to \mathbb{R}$ a density function.

$$d_{\mathcal{M},f,p}(x,y) = \inf_{\Gamma:x\to y} \int_{\Gamma} \frac{1}{f^{(p-1)/d}}$$



- **Convergence of persistence diagrams:**

$$\boxed{\mathrm{dgm}\Big(\mathrm{Rips}\big(\mathbb{X}_n, C(n,p,d)\, d_{\mathbb{X}_n,p}\big)\Big) \xrightarrow[n\to\infty]{B} \mathrm{dgm}\Big(\mathrm{Rips}\big(\mathcal{M}, d_{\mathcal{M},f,p}\big)\Big)}$$

Properties of Fermat distance

- **Convergence of persistence diagrams:**

$$\mathrm{dgm}\Big(\mathrm{Rips}\big(\mathbb{X}_n, C(n,p,d)d_{\mathbb{X}_n,p}\big)\Big) \xrightarrow[n\to\infty]{B} \mathrm{dgm}\Big(\mathrm{Rips}\big(\mathcal{M}, d_{\mathcal{M},f,p}\big)\Big)$$

- **Less sensitiveness to the embedding:**

$$\mathrm{Rips}_\epsilon\big(\mathcal{M}, d_{\mathcal{M},f,p}\big) \simeq \mathcal{M} \qquad \forall \epsilon < \mathrm{conv}(\mathcal{M}, d_{\mathcal{M},f,p})$$

Properties of Fermat distance

- **Convergence of persistence diagrams:**

$$\mathrm{dgm}\Big(\mathrm{Rips}\big(\mathbb{X}_n, C(n,p,d)\, d_{\mathbb{X}_n,p}\big)\Big) \xrightarrow[n\to\infty]{B} \mathrm{dgm}\Big(\mathrm{Rips}\big(\mathcal{M}, d_{\mathcal{M},f,p}\big)\Big)$$

- **Less sensitiveness to the embedding:**

$$\mathrm{Rips}_\epsilon\big(\mathcal{M}, d_{\mathcal{M},f,p}\big) \simeq \mathcal{M} \qquad \forall \epsilon < \mathrm{conv}(\mathcal{M}, d_{\mathcal{M},f,p})$$

- **Robustness to outliers:**

Properties of Fermat distance

- **Convergence of persistence diagrams:**

$$\operatorname{dgm}\Big(\operatorname{Rips}\big(\mathbb{X}_n, C(n,p,d)\, d_{\mathbb{X}_n,p}\big)\Big) \xrightarrow[n\to\infty]{B} \operatorname{dgm}\Big(\operatorname{Rips}\big(\mathcal{M}, d_{\mathcal{M},f,p}\big)\Big)$$

- **Less sensitiveness to the embedding:**

$$\operatorname{Rips}_\epsilon\big(\mathcal{M}, d_{\mathcal{M},f,p}\big) \simeq \mathcal{M} \quad \forall \epsilon < \operatorname{conv}(\mathcal{M}, d_{\mathcal{M},f,p})$$

- **Robustness to outliers:** $\mathbb{X}_n \subseteq \mathcal{M}$ sample, $Y \subseteq \mathbb{R}^D \setminus \mathcal{M}$ outliers.

$$\operatorname{dgm}_k\Big(\operatorname{Rips}_{<\delta^p}\big(\mathbb{X}_n \cup Y, d_{\mathbb{X}_n \cup Y,p}\big)\Big) = \operatorname{dgm}_k\Big(\operatorname{Rips}_{<\delta^p}\big(\mathbb{X}_n, d_{\mathbb{X}_n,p}\big)\Big)$$

for some $\delta > 0$* and all degree $k > 0$.

* Here, for p large enough $\delta^p > \operatorname{diam}(\mathbb{X}_n, d_{\mathbb{X}_n,p})$.

References

- **Preprint:** X. Fernandez, E. Borghini, G. Mindlin, P. Groisman. *Intrinsic persistent homology via density-based metric learning.* arXiv:2012.07621 (2020)
- **Code:** https://github.com/ximenafernandez/intrinsicPH
- **Python library:** fermat.
- **Tutorial:** *Intrinsic persistent homology.*
  YouTube https://www.youtube.com/watch?v=1lP9ndiM60o.

ximena.l.fernandez@durham.ac.uk @pi_ene

References

- **Preprint:** X. Fernandez, E. Borghini, G. Mindlin, P. Groisman. *Intrinsic persistent homology via density-based metric learning.* arXiv:2012.07621 (2020)
- **Code:** https://github.com/ximenafernandez/intrinsicPH
- **Python library:** fermat.
- **Tutorial:** *Intrinsic persistent homology.* https://www.youtube.com/watch?v=1lP9ndiM6Oo.

ximena.l.fernandez@durham.ac.uk @pi_ene

THANKS FOR YOUR ATTENTION!