MATH412/COMPSCI434/MATH713
Fall 2025
Topological Data Analysis

Topic 8: Summary of Persistence Diagrams

Instructor: Ling Zhou

Persistence-based Framework

Input data:
A collection of point sets, graphs, 2D/3D images, arbitrary compacts, etc

→

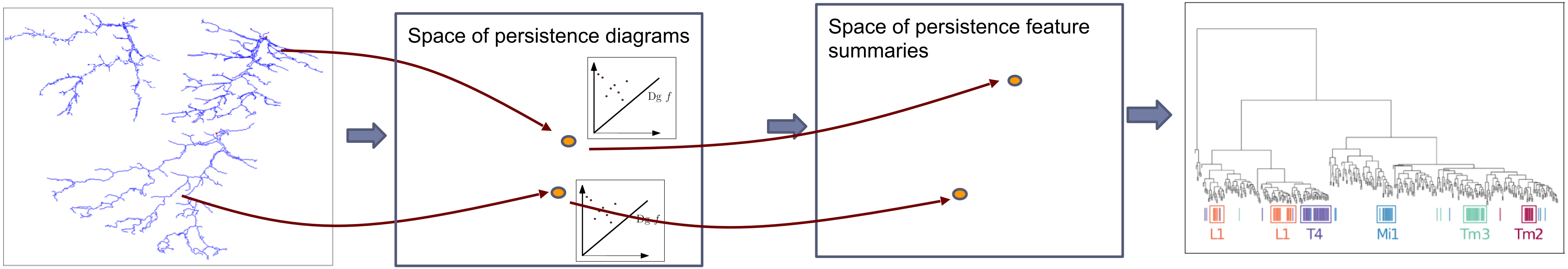
Persistence diagrams:
A collection of persistence diagrams

→

Persistence feature space:
A collection of persistence-based features (e.g, vectors)

→

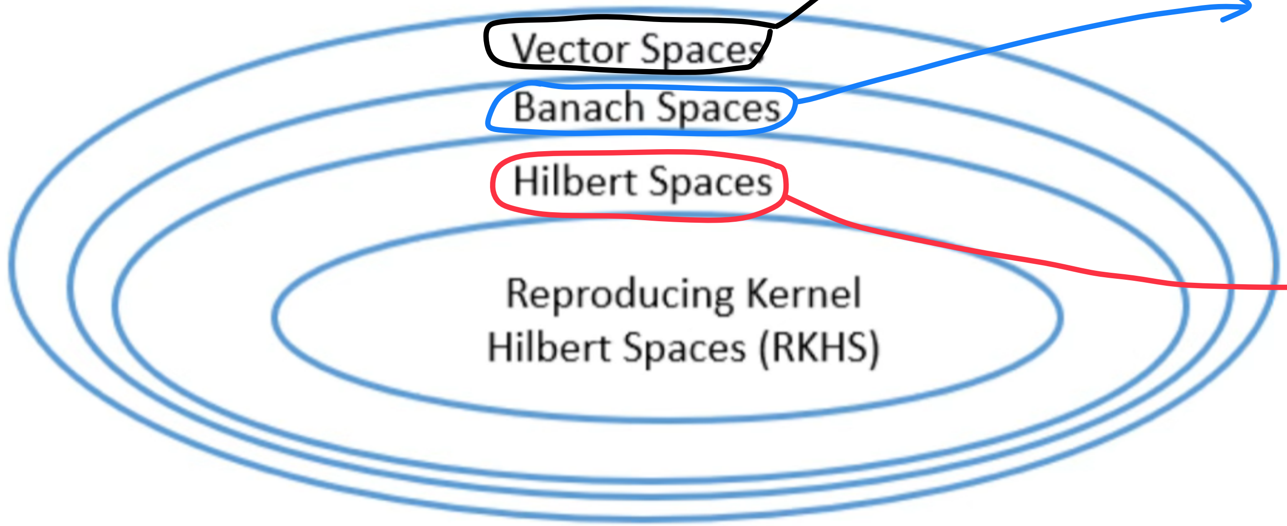
Further data analysis:
Feed to data analysis pipelines, say SVM, CNN, etc



Adapted from [Li, et al, PLOS One 2017]

What is a ‘good’ space?

- Space of persistence diagrams (equipped with bottleneck or Wasserstein distance) is NOT ‘good’
  - does not have nice structure: not linear, no inner product
  - Cannot do statistics



add & scale elements

measure length of elements
(using norm)

measure similarity
(using inner product)

Persistence feature representation

- Persistence landscapes
  - [Bubenik 2012]
- Persistence scale space kernel
  - [Reininghause et al., 2014]
- Persistence images
  - [Adams et al., 2015, 2017]
- Persistence weighted Gaussian kernel
  - [Kusano et al., 2017]
- Sliced Wasserstein kernel
  - [Carriere et al., 2017]
- Persistence Fisher kernel
  - [Le and Yamada 2018]
- ....

- Read: slides

Kernel-based method

Kernels

- Given a topological space X, a kernel is a function $k: X \times X \to R$ such that there exist a Hilbert space $\mathscr{H}$ and a feature map $\Phi: X \to \mathscr{H}$ for which $k\left(x, y\right) = \langle \Phi(x),\ \Phi(y) \rangle_{\mathscr{H}}$.

$$\begin{tikzcd}
X \times X \arrow[r, "k"] \arrow[d, "\Phi \times \Phi"'] & \mathbb{R} \\
\mathcal{H} \times \mathcal{H} \arrow[ur, "\langle\,,\,\rangle"'] &
\end{tikzcd}$$

Kernels

- In practice, we often use positive semi-definite kernels:

$\forall x_1, \cdots, x_n \in X$ & $a_1, \cdots, a_n \in \mathbb{R}$ s.t. $\sum_i a_i = 0$, we have

$$\sum_{i,j} a_i a_j k(x_i, x_j) \geq 0$$

(negative semi-definite if

$\forall x_1, \cdots, x_n \in X$ & $a_1, \cdots, a_n \in \mathbb{R}$ s.t. $\sum_i a_i = 0$, we have

$$\sum_{i,j} a_i a_j k(x_i, x_j) \leq 0$$
)

- A positive semi-definite kernel induces a (pseudo-)metric:
  - $d_k^2(x,\ y) = \left\langle x - y, x - y \right\rangle = k(x,\ x) + k\left(y,\ y\right)\ - 2k\left(x, y\right)$

Examples of kernel-based methods

- [Reininghaus et al. 2015]: Persistence scale space kernel (PSSK)

$$k_\sigma(D,E) = \langle \Phi_\sigma(D), \Phi_\sigma(E)\rangle_{\mathcal{L}^2(\Omega)} = \frac{1}{8\pi\sigma} \sum_{y\in D;\ z\in E} [e^{-\frac{\|y-z\|^2}{8\sigma}} - e^{-\frac{\|y-\bar{z}\|^2}{8\sigma}}].$$

- [Kusano et al. 2017]: Persistence weighted Gaussian kernel (PWGK)

$$k_G^{\omega_{arc}}(x,y) = \omega_{arc}(x)\omega_{arc}(y)k_G(x,y)$$

$$\omega_{arc}(x) = \arctan\left(C \cdot pers(x)^p\right) \qquad k_G(x,y) = e^{-\frac{||x-y||^2}{2\tau^2}}$$

- [Carriere et al., 2017]: Sliced Wasserstein kernel (SW)

$$k_{SW}(D,\ E) := \ e^{-\frac{d_{SW}(D,\ E)}{2\tau^2}}$$

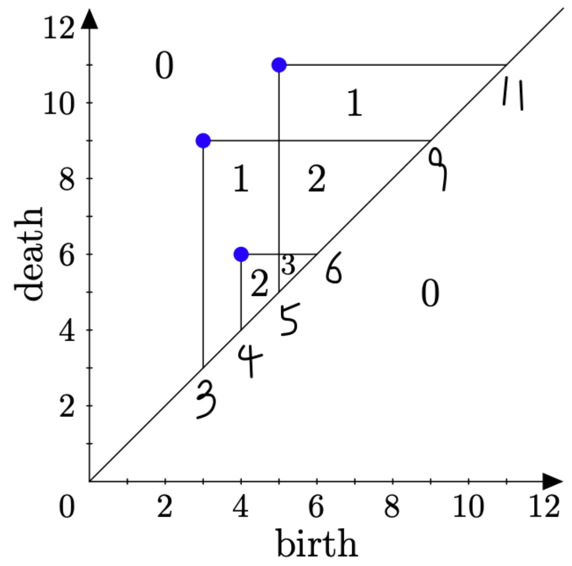
- See comparison in the paper

Persistence landscape

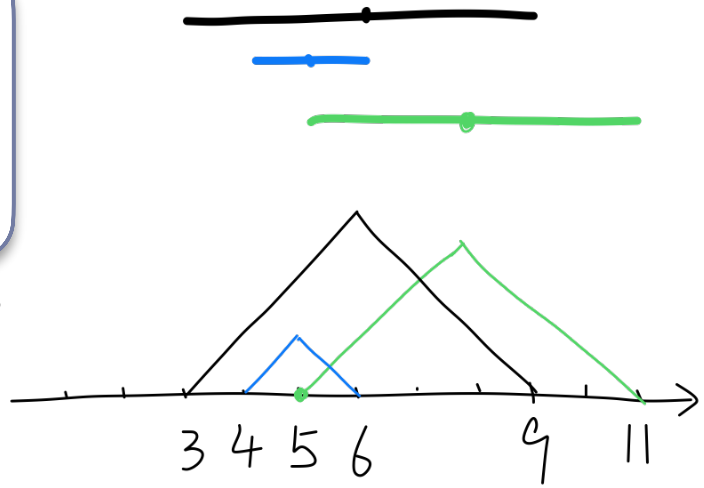
Peter Bubenik PETER.BUBENIK@GMAIL.COM

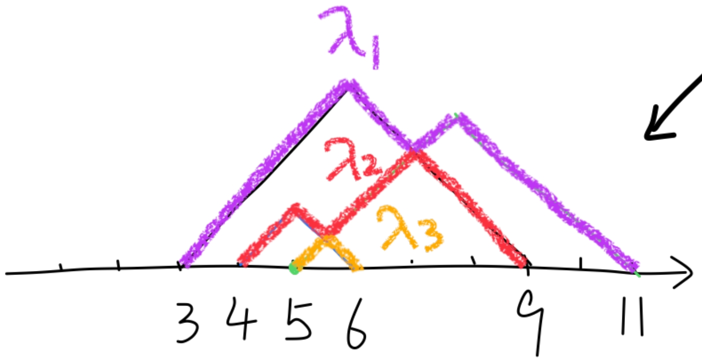
Persistence landscapes

- Map persistence diagrams to a function space [Bubenik 2012]



make each interval [b,d] into a 'tent' whose
- top lies at (d+b)/2 and
- height is (d-b)/2





reorganize the tents into 'layers' based on the height:
- highest
- second highest
- etc

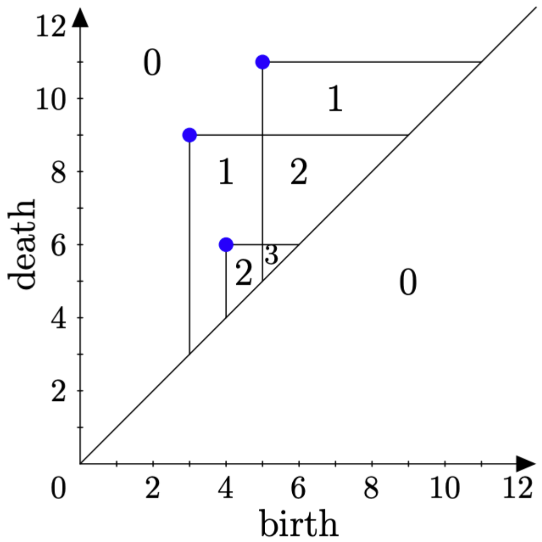
Persistence landscapes

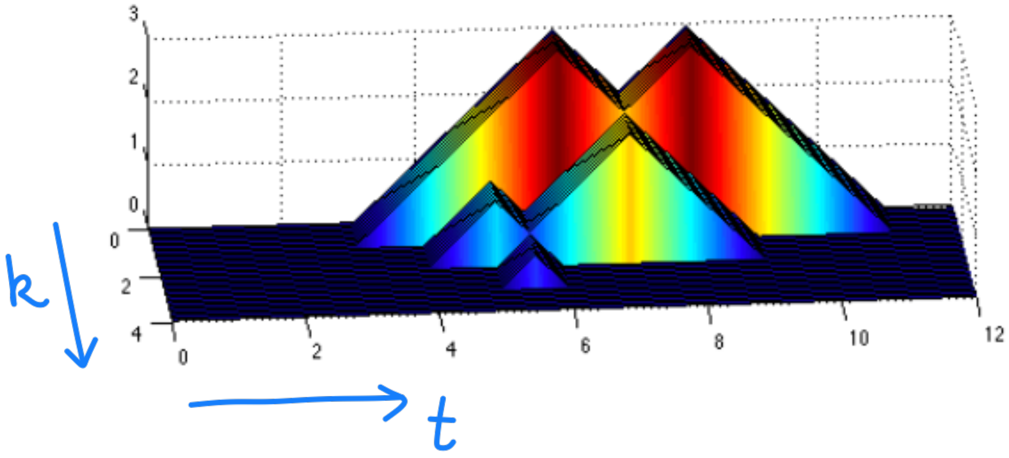
- Map persistence diagrams to a function space [Bubenik 2012]

Definition 13.1 (Persistence landscape). Given a finite persistence diagram $D = \{(b_i, d_i)\}_{i\in[1,n]}$ from $\mathbb{D}$, the *persistence landscape w.r.t.* D is a function $\lambda_D : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ where

$$\lambda_D(k,t) := k\text{-th largest value of } [\min\{t-b_i, d_i-t\}]_+ \text{ for } i \in [1,n].$$

Here, $[c]_+ = max(c, 0)$.





Metric and properties for persistence landscapes

- p-norm: $$\|\lambda_D\|_p^p = \sum_{k=1}^{\infty} \|\lambda_D(k,\cdot)\|_p^p.$$

(PD, Λ_p) is Banach $(1 \leq p < \infty)$
It is Hilbert when $p=2$

- p-landscape distance:

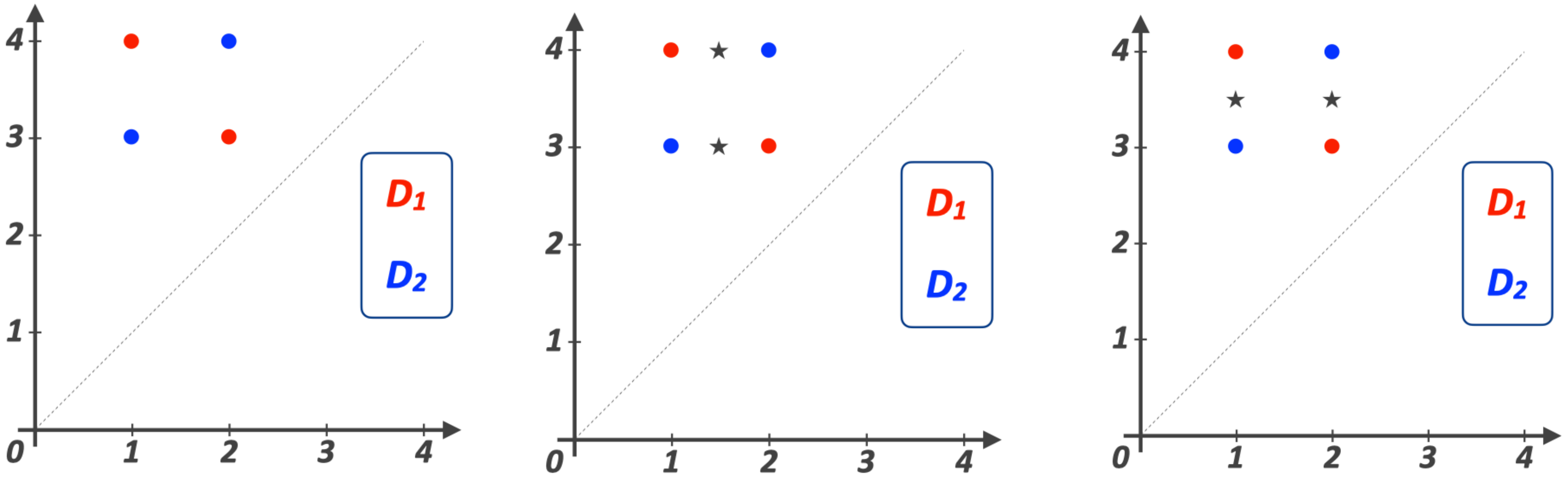
$$\Lambda_p(D_1, D_2) = \|\lambda_{D_1} - \lambda_{D_2}\|_p$$

Claim *Given a persistence diagram D, let λ_D be its persistence landscape. Then from λ_D one can uniquely recover the persistence diagram D.*

Theorem *For persistence diagrams D and D', $\Lambda_\infty(D, D') \leq d_B(D, D')$.*

Statistics for TDA

- There is no well-defined notion of mean for persistence diagrams



Courtesy of Fugacci

- Mean landscape of $\lambda_{D_1}, \lambda_{D_2}, \ldots, \lambda_{D_\ell}$: $\overline{\lambda}(k,t) = \frac{1}{n}\sum_{i=1}^{\ell} \lambda_{D_i}(k,t).$

Average of persistence landscapes

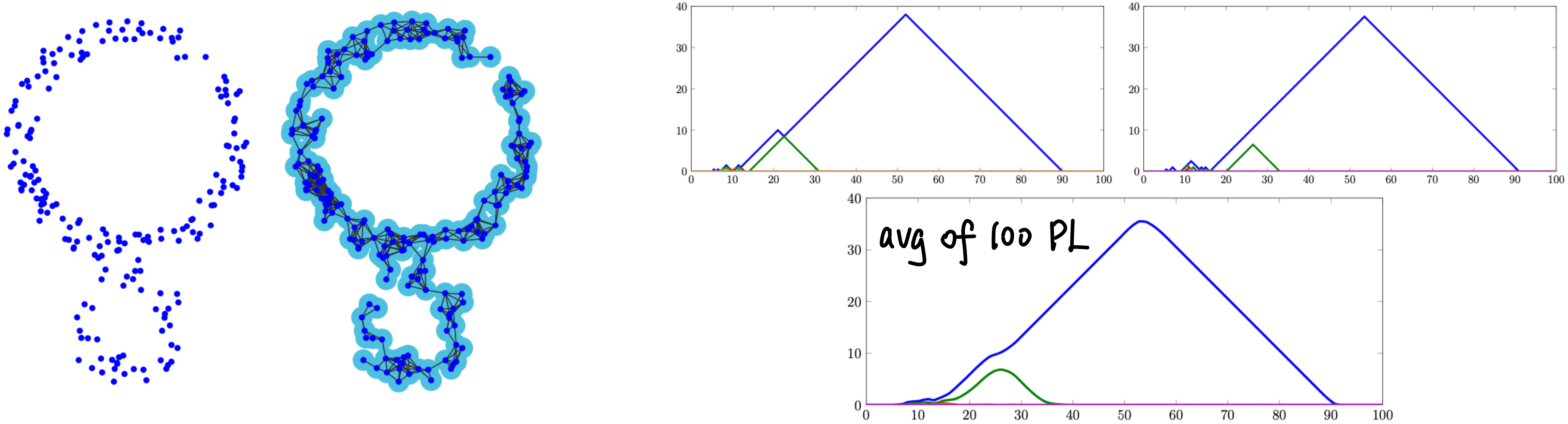


Figure 4: 200 points were sampled from a pair of linked annuli. Here we show the points and a corresponding union of balls and 1-skeleton of the Čech complex. This was repeated 100 times. Next we show two of the degree one persistence landscapes and the mean degree one persistence landscape.

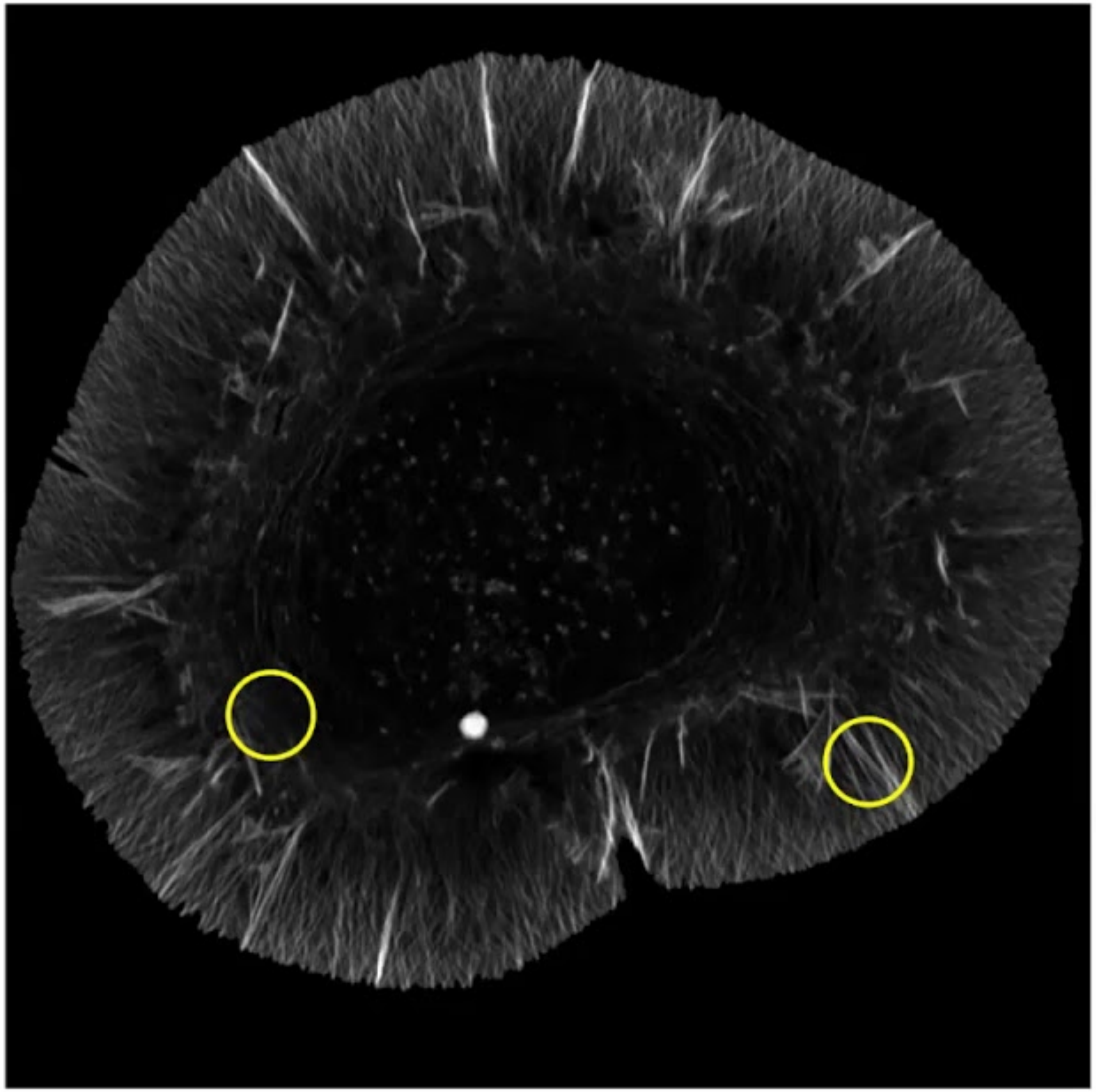
Vectorization of persistence landscapes

- Choose a maximum level k_{max}
- Sample $\lambda_k(t)$ on fixed gird $t_1, \ldots, t_m$ for each $k = 1,\ldots, k_{\text{max}}$
- Flatten sampled value into a vector

$$v(D) = (\lambda_1(t_1), \ldots, \lambda_1(t_m), \lambda_2(t_1), \ldots, \lambda_{k_{\max}}(t_m)) \in \mathbb{R}^{k_{\max} m}.$$

- Read: *the next slide contains a video by Prof. Peter Bubenik who introduced persistence landscapes. The video is about the application of persistence landscapes in biological image.*

e biological images



Persistence image

Persistence Images

B $T(B)$ ρ_B

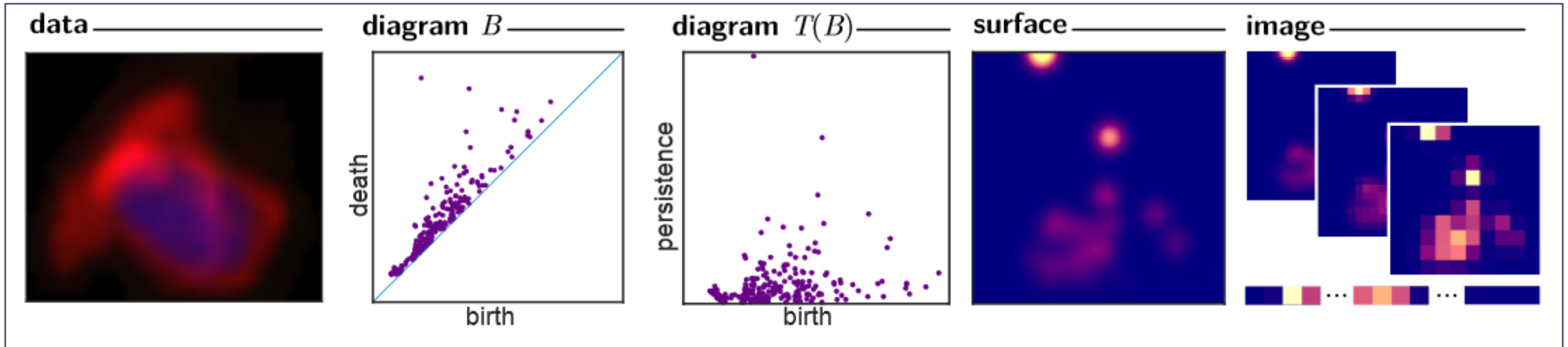


Figure 1: Algorithm pipeline to transform data into a persistence image.

Image taken from [Adams et al., 2017]

- Given a persistence diagram B
  - $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T(x, y) = (x, y - x)$
  - **Persistence surface** $\rho_B : R^2 \to R$ is defined as $\rho_B(z) = \sum_{u \in T(B)} f(u) \cdot \phi_u(z)$

e.g, $\phi_u(z) = \frac{1}{2\pi\sigma^2} e^{-\frac{|z-u|^2}{2\sigma^2}}$

weight function

$T : (\text{birth}, \text{death}) \longmapsto (\text{birth}, \text{lifetime})$

Persistence Images

B $T(B)$ ρ_B I_B

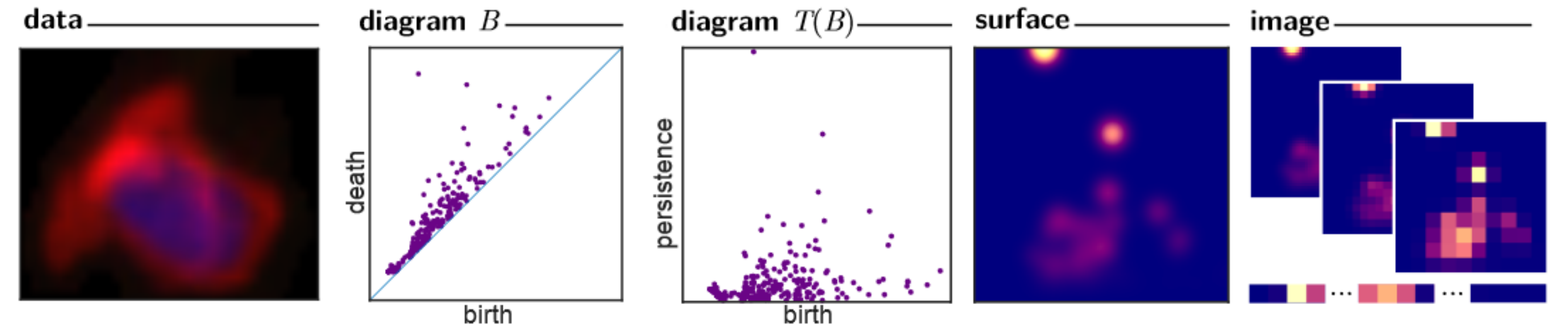


Figure 1: Algorithm pipeline to transform data into a persistence image.

Image taken from [Adams et al., 2017]

- Given a persistence diagram B
  - $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T(x, y) = (x, y - x)$
  - **Persistence surface** ρ_B: $R^2 \to R$ is defined as $\rho_B(z) = \sum_{u \in T(B)} f(u) \cdot \phi_u(z)$
  - **Persistence image** I_B is a discretization of ρ_B: $I_B[p] = \int_p \rho_B dxdy$

Persistent image: vectorization and stability

- Vectorization: flatten the 2D image
- Metric: We can simply use p-norms of vector difference $\|I_D - I_E\|_p$

Theorem 13.3. *Suppose persistence images are computed with the normalized Gaussian distribution with variance σ^2 and weight function $\omega : \mathbb{R}^2 \to \mathbb{R}$. Then the persistence images are stable w.r.t. the 1-Wasserstein distance between persistence diagrams. More precisely, given two finite and bounded persistence diagrams D and E, we have:*

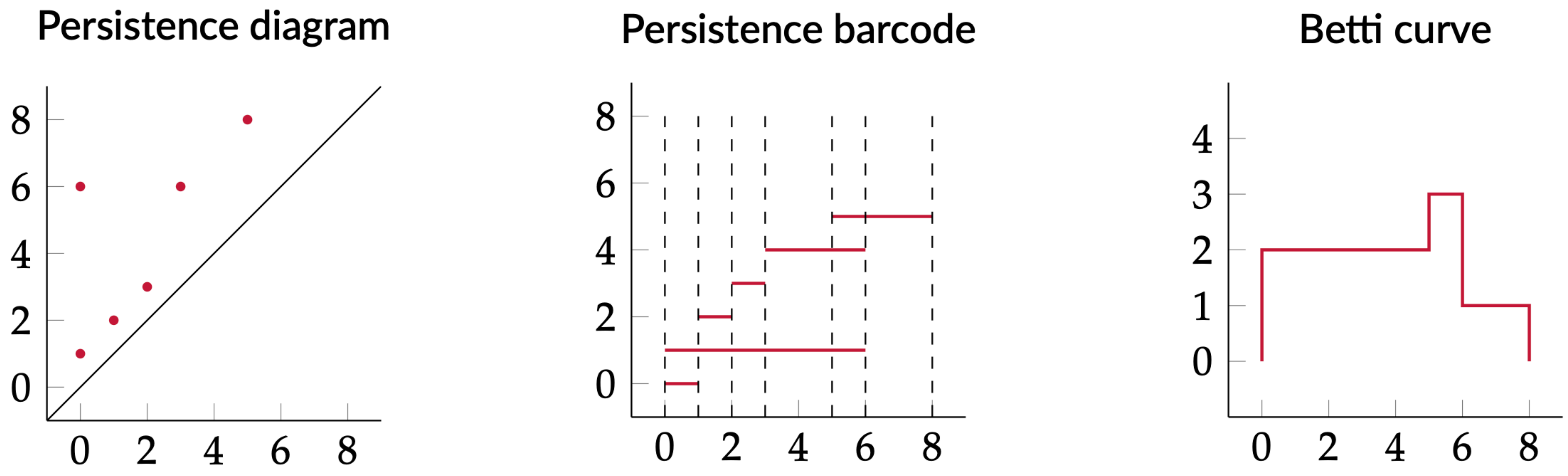
$$\| \mathrm{I}_D - \mathrm{I}_E\|_1 \leq \left(\sqrt{5}|\nabla\omega| + \sqrt{\frac{10}{\pi}} \frac{\|\omega\|_\infty}{\sigma} \right) \cdot \mathrm{d}_{W,1}(D, E).$$

Here, $\nabla\omega$ stands for the gradient of ω, and $|\nabla\omega| = \sup_{z\in\mathbb{R}^2} \|\nabla\omega\|_2$ is the maximum norm of the gradient vector of ω at any point in $\mathbb{R}^2$. The same upper bound holds for $\| \mathrm{I}_D - \mathrm{I}_E\|_2$ and $\| \mathrm{I}_D - \mathrm{I}_E\|_\infty$ as well.

Other methods

Betti curve

$$\beta_k(t) = \dim(H_k(X_t))$$



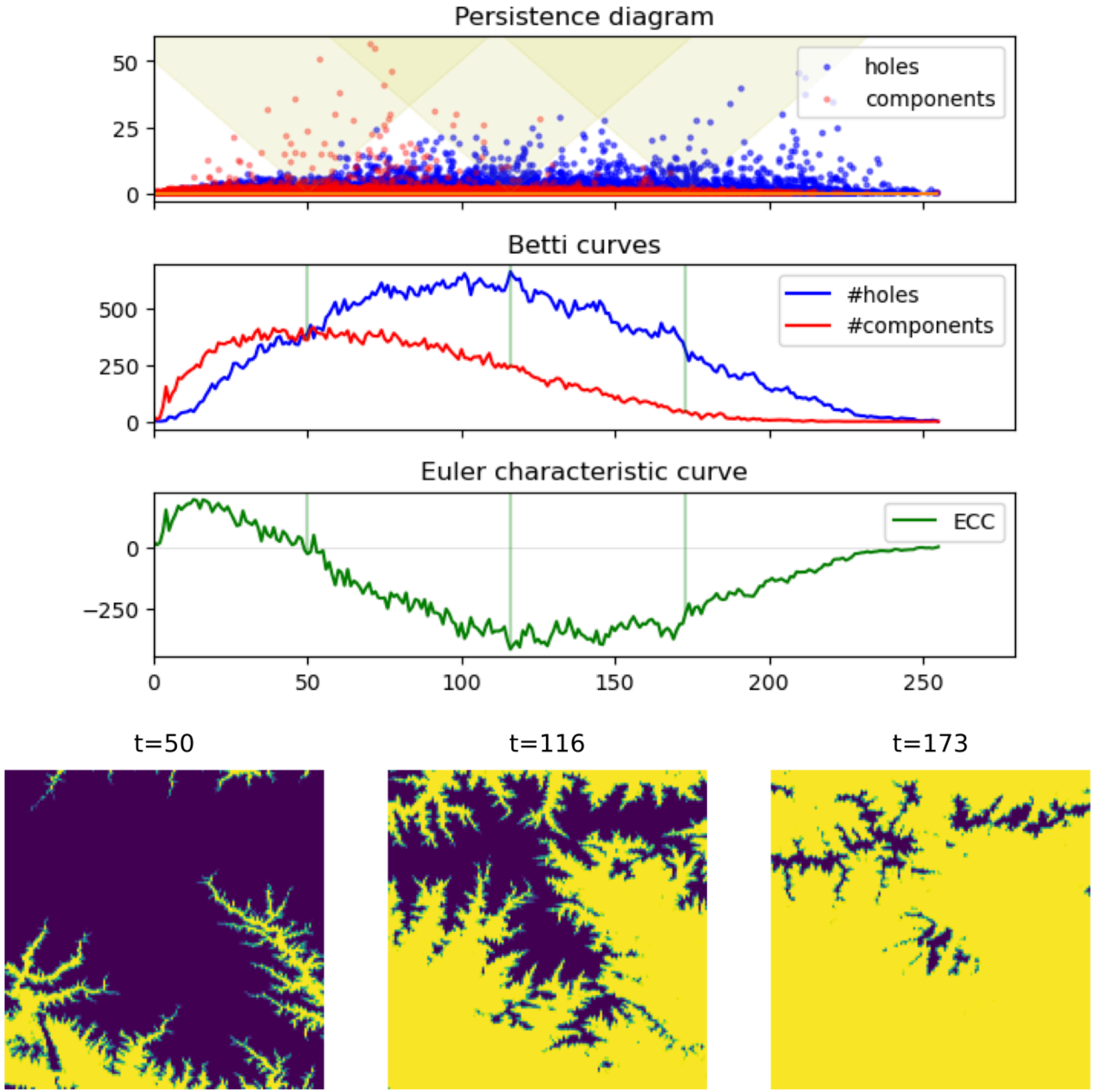
Courtesy of Bastian Rieck

Euler characteristic curve

- [Wang et al. 2023]
- $\chi(X) = \sum (-1)^i \beta_i(X)$
- cubical filtration
- GPU implementation

GPU Computation of the Euler Characteristic Curve for Imaging Data

Fan Wang
Stony Brook University, Stony Brook, US
Hubert Wagner
University of Florida, Gainesville, US
Chao Chen
Stony Brook University, Stony Brook, US



Persistence Entropy

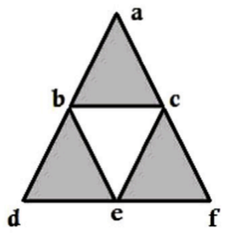
An entropy-based persistence barcode ☆

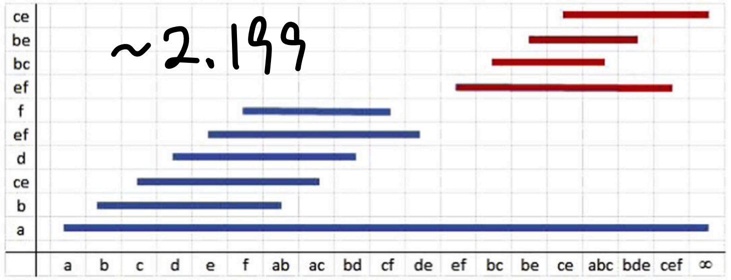
Harish Chintakunta [a], Thanos Gentimis [a], Rocio Gonzalez-Diaz [b,*], Maria-Jose Jimenez [b], Hamid Krim [a]

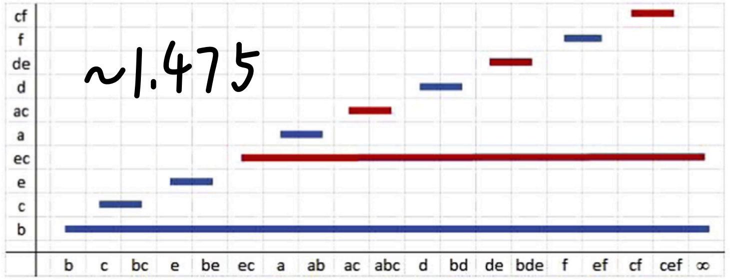
- [Chintakunta et al, 2015]

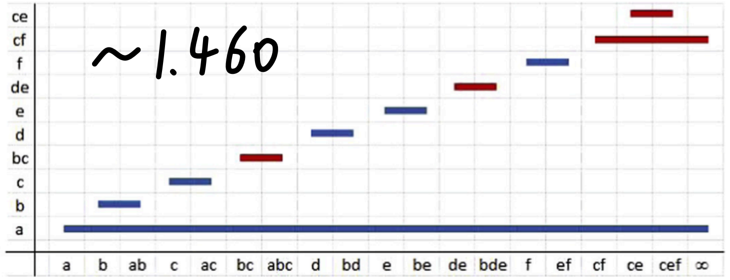
$$E(D) = -\sum_{i \in I} p_i \log(p_i)$$

$$p_i = \frac{(d_i - b_i)}{L_D} \quad \text{and} \quad L_D = \sum_{i \in I} (d_i - b_i).$$









- Entropy measures the 'variation' of bar lengths
- Barcode with more uniform lengths has smaller entropy
- On the left, blue bars are for dim-0 and red ones for dim-1

Packages and group work

- scikit-tda/persim persistence landscape
- scikit-tda/persim persistence image
- scikit-tda/persim persistent entropy

- Practice using persistence landscape + PCA + SVC to do classification here: https://colab.research.google.com/drive/1cZAZp4sGrfkicDF4220XqQEduqE2Ez5f?usp=sharing