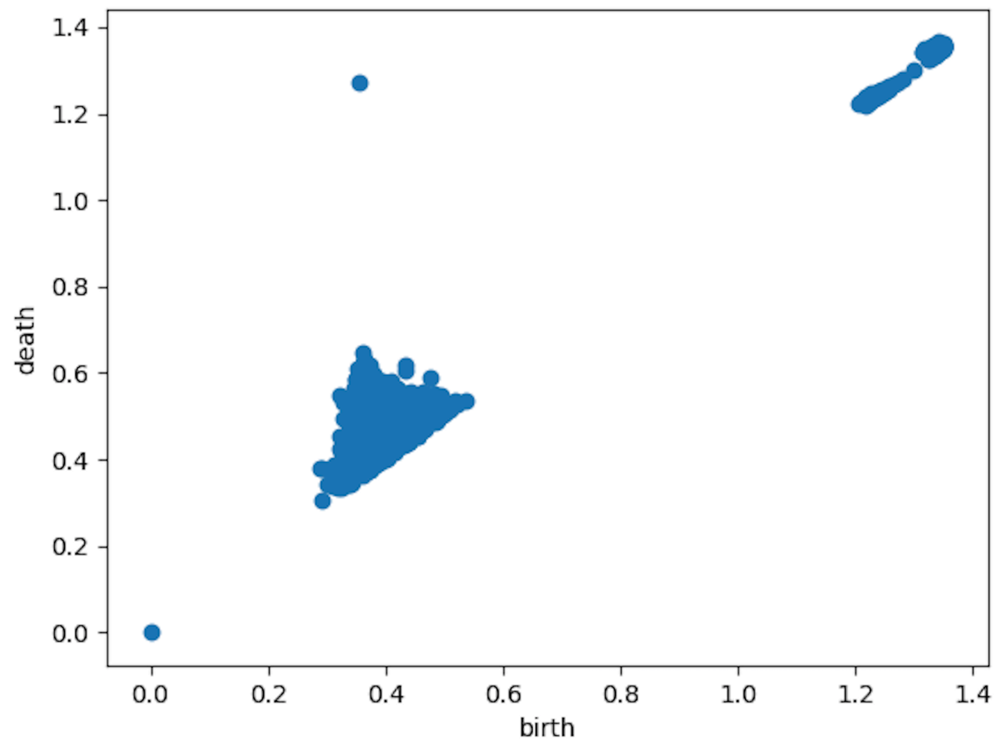
Survey: From persistent homology to machine learning feature vectors

Henry Adams
Colorado State University
ICERM TRIPODS Bootcamp on
"Topology and Machine Learning"
2018



Bottleneck distance

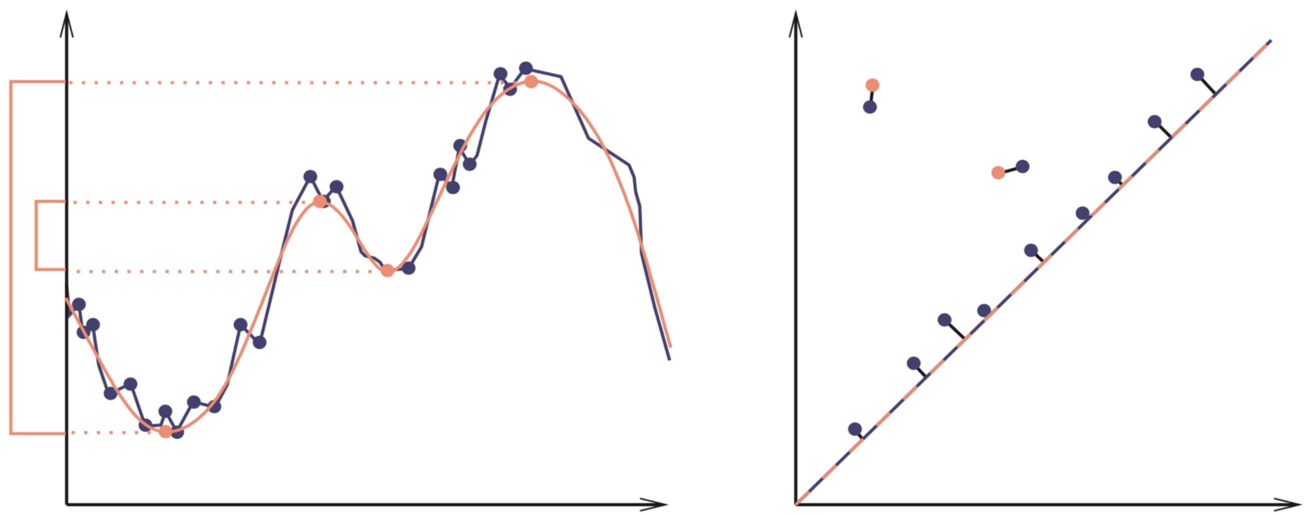
Discrete Comput Geom 37:103–120 (2007)
DOI: 10.1007/s00454-006-1276-5

Discrete & Computational
Geometry

Stability of Persistence Diagrams*

2007

David Cohen-Steiner,[1] Herbert Edelsbrunner,[2] and John Harer[3]



(a) (b)

Fig. 2. (a) Two close functions, one with many and the other with just four critical values. (b) The persistence diagrams of the two functions, and the bijection between them.

Wasserstein distance

Probability measures on the space of persistence diagrams

2011

Yuriy Mileyko[1], Sayan Mukherjee[2] and John Harer[1]

[1] Departments of Mathematics and Computer Science, Center for Systems Biology, Duke University, 27708, USA

[2] Departments of Statistical Science, Computer Science, and Mathematics, Institute for Genome Sciences & Policy, Duke University, 27708, USA

E-mail: yury@math.duke.edu, sayan@stat.duke.edu, john.harer@duke.edu

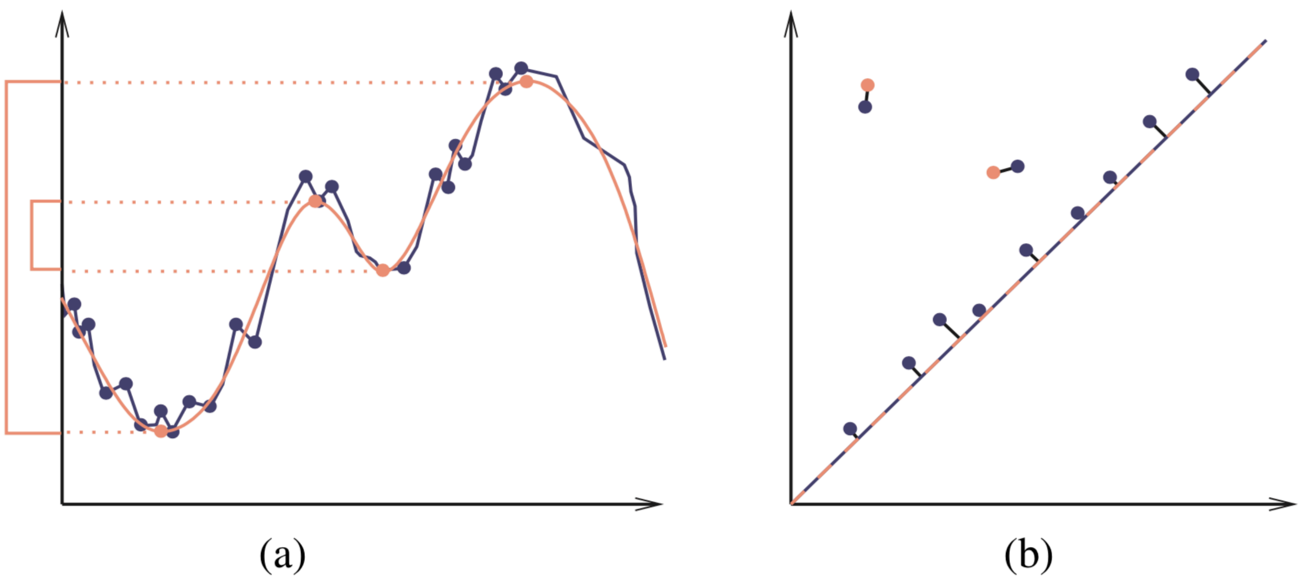


Fig. 2. (a) Two close functions, one with many and the other with just four critical values. (b) The persistence diagrams of the two functions, and the bijection between them.

Computation time

Geometry Helps to Compare Persistence Diagrams

Michael Kerber * Dmitriy Morozov † Arnur Nigmetov ‡

2017

Distance Matrix	Accuracy (Noise 0.05)	Time (Noise 0.05)	Accuracy (Noise 0.1)	Time (Noise 0.1)
PD, H_0, L^1	96.0%	37346s	96.0%	42613s
PD, H_0, L^2	91.3%	24656s	91.3%	25138s
PD, H_0, L^∞	60.7%	1133s	63.3%	1149s
PD, H_1, L^1	100%	657s	96.0%	703s
PD, H_1, L^2	100%	984s	97.3%	1042s
PD, H_1, L^∞	81.3%	527s	66.7%	564s

Possible goals

Problem Statement: How can we represent a persistence diagram so that:

(i) the output of the representation is a vector in $\mathbb{R}^n$, (kernels are also very useful!)

(ii) the representation is stable with respect to input noise,

(iii) the representation is efficient to compute,

(iv) the representation maintains an interpretable connection to the original PD, and

(v) the representation allows one to adjust the relative importance of points in different regions of the PD?

Persistence Landscapes

Journal of Machine Learning Research 16 (2015) 77-102 Submitted 7/14; Published 1/15

2015

Statistical Topological Data Analysis using Persistence Landscapes

Peter Bubenik PETER.BUBENIK@GMAIL.COM

Computation time is much faster!

Distance Matrix	Accuracy (Noise 0.05)	Time (Noise 0.05)	Accuracy (Noise 0.1)	Time (Noise 0.1)
PL, H_0, L^1	92.7%	29s	96.7%	33s
PL, H_0, L^2	77.3%	29s	82.0%	34s
PL, H_0, L^∞	60.7%	2s	63.3%	2s
PL, H_1, L^1	83.3%	36s	80.7%	48s
PL, H_1, L^2	83.3%	50s	66.7%	69s
PL, H_1, L^∞	74.7%	8s	66.7%	9s

Persistence Landscapes

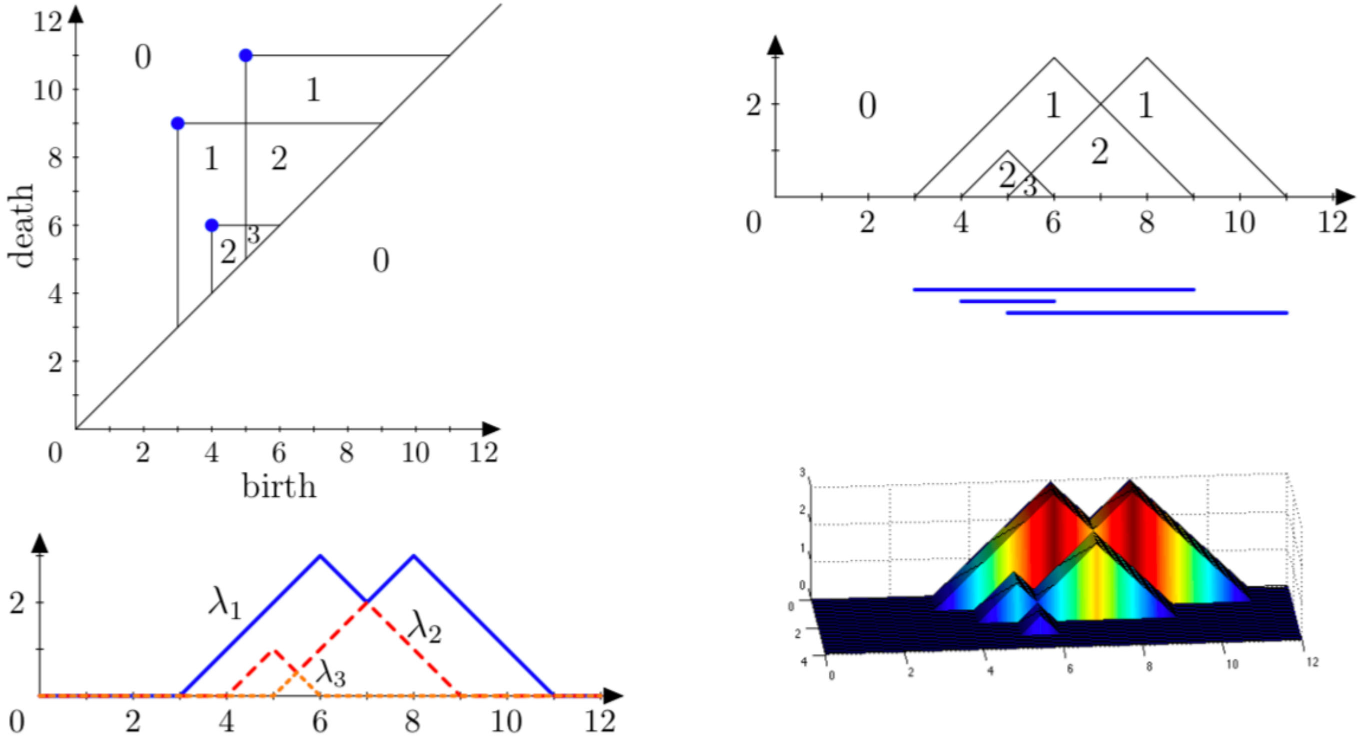


Figure 2: Persistence landscapes for the homology in degree 1 of the example in Figure 1. For the rank function (top left) and rescaled rank function (top right) the values of the functions on the corresponding region are given. The top left graph also contains the three points of the corresponding persistence diagram. Below the top right graph is the corresponding barcode. We also have the corresponding persistence landscape (bottom left) and its 3d-version (bottom right). Notice that λ_1 gives a measure of the dominant homological feature at each point of the filtration.

Persistence Landscapes

Means are unique!

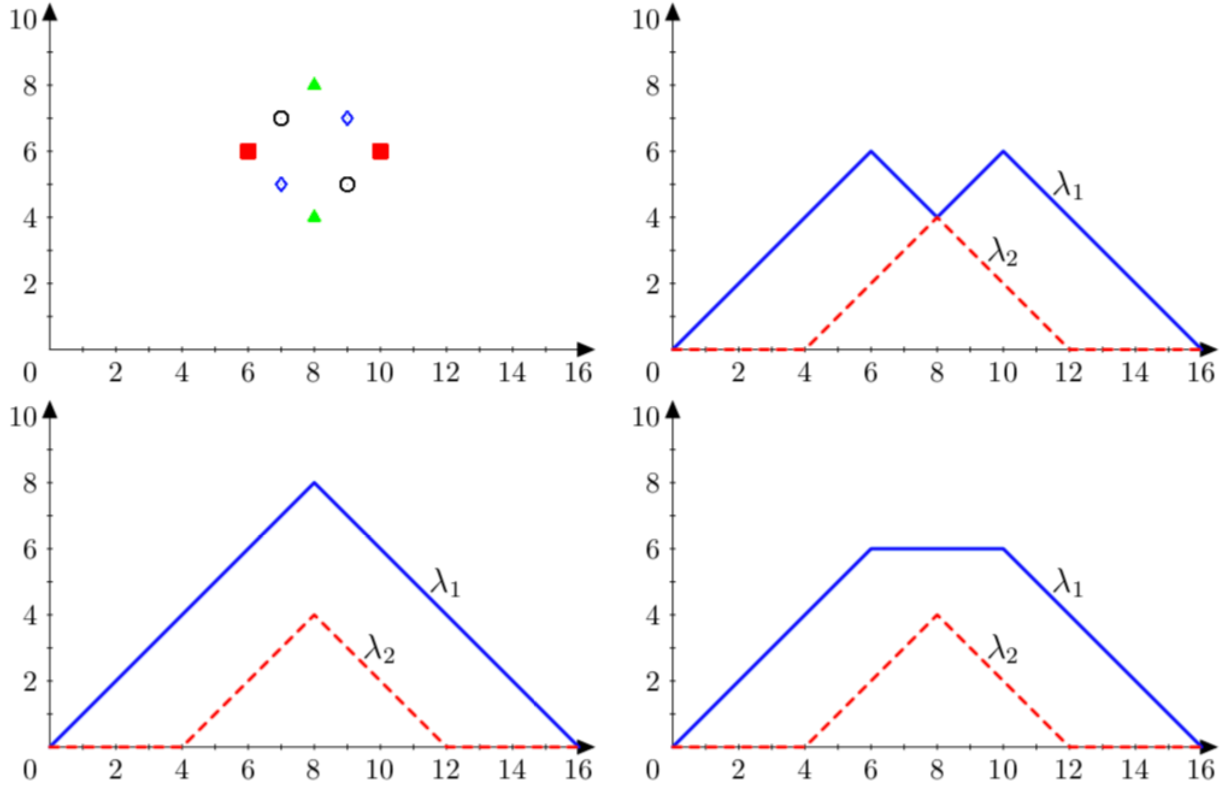


Figure 3: Means of persistence diagrams and persistence landscapes. Top left: the rescaled persistence diagrams $\{(6,6),(10,6)\}$ and $\{(8,4),(8,8)\}$ have two (Fréchet) means: $\{(7,5),(9,7)\}$ and $\{(7,7),(9,5)\}$. In contrast their corresponding persistence landscapes (top right and bottom left) have a unique mean (bottom right).

Persistence Landscapes

ELSEVIER

Journal of Symbolic Computation

Volume 78, January–February 2017, Pages 91-114

A persistence landscapes toolbox for topological statistics ☆

Peter Bubenik [a] ✉, Paweł Dłotko [b, c] ✉

⊞ **Show more**

https://doi.org/10.1016/j.jsc.2016.03.009

Get rights and content

2017

Computations: The Persistence Landscapes Toolbox:
https://www.math.upenn.edu/~dlotko/persistenceLandscape.html

Statistical Analysis of Persistence Intensity Functions

Yen-Chi Chen **Daren Wang** **Alessandro Rinaldo** **Larry Wasserman**
Carnegie Mellon University
Department of Statistics

2015

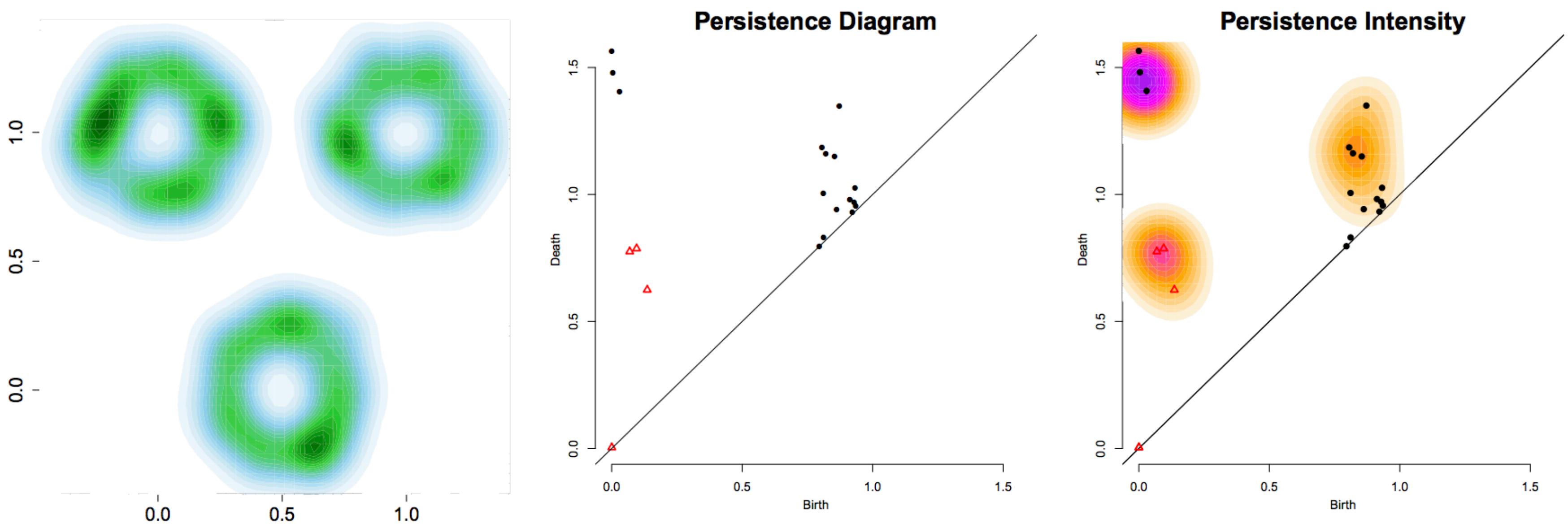
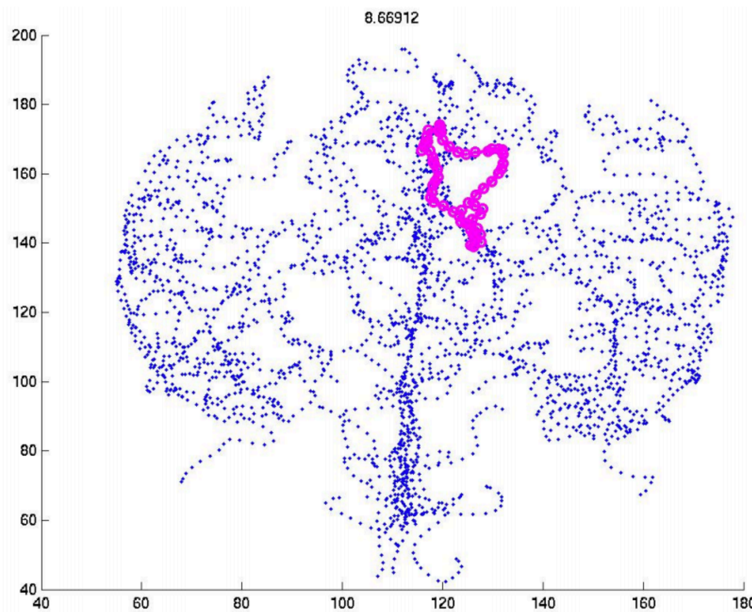


Figure 1: An example of a persistence diagram and the smoothed persistence intensity estimator constructed from a density estimator. Left: the density estimator. Middle: the persistence diagram. Each black dot is a 0-dimensional topological feature and each red triangle is a 1-dimensional topological feature. Right: the smoothed persistence intensity estimator. Note that in this case we only use the topological feature of dimension 0 to compute the intensities. i.e. the connected components.

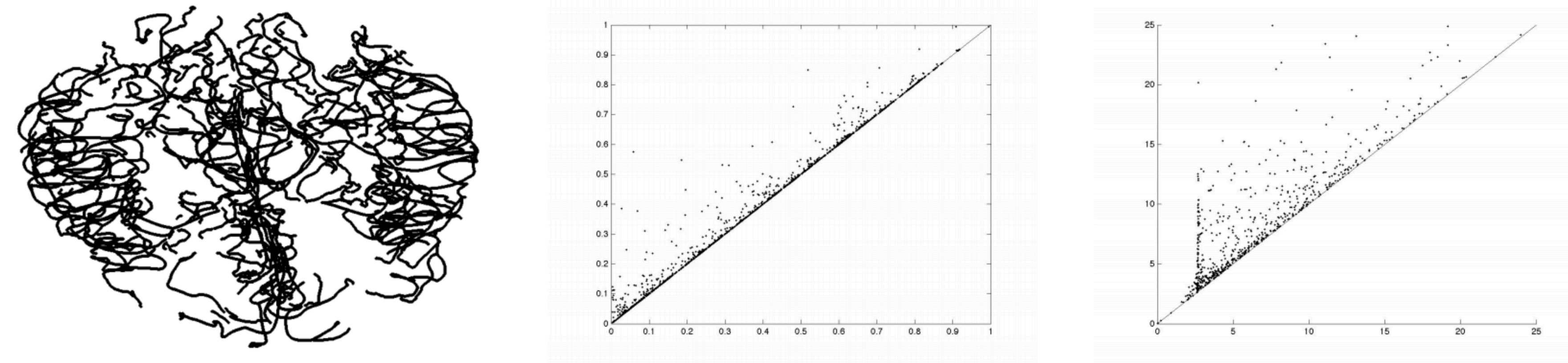
- Topological features can be added to improve state-of-the-art classifiers
- The non-persistent features also matter!



Persistent homology analysis of brain artery trees

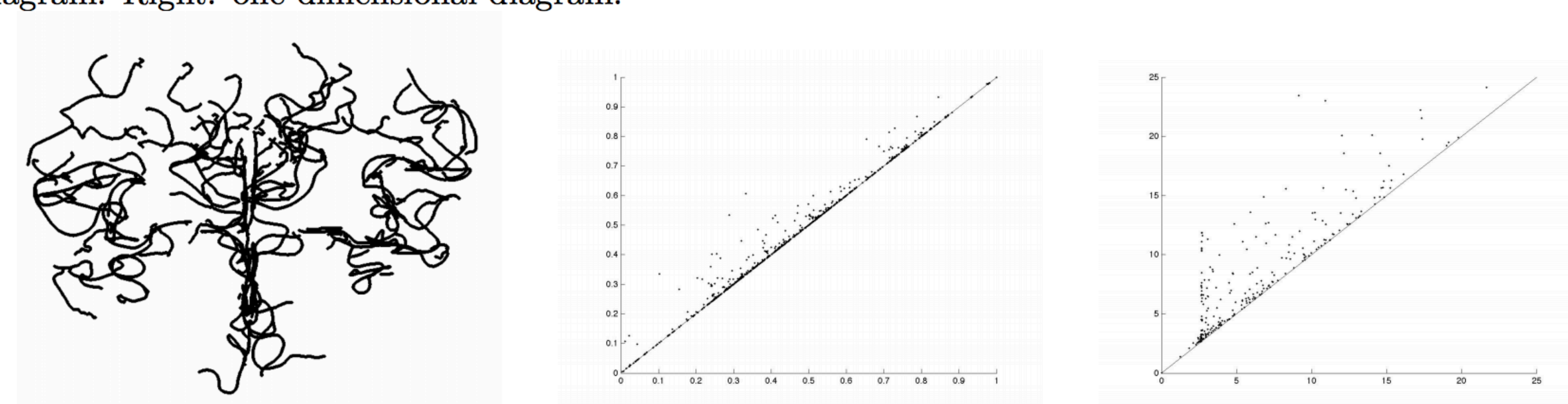
Paul Bendich,* J.S. Marron,† Ezra Miller,* Alex Pieloch,* and Sean Skwerer†

23 November 2014



(a) Brain tree (b) Dgm_0 (c) Dgm_1

Figure 13: Persistent homology data objects from a 24-year old. Left: brain tree. Middle: zero-dimensional diagram. Right: one-dimensional diagram.



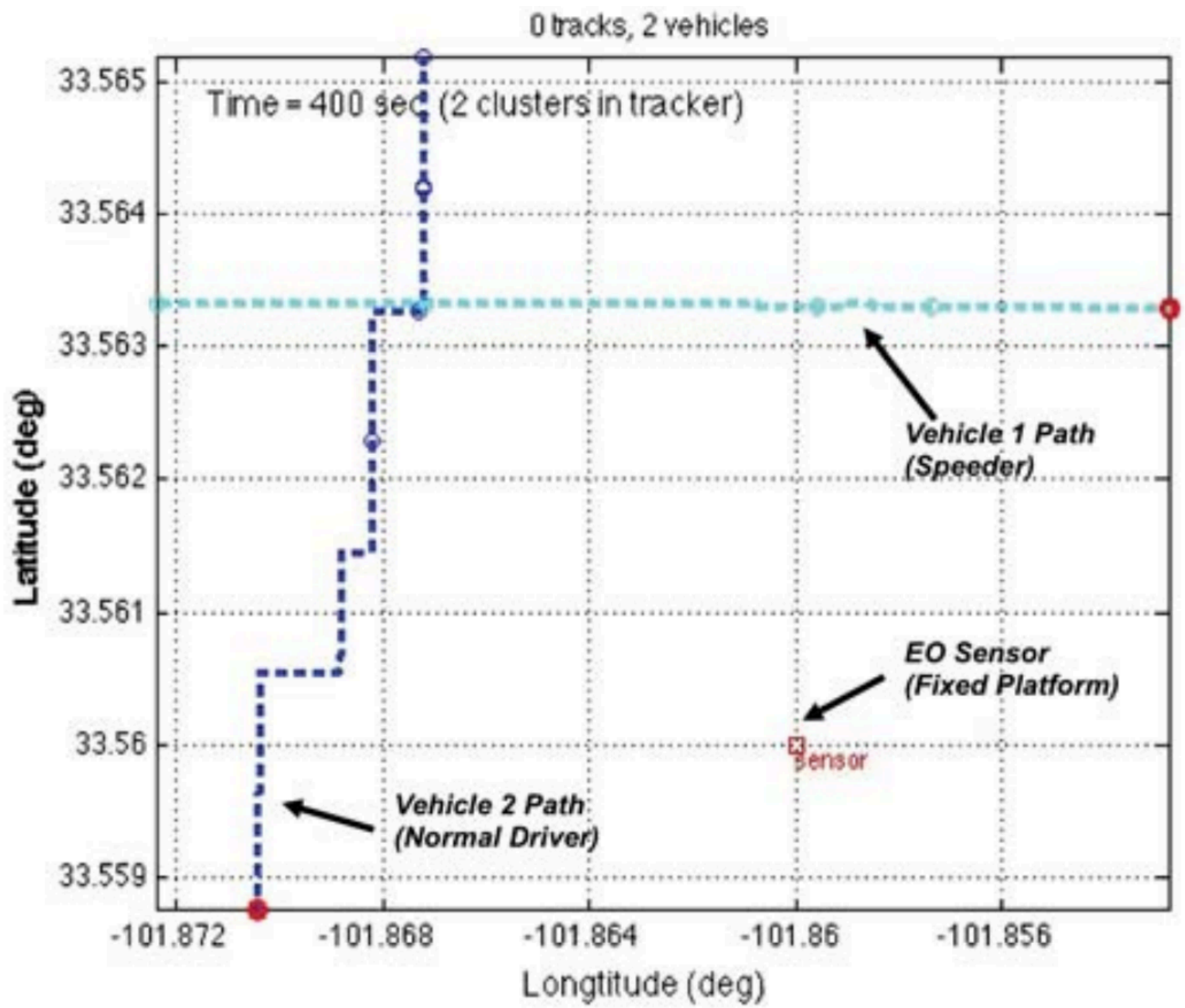
(a) Brain tree (b) Dgm_0 (c) Dgm_1

Figure 14: Persistent homology data objects from a 68-year old. Left: brain tree. Middle: zero-dimensional diagram. Right: one-dimensional diagram.

Topological and Statistical Behavior Classifiers for Tracking Applications

Paul Bendich*, Sang Chin†‡§, Jesse Clarke†, Jonathan DeSena†, John Harer *, Elizabeth Munch *, Andrew Newman †, David Porter †, David Rouse †, Nate Strawn *†, Adam Watkins †

2016



A Stable Multi-Scale Kernel for Topological Machine Learning

Jan Reininghaus, Stefan Huber
IST Austria

Ulrich Bauer
IST Austria, TU München

Roland Kwitt
University of Salzburg, Austria

2015

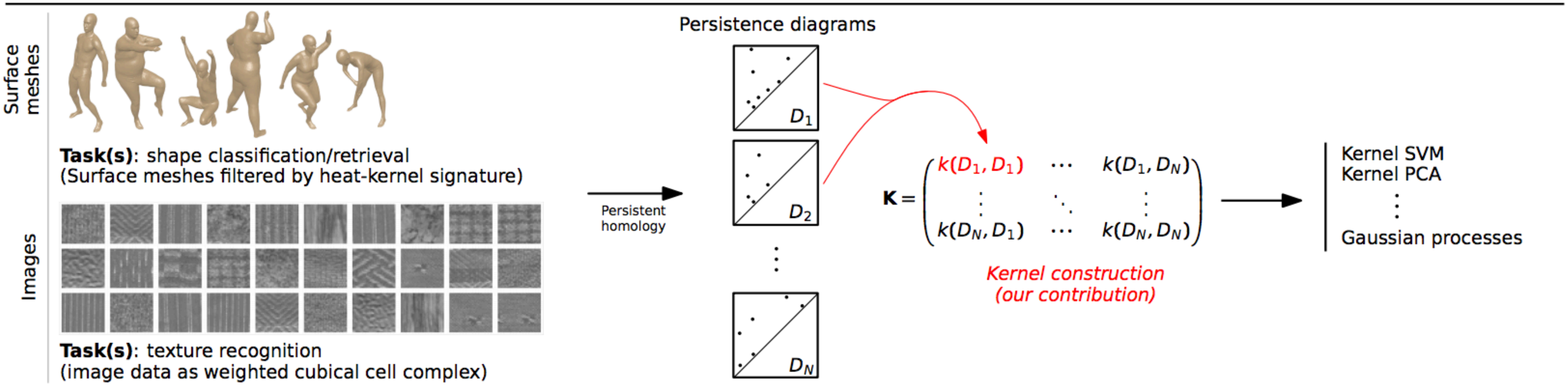


Figure 1: Visual data (*e.g.*, functions on surface meshes, textures, etc.) is analyzed using persistent homology [11]. Roughly speaking, persistent homology captures the birth/death times of topological features (*e.g.*, connected components or holes) in the form of *persistence diagrams*. Our contribution is to define a *kernel for persistence diagrams* to enable a theoretically sound use of these summary representations in the framework of kernel-based learning techniques, popular in the computer vision community.

EUROGRAPHICS 2015 / Mirela Ben-Chen and Ligang Liu
(Guest Editors)

Volume 34 (2015), Number 5

Stable Topological Signatures for Points on 3D Shapes

Mathieu Carrière[1] Steve Y. Oudot[1] Maks Ovsjanikov[2]

[1]INRIA Saclay [2]LIX, École Polytechnique

2015

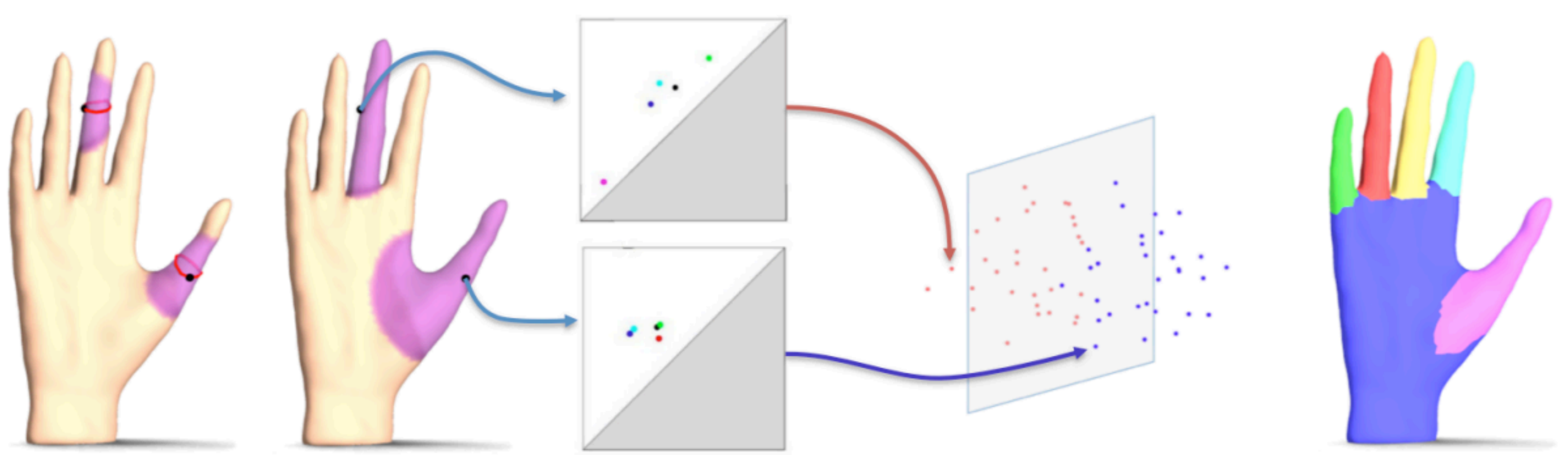


Figure 1: *Our signatures characterize a point x by the birth (leftmost) and death (second left) of the topological features (here, non-trivial loops) in the neighborhood of x in a provably stable way. We show how to compactly encode these events in a vector without losing the stability properties. This is useful in a variety of contexts, including shape segmentation and labeling (as shown on the right) using linear classifiers such as SVM.*

On the stability of persistent entropy and new summary functions for Topological Data Analysis

N. Atienza, R. Gonzalez Diaz, M. Soriano Trigueros

June 7, 2018

Definition 2.12 (Persistent entropy)**.** Consider a persistence barcode $A = \{[x_i^a, y_i^a]\}_{i=1}^{n_a}$ where $\max_i\{y_i\} < \infty$. Persistent entropy of A is:

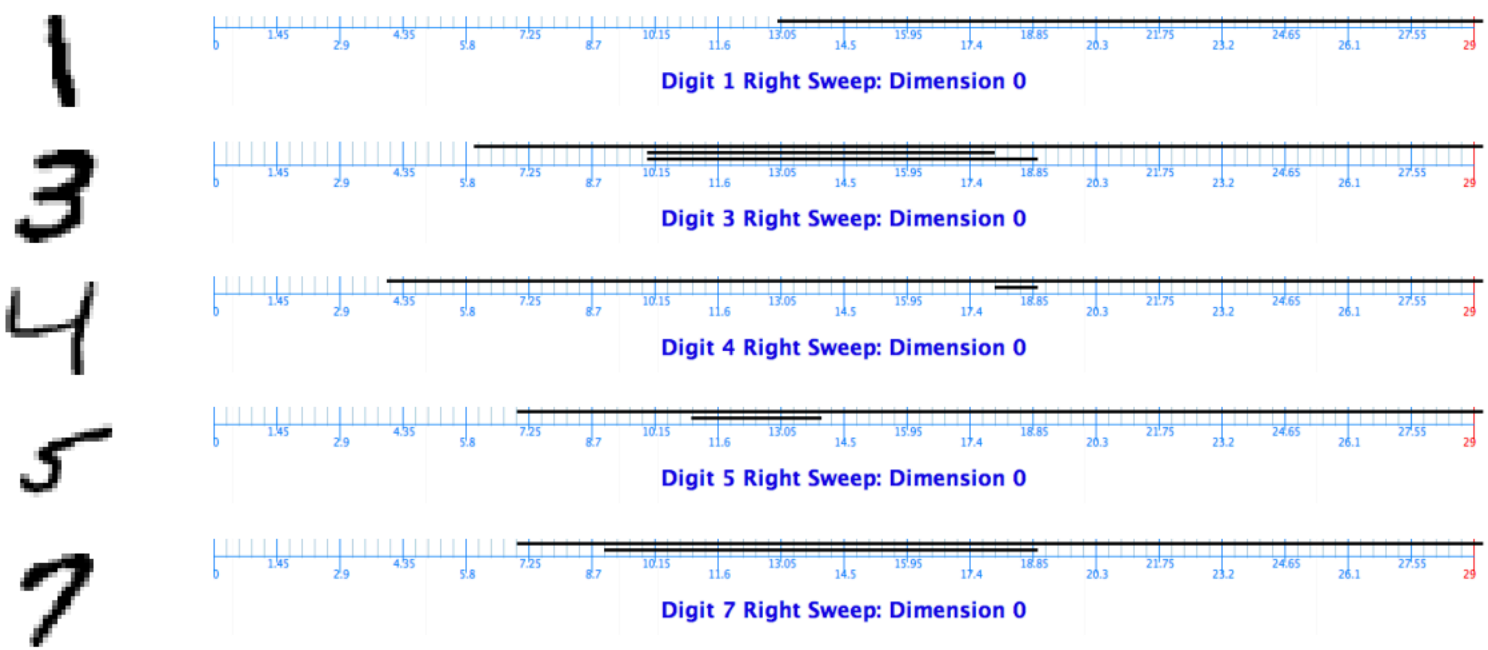
$$E(A) = -\sum_{i}^{n_a} \frac{\ell_i^a}{L_a} \log\left(\frac{\ell_i^a}{L_a}\right),$$

where $\ell_i^a = y_i^a - x_i^a$ and $L_a = \ell_1^a + \cdots + \ell_{n_a}^a$.

The Ring of Algebraic Functions on Persistence Bar Codes

Aaron Adcock   Erik Carlsson   Gunnar Carlsson *

April 3, 2013



Found Comput Math
https://doi.org/10.1007/s10208-018-9379-y
FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

CrossMark

Tropical Coordinates on the Space of Persistence Barcodes

Sara Kališnik[1,2]

2017

Received: 1 April 2016 / Revised: 6 June 2017 / Accepted: 22 December 2017

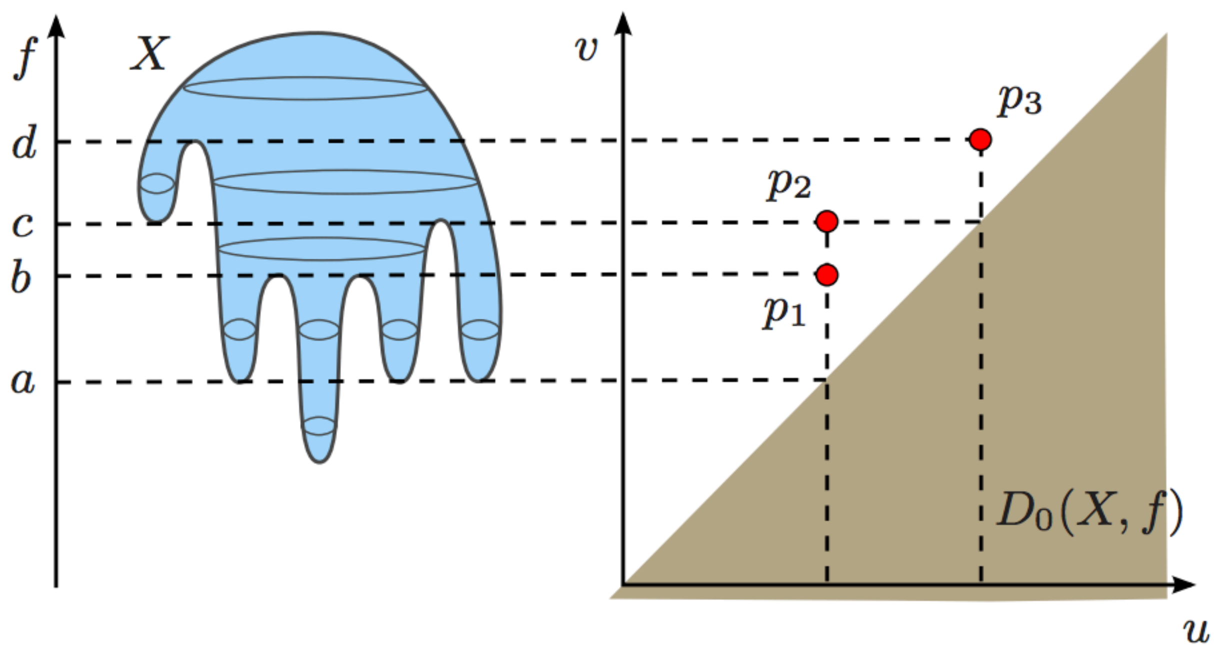
$$a \oplus b := \min(a, b) \text{ and } a \odot b := a + b.$$

Comparing persistence diagrams through complex vectors

Barbara Di Fabio[1] and Massimo Ferri[1,2]

[1] ARCES, University of Bologna, Italy
barbara.difabio@unibo.it,
[2] Department of Mathematics, University of Bologna, Italy
massimo.ferri@unibo.it

2015



Comparing persistence diagrams through complex vectors

Barbara Di Fabio[1] and Massimo Ferri[1,2]

[1] ARCES, University of Bologna, Italy
barbara.difabio@unibo.it,
[2] Department of Mathematics, University of Bologna, Italy
massimo.ferri@unibo.it

2015

- $R : \overline{\Delta^+} \to \mathbb{C}$, with $R(u, v) = u + iv$,
- $S : \overline{\Delta^+} \to \mathbb{C}$, with $S(u, v) = \begin{cases} \frac{v-u}{\alpha\sqrt{2}} \cdot (u + iv), & \text{if } (u, v) \neq (0, 0) \\ (0, 0), & \text{otherwise} \end{cases}$,
- $T : \overline{\Delta^+} \to \mathbb{C}$, with $T(u, v) = \frac{v-u}{2} \cdot (\cos\alpha - \sin\alpha + i(\cos\alpha + \sin\alpha))$,

where $\alpha = \sqrt{u^2 + v^2}$.

Deep Learning with Topological Signatures

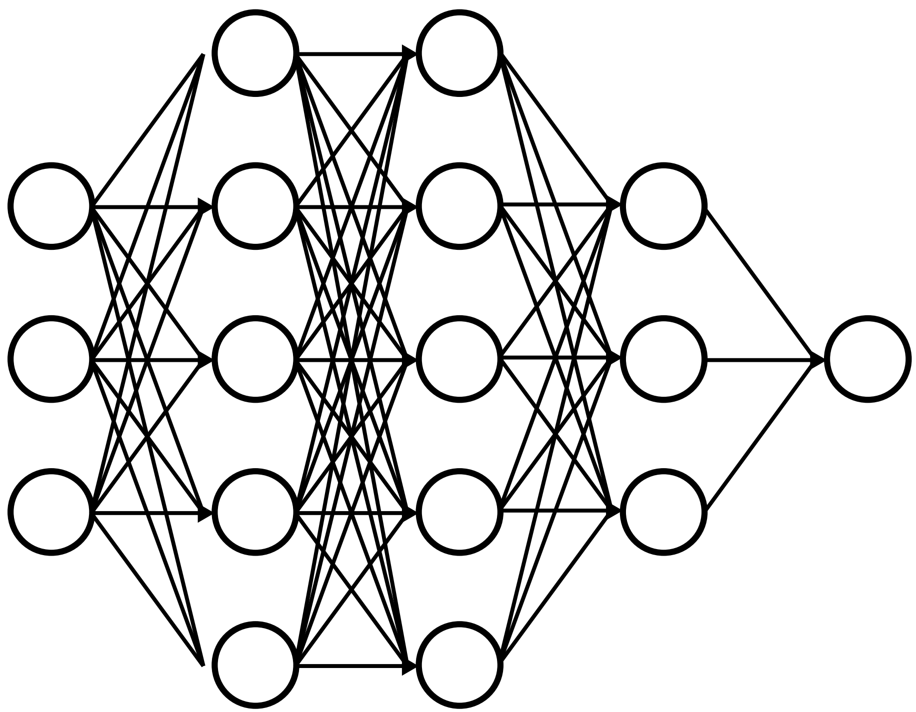
Christoph Hofer
Department of Computer Science
University of Salzburg, Austria
chofer@cosy.sbg.ac.at

Roland Kwitt
Department of Computer Science
University of Salzburg, Austria
Roland.Kwitt@sbg.ac.at

Marc Niethammer
UNC Chapel Hill, NC, USA
mn@cs.unc.edu

Andreas Uhl
Department of Computer Science
University of Salzburg, Austria
uhl@cosy.sbg.ac.at

2018



Deep Learning with Topological Signatures

Christoph Hofer
Department of Computer Science
University of Salzburg, Austria
chofer@cosy.sbg.ac.at

Roland Kwitt
Department of Computer Science
University of Salzburg, Austria
Roland.Kwitt@sbg.ac.at

Marc Niethammer
UNC Chapel Hill, NC, USA
mn@cs.unc.edu

Andreas Uhl
Department of Computer Science
University of Salzburg, Austria
uhl@cosy.sbg.ac.at

2018

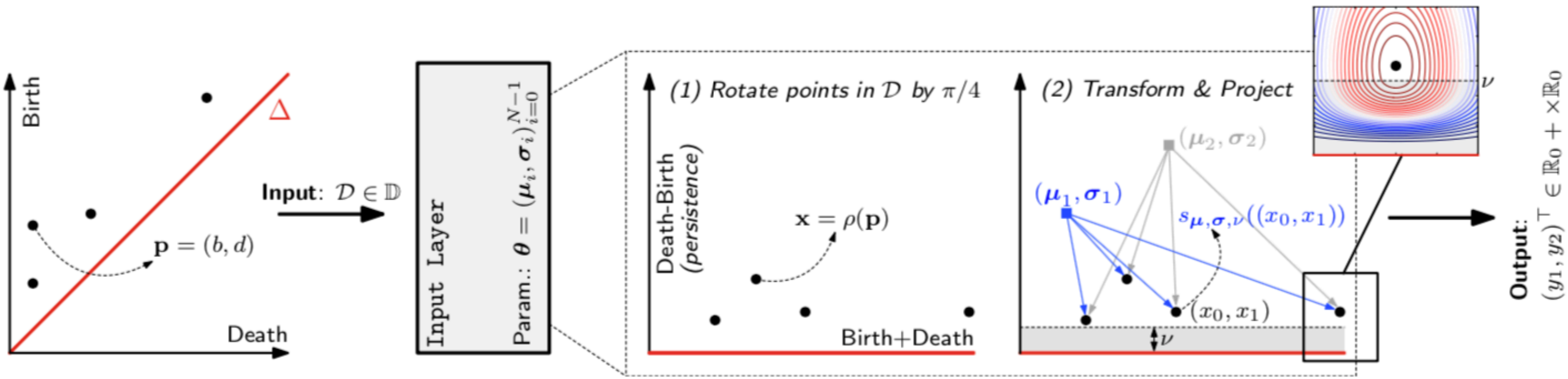


Figure 1: Illustration of the proposed network *input layer* for topological signatures. Each signature, in the form of a persistence diagram $\mathcal{D} \in \mathbb{D}$ (*left*), is projected w.r.t. a collection of *structure elements*. The layer's learnable parameters $\boldsymbol{\theta}$ are the locations $\boldsymbol{\mu}_i$ and the scales $\boldsymbol{\sigma}_i$ of these elements; $\nu \in \mathbb{R}^+$ is set a-priori and meant to discount the impact of points with low persistence (and, in many cases, of low discriminative power). The layer output $\mathbf{y}$ is a concatenation of the projections. In this illustration, $N = 2$ and hence $\mathbf{y} = (y_1, y_2)^\top$.

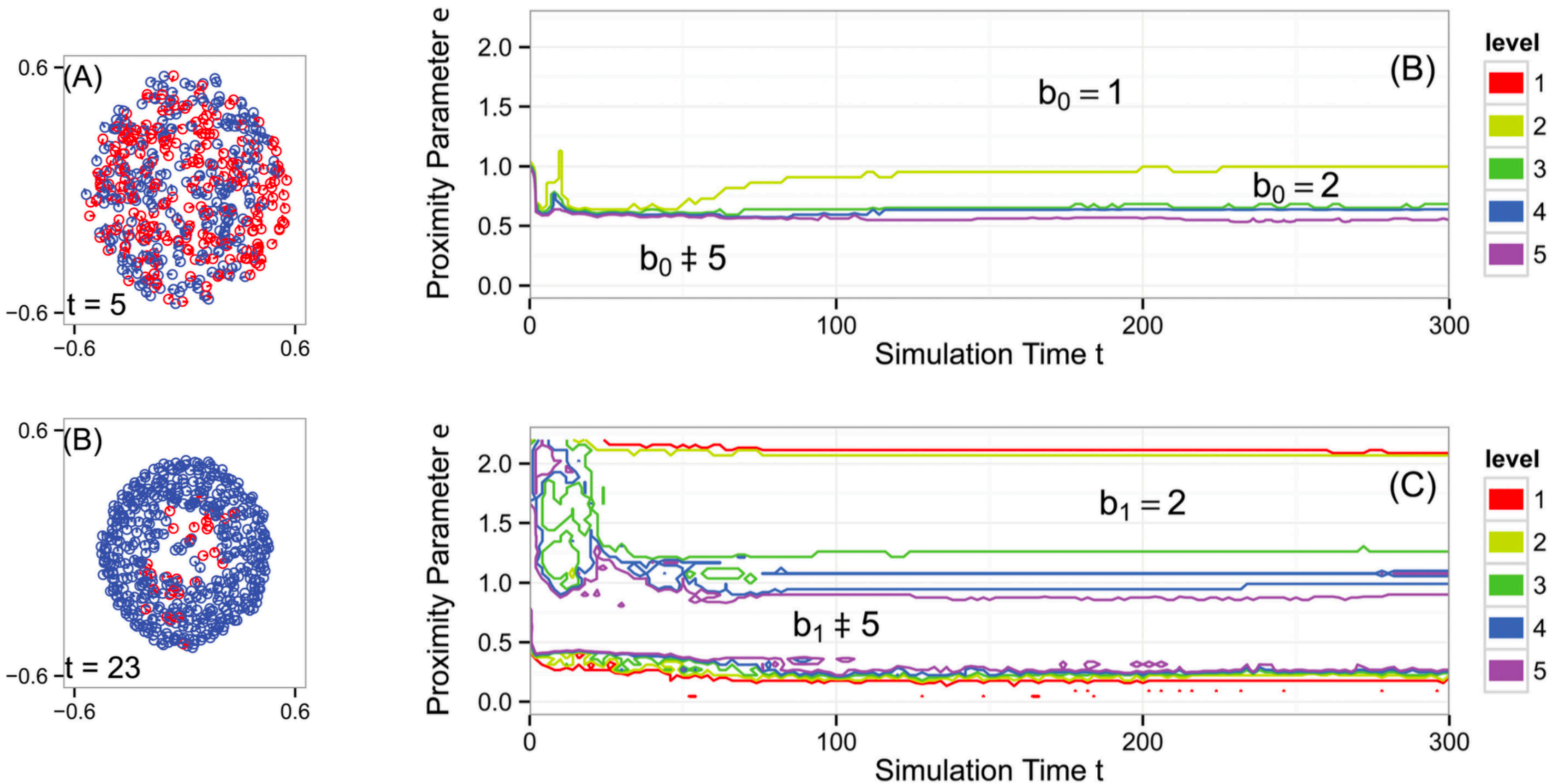
RESEARCH ARTICLE

Topological Data Analysis of Biological Aggregation Models

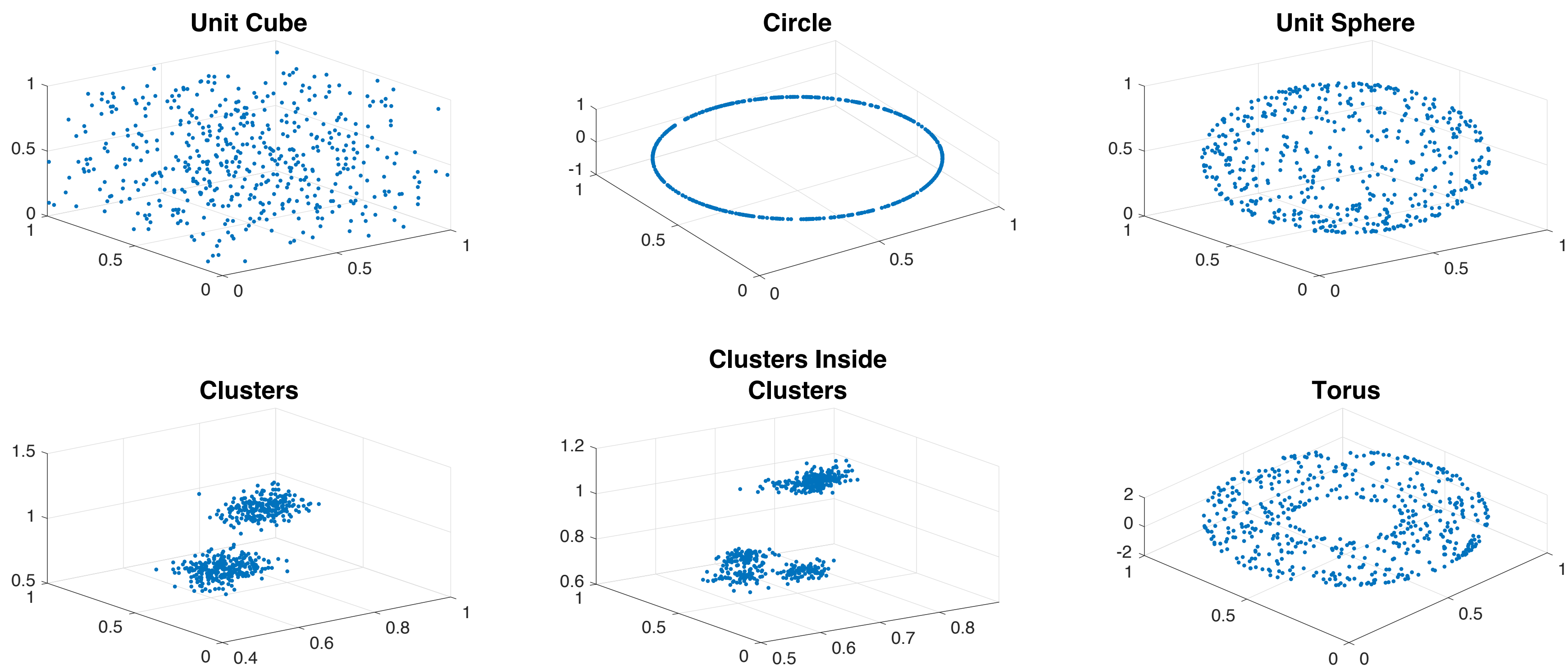
2015

Chad M. Topaz*, Lori Ziegelmeier, Tom Halverson

Department of Mathematics, Statistics, and Computer Science, Macalester College, Saint Paul, Minnesota, United States of America



Survey: From persistent homology to machine learning feature vectors



Distance Matrix	Accuracy (Noise 0.05)	Time (Noise 0.05)	Accuracy (Noise 0.1)	Time (Noise 0.1)
PD, H_0, L^1	96.0%	37346s	96.0%	42613s
PD, H_0, L^2	91.3%	24656s	91.3%	25138s
PD, H_0, L^∞	60.7%	1133s	63.3%	1149s
PD, H_1, L^1	100%	657s	96.0%	703s
PD, H_1, L^2	100%	984s	97.3%	1042s
PD, H_1, L^∞	81.3%	527s	66.7%	564s
PL, H_0, L^1	92.7%	29s	96.7%	33s
PL, H_0, L^2	77.3%	29s	82.0%	34s
PL, H_0, L^∞	60.7%	2s	63.3%	2s
PL, H_1, L^1	83.3%	36s	80.7%	48s
PL, H_1, L^2	83.3%	50s	66.7%	69s
PL, H_1, L^∞	74.7%	8s	66.7%	9s
PI, H_0, L^1	93.3%	9s	95.3%	9s
PI, H_0, L^2	92.7%	9s	95.3%	9s
PI, H_0, L^∞	94.0%	9s	96.0%	9s
PI, H_1, L^1	100%	17s	95.3%	18s
PI, H_1, L^2	100%	17s	96.0%	18s
PI, H_1, L^∞	100%	17s	96.0%	18s