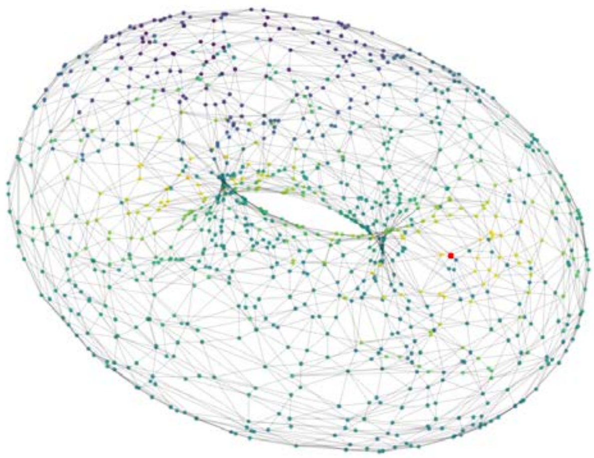
# Topological Data Analysis and Persistence Theory

NSF / CBMS Conference

Valdosta State University

August 8-12, 2022

Peter Bubenik, University of Florida



## Lecture 8 : Applications of TDA

Outline: 1. Curvature

joint with Michael Hull, Dhruv Patel, and Benjamin Whittle

2. Biological Images

joint with Parker Edwards, Kristen Skruber, Nikola Milicevic, and Eric Vitriol

3. Biological Video

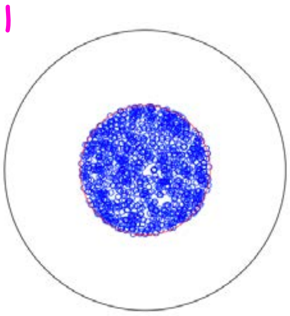
joint with Ashleigh Thomas, Kathleen Bates, Alex Elchesen, Iryna Hartsock, and Hang Lu

Please interrupt me !!!

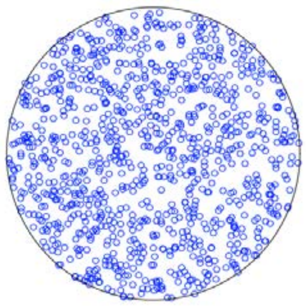
# 1. Learning Curvature of Surfaces from Sampled Points

Sample points from unit disk on a surfaces of constant curvature

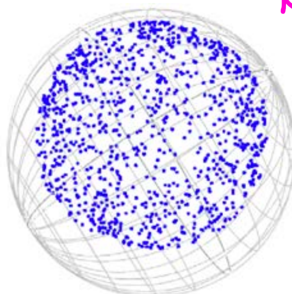
K = -1



K = 0

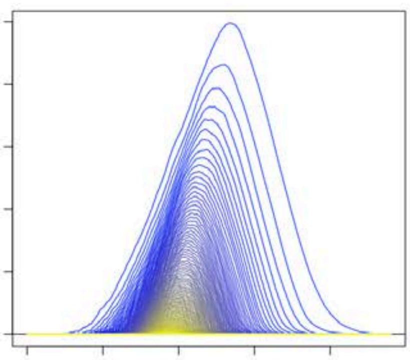


k = 1

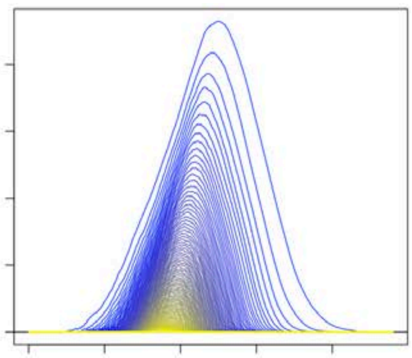


Use pairwise distances to compute persistent homology.

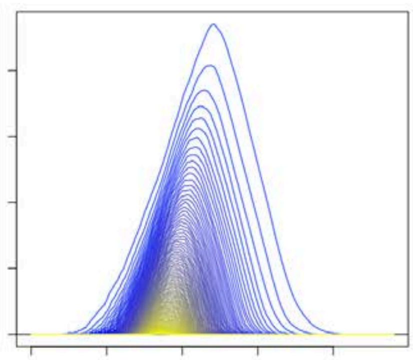
Repeat. Compute Average Persistence Landscapes.



K = -1



K = 0



K = 1

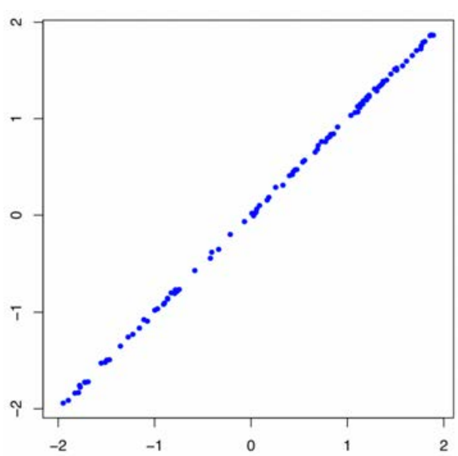
Training data:

K = -2, -1.96, ..., 0, ... 1.96, 2

Testing data

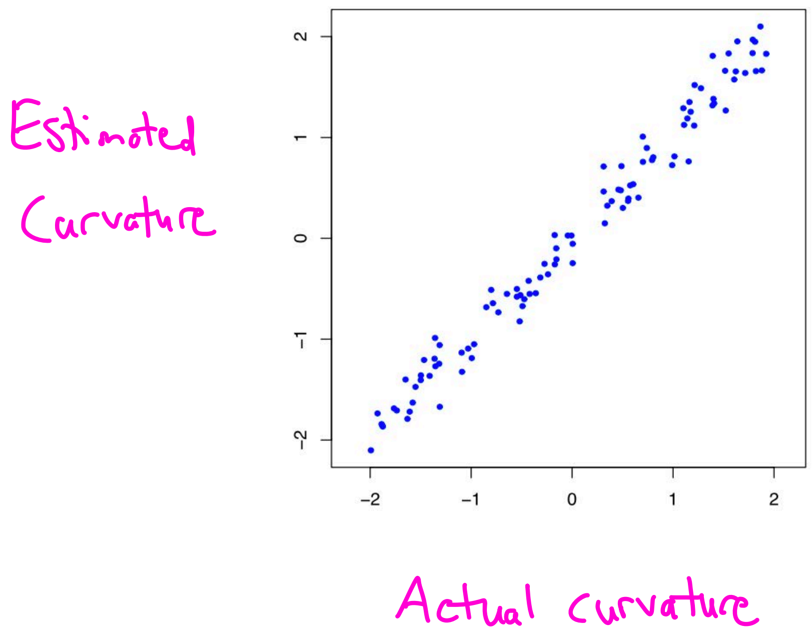
100 random values of K ~ [-2, 2]

Estimated Curvature



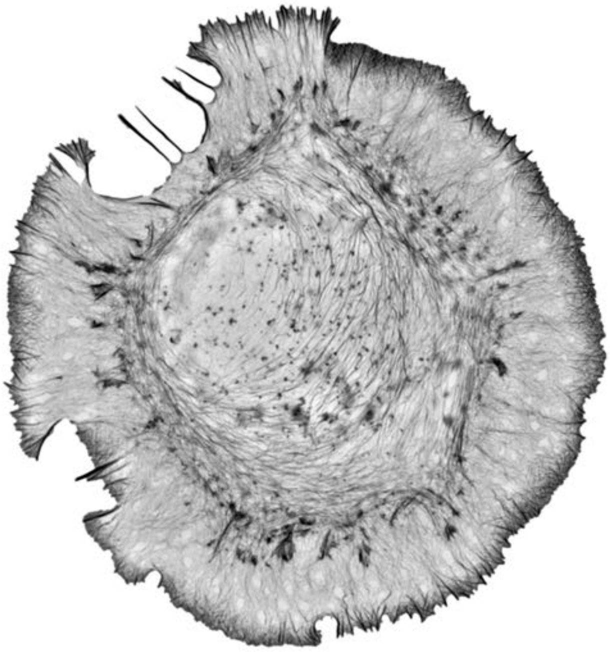

Actual curvature

Use only ordinals of pairwise distances:



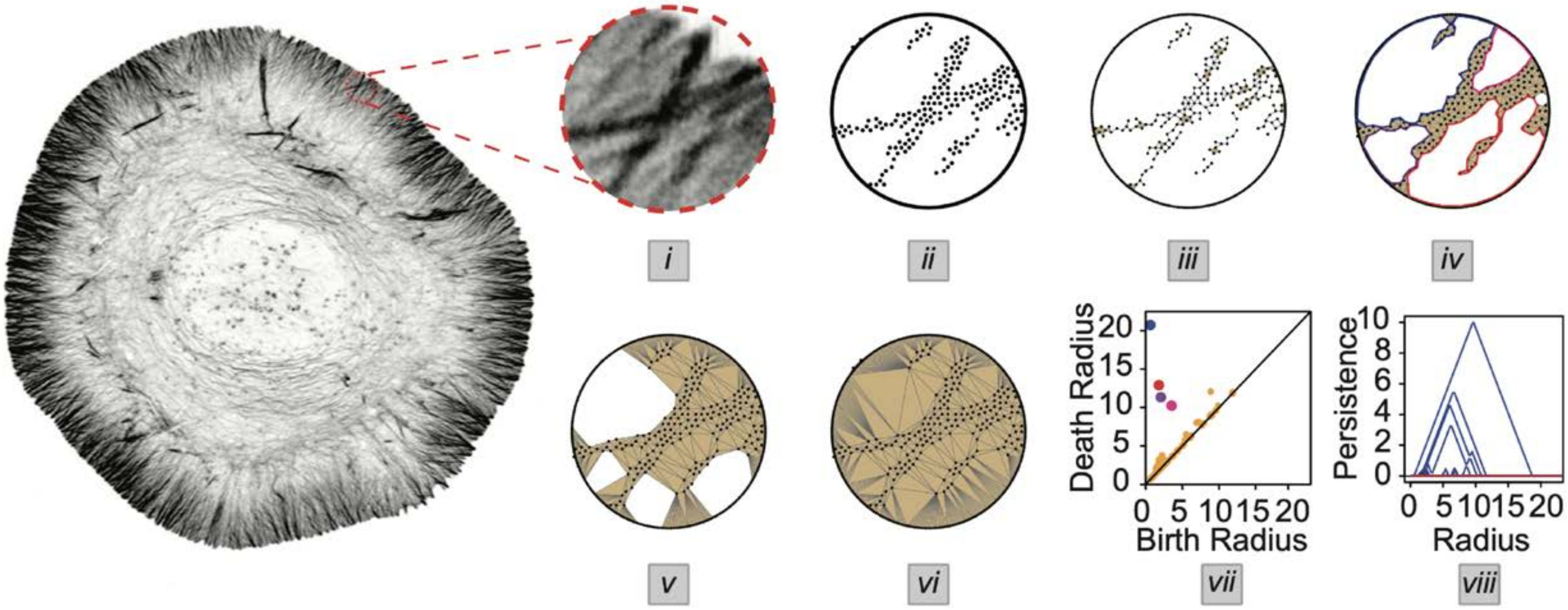
# 2. TDA for Biological Images : TDAExplore

Goal: Understand the dynamics of the actin cytoskeleton.



Show
animated gif

## 2.1 Persistence Landscapes from Image Patches



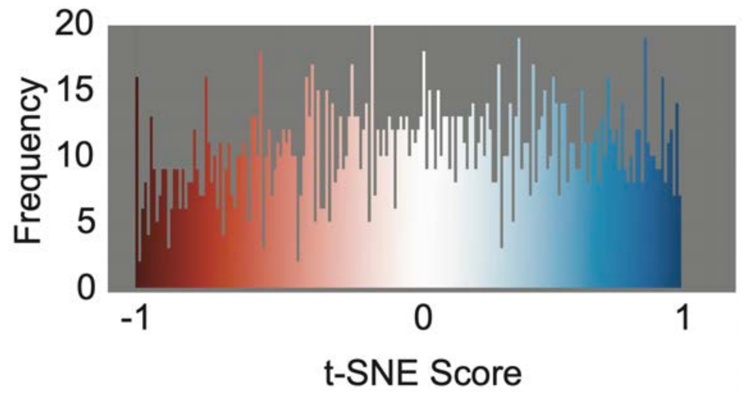
## 2.2 Unsupervised Learning

Compute Persistence Landscape for each Patch

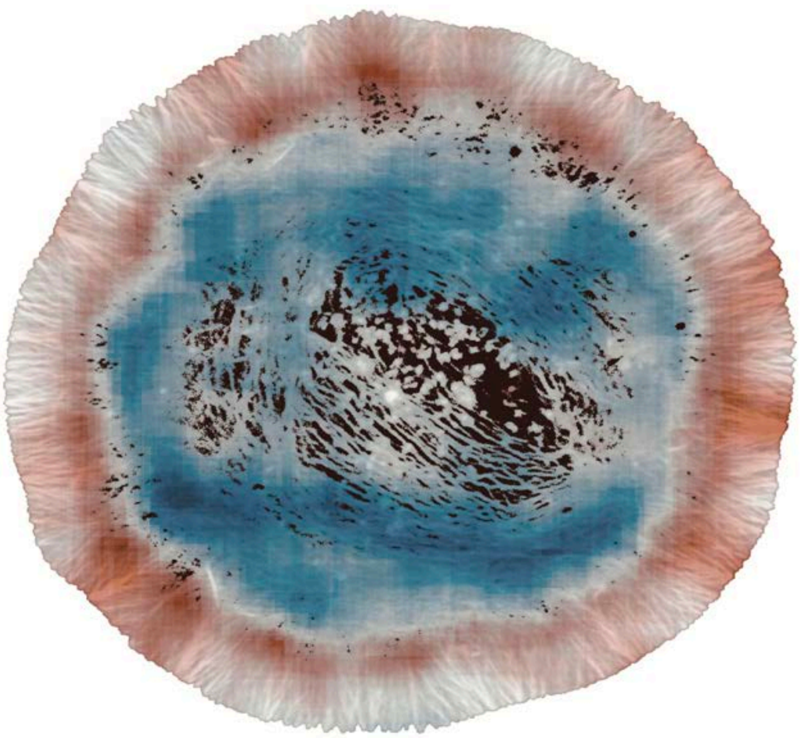
Project to one dimension

Each patch gets a score

Histogram of scores →




Give each pixel the average score of the patches that contain it.



## 2.3 Supervised Learning

Split data into training images and testing images. (images from 2 classes.)

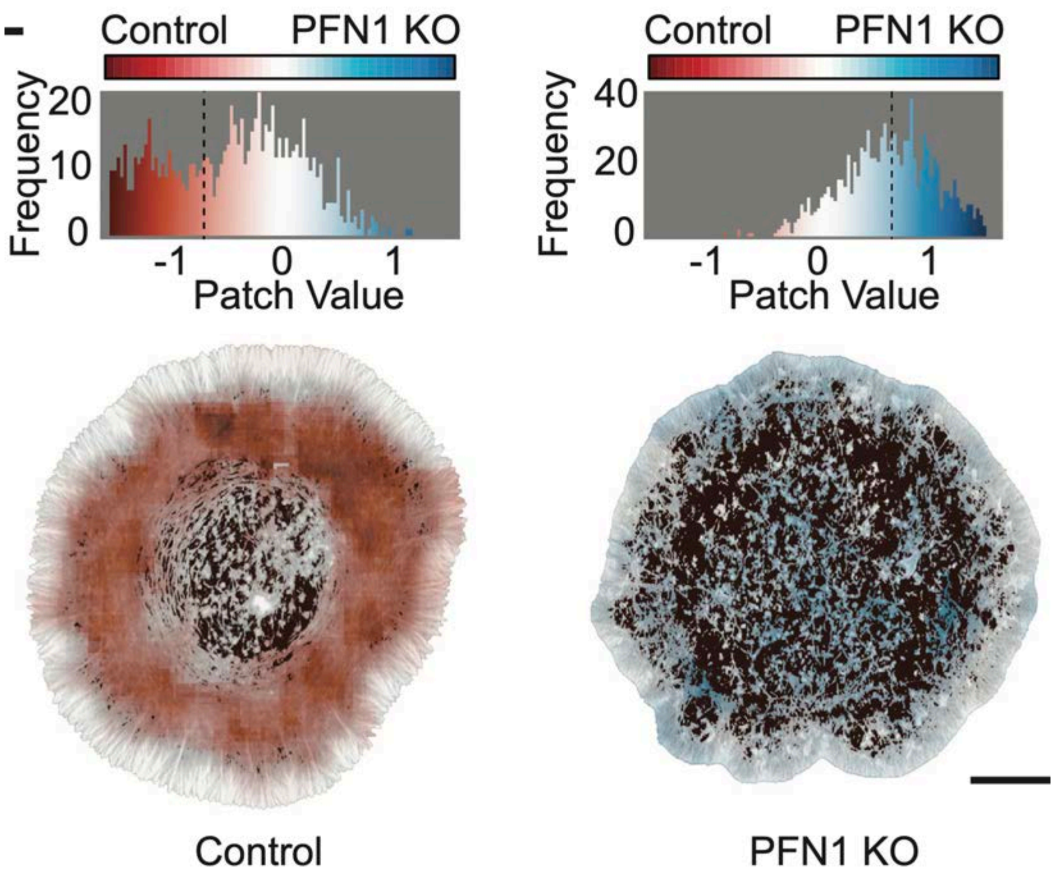
Compute Persistence Landscape for patches in training images.

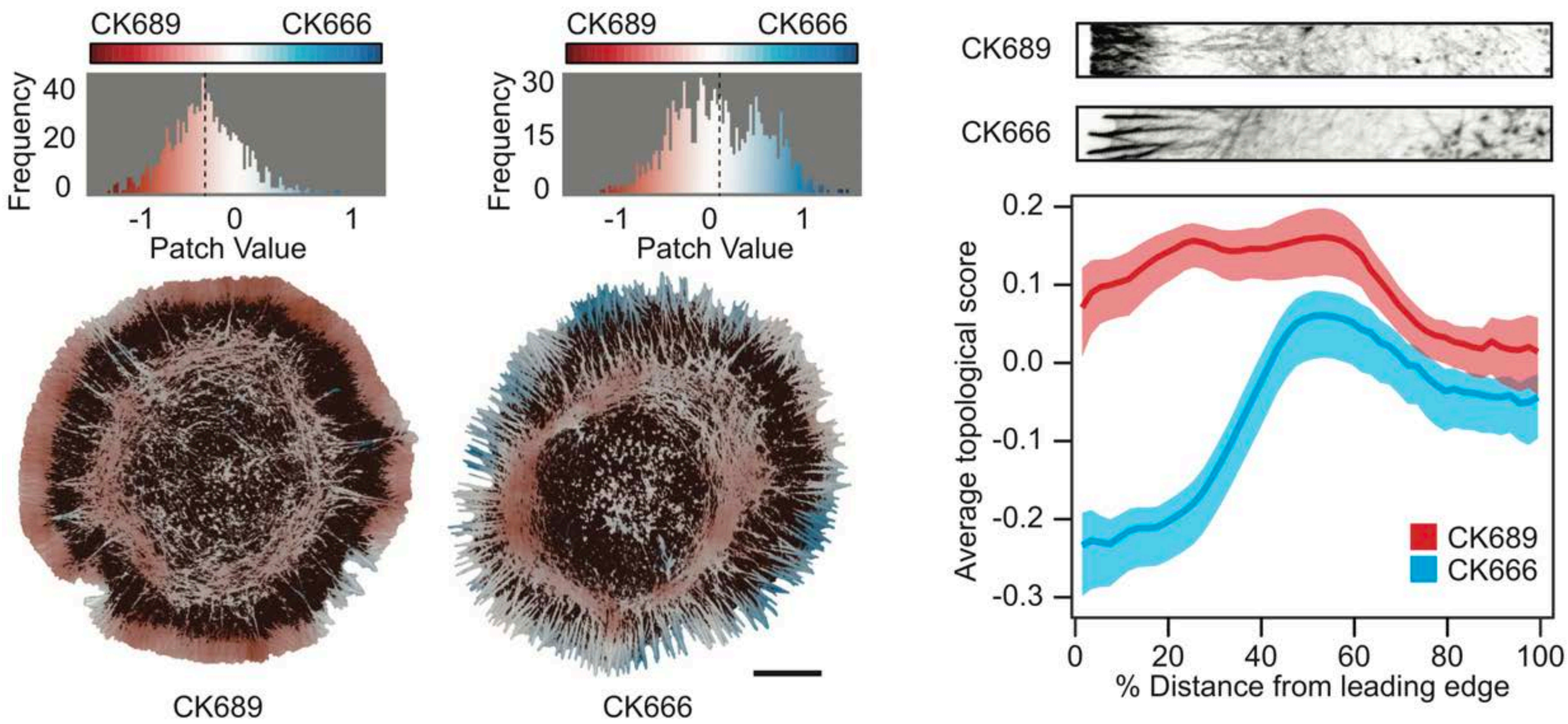
Assign these vectors/points a value of ±1 (depending on the class of the image.)

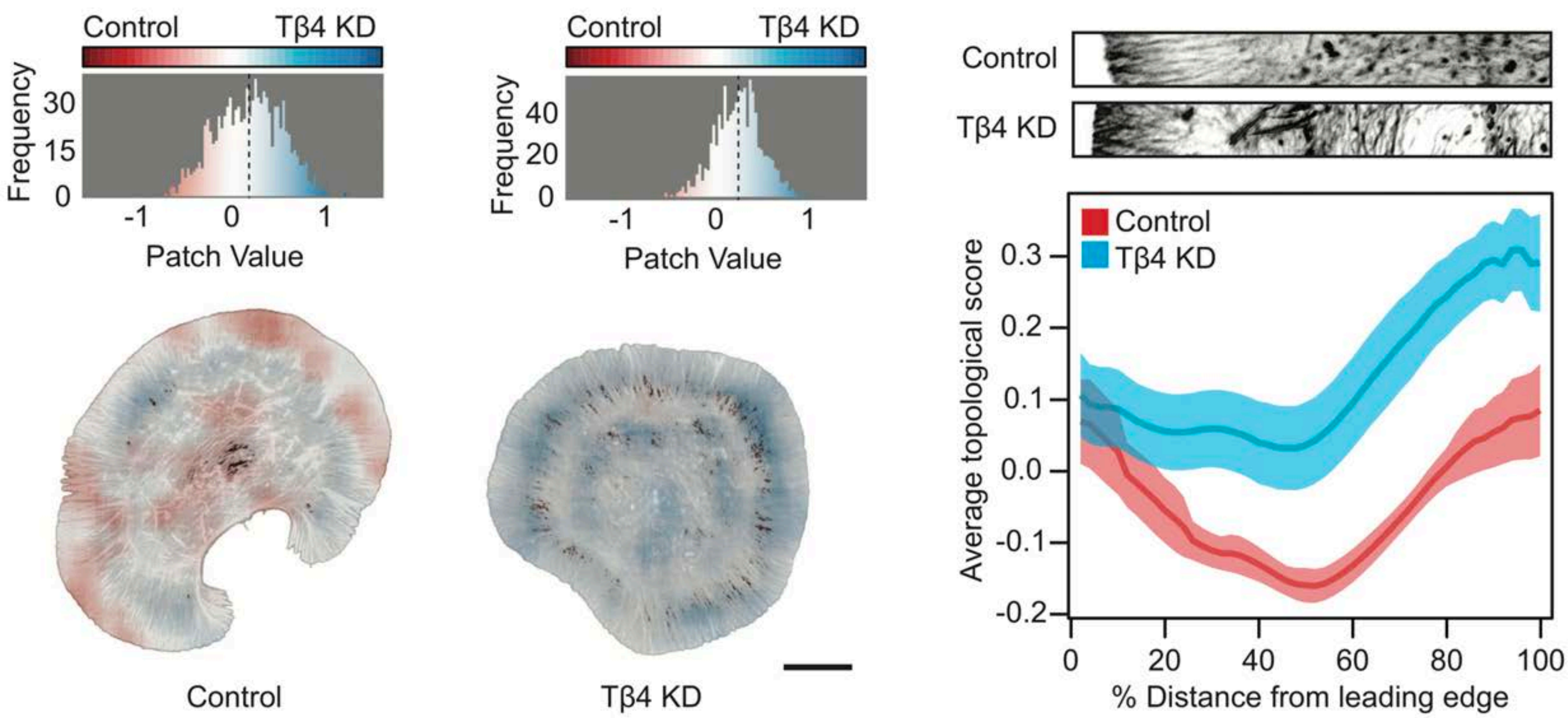
Apply Support Vector Regression.

Use the SVR model to assign a score to the persistence landscape of each patch of the testing images.

Give each pixel the average score of the patches that contain it.




CK689
CK666
Frequency
40
20
0
-1
0
1
Patch Value
CK689
CK666
Frequency
30
15
0
-1
0
1
Patch Value
CK689
CK666
CK689
CK666
Average topological score
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
20
40
60
80
100
% Distance from leading edge
CK689
CK666


Control
Tβ4 KD
Frequency
30
15
0
-1
0
1
Patch Value
Control
Tβ4 KD
Frequency
40
20
0
-1
0
1
Patch Value
Control
Tβ4 KD
Control
Tβ4 KD
Control
Tβ4 KD
Average topological score
0.3
0.2
0.1
0.0
-0.1
-0.2
0
20
40
60
80
100
% Distance from leading edge

# 3. TDA for Biological Videos

Example : Video

## 3.1. Sliding Window Embedding

Suppose we have a sequence $x_1, x_2, x_3, \dots, x_N$, $x_i \in \mathbb{R}^d$

If the sequence is indexed by time, it is a time series.

For example, $x_i$, may be a frame in a video.

Consider a "window" of consecutive $x_i$ : $x_1, x_2, \dots, x_m$

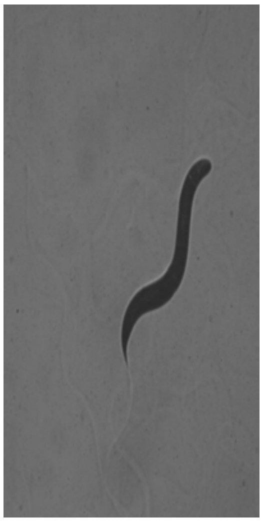
$m << N$. Concatenate to get $(x_1 \dots x_m) \in \mathbb{R}^{md}$

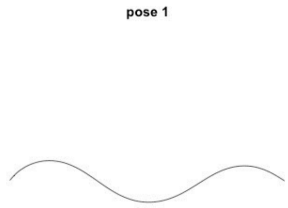
Now "slide" this window across the time series :

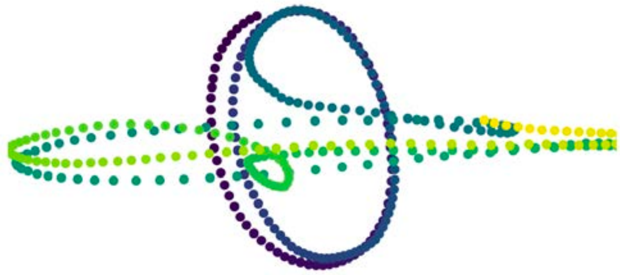
$(x_1, \dots x_m), (x_2, \dots, x_{m+1}), (x_3, \dots x_{m+2}), \dots, (x_{N-m+1}, \dots, x_N) \in \mathbb{R}^{md}$

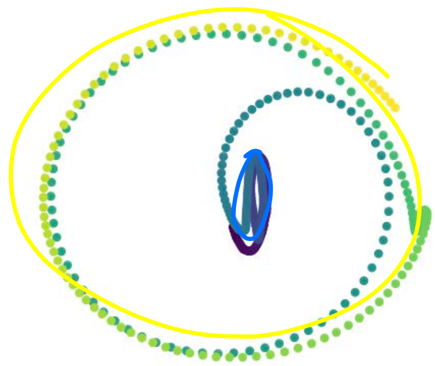
Apply TDA to this point cloud.

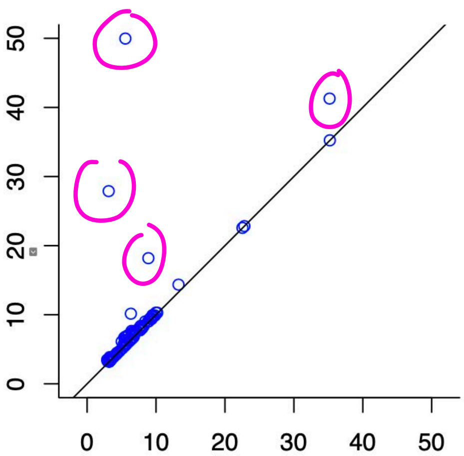
# 3.2 Topological Data Analysis of C. elegans Locomotion

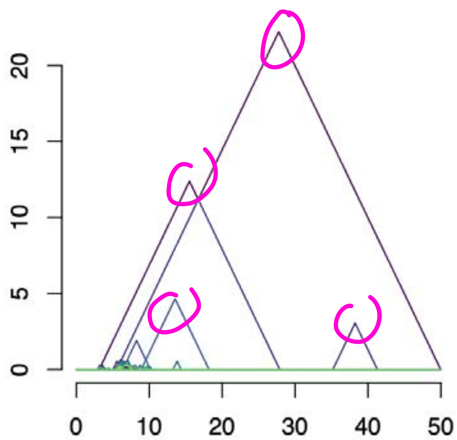








# 3.3 Cycle representatives for Persistent Homology