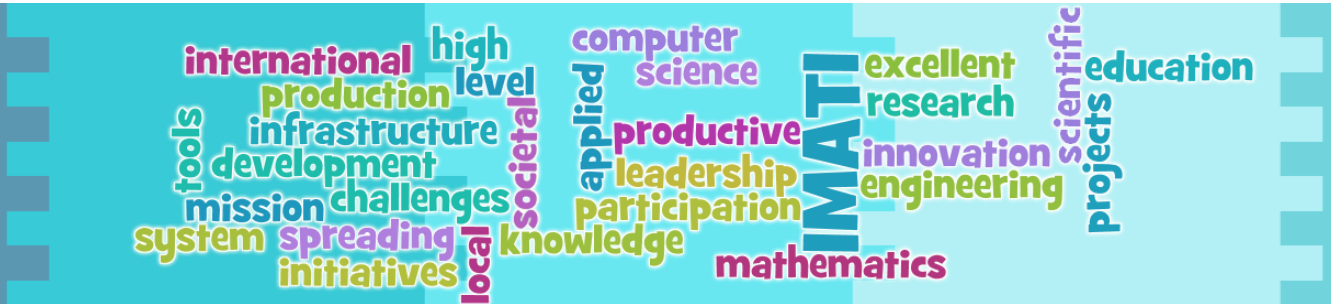
*Topological Data Analysis*

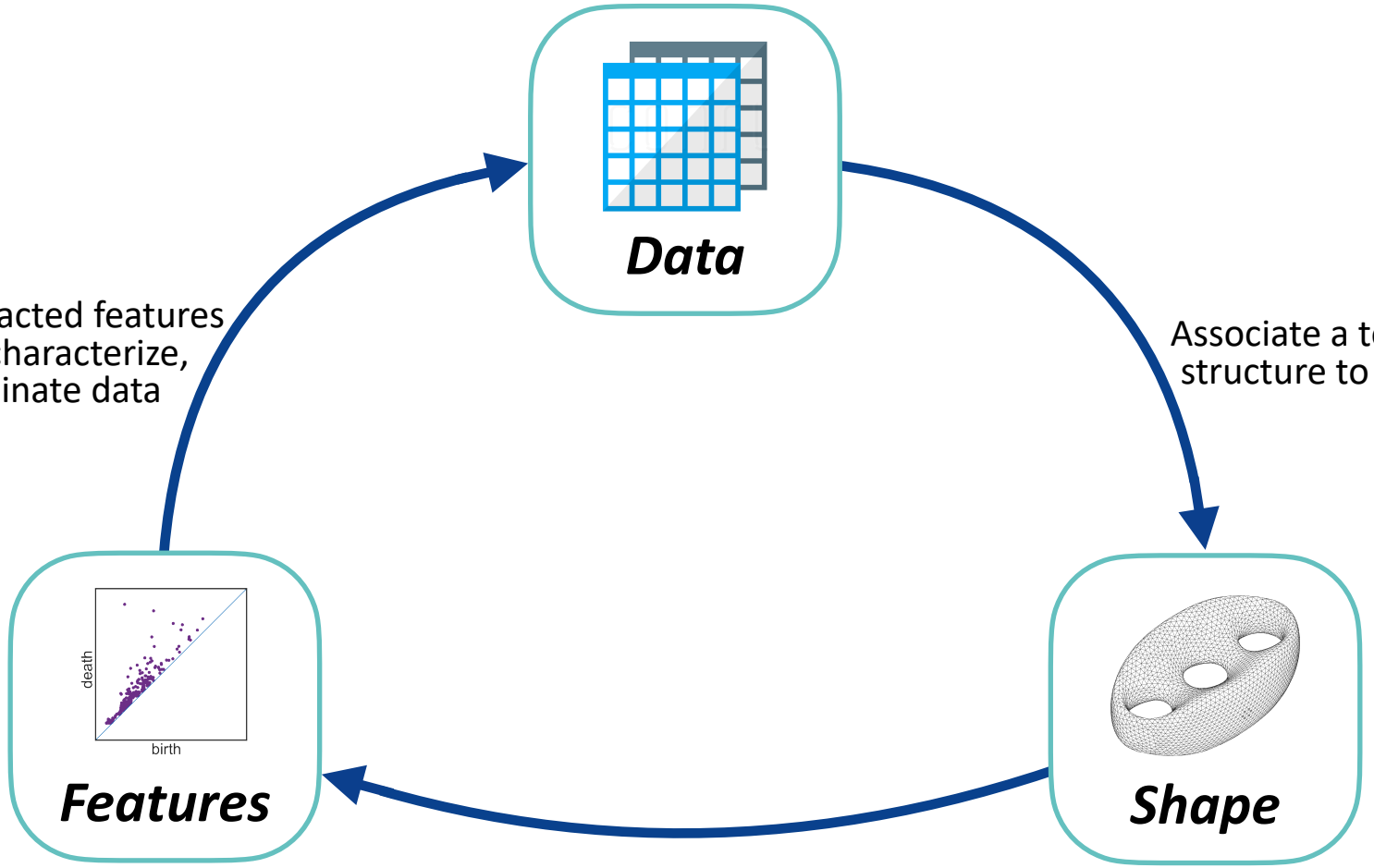
# *Software Packages*

Ulderico Fugacci

**CNR - IMATI**

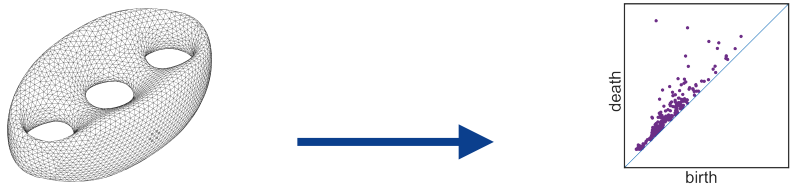


# Topological Data Analysis



# Software Packages

***Topological Data Analysis*** allows for assigning to (almost) ***any dataset*** a collection of features representing a ***topological summary*** of the input data




***Shape*** ***Features***

***Goal:***

*Today, we address one main question:*

- ***What software packages are available for computing persistent homology?***

# Software Packages

*Several software packages for computing persistent homology have been developed:*

- ***javaPlex***
- ***jHoles***
- ***Dionysus***
- ***Perseus***
- ***PHAT***
- ***DIPHA***
- ***Gudhi***
- ***SimpPers***
- ***Ripser***
- ***TDAstats***
- ***…***

# Software Packages

***javaPlex:***

- ***Language:***
  - *Java*
- ***Algorithms:***
  - *Standard*, *Dual*, *Zigzag*
- ***Coefficient Fields:***
  - $\mathbb{Q}$ , $\mathbb{Z}_p$
- ***Homology:***
  - *Simplicial*, *Cellular*
- ***Accepted Inputs:***
  - *Simplicial complexes*, *Zigzag*, *CW complexes*
- ***Computed Filtrations:***
  - *Vietoris-Rips complexes*, *(parametrized) Witness complexes*
- ***Visualization:***
  - *Persistence Barcodes*
- ***Additional Features:***
  - *Homology Generators*

# Software Packages

**jHoles:**

- ***Language:***
  - *Java*
- ***Algorithms:***
  - *Standard (uses javaPlex)*
- ***Coefficient Fields:***
  - $\mathbb{Z}_2$
- ***Homology:***
  - *Simplicial*
- ***Accepted Inputs:***
  - -
- ***Computed Filtrations:***
  - *Weight Rank Clique filtration*
- ***Visualization:***
  - -
- ***Additional Features:***
  - -

# Software Packages

**Dionysus:**

- ***Language:***
  - *C++ (with Python bindings)*
- ***Algorithms:***
  - *Standard, Dual, Zigzag*
- ***Coefficient Fields:***
  - $\mathbb{Z}_2$ *(standard, zigzag),* $\mathbb{Z}_p$ *(dual)*
- ***Homology:***
  - *Simplicial*
- ***Accepted Inputs:***
  - *Simplicial complexes, Zigzag*
- ***Computed Filtrations:***
  - *Vietoris-Rips complexes, Alpha-Shapes, Čech complexes*
- ***Visualization:***
  - -
- ***Additional Features:***
  - *Vineyards, Circle-Valued functions, Homology Generators*

# Software Packages

**Perseus:**

- ***Language:***
  - *C++ (with Python bindings)*
- ***Algorithms:***
  - *Standard, Morse reductions*
- ***Coefficient Fields:***
  - $\mathbb{Z}_2$
- ***Homology:***
  - *Simplicial, Cubical*
- ***Accepted Inputs:***
  - *Simplicial complexes, Cubical complexes*
- ***Computed Filtrations:***
  - *Vietoris-Rips complexes, Lower Star of Cubical complexes*
- ***Visualization:***
  - *Persistence Diagrams*
- ***Additional Features:***
  - *Weighted Points for VR*

# Software Packages

**PHAT:**

- ***Language:***
  - *C++ (with Python bindings)*
- ***Algorithms:***
  - *Standard, Dual, Twist, Chunk, Spectral Sequences*
- ***Coefficient Fields:***
  - $\mathbb{Z}_2$
- ***Homology:***
  - *Simplicial, Cubical*
- ***Accepted Inputs:***
  - *Boundary Matrices*
- ***Computed Filtrations:***
  - -
- ***Visualization:***
  - -
- ***Additional Features:***
  - -

# Software Packages

**DIPHA:**

- ***Language:***
  - *C++ (with Python bindings)*
- ***Algorithms:***
  - *Dual, Twist, Distributed*
- ***Coefficient Fields:***
  - $\mathbb{Z}_2$
- ***Homology:***
  - *Simplicial, Cubical*
- ***Accepted Inputs:***
  - *Boundary Matrices*
- ***Computed Filtrations:***
  - *Vietoris-Rips complexes,*
    *Lower Star of Cubical complexes*
- ***Visualization:***
  - *Persistence Diagrams*
- ***Additional Features:***
  - -

# Software Packages

**Gudhi:**

- ***Language:***
  - *C++ (with Python bindings)*
- ***Algorithms:***
  - *Dual, Annotation, Multifield*
- ***Coefficient Fields:***
  - $\mathbb{Z}_p$
- ***Homology:***
  - *Simplicial, Cubical*
- ***Accepted Inputs:***
  - *Simplicial complexes*
- ***Computed Filtrations:***
  - *Vietoris-Rips complexes, Alpha-Shapes, Witness complexes, Lower Star of Cubical complexes*
- ***Visualization:***
  - *Persistence Diagrams, Persistence Barcodes*
- ***Additional Features:***
  - *Bottleneck distance, Wasserstein distance*

# Software Packages

**SimpPers:**

- ***Language:***
  - *C++*
- ***Algorithms:***
  - *Simplicial Maps*
- ***Coefficient Fields:***
  - $\mathbb{Z}_2$
- ***Homology:***
  - *Simplicial*
- ***Accepted Inputs:***
  - *Maps of simplicial complexes*
- ***Computed Filtrations:***
  - -
- ***Visualization:***
  - -
- ***Additional Features:***
  - -

# Software Packages

**Ripser:**

- **Language:**
  - C++ (with Python bindings)
- **Algorithms:**
  - Dual, Twist
- **Coefficient Fields:**
  - $\mathbb{Z}_p$
- **Homology:**
  - Simplicial
- **Accepted Inputs:**
  - Point Clouds, Distance Matrices
- **Computed Filtrations:**
  - Vietoris-Rips complexes
- **Visualization:**
  - Persistence Diagrams
- **Additional Features:**
  - Representative Cocycles (through Persim: Bottleneck distance, modified Gromov–Hausdorff distance, Sliced Wasserstein kernel, Heat kernel, Persistence Images)

# Software Packages

**TDAstats:**

- **Language:**
  - R
- **Algorithms:**
  - Dual, Twist (uses Ripser)
- **Coefficient Fields:**
  - $\mathbb{Z}_p$
- **Homology:**
  - Simplicial
- **Accepted Inputs:**
  - Point Clouds
- **Computed Filtrations:**
  - Vietoris-Rips complexes
- **Visualization:**
  - Persistence Diagrams, Persistence Barcodes
- **Additional Features:**
  - -

# Software Packages

***Computation Times:***

| Data set | (a) Computations on cluster: wall-time seconds | | | | | |
|---|---|---|---|---|---|---|
| | **eleg** | **Klein** | **HIV** | **drag 2** | **random** | **genome** |
| Size of complex | $4.4 \times 10^6$ | $1.1 \times 10^7$ | $2.1 \times 10^8$ | $1.3 \times 10^9$ | $3.1 \times 10^9$ | $4.5 \times 10^8$ |
| Max. dim. | 2 | 2 | 2 | 2 | 8 | 2 |
| JAVAPLEX (st) | 84 | 747 | - | - | - | - |
| DIONYSUS (st) | 474 | 1,830 | - | - | - | - |
| DIPHA (st) | 6 | 90 | 1,631 | 142,559 | - | 9,110 |
| PERSEUS | 543 | 1,978 | - | - | - | - |
| DIONYSUS (d) | 513 | 145 | - | - | - | - |
| DIPHA (d) | 4 | 6 | 81 | 2,358 | 5,096 | 232 |
| GUDHI | 36 | 89 | 1,798 | 14,368 | - | 4,753 |
| RIPSER | 1 | 1 | 2 | 6 | 349 | 3 |

Table from [Otter et al. 2017]

# Software Packages

***Computation Times:***

| Data set | (b) Computations on cluster: CPU seconds | | | | | |
|---|---|---|---|---|---|---|
| | **eleg** | **Klein** | **HIV** | **drag 2** | **random** | **genome** |
| Size of complex | $4.4 \times 10^6$ | $1.1 \times 10^7$ | $2.1 \times 10^8$ | $1.3 \times 10^9$ | $3.1 \times 10^9$ | $4.5 \times 10^8$ |
| Max. dim. | 2 | 2 | 2 | 2 | 8 | 2 |
| JAVAPLEX (st) | 284 | 1,031 | - | - | - | - |
| DIONYSUS (st) | 473 | 1,824 | - | - | - | - |
| DIPHA (st) | 68 | 1,360 | 25,950 | 1,489,615 | - | 130,972 |
| PERSEUS | 542 | 1,974 | - | - | - | - |
| DIONYSUS (d) | 513 | 145 | - | - | - | - |
| DIPHA (d) | 39 | 73 | 1,276 | 37,572 | 79,691 | 3,622 |
| GUDHI | 36 | 88 | 1,794 | 14,351 | - | 4,764 |
| RIPSER | 1 | 1 | 2 | 5 | 348 | 2 |

Table from [Otter et al. 2017]

# Software Packages

***Computation Times:***

| Data set | (c) Computations on shared-memory system: wall-time seconds | | | | | |
|---|---|---|---|---|---|---|
| | **eleg** | **Klein** | **HIV** | **drag 2** | **genome** | **fract r** |
| Size of complex | $3.2 \times 10^8$ | $1.1 \times 10^7$ | $2.1 \times 10^8$ | $1.3 \times 10^9$ | $4.5 \times 10^8$ | $2.8 \times 10^9$ |
| Max. dim. | 3 | 2 | 2 | 2 | 2 | 3 |
| JAVAPLEX (st) | 13,607 | 1,358 | 43,861 | - | 28,064 | - |
| PERSEUS | - | 1,271 | - | - | - | - |
| DIONYSUS (d) | - | 100 | 142,055 | 35,366 | - | 572,764 |
| DIPHA (d) | 926 | 13 | 773 | 4,482 | 1,775 | 3,923 |
| GUDHI | 381 | 6 | 177 | 1,518 | 442 | 4,590 |
| RIPSER | 2 | 1 | 2 | 5 | 3 | 1,517 |

Table from [Otter et al. 2017]

# Software Packages

***Memory Usage:***

| Data set | (a) Computations on cluster | | | | | |
|---|---|---|---|---|---|---|
| | **eleg** | **Klein** | **HIV** | **drag 2** | **random** | **genome** |
| Size of complex | $4.4 \times 10^6$ | $1.1 \times 10^7$ | $2.1 \times 10^8$ | $1.3 \times 10^9$ | $3.1 \times 10^9$ | $4.5 \times 10^8$ |
| Max. dim. | 2 | 2 | 2 | 2 | 8 | 2 |
| JAVAPLEX (st) | <5 | <15 | >64 | >64 | >64 | >64 |
| DIONYSUS (st) | 1.3 | 11.6 | - | - | - | - |
| DIPHA (st) | 0.1 | 0.2 | 2.7 | 4.9 | - | 4.8 |
| PERSEUS | 5.1 | 12.7 | - | - | - | - |
| DIONYSUS (d) | 0.5 | 1.1 | - | - | - | - |
| DIPHA (d) | 0.1 | 0.2 | 1.8 | 13.8 | 9.6 | 6.3 |
| GUDHI | 0.2 | 0.5 | 8.5 | 62.8 | - | 21.5 |
| RIPSER | 0.007 | 0.02 | 0.06 | 0.2 | 24.7 | 0.07 |

Table from [Otter et al. 2017]

# Software Packages

***Memory Usage:***

| Data set | (b) Computations on shared-memory system | | | | | |
|---|---|---|---|---|---|---|
| | **eleg** | **Klein** | **HIV** | **drag 2** | **genome** | **fract r** |
| Size of complex | $3.2 \times 10^8$ | $1.1 \times 10^7$ | $2.1 \times 10^8$ | $1.3 \times 10^9$ | $4.5 \times 10^8$ | $2.8 \times 10^9$ |
| Max. dim. | 3 | 2 | 2 | 2 | 2 | 3 |
| JAVAPLEX (st) | <600 | <15 | <700 | >700 | <700 | >700 |
| PERSEUS | - | 11.7 | - | - | - | - |
| DIONYSUS (d) | - | 1.1 | 16.8 | 134.2 | - | 268.5 |
| DIPHA (d) | 31.2 | 0.9 | 17.7 | 109.5 | 37.3 | 276.1 |
| GUDHI | 15.4 | 0.5 | 10.2 | 62.8 | 21.4 | 134.8 |
| RIPSER | 0.2 | 0.03 | 0.07 | 0.2 | 0.07 | 155 |

Table from [Otter et al. 2017]

# Software Packages

***Supported Maximal Size:***

| JAVAPLEX | DIONYSUS | | DIPHA | | PERSEUS | GUDHI | RIPSER |
|---|---|---|---|---|---|---|---|
| **st** | **st** | **d** | **st** | **d** | **st** | **d** | **d** |
| $4.5 \cdot 10^8$ | $1.1 \cdot 10^7$ | $2.8 \times 10^9$ | $1.3 \cdot 10^9$ | $3.4 \cdot 10^9$ | $1 \cdot 10^7$ | $3.4 \cdot 10^9$ | $3.4 \cdot 10^9$ |

Table from [Otter et al. 2017]

# Bibliography

## General References:

- ***Books on TDA:***


  - H. Edelsbrunner, J. Harer. ***Computational topology: an introduction***. American Mathematical Soc., 2010.
  - R. W. Ghrist. ***Elementary applied topology***. Seattle: Createspace, 2014.


- ***Papers on TDA:***


  - G. Carlsson. ***Topology and data***. Bulletin of the American Mathematical Society 46.2, pages 255-308, 2009.


## Today's References:

- ***Software Packages:***


  - N. Otter, M.A. Porter, U. Tillmann, P. Grindrod, H.A. Harrington. ***A roadmap for the computation of persistent homology.*** EPJ Data Science, 6.1, 2017.