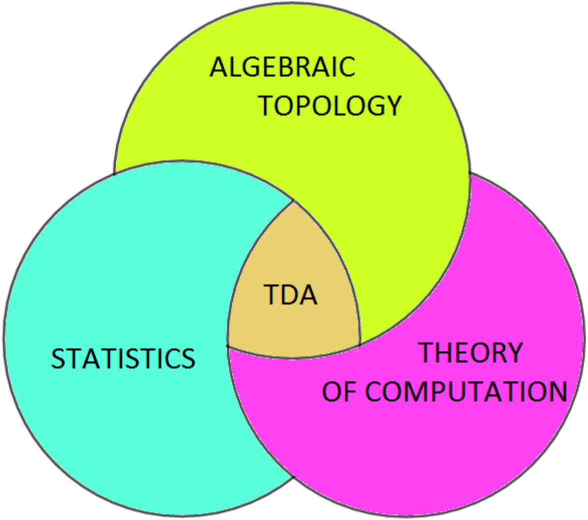
Persistent Homology and Entropy

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

June 2018

Topological Data Analysis

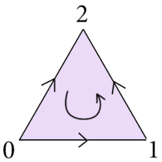


Data Structure

directed multigraphs

↓

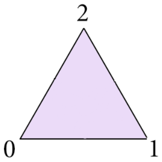
simplicial sets



undirected graphs

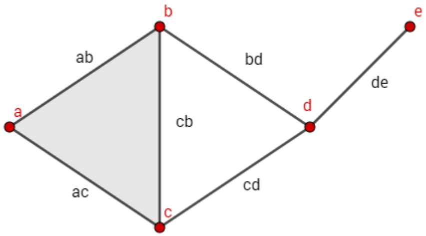
↓

abstract simplicial complexes



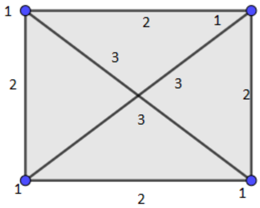
Abstract simplicial complexes are more suitable for computations.

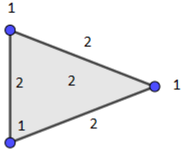
Simplicial Complexes

	Simplicial Complex
2-simplexes	$\{a, b, c\}$
1-simplexes	$\{a, b\}, \{b, c\}, \{a, c\}$ $\{b, d\}, \{c, d\}, \{d, e\}$
0-simplexes	$\{a\}, \{b\}, \{c\}, \{d\}, \{e\}$
	

Filtration

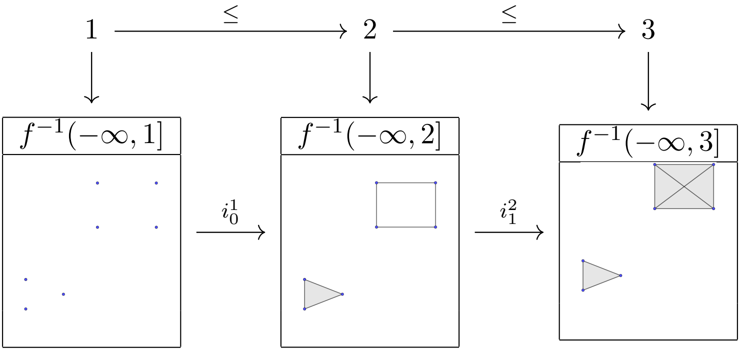
- Fix a simplicial complex SC.
- Define the **filter function** $f : SC \to \mathbb{R}$ s.t. $\sigma \lhd \tau \Rightarrow f(\sigma) \leq f(\tau)$.
- A **filtration** is the assignment $a \mapsto f^{-1}(-\infty, a]$.


1
2
1
3
2
3
2
3
1
2
1


1
2
2
2
1
1
2

Filtration

Each simplicial complex is related with the following ones by inclusion



Filtration (Abstract Concept)

- $\mathbb{K}$ are all subcomplexes of a simplicial complex seen as a category with morphism the inclusion.
- $\mathbb{P}$ is a poset seen as a category with morphism $\leq$.

Filtration

A filtration is a functor $F : \mathbb{P} \to \mathbb{K}$

Filtration (Abstract Concept)

- $\mathbb{K}$ are all subcomplexes of a simplicial complex seen as a category with morphism the inclusion.
- $\mathbb{P}$ is a poset seen as a category with morphism $\leq$.

Filtration

A filtration is a functor $F : \mathbb{P} \to \mathbb{K}$

In practice, $\mathbb{P}$ will be a subset of $\mathbb{R}$.

Persistent Homology

Let $\mathbf{Vec}$ be the category of vector space over a field k.

Persistent Homology

Let $\mathbf{Vec}$ be the category of vector space over a field k.

Persistence Module

A persistence module is a functor $M : \mathbb{P} \to \mathbf{Vec}$

Persistent Homology

Let $\mathbf{Vec}$ be the category of vector space over a field k.

Persistence Module

A persistence module is a functor $M : \mathbb{P} \to \mathbf{Vec}$

We call n-persistent homology to the functor

$$M = H_n(-, k) \circ F$$

Persistent Homology

Let $\mathbf{Vec}$ be the category of vector space over a field k.

Persistence Module

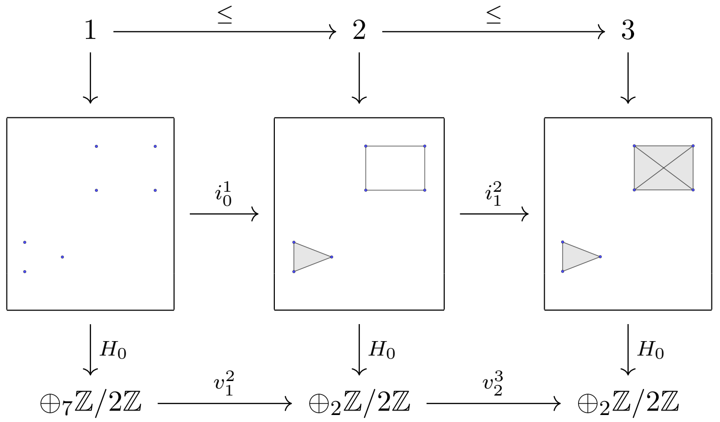
A persistence module is a functor $M : \mathbb{P} \to \mathbf{Vec}$

We call n-persistent homology to the functor

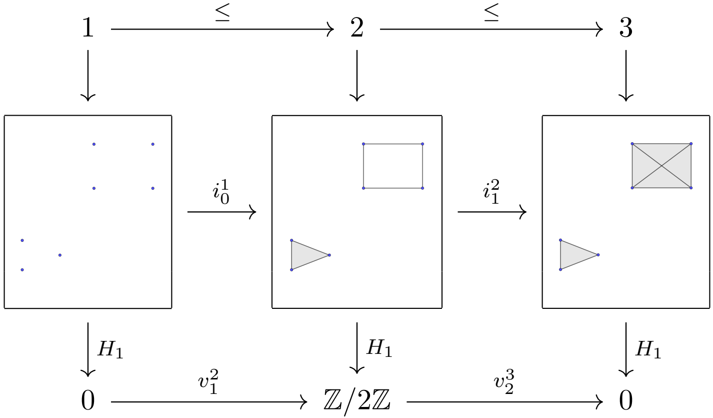
$$M = H_n(-, k) \circ F$$

Where H_n is the n-homology and usually $k = \mathbb{Z}/2\mathbb{Z}$.

Persistent Homology



Persistent Homology



Barcodes & Persistent Diagrams

If J is an interval in $\mathbb{R}$, a **interval module** is k^J is:

$$v_t = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{if } t \in J \\ 0 & \text{otherwise.} \end{cases} \qquad v_t^l = \begin{cases} Id & \text{if } [t,l] \subset J \\ 0 & \text{otherwise.} \end{cases}$$

Barcodes & Persistent Diagrams

If J is an interval in $\mathbb{R}$, a **interval module** is k^J is:

$$v_t = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{if } t \in J \\ 0 & \text{otherwise.} \end{cases} \qquad v_t^l = \begin{cases} Id & \text{if } [t,l] \subset J \\ 0 & \text{otherwise.} \end{cases}$$

Theorem (simplified)

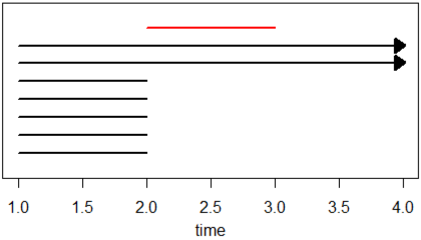
If M is a *nice* persistent homology module in $\mathbb{R}$ then

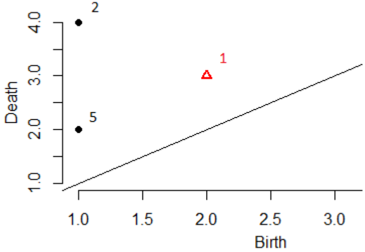
$$M \cong \oplus_{i=1}^{n} k^{J_i}$$

where $J_i = (x_i, y_i)$

Barcodes & Persistent Diagrams

We can represent the multiset $\{(x_i, y_i)\}_{i=1}^{n}$ in several ways

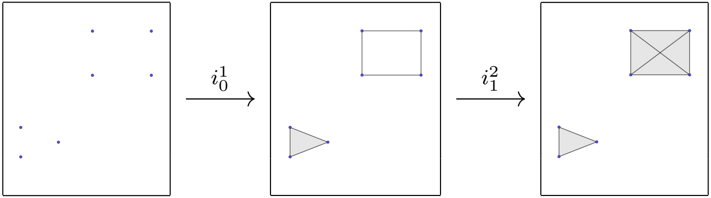




Short bars and points near the diagonal are usually considered noise

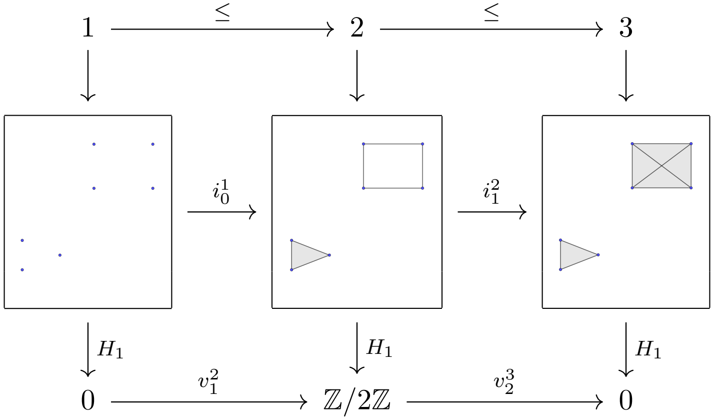
Persistent Homology

$$1 \xrightarrow{\leq} 2 \xrightarrow{\leq} 3$$



$$\oplus_7 \mathbb{Z}/2\mathbb{Z} \xrightarrow{v_1^2} \oplus_2 \mathbb{Z}/2\mathbb{Z} \xrightarrow{v_2^3} \oplus_2 \mathbb{Z}/2\mathbb{Z}$$

Persistent Homology



Barcodes & Persistent Diagrams

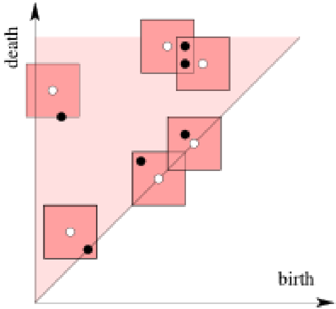


Figure: Computational Topology: An Introduction. H. Edelsbrunner, J. Harer.

Barcodes & Persistent Diagrams

- A, B persistent diagrams,
- $M \subset A \times B$ is a partial matching if satisfies
  - $\forall$ in A there is at most one β with $(\alpha, \beta) \in M$
  - $\forall \beta$ there is at most one α with $(\alpha, \beta) \in M$

There exist a δ-matching if:

- for all pairs (α, β), $||\alpha - \beta||_\infty \leq \delta$
- for all α unmatched, $|\alpha_2 - \alpha_1| \leq \delta$. Idem for β.

Stability Theorem

Bottleneck Distance

Let A, B be two persistent diagrams, then

$$d(A,B) = \inf\{\delta | \text{ exists a } \delta\text{-matching}\}$$

is their bottleneck distance

Stability theorem (simplified version)

Let K be a simplicial complex and $f, g : K \to \mathbb{R}$ filter functions. If A, B are the corresponding persistent diagrams,

$$d(A,B) \leq ||f - g||_\infty$$

Statistical disadvantage

Problems of barcodes:

- There are not obvious algebraic operation defined on them,

Statistical disadvantage

Problems of barcodes:

- There are not obvious algebraic operation defined on them,
- They do not have unique Fréchet mean,

Statistical disadvantage

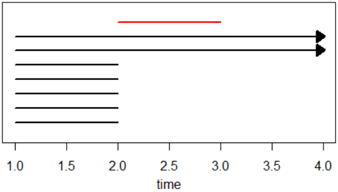
Problems of barcodes:

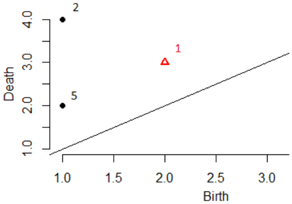
- There are not obvious algebraic operation defined on them,
- They do not have unique Fréchet mean,
- They are uncomfortable for applying the null hypothesis significance test and confidence intervals.

We can associate to each barcode a **real number** and use it to obtain statistical information.

We can associate to each barcode a **real number** and use it to obtain statistical information.

In some applications, counting the number of bars/points seems to work.





8

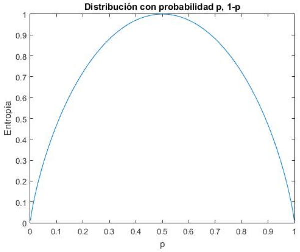
This is not **stable** respect to noise

Our research

Definition

given a finite random distribution $A = \{p_i : \sum_{i=1}^{n} p_i = 1\}$ we define its shannon entropy as

$$E(A) = -\sum_{i=1}^{n} p_i \log(p_i).$$

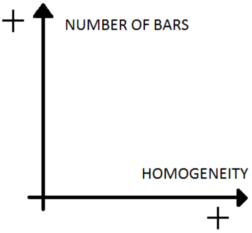


Our Research

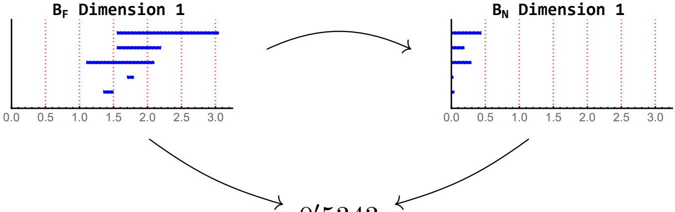
Persistent Entropy

Let $\{\ell_i\}$ be the length of the bars and $L = \ell_1 + \ldots + \ell_n$. The persistent entropy of a barcode is

$$E = -\sum_{i=1}^{n} \frac{\ell_i}{L} \log\left(\frac{\ell_i}{L}\right).$$



Our Research



$$0'5343$$

Results

Stability Theorem

Let K be a simplicial complex and let $f, g : K \to \mathbb{R}$ be two monotonic functions. Let A, B be their barcodes.

$$||f - g||_\infty \leq \delta \Rightarrow |E(A) - E(B)| \leq \frac{4\delta}{\ell_{\max}} \left[\log(n_{\max}) - \log\left(\frac{4\delta}{\ell_{\max}}\right)\right].$$

- n_{max} is the maximum number of bars of the barcodes.
- $\ell_{\max}$ is the maximum normalized length of the bars.

N. Atienza et al. Stability of persistent entropy and new summary functions for TDA.

Applications

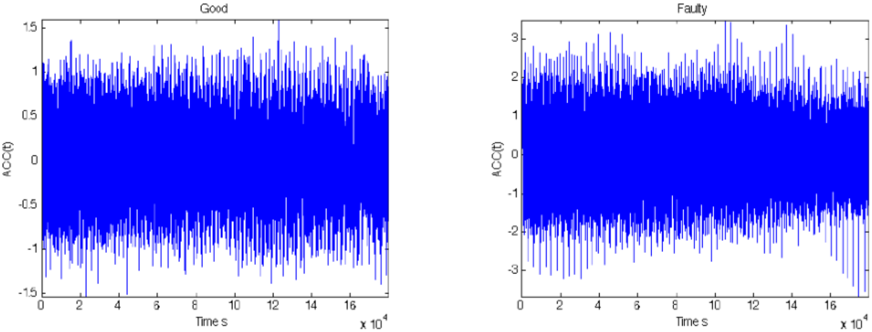
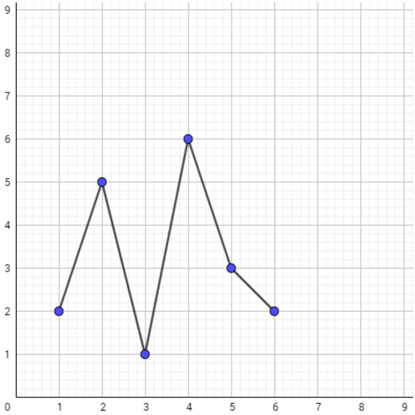


Figure: R. Gonzalez-Diaz et al. A new topological entropy-based approach for measuring similarities among piecewise linear functions.

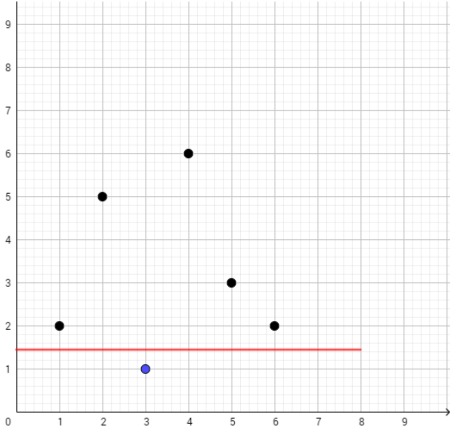
Applications

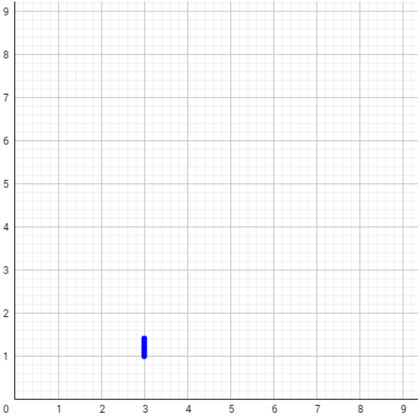


1. Vertex $\{p_i = (x_i, y_i)\}$.
   Edges $\{(p_i, p_{i+1})\}$.
2. Define the filter function
   $f(p_i) = y_i$
   $f(p_i, p_{i+1}) = \max\{y_i, y_{i+1}\}$

Applications

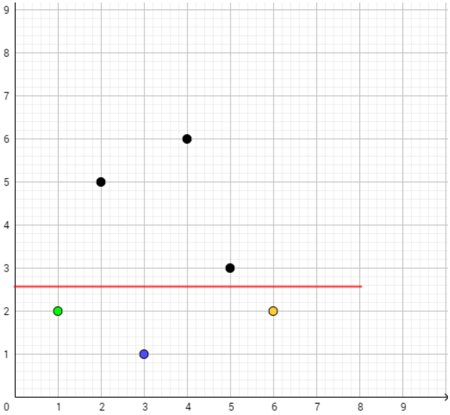
$f^{-1}(-\infty, 1'5)$

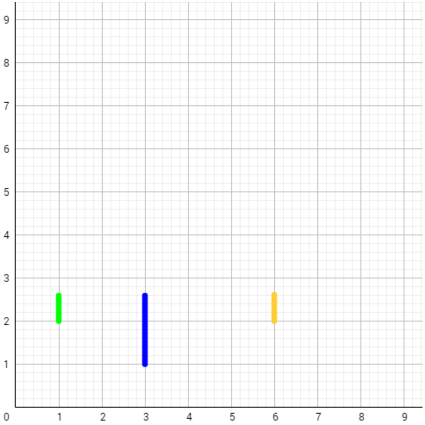




Applications

$f^{-1}(-\infty, 2'7)$

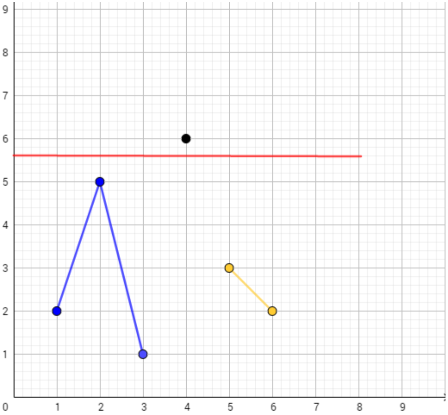


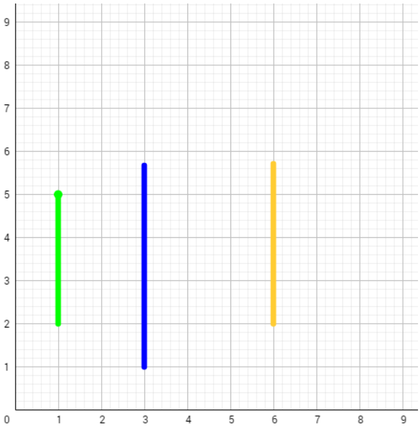


Applications

Applications

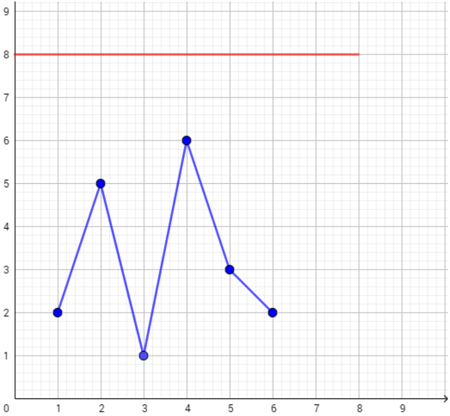
$f^{-1}(-\infty, 5'7)$

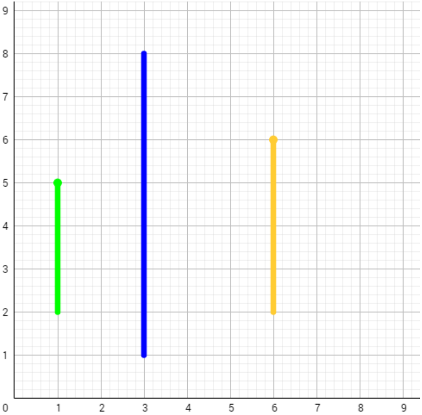




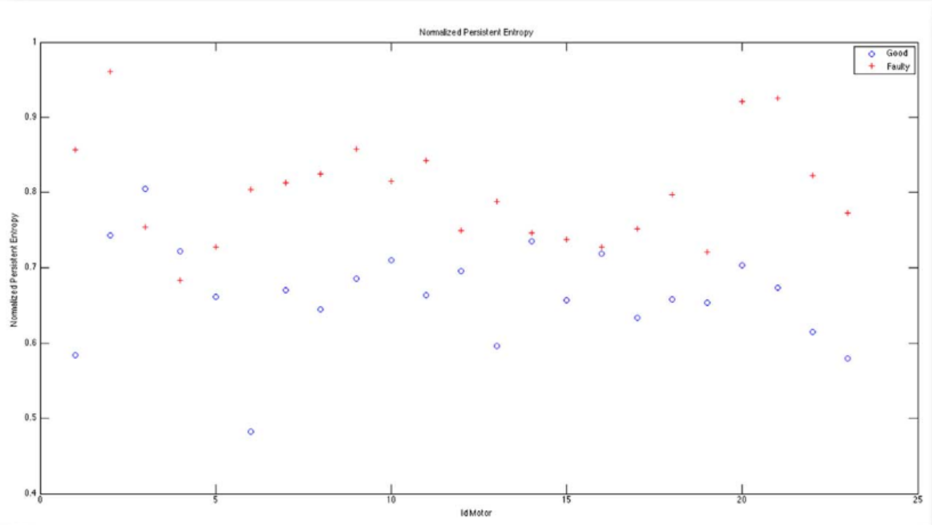
Applications

$f^{-1}(-\infty, 8)$





Our Research



Delving into TDA

- Developing representation and stability results for new posets, specially $\mathbb{R}^n$.
- Finding computational efficient functors, apart from homology, suitable for applications. (E.g. A_∞-coalgebras).
- Understanding the homotopy types of most common filtrations. (E.g. What are the homotopy types of the Vietoris-Rips complex of $\mathbb{S}^n$?)

Bibliography

Email: msoriano4@us.es

Computational Topology.

- H. Edelsbrunner, J.L. Harer. Computational Topology: an introduction.

Persistent modules.

- S.Y. Oudot. Persistence Theory: From Quiver Representations to Data Analysis.
- F. Chazal, V. de Silva, M. Glisse, S.Y. Oudot. The Structure and Stability of Persistence Modules.

Topological data analysis.

- G. Carlsson. Topology and Data
- R. Ghrist. Barcodes: the persistent topology of data.


THANK YOU
TEŞEKKÜR
הדות
감사합니다
GRACIAS
ĎAKUJEM
DĚKUJI
DANKE
спасибо
ХВАЛА
TERIMA KASIH.
NGIYABONGA
הדות
KIITOS
DANKE
ARIGATÔ
OBRIGADO
TEŞEKKÜR
GRACIAS
DHANYAVĀD
TAK
감사합니다
DHANYAVĀD
KÖSZÖNÖM
ευχαριστώ
MERCI
ARIGATÔ
спасибо
GRÀCIES
THANK YOU
SHUKRAN
XIÈXIÈ
DĚKUJI
TAK
ευχαριστώ
DĚKUJI
MERCI
GRAZIE
OBRIGADO
GRACIAS
XIÈXIÈ
הדות
DANKE
TAK
GRÀCIES
KIITOS
감사합니다
TERIMA KASIH.
HVALA
THANK YOU
TERIMA KASIH.
KOP KHUN
DANKE
KOP KHUN
DHANYAVĀD
TACK
KIITOS
감사합니다
OBRIGADO
MERCI
GRAZIE
TEŞEKKÜR
KÖSZÖNÖM
DANKE
TERIMA KASIH.
ХВАЛА
THANK YOU
ĎAKUJEM
KOP KHUN
감사합니다
ARIGATÔ
XIÈXIÈ
THANK YOU
GRAZIE
MERCI
GRACIAS
ХВАЛА
NGIYABONGA
SHUKRAN
TACK