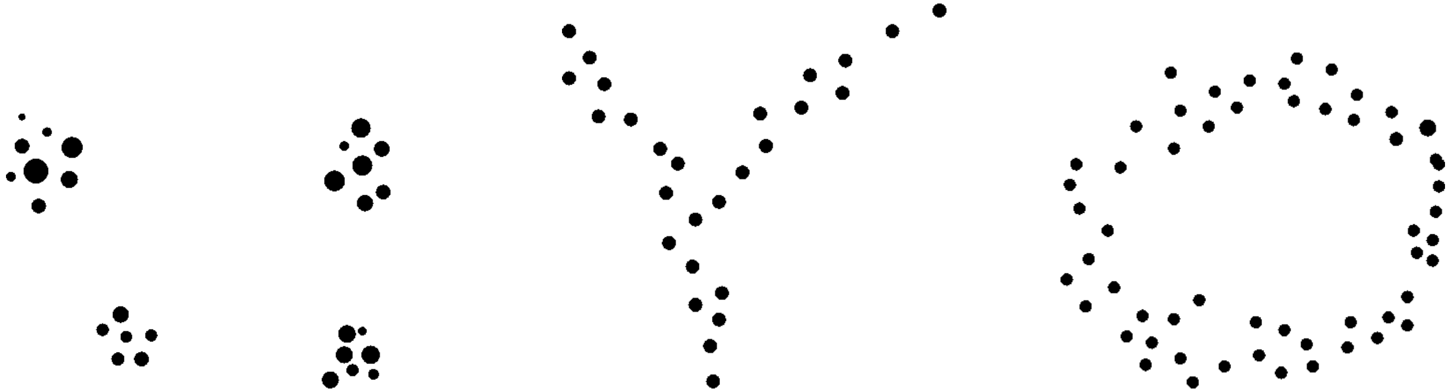
Challenges in persistence homology and applications

Ulrike Tillmann
Centre for Topological Data Analysis, Oxford and
Isaac Newton Institute, Cambridge

Kavli Institute, Tokyo
July 2025

Topological Data Analysis TDA

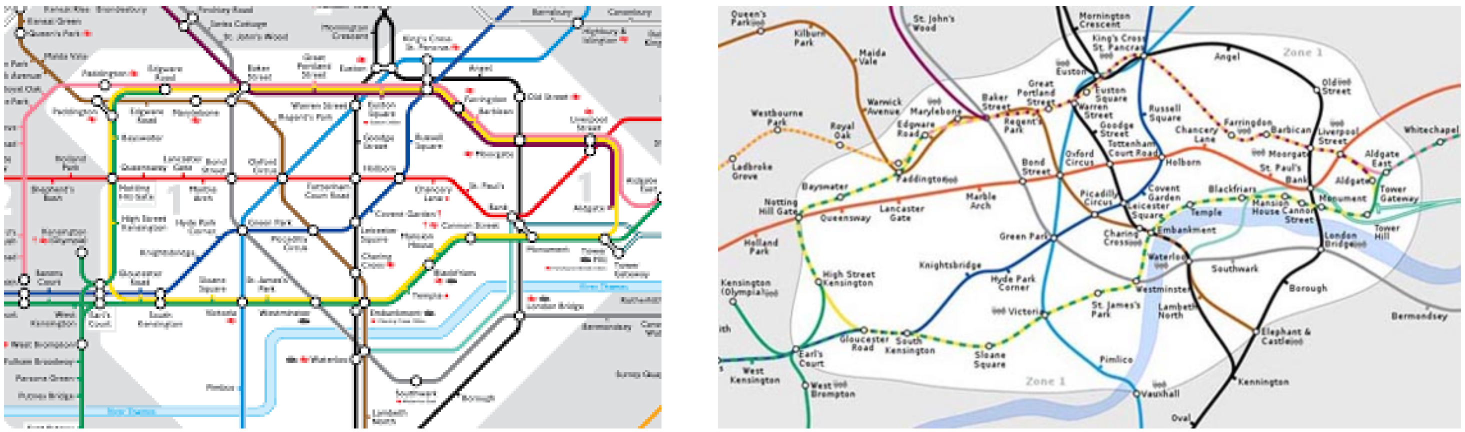
Shape and data



Clusters Y-formation Circle

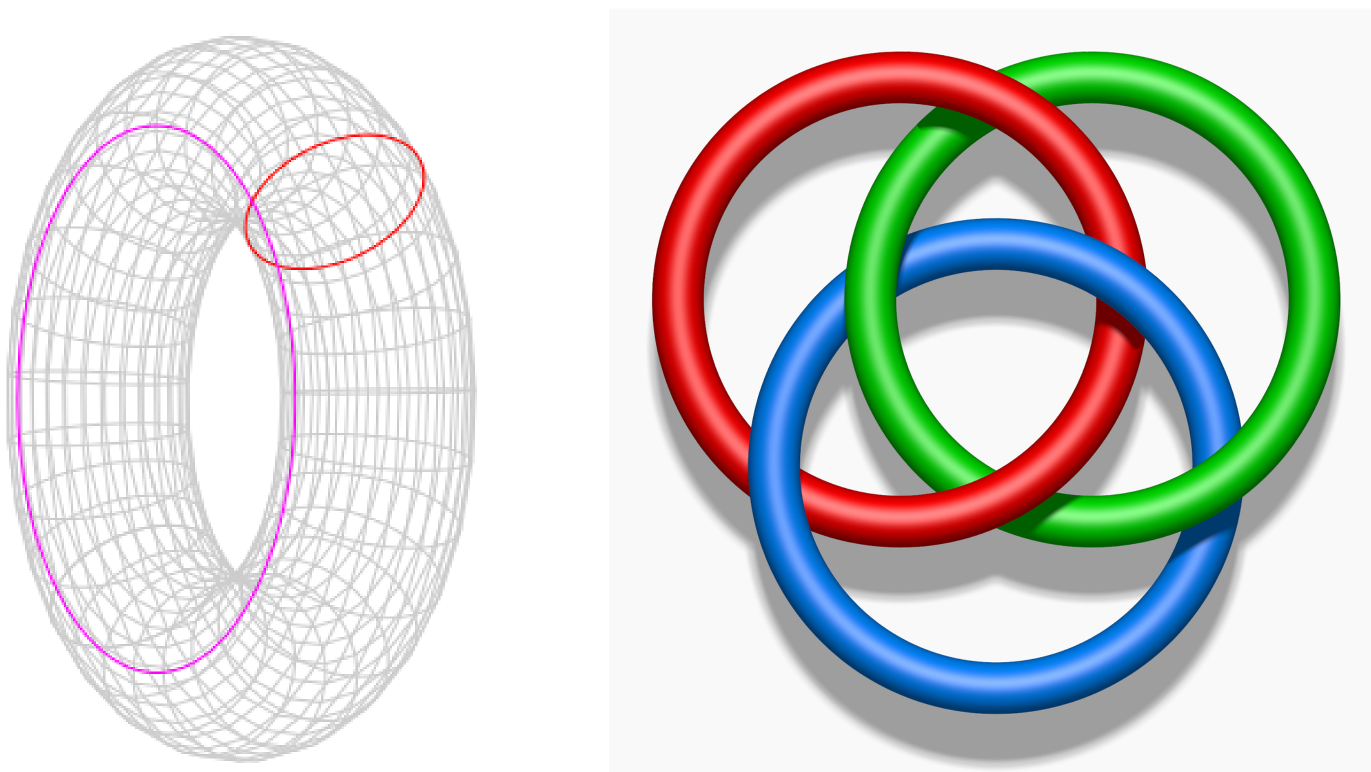
Why topology?

1. Suppressing information.



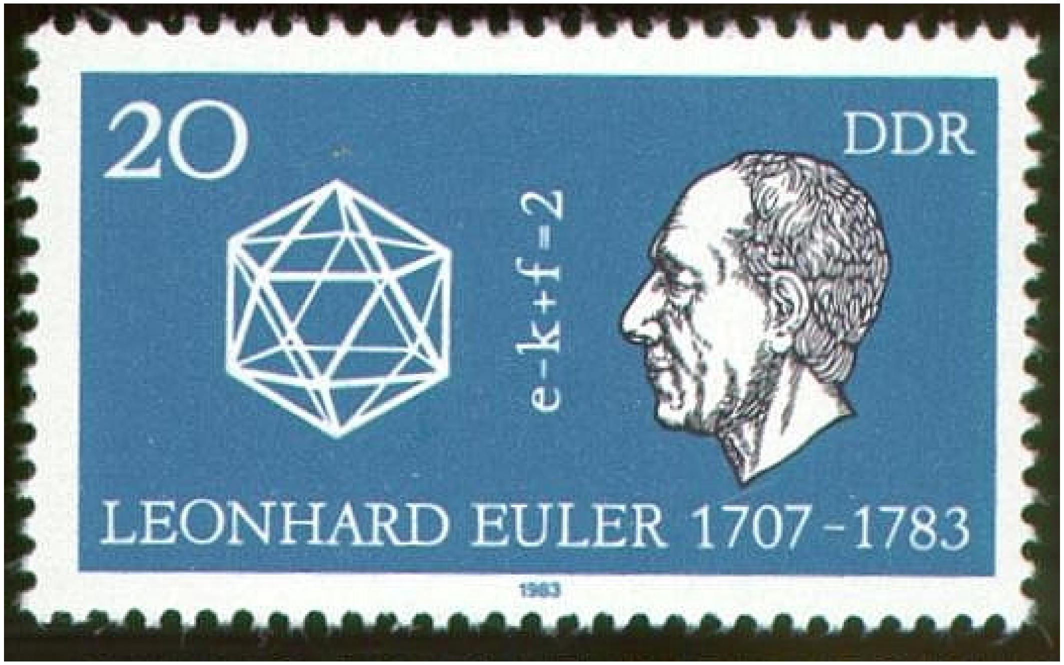
A topological and a geometric map

2. Higher dimensional information.



Torus and Borromean rings

3. Computable signatures.



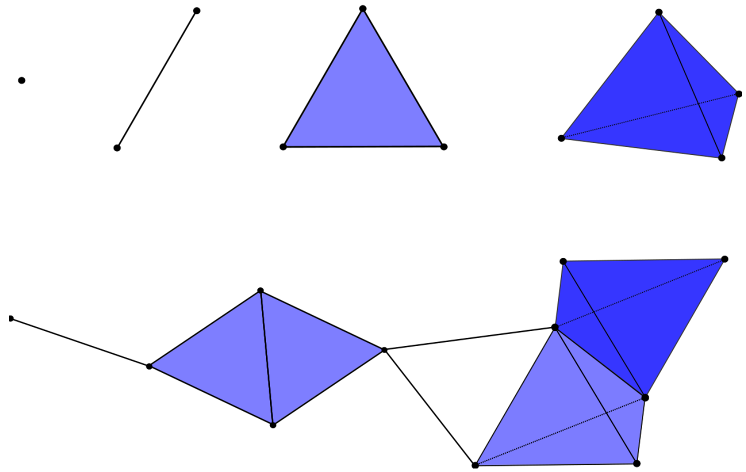
Euler characteristic:

$$\chi = \sharp\text{vertices} - \sharp\text{edges} + \sharp\text{faces}$$

Homology

(Persistent homology)

Combinatorial spaces K



Simplicial complex: K is a combinatorial representation of a topological space built from
$K_0 =$ vertices
$K_1 =$ edges
$K_2 =$ triangles
$K_3 =$ tetrahedra etc.

Homology of K

Chain complex (over a field $\mathbb{F}$)

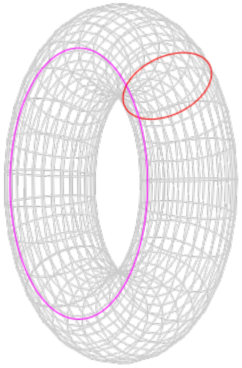
$$\ldots \mathbb{F}[K_{n+1}] \xrightarrow{d_{n+1}} \mathbb{F}[K_n] \xrightarrow{d_n} \mathbb{F}[K_{n-1}] \cdots \longrightarrow \mathbb{F}[K_1] \xrightarrow{d_1} \mathbb{F}[K_0]$$

Key-observation:
The boundary of a boundary is empty: $d_n \circ d_{n+1} = 0$

n-th **Homology group**: $H_n(K) := Ker(d_n)/Im(d_{n+1})$
n-th **Betti number**:

$$\beta_n := \text{dim} H_n(K) = \text{dim} Ker(d_n) - \text{dim} Im(d_{n+1})$$

$$\chi(K) = \beta_0 - \beta_1 + \beta_2 - \ldots$$



$\beta_0 = 1,\ \beta_1 = 2,\ \beta_2 = 1, \quad \chi = 0$

Functoriality: $L \subset K$ a subcomplex

induced map on chain complexes:

$$\mathbb{F}[L_n] \hookrightarrow \mathbb{F}[K_n]$$

induced map on homology:

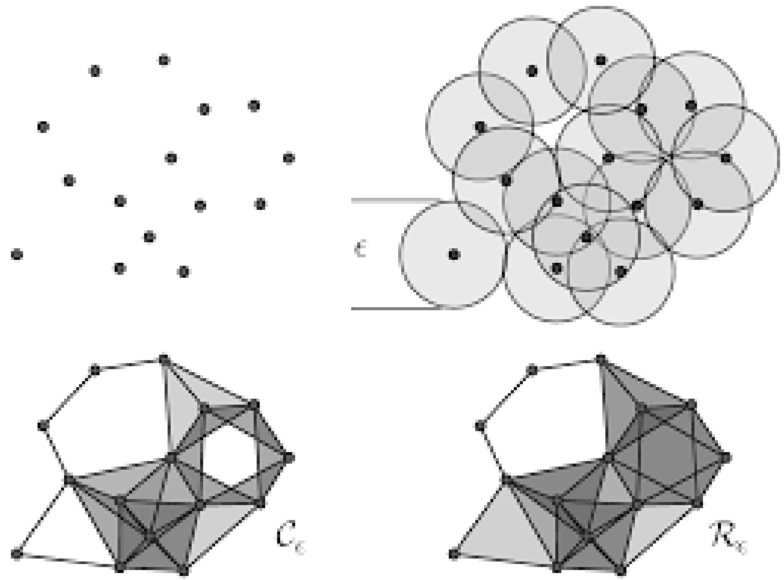
$$H_n(L) \longrightarrow H_n(K)$$

Henri Poincaré 1854–1912

Emmy Noether 1882–1935

From data to topological spaces

Topology of the ϵ-cloud



Čech-complex and Vietoris-Rips complex:

$K_0 =$ vertices
$K_1 =$ edges between point of distance $< \epsilon$
$K_2 = \ldots$

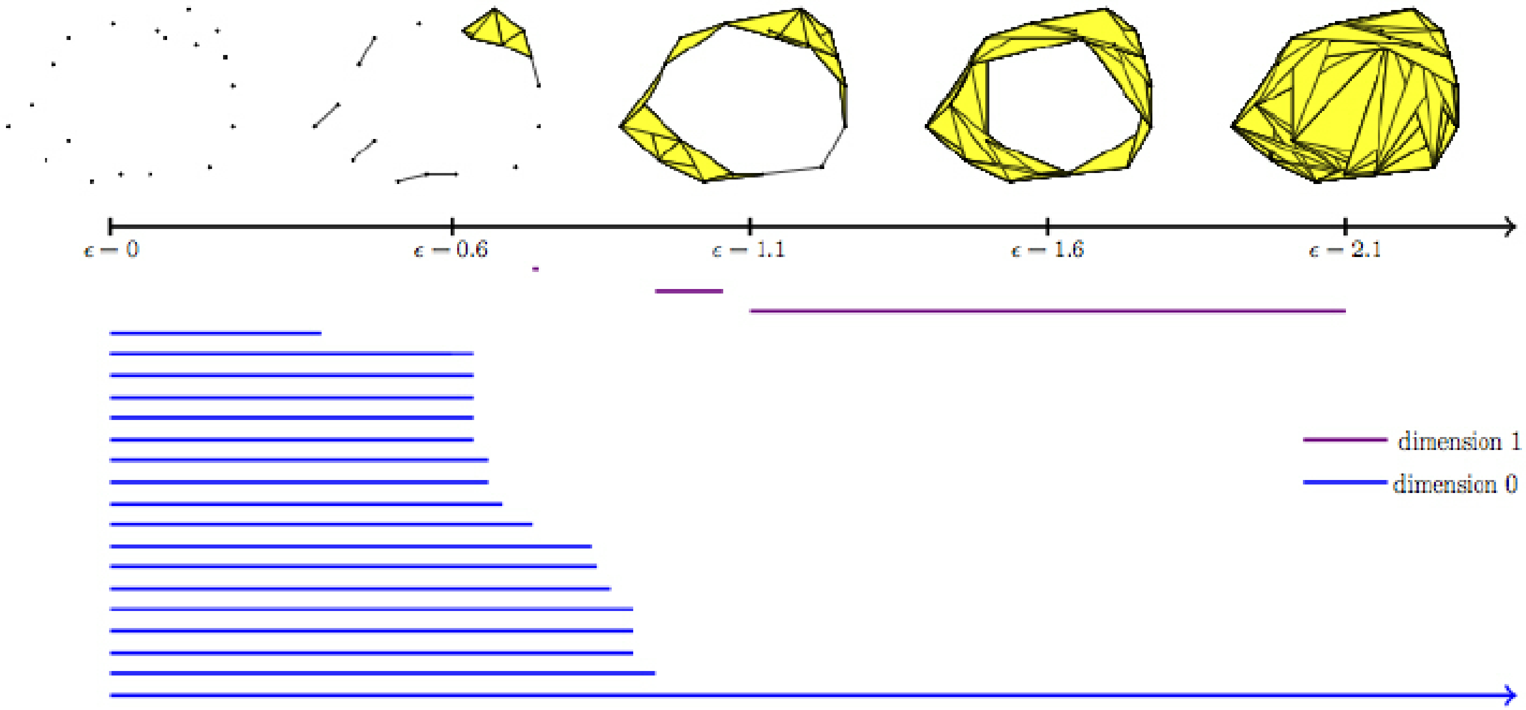
Nerve Theorem

For any point cloud in $\mathbb{R}^d$, its ϵ-cloud has the same homology as its ϵ-Čech-complex $\mathcal{C}_\epsilon$.

Also note:

$$\cdots \subset \mathcal{C}_\epsilon \subset \mathcal{R}_\epsilon \subset \mathcal{C}_{2\epsilon} \subset \ldots$$

Persistent homology



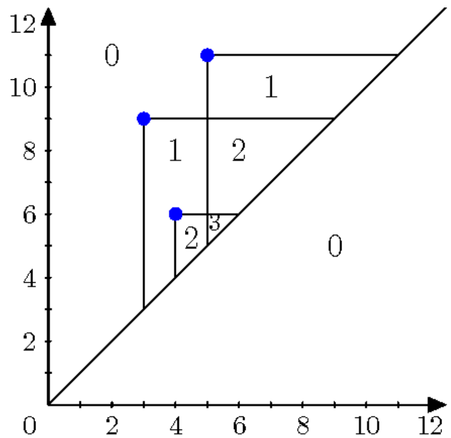
Source: Otter et al.

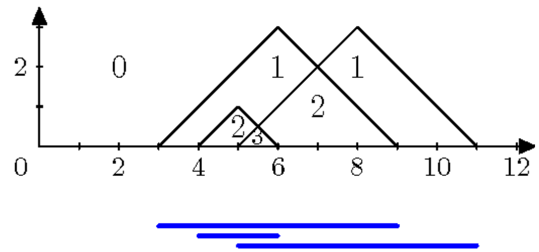
META THEOREMS

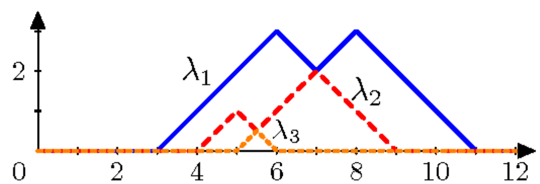
Existence of Barcodes [Carlsson-Zomorodian 2005]:
One can always find barcodes representing persistent homology (in a unique way upto permutation).

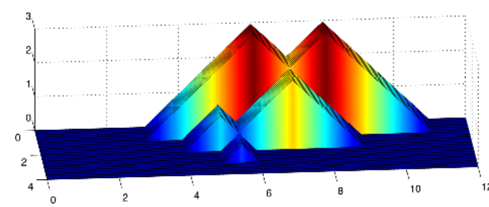
Stability Theorem [Cohen-Steiner-Edelsbrunner-Harer 2007]:
Small changes of input data result in small changes of output statistics.

Persistence diagrams and landscapes









Bubenik, JMLR 2015

- Equivalent information
- Advantages: (1) averages; (2) standard norms
- Properties: Law of Large Numbers, Central Limit Theorem

Computing PH

Survey

Otter, Porter, Tillmann, Grindrod, Harrington,
A Roadmap for the computation of persistent homology,
EPJ Data Science (2017),

Statistics

Fasy, Kim, Lecci, Maria, Rouvreau
TDA: Statistical Tools for Topological Data Analysis,
Introduction to the R package TDA (2019)

New improved software – inspired by Morse theory

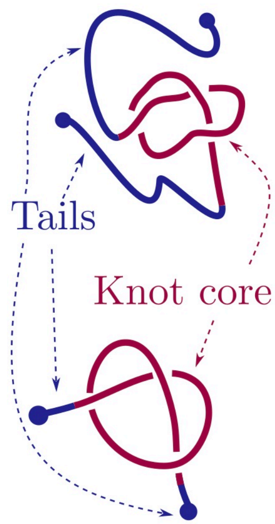
Bauer: Ripser (2019)

Identification of cycles

Henselman: Eirene

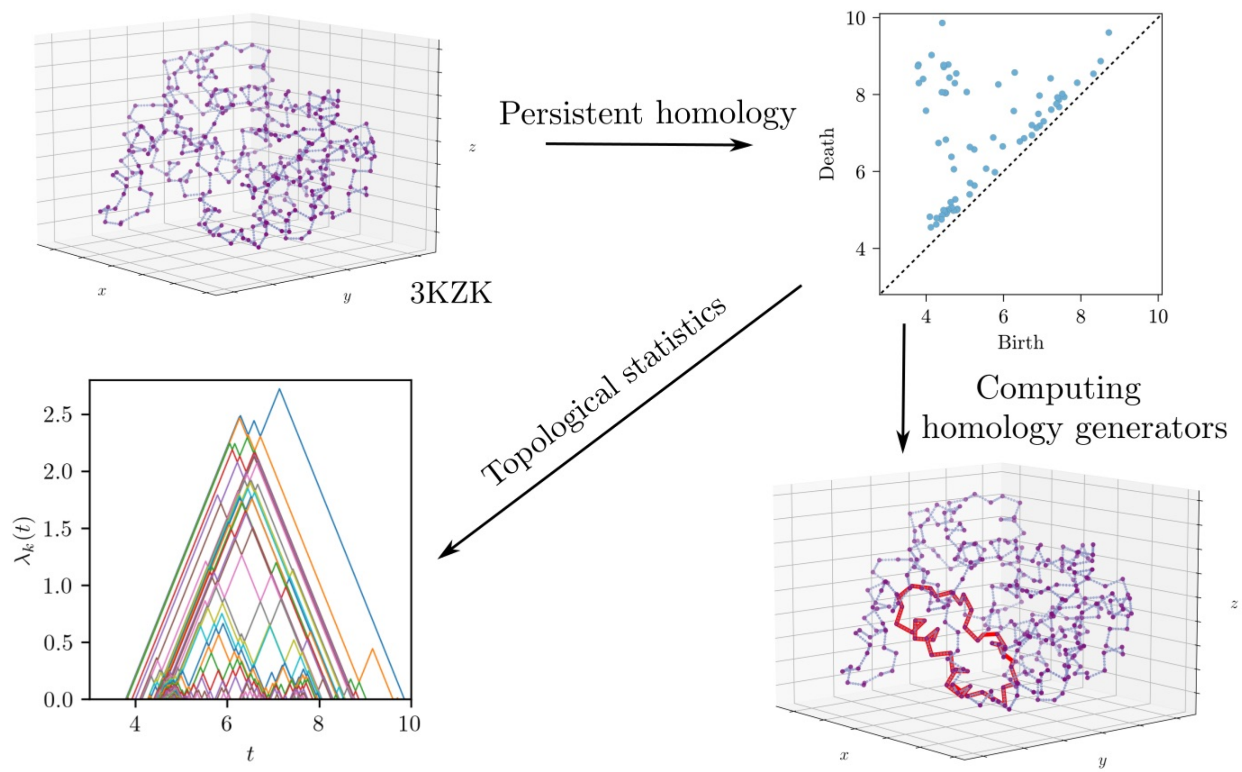
Applications of PH

1. Application: Knotted proteins



Knot depth: $D(Knot) = \frac{L(Tail_1)L(Tail_2)}{L(Knot)}$

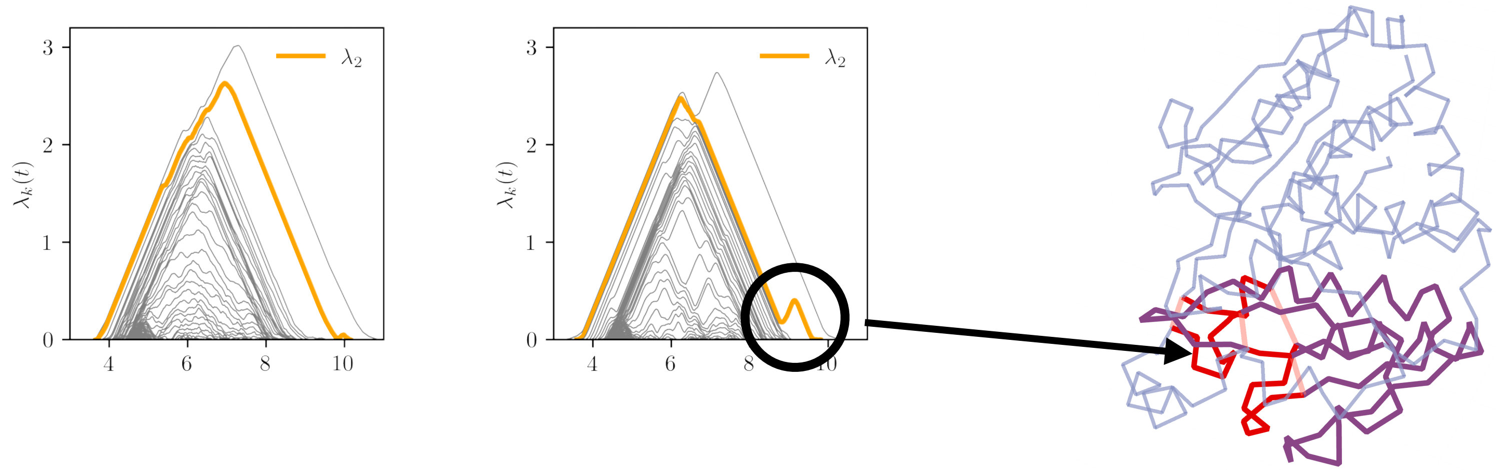
Protein-databank: PDB (aminoacid-chains: 3KZK, 4JQO, ..)
Knotted proteins: KnotProt (over 1000; structural stability (?))
Trefoil knots : most



TDA - Protein - Pipeline

Protein: C_α carbon point cloud — Persistence Diagram

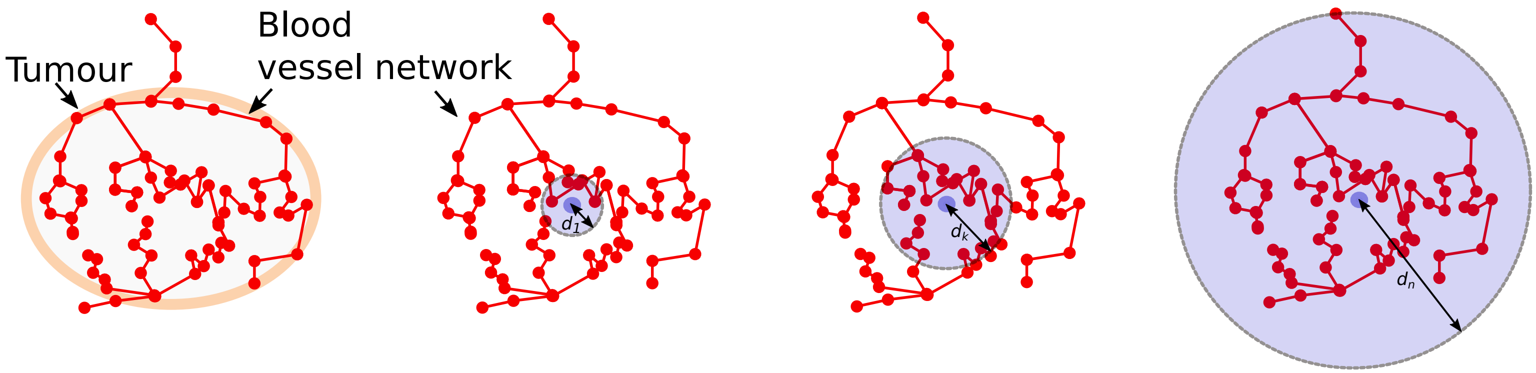
PH Landscape — Identification of cycles

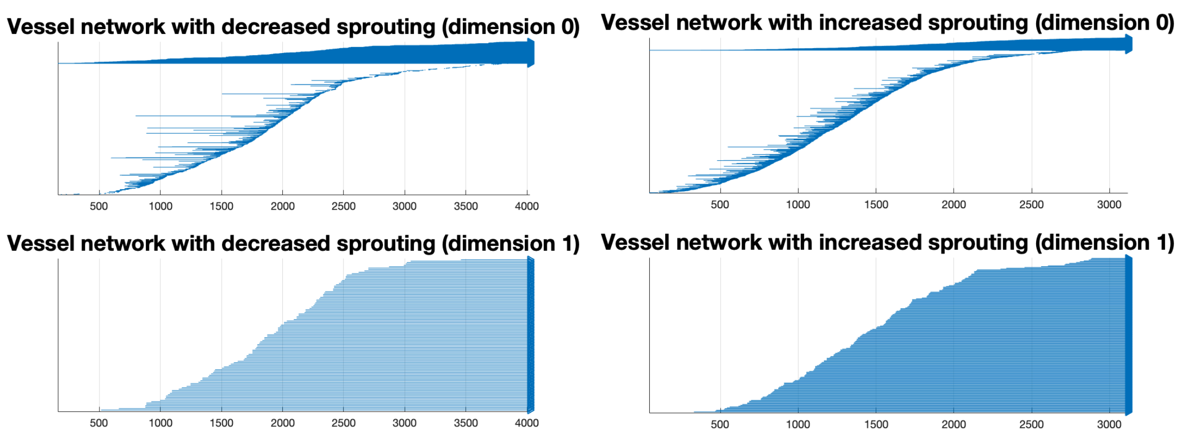


Right: knotted 3KZK overlaid with un-knotted 4JQO
Left: landscape averages: λ_2 in yellow
Right: extra cycle for 3KZK: contains critical crossing

Benjamin-Mukta-Moryoussef-Uren-Harrington-Tillmann-Barbensi,
J R Soc Interface 2023

2. Application: Blood vessels in tumors





Tortuosity (H_0) and cycles (H_1) Stolz et al. *IMA* 2019

An Introduction to Topological Data Analysis for Physicists: From LGM to FRBs

Jeff Murugan & Duncan Robertson

Laboratory for Quantum Gravity & Strings
Department of Mathematics and Applied Mathematics
University of Cape Town

April 26, 2019

Published for SISSA by Springer

Received: January 24, 2019
Accepted: February 25, 2019
Published: March 11, 2019

Topological data analysis for the string landscape

Alex Cole and Gary Shiu

Department of Physics, University of Wisconsin,
Madison, WI 53706, U.S.A.

E-mail: acole4@wisc.edu, shiu@physics.wisc.edu

Abstract: Persistent homology computes the multiscale topology of a data set by using a sequence of discrete complexes. In this paper, we propose that persistent homology may be a useful tool for studying the structure of the landscape of string vacua. As a scaled-down

Procedure to Reveal the Mechanism of Pattern Formation Process by Topological Data Analysis

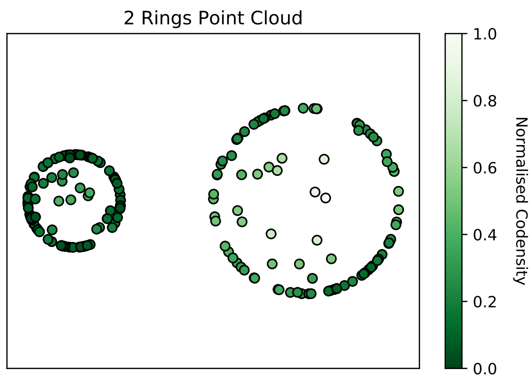
Yoh-ichi Mototake[a], Masaichiro Mizumaki[b], Kazue Kudo[c,d], Kenji Fukumizu[e]

[a] *Graduate School of Social Data Science, Hitotsubashi University, 2-1, Naka, Kunitachi, 186-8601, Tokyo, Japan*
[b] *Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto, 860-8555, Kumamoto, Japan*
[c] *Department of Computer Science, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, 112-8610, Tokyo, Japan*
[d] *Department of Computer and Mathematical Sciences, Tohoku University, 6-3-09 Aoba, Aramaki-aza Aoba-ku, Sendai, 980-8579, Miyagi, Japan*
[e] *The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, 190-8562, Tokyo, Japan*

Hiraoka group at Kyoto University

Okinawa Institute of Science and Technology (24.06-08.08.2025)
TSVP Symposium: Representation Theory and TDA

Multiparameter PH and Noise



Multiparameter persistent homology:

$$MPH : (\mathbb{R}^n, \leq) \xrightarrow{K} \triangle - \text{Complex} \xrightarrow{H_p} \text{Vect}_{\mathbb{F}}$$

Multiparameter persistence modules = functors of categories:

$$V : (\mathbb{R}^n, \leq) \longrightarrow \text{Vect}_{\mathbb{F}}$$

- Interleaving distance well-defined (NP-hard for $n > 1$)
- Stability holds
- No analogue of barcodes (Gabriel's theorem)

Carlsson-Zomorodian, Lesnick-Wright, Miller, Harrington-Otter-Schenck-Tillmann, ...

Fix a multiparameter module $V : (\mathbb{R}^n, \leq) \longrightarrow \mathsf{Vect}_{\mathbb{F}}$

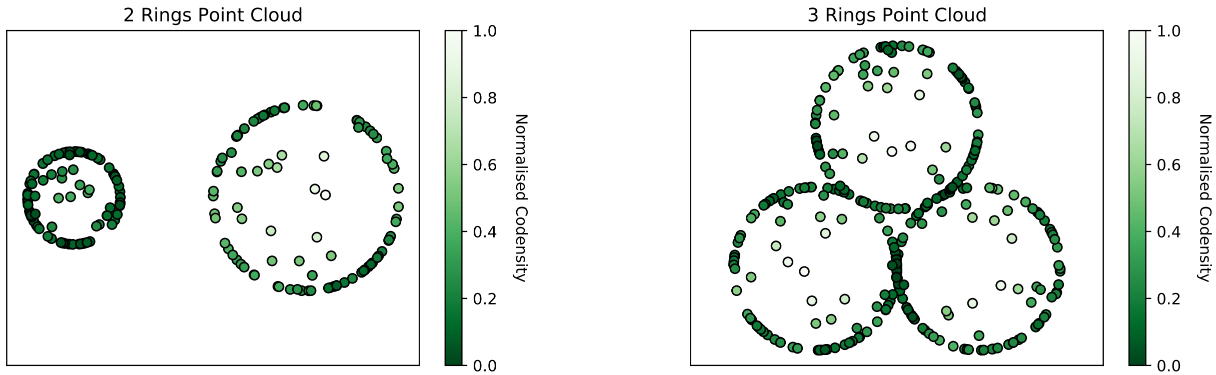
Rank invariant:
$\mathrm{rank}(\bar{s}, \bar{s} + \bar{t}) := \dim \mathrm{Im}(V(\bar{s}) \to V(\bar{s} + \bar{t}))$

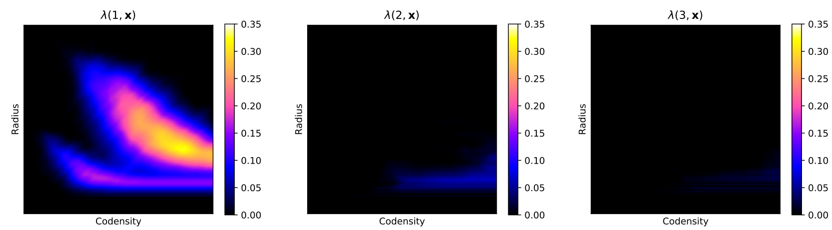
Multiparameter persistence landscape:
$\lambda : \mathbb{N} \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ defined by

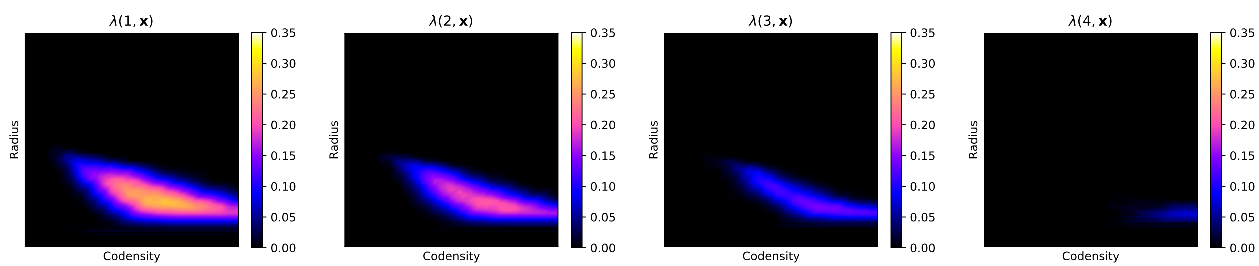
$$\lambda(k, \bar{x}) = \sup\{\epsilon \geq 0 : \mathrm{rank}(\bar{x} - \epsilon\bar{1}, \bar{x} + \epsilon\bar{1}) \geq k\}$$

- Stability
- Strong Law of Large Numbers
- Central Limit Theorem

$\rightsquigarrow$ approximate normal distributions and confidence intervals



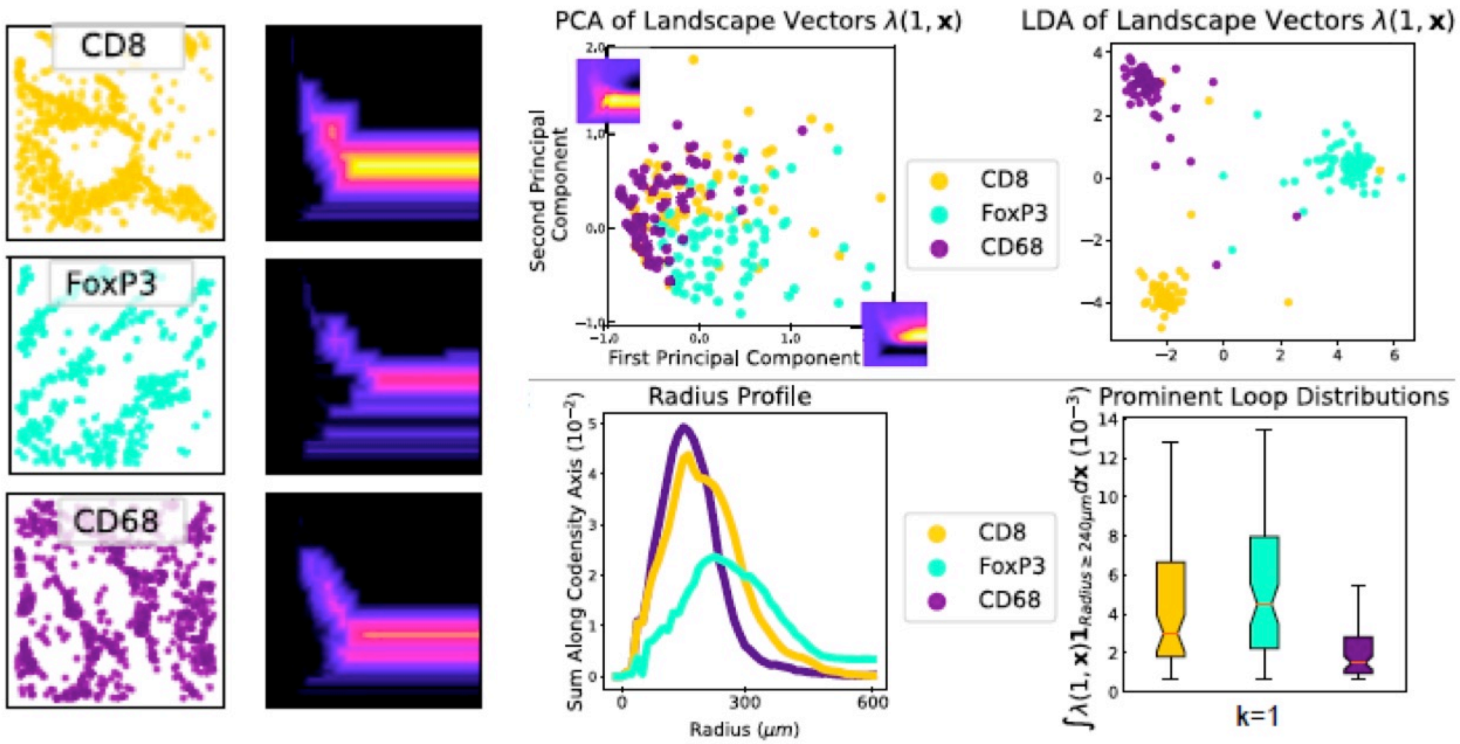




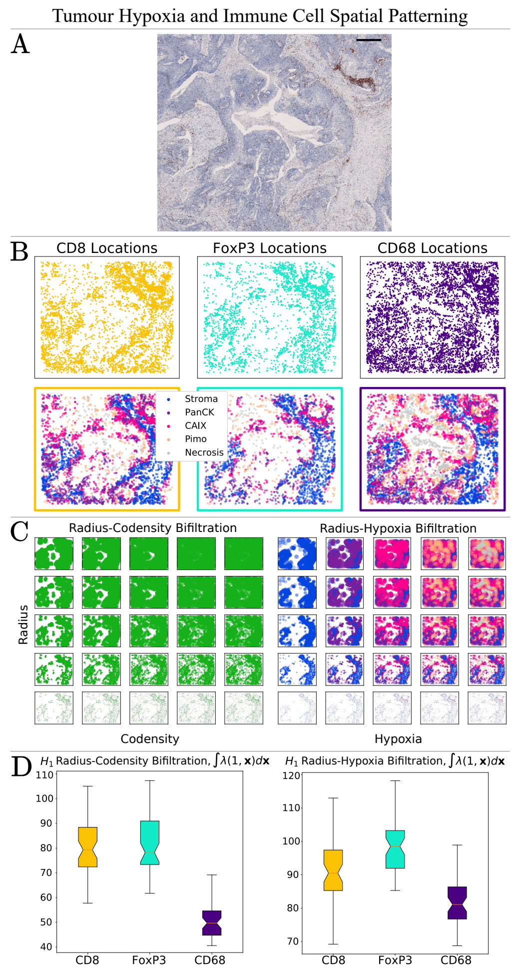
Source: Vipond et al

Applications of MPH

4. Application: immune cells in tumours



Head-Neck Tumor: T-cells (CD8, FoxP3), macrophages (CD68)
PCA: Principle Component Analysis
LDA: Linear Discriminant Analysis



Multi-parameter persistence: density/radius $\sim$ hypoxia/radius

References:

Vipond: Multiparameter Persistence Landscapes
. *J. Mach. Learn. Res 2020*

Vipond, Bull, Macklin, Tillmann, Pugh, Byrne, Harrington: Multiparameter persistence homology landscapes identify spatial patterns of immune cells in tumors *PNAS 2021*

Benjamin, ..., Tillmann, Harrington, Bull: Multiscale topology classifies cells in subcellular spatial transcriptomics *Nature 2024*

Theory

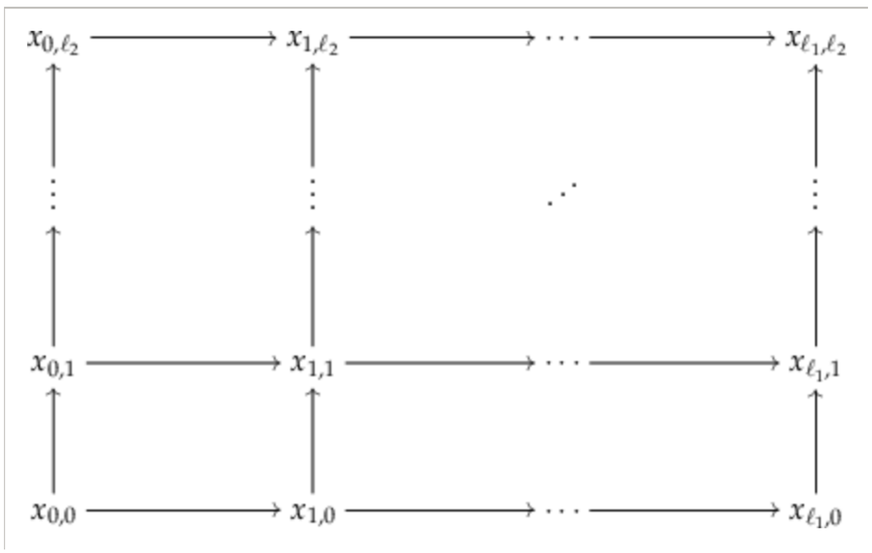
Quiver representations

$Q = (Q_0, Q_1)$ – a quiver with vertices Q_0 and arrows Q_1

Example: A_l ordinary and zig-zag

$$x_0 \overset{e_1}{\text{———}} x_1 \overset{e_2}{\text{———}} \cdots \overset{e_{\ell-1}}{\text{———}} x_{\ell-1} \overset{e_\ell}{\text{———}} x_\ell.$$

Example: multiparameter grid



$\mathcal{C}_Q$ associated category

Definition:
A quiver representation is a functor $V : \mathcal{C}_Q \to \mathsf{Vect}_{\mathbb{F}}$

V_x f.d. vector space for $x \in Q_0$
V_e linear map for $e \in Q_1$

Example: Identity representation on subquiver $Q' \subset Q$

$$I[Q']_x = \begin{cases} \mathbb{F} & \text{if } x \in Q'_0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad I[Q']_e = \begin{cases} \mathrm{id}_{\mathbb{F}} & \text{if } e \in Q'_1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Example: Subrepresentation generated at x

$< V_x >$ smallest subrepresentation of V containing V_x

If Q is finite,
there is a unique decomposition into indecomposables:

$$V \simeq \bigoplus_{I \in \mathrm{Ind}_Q} d_V(I).I$$

Gabriel's Theorem:
Ind_Q is finite if and only if the underlying graph of Q is a finite union of simply laced Dynkin diagrams (type A, D, E).

Ordinary and zigzag PH are well behaved.
MPH is not!

$==== >$ search for good invariants for MPH

Harder-Narasimhan filtration

$\mathcal{C}$ abelian category

stability condition:
$Z : K(\mathcal{C}) \to \{z \in \mathbb{C} | \, Re(z) > 0\}$ group homomorphisms

slope:
$\mu_Z(V) := \frac{ImZ(V)}{ReZ(V)}$ for $V \neq 0$

V is Z-**semi-stable** if $\mu_Z(V') \leq \mu_Z(V)$ for all $V' \subset V$
V is Z-**stable** if $\mu_Z(V') < \mu_Z(V)$ for all $V' \subset V$

Theorem:
Harder-Narasimhan filtration of V along Z

If $\mathcal{C}$ is both Noetherian and Artinian, then every non-zero $V \in \mathcal{C}$ admits a unique filtration $V^\bullet$:

$$0 = V^0 \subset V^1 \subset \cdots \subset V^n = V$$

with semi-stable quotients $S^k = V^k / V^{k-1}$ such that

$$\mu_Z(S^1) > \mu_Z(S^2) > \cdots > \mu_Z(S^n)$$

$\mathcal{C} = \mathrm{Rep}(Q)$ satisfies hypothesis for finite Q

$\underline{\dim} : K(\text{Rep}(Q)) \to \mathbb{Z}^{Q_0}$

(This is an isomorphisms for acyclic quivers)

If Z factors through $\underline{\dim}$ then Z is determined by $\alpha, \beta : Q_0 \to \mathbb{R}$:

$$Z(V) = \Sigma_{x \in Q_0}(\beta(x) + i\alpha(x)). \dim V_x$$

may assume $\beta(x) = 1$ for all $x \in Q_0$;
so Z determined by **central charge** α;
slope simplifies to

$$\mu_\alpha(V) = \frac{\Sigma_{x \in Q_0}\alpha(x) \dim V_x}{\Sigma_{x \in Q_0} \dim V_x}$$

Definition: HN-type
$HN[V;\alpha] := (\underline{\dim} S^1, \underline{\dim} S^2, \ldots, \underline{\dim} S^n)$

α is **complete** if

$$HN[V;\alpha] = HN[W;\alpha] \iff V \simeq W$$

Proposition:
If α is complete then every indecomposable representation in Rep(Q) is α-stable

Example: Euler
$\epsilon : Q_0 \to \mathbb{R}$ with $\epsilon(x) = 1 - \sharp$(in-arrows)

Example: skyscraper
$\delta_y : Q_0 \to \mathbb{R}$ with $\delta_y(x) = \delta_{xy}$

Quivers of type $\mathbb{A}$

Ordinary persistence: $\mathbb{A}_n : x_0 \to \cdots \to x_{n-1}$

Euler $\epsilon = \delta_{x_0}$

Theorem:
Let V be a persistence module and let $j_1 < \ldots < j_\ell$ be the collection of all indices j in $\{0, \ldots, n-1\}$ satisfying $I[0, j] \in \mathrm{Ind}_Q$, ie $[0, j] \in \mathrm{Bar}(V)$.
Then $V^\bullet$ has length $\ell + 1$, and for each integer $0 < k \leq \ell$ the quotient $S^k := V^k / V^{k-1}$ satisfies

$$\dim(S^k)_{x_i} = \begin{cases} d_{0,j_k} & i \in [0, j_k], \\ 0 & \text{otherwise.} \end{cases}$$

δ_{x_0} is not complete!

Kinser (2022) determined all complete central charges

Corollary:
For persistence modules, a central charge

$$\alpha : Q_0 \to \mathbb{R}$$

is complete if and only if the inequality $\alpha(x_i) > \alpha(x_{i+1})$ holds for all $i \in \{0, 1, \ldots, \ell - 1\}$.

Complete charges have been computed for $\mathbb{A}$, $\mathbb{D}$ and $\mathbb{E}$ quivers.

Skyscraper invariant

Proposition:
Let $0 = V^0 \subsetneq \cdots \subsetneq V^n = V$ be the HN filtration of V **along** δ_x**.**
If j is the smallest index for which $V_x^j = V_x$, then:
(1) either $j = n$ or $j = n - 1$, and
(2) for every $1 \leq k \leq j$, we have $V^k = < V_x^k >$.

Proof: (1)

$$\mu_{\delta_x}(W) = \frac{\dim W_x}{\Sigma_{y \in Q_0} \dim W_y}$$

$V_x^j = V_x$. So $\mu_{\delta_x}(S^{j+1}) = \cdots = \mu_{\delta_x}(S^n) = 0$.
But these slopes are strictly decreasing.

(2) By induction.
Use that V^1 is semi-stable and $< V_x^1 >$ is a submodule.
Comparing slopes forces $V^1 = < V_x^1 >$.

Definition:

Skyscraper invariant $\delta_\bullet$ on Rep(Q) assigns to each representation V the collection of HN types along skyscraper central charges at all of the vertices:

$$\delta_V = \{HN[V; \delta_x] \mid x \in Q_0\}$$

Proposition
Skyscraper is complete on $\text{Rep}_{\mathbb{A}_n}$.

Rank invariant

Definition:
Rank invariant is the map $\rho_V : Q_0 \times Q_0 \to \mathbb{N}$ given by

$$\rho_V(x, y) = \dim < V_x >_y$$

This generalises Carlsson-Zomorodian's (2007) invariant for MPH to any quiver.

Theorem: The skyscraper invariant is strictly stronger than the rank invariant.

Proof: (1) $\rho(x, y) = \Sigma_{k=1}^{j} \dim S_y^k$ for j as in the Proposition. So the rank invariant can be recovered.
(2) By example (on a commutative square)

Generalised rank invariant

Definition:

Generalised rank Invariant is the map GRI that for each sub-quiver $Q' \subset Q$ assigns the rank of the canonical limit-to-colimit map,

$$\mathrm{rk}(\varinjlim V|_{Q'} \longrightarrow \varprojlim V|_{Q'})$$

This generalises Kim and Mémoli's (2021) invariant for MPH to any quiver.

Theorem:

The skyscraper and the GRI are not comparable.

Restricting the category

$\mathsf{Rep}_{id}(Q)$: the category generated by identity representations $I[Q']$ on any subquiver $Q' \subset Q$.

Theorem:
GRI is complete on $\mathsf{Rep}_{id}(Q)$

This generalises Kim and Mémoli's (2021) theorem for MPH to any quiver.

Problem:
Determine a maximal subcategory on which the skyscraper invariant is complete.

$Q \subset \mathbb{Z}^d$ grid quiver of shape $L = (\ell_1, \ell_2, \ldots, \ell_d)$

$R = [a_1, b_1] \times \cdots \times [a_d, b_d]$ rectangle

$\text{Rep}_{rec}(Q)$: subcategory of direct sums of $I[R]$ for R and rectangle in Q

Theorem:
A central charge $\alpha \notin \mathcal{H}$ is complete on $\text{Rep}_{rec}(Q)$ if and only if it satisfies the inequality $\alpha \circ s(e) > \alpha \circ t(e)$ for each edge $e \in Q_1$.

Proof:
Uses lattice theory and max-flow/min-cut theorem.

References

Fersztand, Nanda, Tillmann : Harder-Narasimhan filtrations and zigzag persistence, *Advances in Applied Mathematics 2024*

Fersztand, Jacquard, Nanda, Tillmann : Harder–Narasimhan filtrations of persistence modules, *Transactions LMS 2024*

Fersztand : Harder-Narasimhan filtrations of persistence modules: metric stability, *arXiv 2025*

Fersztand, Jendrysiak : Computing the skyscraper invariant, *forth coming*

"Data has shape, and shape matters"

Gunnar Carlsson

EPSRC
Engineering and Physical Sciences
Research Council

The Alan Turing Institute

UNIVERSITY OF OXFORD

Centre for Topological Data Analysis