HELMHOLTZ MUNICH

AIH Institute of AI for Health

Topological Machine Learning: The (W)Hole Truth

Lecture 6

Bastian Rieck (@Pseudomanifold)

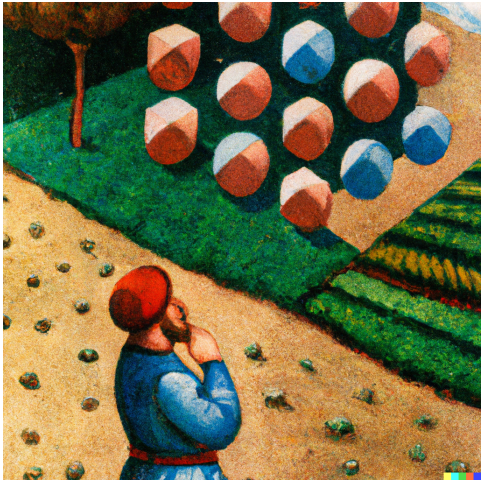
Preliminaries

Do you have feedback or any questions? Write to bastian.rieck@helmholtz-muenchen.de or reach out to @Pseudomanifold on Twitter. You can find the slides and additional information with links to more literature here:

`https://heidelberg.topology.rocks`

In this lecture

Reaping the fruits of our labour



Some interesting case studies of topological data analysis 'in the wild,' with a focus on the natural sciences (and somewhat biased towards the life sciences).

Antimicrobial Resistance Prediction

Antimicrobial resistance

Relevance

- *It's time to confront the pandemic of antibiotic resistance*[1]
- *The World's Next Big Health Emergency Is Already Here*[2]
- *Millions are dying from drug-resistant infections, global report says*[3]

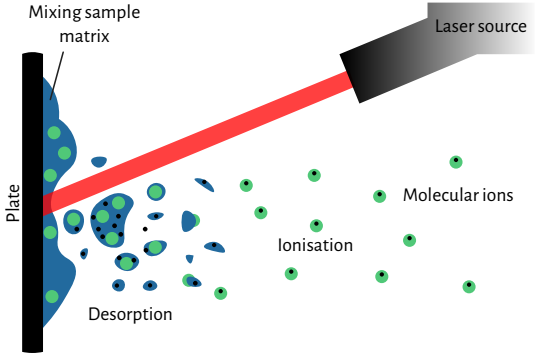
1 https://www.ft.com/content/da746047-ecbc-4c4f-b95d-401421ce13c1

2 https://www.bloomberg.com/opinion/articles/2022-01-27/after-covid-antimicrobial-resistance-is-the-world-s-biggest-health-emergency

3 https://www.bbc.com/news/health-60058120

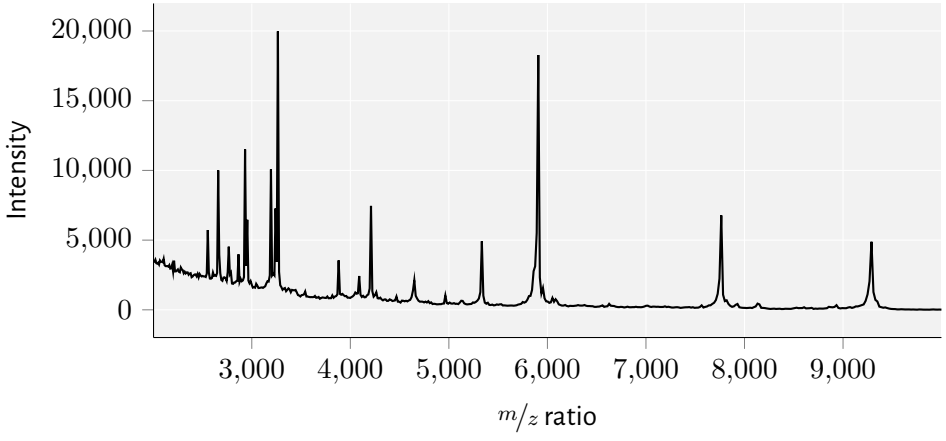
MALDI-TOF mass spectrometry

Matrix-**a**ssisted **l**aser **d**esorption **i**onisation **t**ime-**o**f-**f**light mass spectrometry

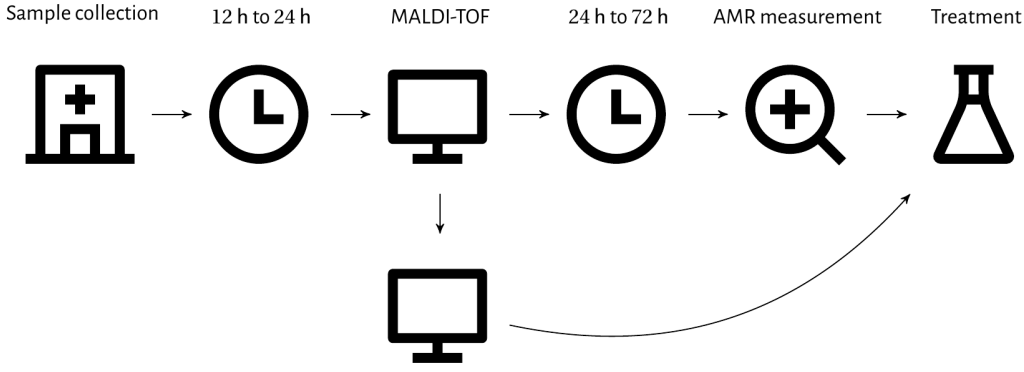


- Obtain a quick overview of sample microbial composition.
- Spectra are known to be *highly characteristic* of a microbial species.
- Standard tool for *species identification* in clinical practice.

Example


20,000
15,000
10,000
5,000
0
Intensity
3,000
4,000
5,000
6,000
7,000
8,000
9,000
m/z ratio

Antimicrobial treatment workflow



A tale of two pre-processing pipelines

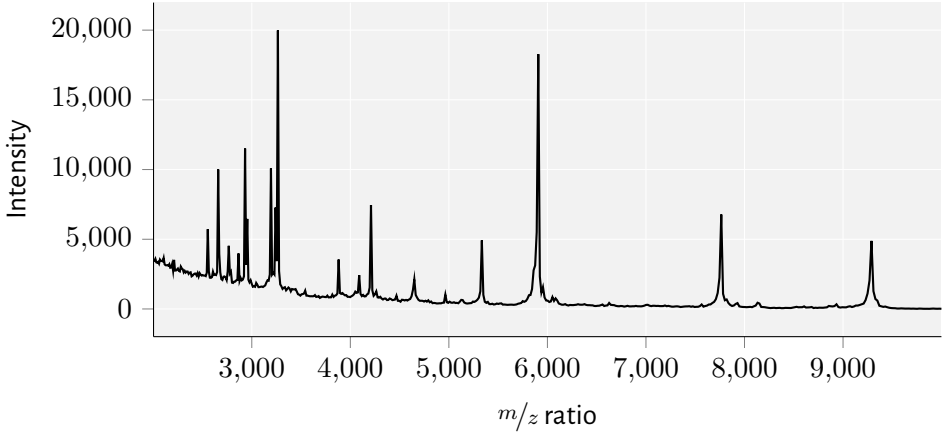
1. Variance stabilisation
2. Smoothing
3. Baseline removal
4. Intensity calibration
5. Intensity trimming

State-of-the-art: S. Gibb and K. Strimmer, '`MALDIquant`: a versatile R package for the analysis of mass spectrometry data', *Bioinformatics* 28.17, 2012, pp. 2270–2271

Treat spectrum as function and use the *persistence* of critical points as a proxy for the heights of a peak.

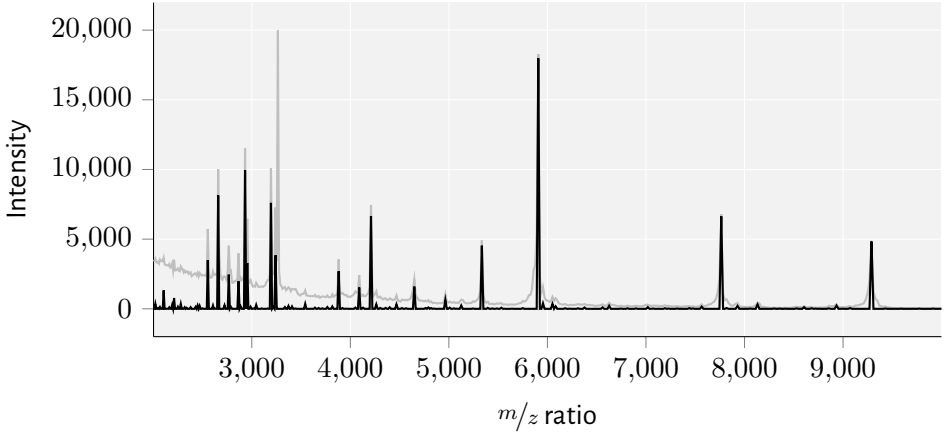
Example

Topological pre-processing



Example

Topological pre-processing



Using spectra for machine learning tasks

- Choose bin size to obtain fixed-sized feature vectors
- Use ‘sparse’ representation based on tuples

The latter part is reminiscent of M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov and A. J. Smola, ‘Deep Sets’, *Advances in Neural Information Processing Systems 30*, ed. by I. Guyon et al., Red Hook, NY, USA: Curran Associates, Inc., 2017, pp. 3391–3401.

Towards antimicrobial resistance prediction

PIKE: Peak Information Kernel

For two spectra S and S', with m/z values x_i and x'_i and intensities λ_i and λ'_i, respectively, we calculate the following expression:

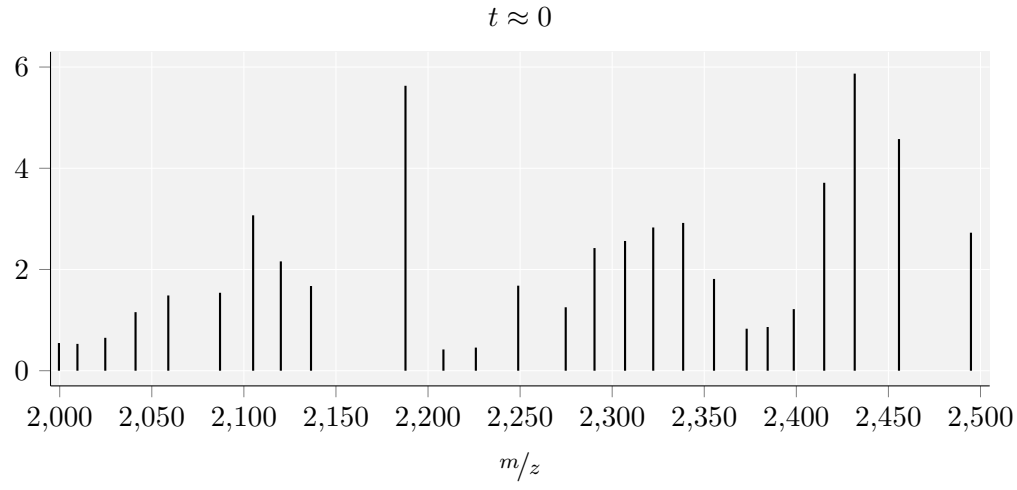
$$\mathsf{k}_t(S, S') = \frac{1}{2\sqrt{2\pi t}} \sum_{i,j} \lambda_i \lambda'_j \exp\left(-\frac{\left(x_i - x'_j\right)^2}{8t} \right)$$

Properties

- C^∞ function with closed-form gradient expression
- Calculate similarity based on peak 'distances'
- Interactions between peaks are captured
- Single parameter $t \in \mathbb{R}$ controls smoothing
- Can be easily integrated into *any* kernel-based model: SVM, Gaussian Processes, ...

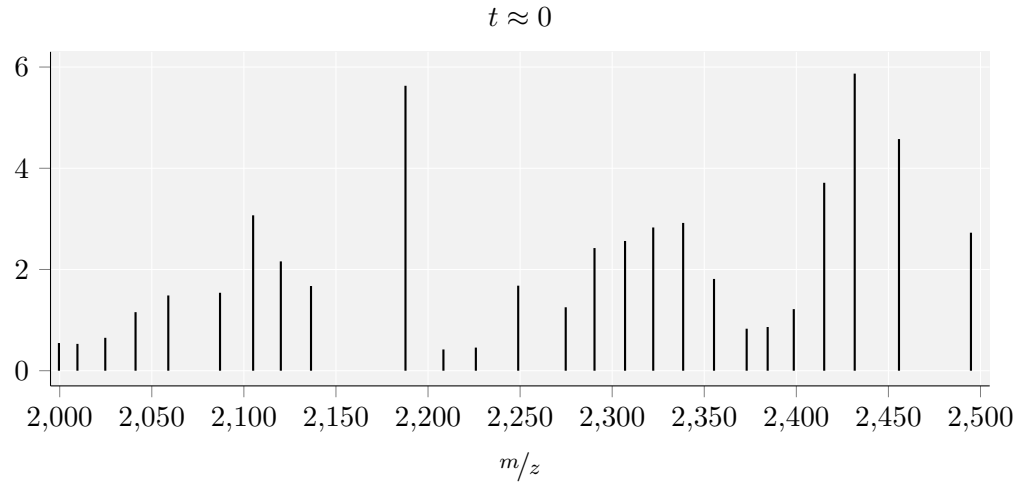
Example

Varying t



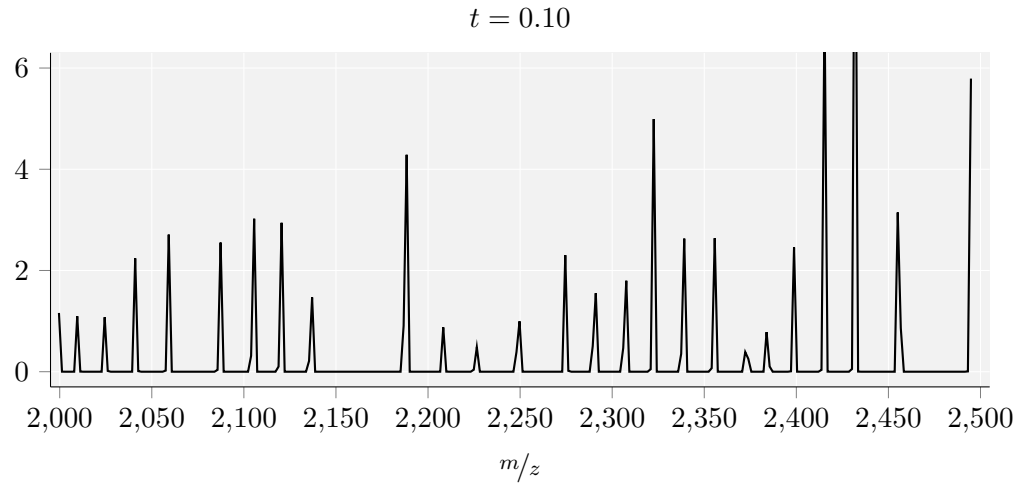
Example

Varying t



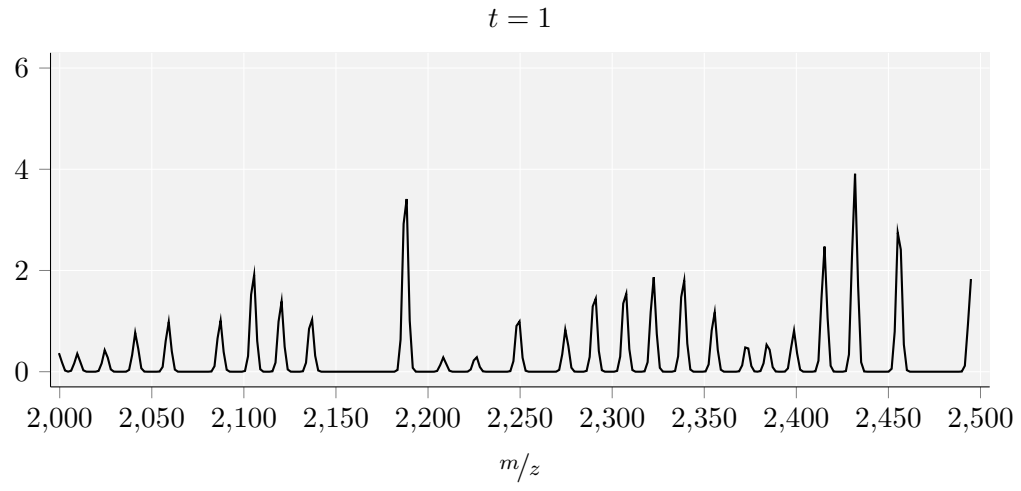
Example

Varying t



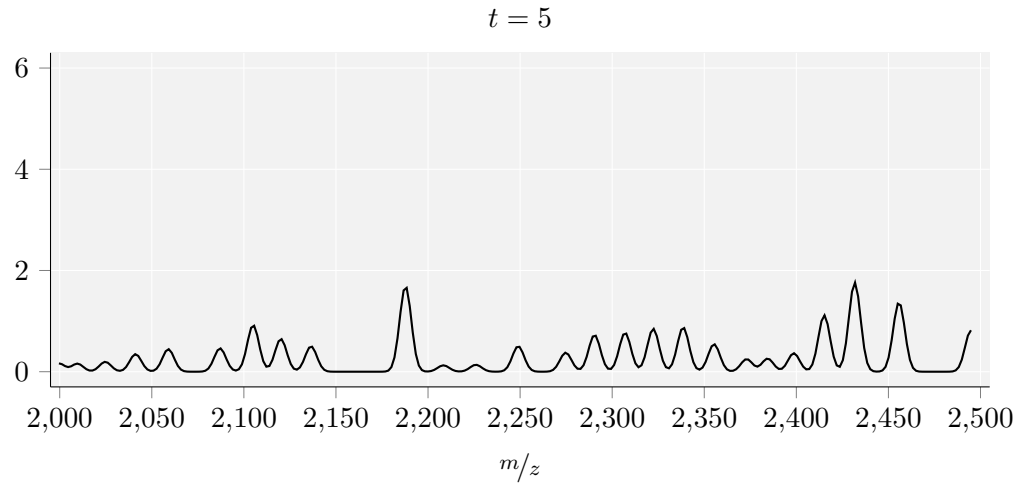
Example

Varying t



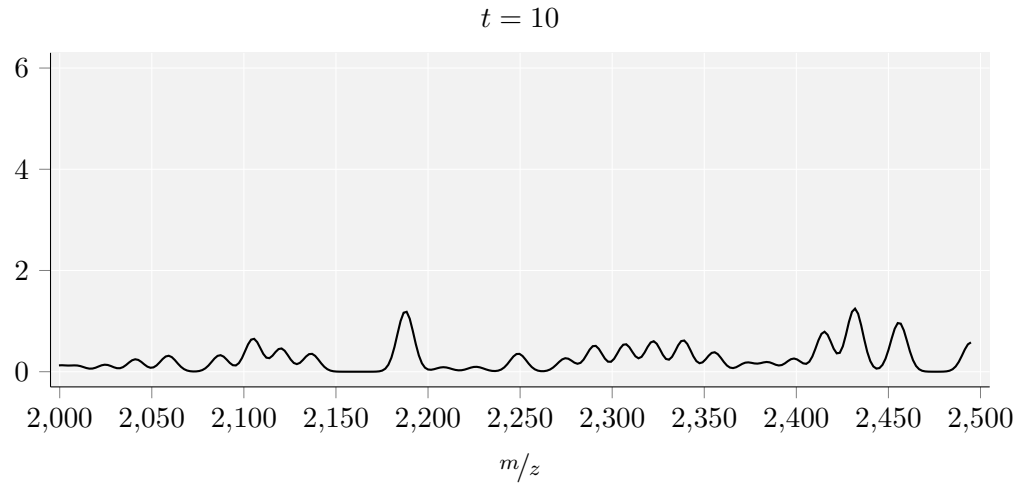
Example

Varying t



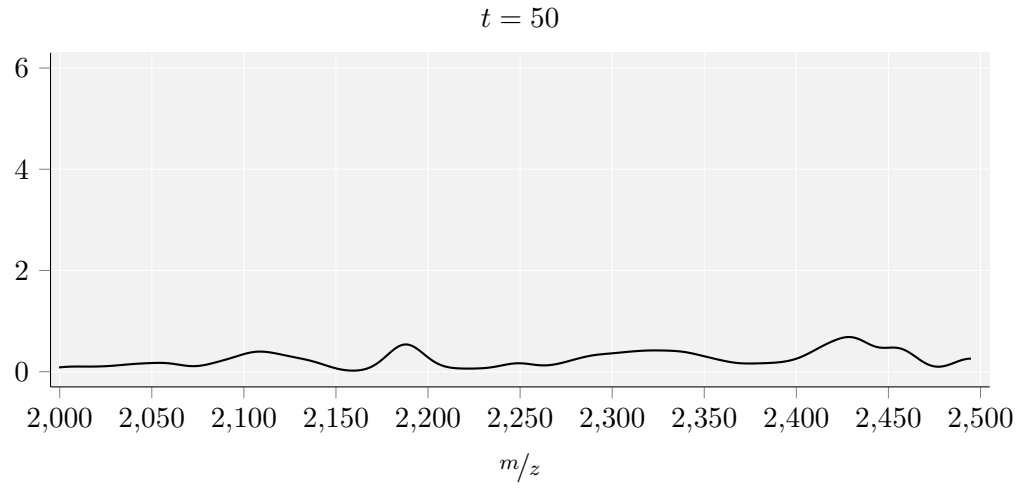
Example

Varying t



Example

Varying t



Data set

Species	Antibiotic	Samples	% resistant
E. coli	amoxicillin / clavulanic acid	1043	28.9
	ceftriaxone	1060	20.4
	ciprofloxacin	1051	29.7
K. pneumoniae	ceftriaxone	597	15.1
	ciprofloxacin	596	16.8
	piperacillin / tazobactam	576	13.9
S. aureus	amoxicillin / clavulanic acid	973	13.7
	ciprofloxacin	987	14.7
	penicillin	941	71.4

Publication

C. Weis*, M. Horn*, **B. Rieck***, A. Cuénod, A. Egli and K. Borgwardt, 'Topological and kernel-based microbial phenotype prediction from MALDI-TOF mass spectra', *Bioinformatics* 36.Supplement_1, 2020, pp. i30–i38

GP-PIKE: superior performance

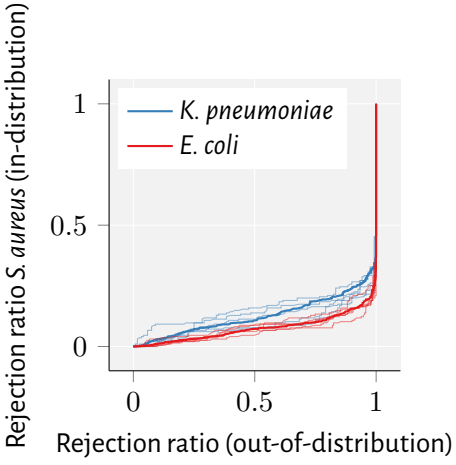
AUPRC

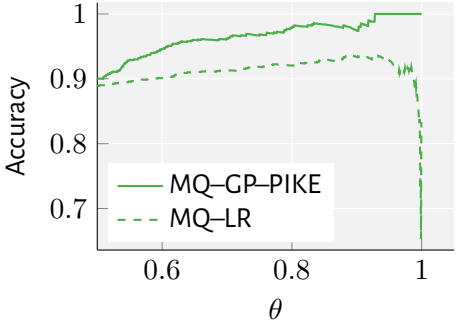
Species	Antibiotic	LR	GP–RBF	GP–PIKE
E. coli	amoxicillin / clavulanic acid	41.0 ± 7.4	32.5 ± 8.5	**47.1 ± 3.9**
	ceftriaxone	63.2 ± 6.1	46.3 ± 24.0	**70.6 ± 3.2**
	ciprofloxacin	61.4 ± 8.5	34.7 ± 10.7	**68.0 ± 3.0**
K. pneumoniae	ceftriaxone	58.2 ± 9.8	58.7 ± 25.3	**77.0 ± 6.8**
	ciprofloxacin	41.7 ± 9.8	30.9 ± 13.5	**54.6 ± 10.1**
	piperacillin / tazobactam	31.6 ± 6.8	13.8 ± 0.0	**56.5 ± 9.7**
S. aureus	amoxicillin / clavulanic acid	52.9 ± 3.9	13.9 ± 0.0	**69.2 ± 9.2**
	ciprofloxacin	34.1 ± 3.3	23.3 ± 11.9	**39.4 ± 6.6**
	penicillin	79.7 ± 3.3	74.2 ± 3.2	**83.2 ± 3.5**

Advantages & disadvantages

Sparse processing and 'built-in' confidence analysis, but insufficient scalability to larger data sets. How does a model fare on larger data sets?

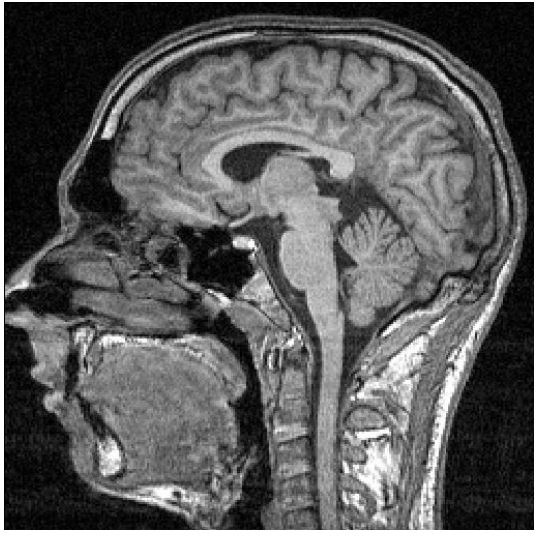
Rejecting 'unsure' samples





Topology-Driven fMRI Data Analysis

fMRI data



Our approach

- ✯ Consider the BOLD activation function f to be a time-varying function on a manifold $\mathcal{M}$
- ✯ Calculate topological features of $\mathcal{M}$ 'measured' via f
- ✯ Obtain stable topological summaries at different resolutions

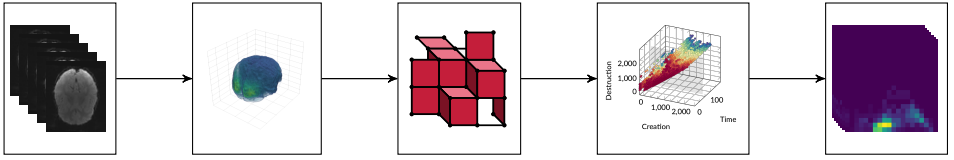
Main advantage of this approach

Working on the 'raw' data; no auxiliary representations necessary! In particular, no *atlas* required (fewer modelling choices in total).

Publication

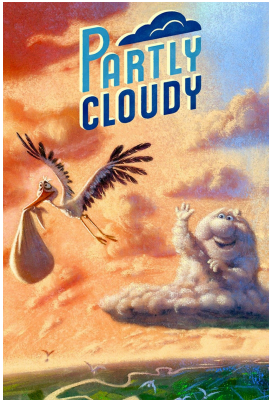
B. Rieck* et al., 'Uncovering the Topology of Time-Varying fMRI Data using Cubical Persistence', *Advances in Neural Information Processing Systems*, ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, vol. 33, Curran Associates, Inc., 2020, pp. 6900–6912, arXiv: `2006.07882 [q-bio.NC]`

Workflow


fMRI stack
fMRI volume
Cubical complex
Persistence diagram
Persistence images
Destruction
2,000
1,000
0
0
1,000
2,000
Creation
0
100
Time

Our data set

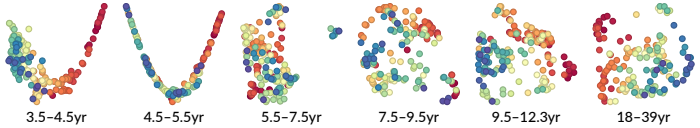
- 155 (122 children, 33 adults) participants are being shown the film ‘Partly Cloudy‘
- *Continuous* stimulation of participants
- 168 time steps
- *No additional information about participants has been provided on purpose*



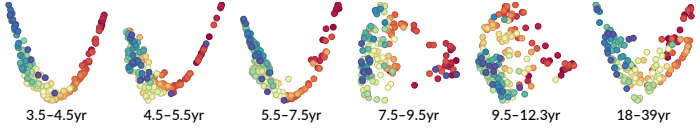
Age prediction based on summary statistics

Method	BM	OM	XM
`baseline-tt`	0.09	0.02	0.24
`baseline-pp`	0.41	0.40	0.40
`tt-corr-tda`	0.17	0.11	0.23
`pp-corr-tda`	0.25	0.27	0.23
`srm`	0.44		
$\|\mathcal{D}\|_1$	0.46	0.67	0.48
$\|\mathcal{D}\|_\infty$	**0.61**	**0.77**	**0.73**

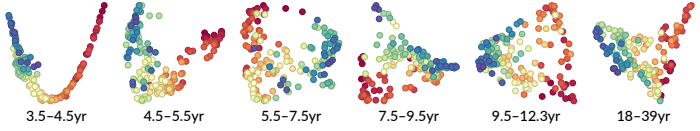
Brain state trajectories



Whole-brain mask



Occipital-temporal mask

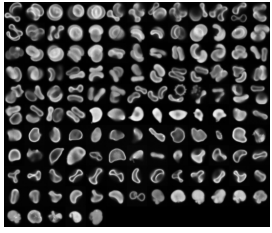


XOR mask

Predicting the Shape of Cells

Cell shape prediction

- Use confocal fluorescence microscopy to obtain images of cells.
- What is the 3D shape of a cell?
- Morphological analysis is crucial for certain pathologies!

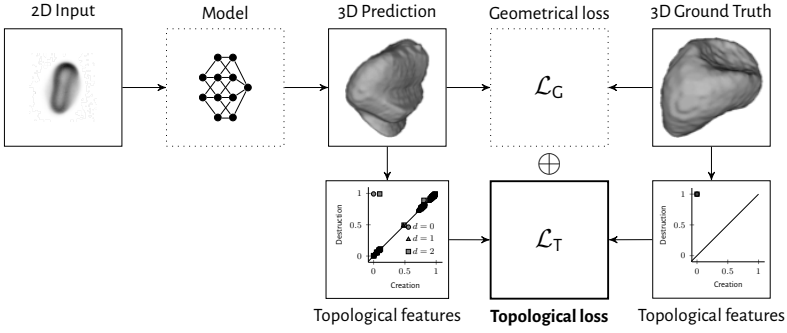


When used properly, RBC [red blood cell] morphology can be a **key tool** for laboratory hematology professionals to recommend appropriate clinical and laboratory follow-up and to select the best tests for definitive diagnosis.

(J. Ford, 'Red blood cell morphology', *International Journal of Laboratory Hematology* 35.3, 2013, pp. 351–357.)

SHAPR

Overview

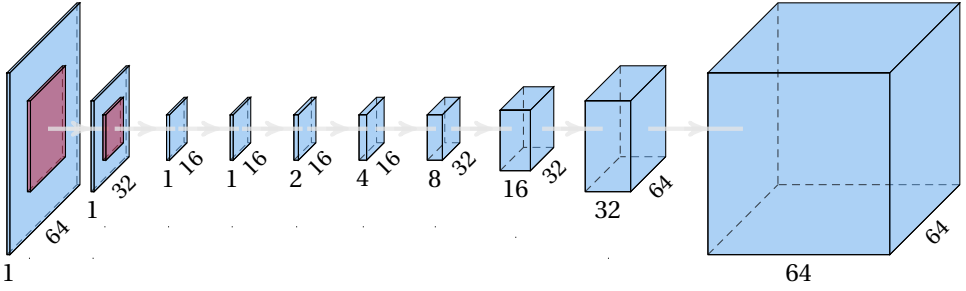


We are trying to solve a complicated *inverse problem*, going from 2D to 3D. This is an ill-defined problem with a large number of potential solutions.

D. J. E. Waibel, S. Atwell, M. Meier, C. Marr and **B. Rieck**, 'Capturing Shape Information with Multi-Scale Topological Loss Terms for 3D Reconstruction', *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, ed. by L. Wang, Q. Dou, P. T. Fletcher, S. Speidel and S. Li, Cham, Switzerland: Springer, 2022, pp. 150–159, arXiv: `2203.01703 [cs.CV]`

SHAPR

Architecture



We are learning a *likelihood function* $f\colon \mathbb{R}^3 \to \mathbb{R}$. Formally, f 'lives' on a voxel grid, assigning each voxel x a value that indicates the likelihood of x being part of the 'true' volume.

SHAPR

Loss function

$$\mathcal{L}_{\mathsf{G}}(f, f') := \frac{2\,\mathcal{L}_{\mathsf{Dice}}(f, f') + \mathcal{L}_{\mathsf{BCE}(f,f')}}{2}$$

$$\mathcal{L}_{\mathsf{Dice}}(f, f') := \frac{2\left|\mathsf{Vol}_f \cap \mathsf{Vol}_{f'}\right|}{\left|\mathsf{Vol}_f\right| + \left|\mathsf{Vol}_{f'}\right|} = \frac{2\mathsf{TP}}{2\mathsf{TP} + \mathsf{FP} + \mathsf{FN}}$$

Intuition

Compare *geometry* of the resulting volumes on a per-voxel basis. Is the reconstructed volume well-aligned with the ground truth one?

SHAPR goes topological

$$\mathcal{L}_{\mathsf{T}}(f, f', q) := \sum_{i=0}^{d} \mathsf{W}_q\left(\mathcal{D}_f^{(i)}, \mathcal{D}_{f'}^{(i)}\right) + \mathsf{pers}\left(\mathcal{D}_{f'}^{(i)}\right)$$

SHAPR goes topological

$$\mathcal{L}_{\mathsf{T}}(f, f', q) := \sum_{i=0}^{d} \mathsf{W}_q\left(\mathcal{D}_f^{(i)}, \mathcal{D}_{f'}^{(i)}\right) + \mathsf{pers}\left(\mathcal{D}_{f'}^{(i)}\right)$$

Loss components

- Aligning the ground truth likelihood f and the predicted likelihood function f'.
- Reducing the geometrical–topological variation of the predicted likelihood function f'.

SHAPR goes topological

$$\mathcal{L}_{\mathsf{T}}(f, f', q) := \sum_{i=0}^{d} \mathsf{W}_q\left(\mathcal{D}_f^{(i)}, \mathcal{D}_{f'}^{(i)}\right) + \mathsf{pers}\left(\mathcal{D}_{f'}^{(i)}\right)$$

Loss components

- Aligning the ground truth likelihood f and the predicted likelihood function f'.
- Reducing the geometrical–topological variation of the predicted likelihood function f'.

We obtain a *combined loss* by choosing $\lambda \in \mathbb{R}_{>0}$ and calculating:

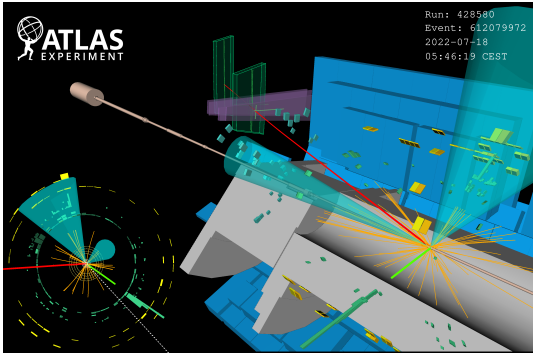
$$\mathcal{L} := \mathcal{L}_{\mathsf{G}} + \lambda\, \mathcal{L}_{\mathsf{T}}$$

Quantitative results

		Red blood cell ($n = 825$)		Nuclei ($n = 887$)	
Metric	$\mathcal{L}_T$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$
1-IoU	✗	0.48	0.49 ± 0.12	0.62	0.62 ± 0.11
	✓	**0.46**	**0.47 ± 0.10**	**0.61**	**0.61 ± 0.11**
Volume	✗	0.31	0.35 ± 0.31	0.34	0.48 ± 0.47
	✓	**0.21**	**0.25 ± 0.24**	**0.32**	**0.43 ± 0.42**
Surface area	✗	0.20	0.24 ± 0.20	0.21	0.27 ± 0.25
	✓	**0.13**	**0.18 ± 0.16**	**0.18**	**0.25 ± 0.24**
Surface roughness	✗	0.35	0.36 ± 0.24	**0.17**	**0.18 ± 0.12**
	✓	**0.24**	**0.29 ± 0.22**	0.18	0.19 ± 0.13

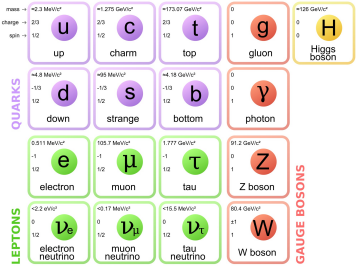
Jet Tagging

ATLAS



- Largest particle detector experiment at CERN
- Searching for ‘exotic particles’

Standard model



We never observe individual particles but collimated sprays of particles, known as *jets*. By *tagging* the jet, i.e. classifying it, we hope to understand which particle initiated the jet.

Why topology?

- Topological features are *invariant* under rotations and translations of point clouds.
- They are invariant to permutations.
- **Open question:** Do they satisfy Lorentz invariance?

What we already know

If particles are displaced by some small distance, the Gromov–Hausdorff distance between the original and *boosted* point cloud remains small, thus bounding the distance between persistence diagrams.

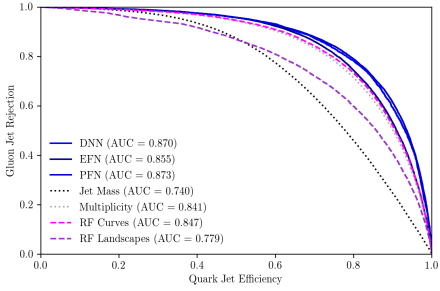
Preliminary results

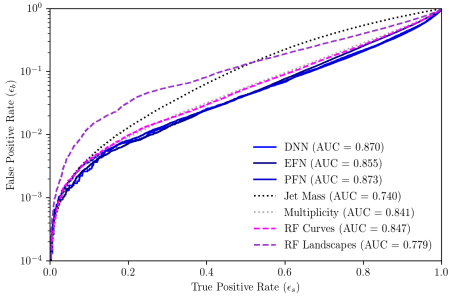
Data and model

- Use quark–gluon tagging data set
- Each jet consists of a point cloud of particles, consisting of a transverse momentum p_T measurement, a rapidity, and azimuthal angle
- Use 375K jets for training, 75K jets for validation, and 50K jets for testing.
- Use *random forest* classifier with vectorised Betti curves and persistence landscapes

Preliminary results

Performance curves





The End

Thank you very much for having me!

Some things to do

- Join the ‘Geometry and Topology in ML’ Slack community!
- Join AATRN and help us build a community