Uncovering the Topology of Time-Varying fMRI Data using Cubical Persistence

Bastian Rieck†, Tristan Yates†, Christian Bock, Karsten Borgwardt, Guy Wolf, Nick Turk-Browne‡, Smita Krishnaswamy‡
†: Equal contributions
‡: Joint supervision

NEURAL INFORMATION PROCESSING SYSTEMS

Abstract

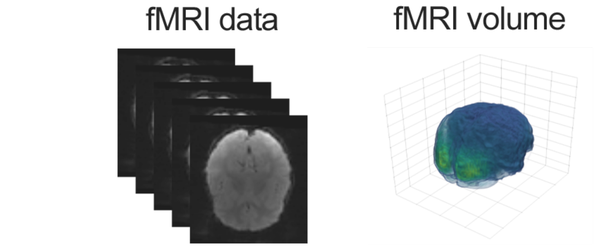
Functional magnetic resonance imaging (fMRI) is a crucial technology for gaining insights into cognitive processes in humans. Data amassed from fMRI measurements result in volumetric data sets that vary over time and are challenging to analyse. We present a novel topological approach that encodes each time point in an fMRI data set as a persistence diagram of **topological features**, i.e. high-dimensional voids present in the data. Here, we apply both clustering and trajectory analysis techniques to a group of participants watching the movie 'Partly Cloudy'. We observe significant differences in both brain state trajectories and overall topological activity between adults and children watching the same movie.

Challenges in fMRI data analysis

- Measurements are inherently **noisy**
- A high degree of **variability**
- Relevant features occur at **multiple scales**

Our approach: Cubical Persistent Homology

We represent an fMRI stack as a time-varying volume (4D), which we 'convert' into a cubical complex. The fMRI activation function values become scalar values at the vertices of the cubical complex, and we extend this function to a weight function over the full cubical complex. The cubical complex enables us to calculate **multi-scale topological** features using **persistent homology**.

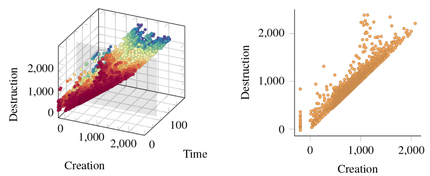


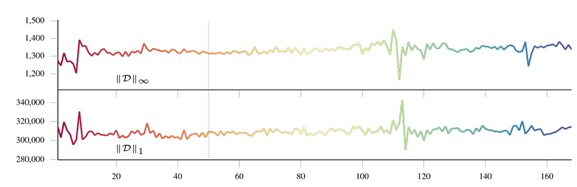
Topological summary statistics for age prediction

Starting from a sequence of time-varying persistence diagrams of one participant, for each diagram slice, we evaluate a scalar-valued statistic $S\colon \mathcal{D} \to \mathbb{R}$, leading to a time series. In this example, we use two statistics, the *infinity norm* $\|\mathcal{D}\|_\infty$ and the p-norm $\|\mathcal{D}\|_p$ of a persistence diagram $\mathcal{D}$:

$$\|\mathcal{D}\|_\infty = \max_{x,y\in\mathcal{D}} |y-x| \quad \text{or} \quad \|\mathcal{D}\|_p = \sqrt[p]{\textstyle\sum_{x,y\in\mathcal{D}} |y-x|^p}$$

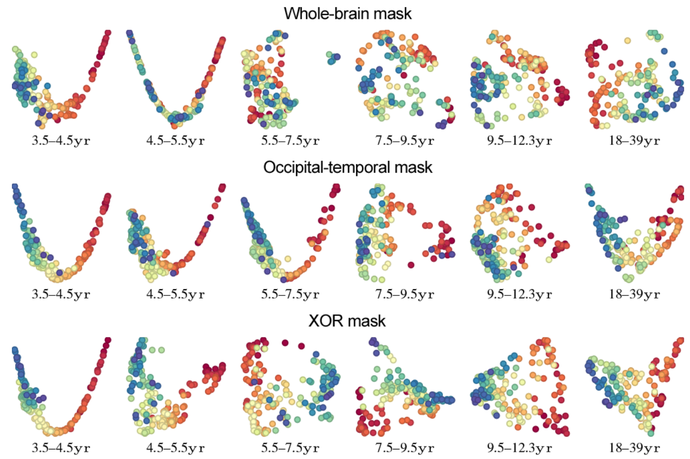
Calculating any one of these norms for a 2D 'slice' of a 3D persistence diagram results in a single scalar value—one for each time step. By collating all summary values, we obtain a **time series** of the topological activity.





Using various **masks** for the volume, i.e. a whole-brain mask (BM), an occipital–temporal mask (OM), and the logical XOR of the two two (XM), we obtain three different types of topological summary statistics time series. We can use the time series to train a ridge regression model to predict the age of non-adult participants. Comparing this to existing methods and some baselines, we find that the infinity norm is, due to its robustness, a suitable choice to obtain high-quality predictions—measured in terms of the correlation coefficient (CC) and the mean squared error (MSE).

Method / Mask	BM		OM		XM	
	CC	MSE	CC	MSE	CC	MSE
BASELINE-TT	0.09	10.15	0.02	13.81	0.24	7.19
BASELINE-PP	0.41	6.23	0.40	6.40	0.40	6.65
SRM	0.44	6.05	—	—	—	—
$\|\mathcal{D}\|_1$	0.46	4.27	0.67	2.95	0.48	4.17
$\|\mathcal{D}\|_\infty$	**0.61**	**3.38**	**0.77**	**2.20**	**0.73**	**2.53**



Brain state trajectories

Next to the static analysis based on summary statistics, we can calculate **trajectories** based on the time-varying topological dissimilarity between participants. Letting Ψ refer to the persistence image calculation, we can evaluate $\left(\mathcal{D}_2^{(i,j)}\right)$ for each time step t_j in our dana set.

This turns the sequence of persistence diagrams of the ith participant into a **matrix** $X^{(i)} \in \mathbb{R}^{m\times r^2}$, where m is the number of time steps and r is the resolution of the persistence diagram, such that the jth row corresponds to the 'unravelled' persistence image of time step t_j. Our participants are stratified in six age-related cohorts, ranging from young children to adults. For each of these cohorts, we now calculate the sample mean $\overline{X_k}$ of the matrices making up a cohort. The resulting six matrices represent the average time-varying topological activity of participants in each cohort. We can **embed** these matrices into a 2D space using PHATE [1]. This results in a set of *brain state trajectories* (for each of the brain masks).

Paper

Code

References

Kevin R. Moon et al. (2019). Visualizing structure and transitions in high-dimensional biological data. In: Nature biotechnology 37, no. 12 (2019): 1482-1492.