Homological scaffold for brain data

Alessandro De Gregorio, Marco Guerra

Supervisor: Prof. Francesco Vaccarino

Politecnico di Torino - DISMA

Motivation and background

Topological data analysis is a recent approach to the analysis of data, based on the intuition that the shape of data carries valuable information within itself. The general pipeline involves retrieving a points cloud jointly with a notion of similarity through pairwise distances and then studying its topological characteristics.

One of the main technique used in TDA is Persistent Homology, a theory that formalizes the concept of "hole" in a manifold at different scale levels and different dimensions. By enriching geometric information arising from a distance or a filtering function one can obtain a sequence of nested subspaces of the shape, and henceforth apply persistent homology to observe the evolution of topological features at different scales.

The homological scaffold is a descriptor of the 1-skeleton of a shape made by the cycles that characterize in the best way the one-dimensional holes across the various levels of the filtration.

We have developed a new approach to the scaffold by computing the minimal cycles across the various scales and their lifespan.

We will present some of the outcomes of the tests we performed with our tool on simulated and real data obtained from fMRI's of the human brain.

Addressed problem

The most essential description of the shape of a geometrical object is given by the holes that cut through it. The intuition underpinning TDA is to evaluate the relevance of local features by measuring their contribution to the global shape of a dataset;.

The key intuition is as follows: if two points are connected by a lone arc, this represents a bridge in the dataset. But if three points are all mutually connected, then at this distance scale the resulting triangle blurs and can be confused for a single point. The large-scale holes that survive after this equivalence is taken are assumed to represent relevant features in the data.

To link triangles and edges, we use the boundary operator

$$\partial_k\ [\sigma_0, ..., \sigma_k]\ \mapsto\ \sum_{i=0}^{k} [\sigma_0, ..., \hat{\sigma_i}, ..., \sigma_k]$$

Its main property is

$$\partial_k\ \partial_{k+1} = 0 \quad \text{Im}\ \partial_{k+1} \subseteq \ker \partial_k$$

The k-th homology group is defined as

$$H_k := \frac{\ker \partial_k}{\text{Im}\ \partial_{k+1}}$$

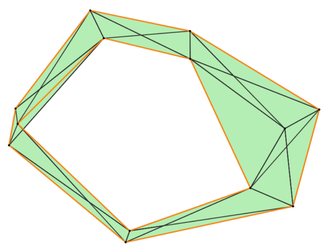


Fig 1. Two homologous cycles.

Persistent homology works by sweeping the range of spatial scales, increasing the scale of interaction between points and computing the corresponding topological description. Then computing the persistence, i.e. sweeping the spatial range increasing connectivity, amounts to repeating the analysis over a filtration of nested simplicial complexes.

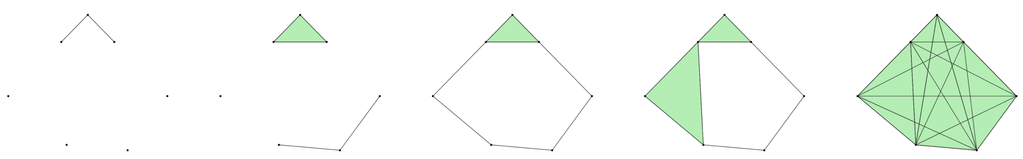


Fig 2. Filtration of simplicial complexes

The homological scaffold [1] gives an aggregate summary of the persistence over the filtration: after we have fixed a representative for each independent homology cycle at each step of the filtration, we build it as a graph on the set of points such that each edge is weighted by the number of times it belongs to a homology cycle, across the filtration. This yields a measure of the relevance of every edge with respect to a global feature of the dataset.

Our work is aimed at fixing a loophole in the previous pipeline: since homology cycles are equivalence classes, there is no obvious way to fix a representative. We have worked to compute at each step a *localization of homology*, i.e. the cycles of minimal total length which generate H_1.

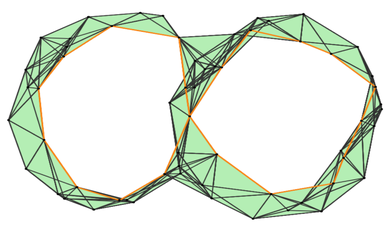


Fig 3. The highlighted cycles, output by the algorithm, form a shortest cycle basis for the first homology group.

Methods

The core of the problem we want to address is choosing a set of cycles so as to span the homology group, while minimizing the total length of the basis,

$$h_1, ..., h_g = \underset{\text{span}\ \{h_i\}=H_1}{\text{argmin}} \sum_i \mu(h_i)$$

Where $\mu(h_i)$ is the length of the shortest representative of the homology class h_i.

We show the general structure of an algorithm as presented in [2], which computes such a basis in polynomial time. Its outline is as follows:

- It computes an annotation of the edges. An annotation is a function $a: E \to \mathbb{Z}_2^g$, that can be linearly extended onto cycles, such that two cycle have the same annotation if and only if they are homologous.
- By computing the annotation we can determine the dimension of H_1, namely g.
- The candidate set of cycles is generated by classical methods involving the spanning tree.
- The selection of each element of the homology basis is performed as follows:

```
procedure CycleBasis(G)
    for i =1 t o g do :
        Initialize S_i = e_i
    end
    ExtendBasis(1,g)
end
```

```
procedure ExtendBasis(i,k)
    if k = 1 then:
        ShortestCycle(S_i)
    else:
        ExtendBasis(i, ⌊k/2⌋ )
        Update(i,k)
        ExtendBasis(i+ ⌊k/2⌋ , ⌈k/2⌉ )
    end
end
```

Fig 4. Pseudocode sketch of the algorithm.

The shortest homology basis needs not be unique, as the graph may contain several cycles of the same length. Therefore, the scaffold is computed summing the incidence of basis cycles onto each edge. In case of a draw between multiple cycles, their incidence is *smeared* over their different representatives, so the approach does not arbitrarily privilege a cycle over others.

Results

The shortest-basis homological scaffold algorithm was applied to a public dataset of neurological images, composed of time-series correlation of brain regions activity, measured through fMRI scans of hemoglobin deoxigenation. Data was provided by the Human Connectome Project.

Distance matrices are obtained from a 1200-points time series on 89 parcels of the human brain, subset of the AAL brain atlas, analyzed on patients in resting state. The metric is expressed as an inverse of correlation values. The input matrix induces a virtually complete weighted graph on the 89 nodes.

Fig 5. Chord diagram of the homological scaffold over a 100-step filtration of the data. Connectivity is reduced from 3916 edges to just 191 while maintaining the essential structure of the non-trivial cycles. Color intensity of an arc reflects its weight in the scaffold. Size and coloring of a node reflect its strength.

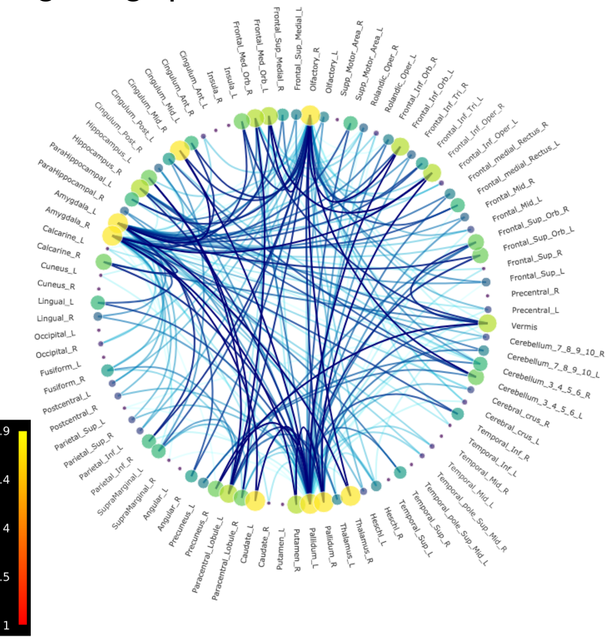
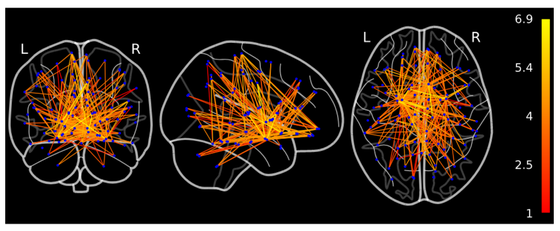


Fig 6. Embedding of the scaffold in a real-world human brain. Weights are in log scale for visualization.



References

[1] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, F. Vaccarino. Homological scaffolds of brain functional networks. J. R. Soc. Interface 2014 11 20140873; DOI: 10.1098/rsif.2014.0873. (2014)

[2] T. Dey et al. Efficient algorithms for computing a minimal homology basis. arXiv:1801.06759 (2018)

[3] T. Kavitha, K. Mehlhorn, D. Michail, and K. Paluch. A faster algorithm for minimum cycle basis of graphs. In 31st International Colloquium on Automata, Languages and Programming, Finland, pages 846–857, (2004)

Data Source: M. Termenon, C. Delon-Martin, A. Jaillard and S. Achard. Reliability of graph analysis of resting state fMRI using test-retest dataset from the human connectome project. Neuroimage 142(15): 172–187, doi : 10.1016/j.neuroimage.2016.05.062, 2016.

Data provided by the Human Connectome Project http://www.humanconnectomeproject.org/data/ Data collection and sharing for this project was provided by the MGH-USC Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J. Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles.

Joint work with: Esther Ibanez Marcelo, Giovanni Petri – ISI Foundation