Survey of TOPDRIM applications of Topological Data Analysis

Matteo Rucco⋆, Adane Letta Mamuye, Marco Piangerelli, Michela Quadrini,
Luca Tesei and Emanuela Merelli,

BigData Lab@Unicam, University of Camerino. Camerino, Italy.
`{name.surname}@unicam.it`
`http://www.topdrim.eu-http://www.jholes.eu`

Abstract. Every moment of our daily life belongs to the new era of
"Big Data". We continuously produce, at an unpredictable rate, a huge
amount of heterogeneous and distributed data. The classical techniques
developed for knowledge discovery seem to be unsuitable for extracting
information hidden in these volumes of data. Therefore, there is the
need to design new computational techniques. In this paper we focus
on a set of algorithms inspired by algebraic topology that are known
as *Topological Data Analysis* (TDA). We briefly introduce the principal
techniques for building topological spaces from data and how these can
be studied by persistent homology. Several case studies, collected within
the TOPDRIM (Topology driven methods for complex systems) FP7-
FET project, are used to illustrate the applicability of these techniques
on different data sources and domains.

Keywords: Topological Data Analysis, Data Mining, Simplicial Com-
plexes, Persistent Homology, Persistent Entropy

1 Introduction

Topology is the branch of mathematics that studies shapes and maps among
them. A topological space is a powerful mathematical concept for describing the
connectivity of a space. Informally, *a topological space is a set of points each of
which equipped with a notion of neighbouring.* One way to represent a topolog-
ical space is by connecting *simple* pieces such that their common intersections
are lower-dimensional pieces of the same kind and are known as *simplices.* Fig-
ure 1 shows the geometrical realisation of simplices: points for 0-simplices, line
segments for 1-simplices, filled triangles for 2-simplices and filled tetrahedra for
3-simplices. They can be extended naturally to n-dimensional objects realising
$(n-1)$-simplices [5].

Recently, a new set of algorithms, identified as Topological Data Analysis
(TDA), has been derived from algebraic topology. These algorithms are designed
for investigating high-dimensional data in a quantitative manner. For example,

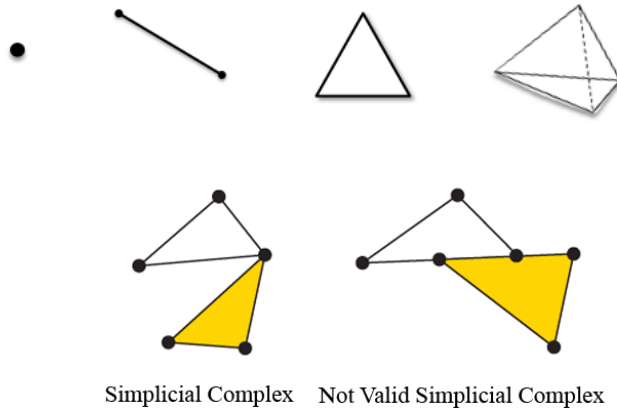---

⋆ Corresponding author

Fig. 1. Top: From left to right, simplices of dimensions 0 (points), 1 (edges), 2 (triangles) and 3 (tetrahedra). Bottom-left: a valid simplicial complex, Bottom-right: a not valid simplicial complex [5].

they have been used for studying the characteristics of functional brain networks at the mesoscopic level [14] as well as for deciphering viral evolution in biological complex systems [3]. Other examples are related to the analysis of sensor networks [4] and immunology [15, 10].

In the context of data mining, TDA can be seen as a new set of tools for performing *exploratory data analysis*. It permits to study high dimensional datasets without dimensionality reductions. It reveals local relationships hidden in the data by transforming them into global objects, which are simplicial complexes.

TDA can be divided in two families: topological data compression and topological data completion [11]. Algorithms for topological data compression aim at representing a collection of higher dimensional data points through a graph. The main algorithm in this area is Mapper [20]. Conversely, topological data completion, based on *persistent homology* [5], completes data to more complex structures, i.e. simplicial complexes, which can be analysed in a more easy way.

In this paper, after a brief introduction of the basic mathematical machinery (Section 2, we focus on the following techniques, based on persistent homology and working on different data sources: Vietoris-Rips [2], Clique Weight Rank Persistent Homology [5, 14] and Piecewise Complex [17]. Vietoris-Rips can be used for pinpointing out the topology of a continuous object by starting from a discrete sampling represented by a point cloud data (Section 3. Examples of continuous objects that can be studied are: the trajectory of an object in a physical space or in a phase space, a geometrical object, and so on. Clique Weight Rank Persistent Homology can be used when the data source is a weighted undirected graph (Section4. Graphs are a powerful tool for representing 2-bodies relationships (a classical example is the world wide web, where a link connects two clients or a client with a server) but this representation does not capture higher dimensional relationships. For this reason, we will use Clique Complexes, which

are the right topological tool for extracting higher dimensional patterns from a graph. For example, in the case of a *sensor network*, a higher dimensional pattern can represent a subset of sensors not directly connected but that are interacting simultaneously, thus determining a loss of performance of the whole system through these interactions. Piecewise Complexes can be used to derive a topological space from discrete signals (Section 5). Their most relevant application is the topological comparison of real length noisy signals. Moreover, in order to study dynamical systems from the data perspective using the three techniques above, there is the need to equip them with statistics. In our works we introduced *persistent entropy* [10]. It is used in Section 4 and Section 5.

We report on a set of applications of topological data analyses performed within the TOPDRIM (TOPology DRIven Methods for complex systems) FP7-FET project. We hope that these applications, belonging to different domains and coping with different problems, can be used by the reader as signposts in his or her personal experience in using topological data analysis.

2 Persistent Homology based algorithms

In the context of topology, homology is an algebraic machinery used for describing a topological space $\mathfrak{C}$ by associating to it a sequence of homology groups. Informally, for any $k \in \mathbb{N}$, the $k-$th *Betti number*, denoted by β_k, represents the rank of the $k-$dimensional associated homology group and counts the number of $k-$dimensional holes characterizing $\mathfrak{C}$. For instance, β_0 is the number of connected components, β_1 counts the number of holes in 2D or tunnels in 3D[1], β_2 can be thought as the number of voids in geometric solids, and so on.

Persistent homology is a method for computing $k-$dimensional holes at different spatial resolutions. More persistent holes are detected over a wide range of length and are more likely to represent true features of the underlying space, rather than artifacts of sampling, noise, or particular choice of parameters. Persistent homology appears as a fundamental tool in Topological Data Analysis. It studies the evolution of $k-$dimensional holes along a sequence of simplicial complexes (i.e. a filtration). The set of intervals representing birth and death times of $k-$dimensional holes along such sequence is called the persistence barcode. $k-$dimensional holes with short lifetimes are informally considered to be "topological noise", and those with a long lifetime are considered to be topological feature associated to the given data (i.e. the filtration). The key idea of Persistent Homology is as follows: First, the space must be represented as a simplicial complex. Second, a filtration of the simplicial complex, that is a nested sequence of increasing subsets (referred above as different spatial resolutions), is computed. More concretely, a filtration of a simplicial complex K is a collection of simplicial complexes $\{K(t)|t \in \mathbb{R}\}$ of K such that $K(t) \subset K(s)$ for $t < s$ and there exists $t_{\max} \in \mathbb{R}$ such that $K_{t_{\max}} = K$. The filtration time (or filter value) of a simplex $\sigma \in K$ is the smallest t such that $\sigma \in K(t)$. Persistent homology

---

[1] nD refers to the $n-$dimensional space $\mathbb{R}^n$.

describes how the homology of a given simplicial complex K changes along filtration. If the same topological feature (i.e., $k-$dimensional hole) is detected along a large number of subsets in the filtration, then it is likely to represent a true feature of the underlying space, rather than artifacts of sampling, noise, or particular choice of parameters. More concretely, a $k-$dimensional Betti interval, with endpoints $[t_{start}, t_{end})$, corresponds to a $k-$dimensional hole that appears at filtration time t_{start} and remains until filtration time t_{end}. The set of intervals representing birth and death times of homology classes is called the *persistence barcode* associated to the corresponding filtration. For more details and a more formal description we refer to [5]. There are currently several software products for the computation of persistent homology:*the plex family, PHAT, jHoles, Perseus, DIPHA, GUDHY, and Dionysus*. For a complete review we refer to [13].

3   Analyzing high dimensional point cloud data sets

Higher dimensional dataset are usually studied with techniques of dimensionality reduction. However sometimes this constrain gives rise to drop out useful information. Conversely, topology can be used for visualizing and exploring high dimensional and complex real-world point cloud data sets in which each data point belongs to $\mathbb{R}^n$. Vietoris-Rips filtration is a versatile tool in topological data analysis and it is used for studying point cloud data. More formally, it is a sequence of simplicial complexes built on a metric space to add topological structure to an otherwise disconnected set of points. It is widely used because it encodes useful information about the topology of the underlying metric space. Two classical examples of abstract simplicial complexes are Čech complexes and Vietoris-Rips complexes (see [5, Chapter III]). Let V be a finite set of points in $\mathbb{R}^n$. The *Čech complex* of V and r denoted by $\check{C}_r(V)$ is the abstract simplicial complex whose simplices are formed as follows. For each subset S of points in V, form a closed ball of radius $r/2$ around each point in S, and include S as a simplex of $\check{C}_r(V)$ if there is a common point contained in all of the balls in S. This structure satisfies the definition of abstract simplicial complex. The *Vietoris-Rips complex* denoted as $VR_r(V)$ is essentially the same as the Čech complex. Instead of checking if there is a common point contained in the intersection of the $(r/2)-$ball around v for all v in S, we may just check pairs adding S as a simplex of $\check{C}_r(V)$ if all the balls have pairwise intersections. We have $\check{C}_r(V) \subseteq VR_r(V) \subseteq \check{C}_{\sqrt{2}r}(V)$. See Fig.2. In our opinion, the homology captured by Vietoris-Rips can be interpreted as geometrical signatures, that together with other features, can be used for training machine learning methods for shape identification and retrieval.

Application 3.1 - Topological Classification of small DC Motors.

Persistent homology can be used for dealing with the comparison of high frequency noisy signals. A new methodology based on signal embedding and applied topology has been proposed in [16]. Signal embedding is a useful tool but in same
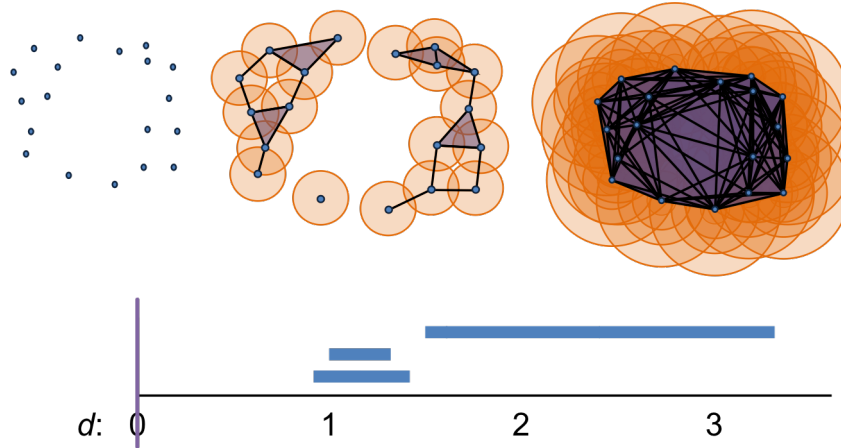
Fig. 2. How to obtain Vietoris-Rips complexes from a point cloud data: starting from a PCD in a metric space, we surround each point with a sphere that grow up simultaneously and equally. The ϵ parameter coincides with the radii of the spheres. Each ϵ gives rise to new intersections. A new intersection of dimension k is equal to a k-1 simplices. For each ϵ persistent homology is computed and in this case graphically represented by the barcode.

case it is a not enough for studying long range noisy signals. Conversely, we argue that an embedded signal in $\mathbb{R}^m$ space can be properly analyzed with topological based techniques. We obtained numerical evidences that our procedure applied to vibrational data acquired from the bench properly classifies small DC motors into two classes: *good* and *faulty*. In order to classify the DC motors we defined a new methodology. The first step to be accomplished is to represent the signal into the $\mathbb{R}^n$ space. The selection of the right parameters (*time-delay* and *minimum embedding dimension*) for performing the embedding of the signal is a crucial point. We suggest to use *mutual information* for deciding the *time-delay* value and *Cao's method* for selecting the embedding dimension. The embedding step maps the signal points into a point cloud data (PCD). PCDs are used for completing the data with simplicial complexes by constructing the Vietoris-Rips complexes (or generally *flag complexes*). *Betti numbers* are computed by persistent homology and they are used for analyzing the features of this topological space:

Step 1 computing *mutual information* of the signal for finding the proper time-delay
Step 2 computing *Cao's method* for deciding the minimum embedding dimension, and to distinguish deterministic data from random data
Step 3 executing the embedding in the new $\mathbb{R}^n$ space
Step 4 transforming point cloud data in simplicial complexes using Vietoris-Rips
Step 5 computing persistent homology and representing Betti numbers via persistent bar-codes
Step 6 statistically analyze the collection of Betti numbers.

We applied our methodology to 76 small DC motors. The computation of *auto mutual information* found two values for the *time-delay* parameter, respectively 5 and 7. Cao's method identified as *minimum dimension* parameter the value $m = 3$ for all DC motors. From each embedded signal we constructed the *witness complexes* with a *max-min* criterion for the selection of PCD. The simplicial complexes are analyzed with persistent homology and the Betti numbers are presented with the persistent barcodes. We classified the DC motors into two classes according to the Betti numbers sequences. We labeled the classes with *good* and *faulty*, and then we compared our results with the classifications performed by an expert operator. From this comparison we can argue that the signals with Betti numbers $\beta_0 = 1$, $\beta_1 = 1$ are good, while the motors with $\beta_0 = 1$, $\beta_{i, i \geq 1} = 0$ are faulty motors. In good motors there are not evidence of periodical behaviors and they are topological equivalent to a closed loop. Conversely, faulty motors are characterized by periodical vibrations in the embedded signal, this behavior is topologically equivalent to a filled *geometrical object* with topological invariant $\beta_0 = 1$.

Application 3.2 - Topological clustering of RNA Secondary Structure Space

Homology is the natural tool for dealing with circular shapes recognition and comparison. This task can be used for satisfying the analysis of shapes with biological meaning, e.g. for studying DNA or RNA sub-motifs. We have employed persistent homology to classify RNA suboptimal secondary structures [6]. Since the lowest free energy common structure is not always the correct structure, topological analysis help to suggest alternative set of conformations that are structurally similar to the lowest free energy structure. Here, RNA suboptimal structures are considered as a point cloud data coming from a sequence; then RNA distance is used to compute the similarity between each point. Since we obtained a distance function on the set of point clouds; Rips filtration is performed over them. We compared different metric spaces and at the end we selected the *tree edit distance* over which we computed the *Vietoris-Rips. Vietoris-Rips* are used for probing the topological space associated to the suboptimal structures. In this ongoing study, all the filtrations are performed by using JavaPlex. This analysis revealed that there is a structure conservation among RNA secondary structures of the same family. Our preliminary result shows that persistent homology is captures important information from secondary structure space of species of different family; it clusters the species minimum free energy structure into different families. The persistent homology analysis shows that structural dissimilarity can be observed even for species that are classified under the same genus. Moreover, we introduce a shape language for representing RNA secondary structures in a non-standard, non-linear way [7]. The main motivation is to propose a new interpretation of RNA folding as a self-adaptability process, within the S[B] paradigm, towards a minimum free energy configuration. An RNA secondary structure is decomposed first by distinguishing between pseudoknot free and pseudonotted sub-structures. For pseudoknot free sub-structures a proper

formal language is defined. To address the representation of pseudoknotted substructures the crucial aspects of RNA irreducible shapes and their associated automatic groups are introduced.

4 Discovering higher dimensional relationships in weighted networks: Clique Weight Rank Persistent Homology

Simplicial complexes can be built also from graphs. Given an undirected graph, a *Clique Complex* is based on the individuation of *cliques*, subsets of the vertex set where each element is connected with all the others. Once that all the cliques are found, and they will become the faces of the complex [14].

Definition 1 (Clique complexes). *Given a graph $G = (V, E)$, where V is the set of vertex and E is the set of edges $E \subseteq V \times V$, a clique complex of G is the simplicial complex $X(G)$ on V whose simplices are all cliques $\sigma \subset V$. For example see Figure 3*

The technique that completes a graph to a simplicial complex and studies its homology is known as Clique Weight Rank Persistent Homology (CWRPH), see Fig.3. CWRPH is implemented by *jHoles* algorithm [1]. The reason to move from a graph to a simplicial complex is that the former is a suitable representation for a collection of two bodies problem, but in case of complex systems (systems of dynamical simultaneously interacting systems) this representation does not handle with higher dimensional structures. A simplicial complex is the natural algebraic representation of a such structures. For example a 2-dim simplicial complex (that is a filled triangle) might be used for representing three components that are interacting simultaneously. It exists if and only if the interaction is not decomposed into a collection of smallest pieces, for example three edges that otherwise would represent three 2-bodies problems. In our opinion, CWRPH can be connected with machine learning techniques, e.g. support vector machine, by defining new kernels based on simplices. This will play a fundamental role in the identification of the most relevant higher dimensional communities within a complex systems. CWRPH has been successfully used for analyzing biological networks. Here we report on four experiments.

Application 4.1 - Topological classification of brain activities.

The first experiment regards the brain and was conducted by Petri et at. [14]. In this experiment the authors converted functional magnetic resonance signals into complex networks that are completed and studied by CWRPH. In the paper the authors study the characteristics of functional brain networks at the mesoscopic level from a novel perspective that highlights the role of inhomogeneities in the fabric of functional connections. This can be done by focusing on the features of a set of topological objectshomological cyclesassociated with the weighted
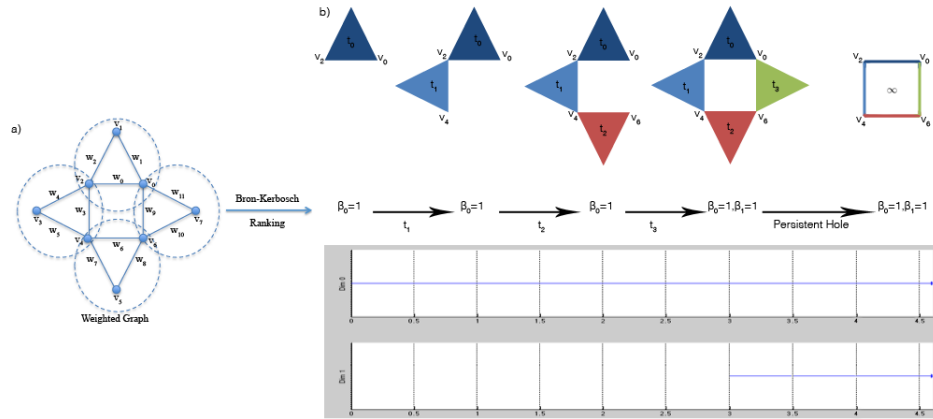
Fig. 3. How CWRPH works: a) Starting from a weighted undirected graph Bron-Kerbosch algorithm is used for listing all the maximal cliques. The weight distribution is categorized and each clique is weighted with the corresponding minimum or maximum category. b) A k-clique is equivalent to a k-1 simplices. The simplicial complex is build incrementally nesting the filtered clique complexes and persistent homology is computed.

functional network. We leverage the detected topological information to define the homological scaffolds, a new set of objects designed to represent compactly the homological features of the correlation network and simultaneously make their homological properties amenable to networks theoretical methods. As a proof of principle, we apply these tools to compare resting- state functional brain activity in 15 healthy volunteers after intravenous infusion of placebo and psilocybinthe main psychoactive component of magic mush- rooms. The results show that the homological structure of the brains functional patterns undergoes a dramatic change post-psilocybin, characterised by the appearance of many transient structures of low stability and of a small number of persistent ones that are not observed in the case of placebo.

Application 4.2 - Topological description of skin cancer.

The second paper by Binchi et al., reported on the application of *jHoles* for studying the evolution of skin cancer [1]. The authors shown how the connectivity of epidermal cells changes in response to a tumor by analyzing an in silico model. The biological network has been derived from the proliferative, differentiated and stratum corneum compartments, and jHoles used for studying variation of the connectivity. Briefly, models for tumor growth and skin turnover are combined with pharmacokinetic (PK) and pharmacodynamic (PD) models to assess the impact of two alternative dosing regimens on efficacy and safety. We studied the evolution of the topology (or the local connectivity). Epidermal cells sequentially pass three compartments, named proliferative (pc), differen-

tiated (dc), and stratum corneum (sc) compartments. We obtained a network representation of the compartments connecting the cells using both their admissible evolution (i.e., proliferative are connected only with differentiated and differentiated with stratum) and their concentration. The homological analysis of the network for the healthy epidermis shows a higher number of holes that means a more spread cells distribution (the Betti numbers sequence: $\beta_0 = 1$ and $\beta_1 = 28698$), due to the presence of the three compartments. After the tumor the topology of network changed and the new sequence of Betti number is $\beta_0 = 1$ and $\beta_1 = 24698$ with a reduced number of holes that means healthy cells disappeared and the network is less connected. Moreover, the authors were able to detect the diffusion direction of the tumor and observed that the intermediate cells disappeared more quickly then the others [1].

Application 4.3 - Topological modeling of human immune system.

The third paper reports on the application of CWRPH to a network based model of the mammal immune system, the so-called *idiotypic network* and simulated with *C-ImmSim*. In the simulator each idiotype (both antigens and antibodies) is represented with a bit-string, in our case of 12 bit length. In our configuration a simulation has a lifespan of 2190 ticks, where a tick=8 hours, and a repertoire of at most 10^{12} antibodies and antigen volume equal to $V = 10\mu L$. An idiotype interacts with each other if and only if their Hamming distance is $11 \leq d(A_j, A_k) \leq 12$. The pair-wise distances are stored in a matrix, the so-called *Affinity matrix*: $J_{i,k}$. From the affinity matrix and the volumes of each antibodies a new weighting function is derived the so-called *coexistence function*. For each simulation we computed the coexistence function and we used the weighted idiotypic network as input for the CWRPH. The persistent barcodes are used for computing both the *persistent entropy* and for identifying the persistent holes and their generators, namely the persistent antibodies that govern the evolution of the idiotypic network during the virgin state, the activation and the *immune memory*. *Persistent entropy* (see figure **??**) is able to recognize the activation of the immune system: the peaks in the charts point out the *immune activation* that is following by a transient that represents the *immune response*. During the immune response the antibodies play a dual role: they can simultaneously elicit and suppress each other. After this transient there is a plateau that represents the persistent *immune network activation* corresponding to the immune memory. Persistent entropy is directly computed from the result of *persistent homology*: the Betti numbers. The analysis of the generators of the homological classes allows to identify the real number of antibodies that have been used: 203 instead of 4096. The analysis of the persistent Betti numbers reveals that there is a subset of antibodies arranged in a 1-dimensional hole that is present both in the *activation state* and in the *memory state*. This 1-dimensional hole is formed by the antibodies Ab_1, Ab_2, Ab_7, Ab_{13}. This hole is formed by the most active antibodies. The removal of this 1-dimensional hole from the barcodes will flatten the entropy, that means this cycle is formed by the most specialized antibodies

for the antigen that has been injected. Both the *approximated von Neumann entropy* and the *persistent entropy* can be thought as *complexity measures* for graphs or for simplicial complexes. The reason is evident in their mathematical definitions: von Neumann entropy depends on the total number of vertices and the degree of linked vertices, while persistent entropy depends on the *topological noise* and by the *persistent topological features*. From this paper another one has been derived and it focuses on the definition of a new general methodology for modeling complex systems. The methodology is based on TDA, Information Theory and formal grammars [10].
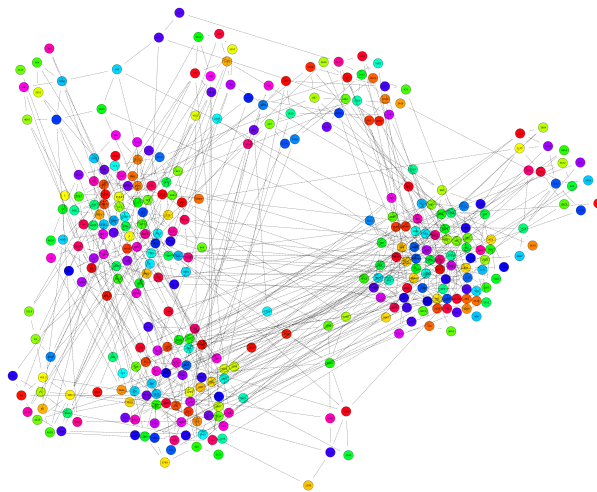

Fig. 4. Example of immune network. The nodes represent the antibodies, a link exists if and only if two antibodies are affine and are interacting. The node color represents the antibodies classes.

Application 3.4 - Topological detection of epileptic seizure.

We conclude by summarizing the fourth paper that describes a methodology based on TDA for capturing when a complex system, represented by a multivariate time series, changes its internal organization. The methodology segments a multivariate time series, i.e. a EEG, and transforms each segment into a simplicial complex. Simplicial complexes are studied by persistent homology and persistent entropy. In order to verify the reliability of the methodology, the authors have analyzed the EEG signals of *PhysioNet* database and they have found numerical evidences that the methodology is able to detect the transition between the pre-ictal and ictal states. The EEG signals used in this study were collected at the Children's Hospital Boston, and they consists of EEG recordings

from pediatric subjects with intractable seizures. Subjects were monitored for up to several days following withdrawal of anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. We applied the procedure described to both signals and we found the optimal size of the segmentation is equal to 120secs, then we segmented the whole EEG track in 30 windows. For each window we computed the partial correlation coefficients and we used as threshold $\theta = 0$. The upper triangular part of each matrices was parsed and saved as edge list, hence the edge list was used as input for *jHoles*. *jHoles* provides the Betti barcodes both in graphical and textual formats, and we used the latter for computing the weighted persistent entropy over each homological dimension (H_0, H_1, H_2, and H_3). We plotted the WH_j values for each matrix (see Figure 5). The two signals have been previously classified
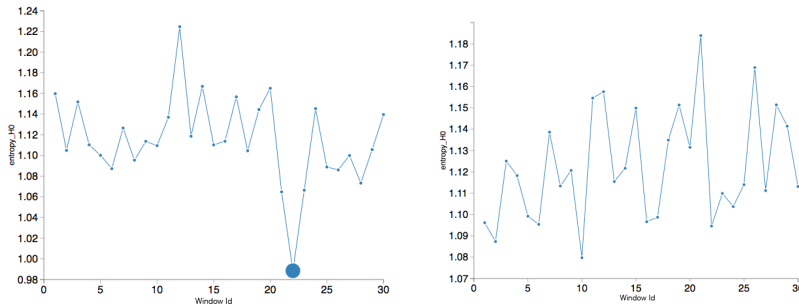

Fig. 5. Weighted Persistent Entropy for the homological group H_0. Left: Weighted persistent entropy for the SIGI, the marked peak corresponds to an ictal state. Right: Weighted persistent entropy for the SIGII. In SIGI a phase transition is well evident.

by the *PhysioNet* users. Respectively, SIGI corresponds to an individual affected by epilepsy, while SIGII belongs to an healthy patient. This classification also is identified by our methodology. The analysis of persistent entropy reveals that in WH_0 of SIGI a phase transition occurs (see the upper picture in Figure 5). The topological interpretation is that among the windows with id = 20, 21 and 22, the number of connected components tends to be one and the topological noise is minimized (all the features are persistent). Before and after this period, the number of connected component is higher and the barcodes are noisy. These three windows correspond exactly to the transition from the pre-ictal state to the ictal state. In both signals, Betti numbers for higher dimensions are present (β_1, β_2, β_3) but in these signals the corresponding barcodes do not change significantly [9]. These results will be exploit to characterize a more general framework for monitoring epilepsy [8].

5 Shaping signals with simplicial complexes

Piecewise linear function (PL) is a powerful mathematical tool largely used for approximating signals. The task of measuring the similarity among piecewise linear functions (PLs) is still an open issue and a solution is strongly required in machine learning methods. The comparison between the area under the curves (AUCs) of discrete digital signals is a weak measure: for each value of AUC a family of infinite signals exists. A PL can be threated as a 1-dimensional filtered simplicial complex (a chain graph) and then studied by persistent homology and persistent entropy. We argue this approach can be a fruitful kernel for new machine learning methods for dealing with the supervised classification of univariate real noisy signals.

Application 5.1 - Topological classification of real lenght noisy signals

Rucco et al., present a novel methodology based on a topological entropy, the so-called persistent entropy, for addressing the comparison among discrete piecewise linear functions [17]. The comparison is certified by the *stability theorem for persistent entropy*. The theorem is used in the implementation of a new algorithm. The algorithm transforms a discrete piecewise linear function into a filtered simplicial complex that is analyzed with persistent homology and persistent entropy. Briefly, the methodology threats each points within the input signal as a 0-simplex filtered by their y value. Two subsequent points are connected by one 1-simplex filtered by $max\{y_1, y2\}$. The resulting simplicial complex is studied by persistent homology and persistent entropy.The authors used this approach for facing the supervised classification problem of real long length signals of DC electrical motors. The quality of classification is stated in terms of the area under receiver operating characteristic curve (AUC=94.52%).

6 Conclusions and future work

Computational topology has played a synergistic role in bringing together research work from computational geometry, algebraic topology, data analysis and many other related scientific areas. Recently, the field has undergone particular growth in the area of TDA. The application of topological techniques to traditional data analysis, which traditionally was mostly based on a statistical setting, has opened up new opportunities. This short review is intended to summarize the contribution of TDA in the study of different types of datasets. We focused on persistent homology based techniques for dealing with point cloud data, complex networks and signals. In details, we reported on Vietoris-Rips, Clique Weight Rank Persistent Homology and Piecewise Complexes techniques and the related applications. We hope that these applications might be used by the reader as signpost in his or her personal experience in using topological data analysis. We remark that here we reported only briefly an incomplete list of our works. We plan to extend this work to include also Q-Analysis, Mapper and how

TDA can be connected to formal grammars, to machine learning for modelling complex systems and to a topological field theory of data. Preliminary results have been introduced in [10, 18, 19, 12].

References

 1. Jacopo Binchi, Emanuela Merelli, Matteo Rucco, Giovanni Petri, and Francesco Vaccarino. jholes: A tool for understanding biological complex networks via clique weight rank persistent homology. *Electronic Notes in Theoretical Computer Science*, 306:5–18, 2014.
 2. Gunnar E. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
 3. Joseph Minhow Chan, Gunnar E. Carlsson, and Raul Rabadan. Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46):18566–18571, 2013.
 4. Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(339-358):24, 2007.
 5. Herbert Edelsbrunner and John Harer. *Computational topology: an introduction.* American Mathematical Soc., 2010.
 6. Adane Mamuye, Emanuela Merelli, and Matteo Rucco. Persistent homology analysis of the rna folding space. In *Proc. 9th EAI Conference on Bio-inspired Information and Communications Technologies (BICT 2015)*, 2015.
 7. Adane Mamuye, Emanuela Merelli, and Luca Tesei. Towards a shape language for interpreting rna folding. In *Proc. 9th EAI Conference on Bio-inspired Information and Communications Technologies (BICT 2015)*, 2015.
 8. Emanuela Merelli and Marco Piangerelli. Rnn-based model for self-adaptive systems, the emergence of epilepsy in the human brain. In *Proceedings of the International Conferenceon Neural Computation Theory and Applications (NCTA-2014)*, 2014.
 9. Emanuela Merelli, Matteo Rucco, Marco Piangerelli, and Daniele Toller. A topological approach for multivariate time series characterization: the epilepsy case study. *Proc. 9th EAI Conference on Bio-inspired Information and Communications Technologies (BICT 2015)*, 2015.
10. Emanuela Merelli, Matteo Rucco, Peter Sloot, and Luca Tesei. Topological characterization of complex systems: Using persistent entropy. *Entropy*, 17(10):6872–6892, 2015.
11. Vidit Nanda and Radmila Sazdanović. Simplicial models and topological inference in biological systems. In *Discrete and Topological Models in Molecular Biology*, pages 109–141. Springer, 2014.
12. Vidit Nanda and Radmila Sazdanović. The topological field theory of data: a program towards a novel strategy for data mining through data language. In *Journal of Physics: Conference Series, 626*, pages 109–141. IOP Publishing, 2015.
13. Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *arXiv preprint arXiv:1506.08903*, 2015.
14. Giovanni Petri, Paul Expert, Federico Turkheimer, Robin Carhart-Harris, David Nutt, Peter J. Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.

15. Matteo Rucco, Filippo Castiglione, Emanuela Merelli, and Marco Pettini. Characterisation of the idiotypic immune network through persistent entropy. In *Proc. of 11th European Conference on Complex Systems (ECCS 2014)*, Springer Proceedings in Complexity, 2015.

16. Matteo Rucco, Enrico Concettoni, Cristina Cristalli, Andrea Ferrante, and Emanuela Merelli. Topological classification of small dc motors. In *1st Int. Forum on Research and Technologies for Society and Industry (RTSI)*, pages 192–197. IEEE, 2015.

17. Matteo Rucco, Rocio Gonzalez-Diaz, Maria Jose Jimenez, Nieves Atienza, Cristina Cristalli, Enrico Concettoni, Andrea Ferrante, and Emanuela Merelli. A new topological entropy-based approach for measuring similarities among piecewise linear functions. *CoRR*, abs/1512.07613, 2015.

18. Matteo Rucco, Emanuela Merelli, Damir Herman, Devi Ramanan, Tanya Petrossian, Lorenzo Falsetti, Cinzia Nitti, and Aldo Salvi. Using topological data analysis for diagnosis pulmonary embolism. *Journal of Theoretical and Applied Computer Science*, 9:41–55, 2015.

19. Matteo Rucco, David Rodrigues, Emanuela Merelli, Jeffrey H Johnson, Lorenzo Falsetti, Cinzia Nitti, and Aldo Salvi. Neural hypernetwork approach for pulmonary embolism diagnosis. *BMC Research Notes*, 8:617, 2015.

20. Gurjeet Singh, Facundo Mémoli, and Gunnar E. Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In *Symposium on Point Based Graphics, Prague, Czech Republic, 2007. Proceedings*, pages 91–100, 2007.