Exploratory Topological Data Analysis of Resting-State fMRI Data

Marta Raquel Azevedo Castro
marta.castro@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2019

Abstract

Algebraic topology offers methods to assess the global properties of networks through the computation of homology groups. When studying networks at a mesoscopic scale, one is often interested in distinguishing important features from noise. This is the subject of persistent homology, whose purpose is to extract long-lived topological features of the data through the gradual construction of the network and computing its homology groups.

In this work, time-series data from 10 resting state functional magnetic resonance images were extracted and used to construct functional connectivity networks, with the aim of investigating the presence of weaker functionally connected areas.

Comparisons of the networks with two randomized versions showed that regions enclosing weaker functional connectivity areas in the original networks, are born in later filtration steps and have shorter persistence intervals. Regions enclosing persistent weaker functional connectivity areas were obtained inspiring further investigations.

Keywords: algebraic topology, persistent homology, functional magnetic resonance image, functional connectivity, barcodes

1. Introduction

Network theory has been commonly used to study functional brain activity using neuroimaging data. The standard approaches from network theory rely on the measures of the strength of the pairwise connections between the microscopic constituent elements, resulting in a weighted graph that is often studied using graph theory.

Algebraic topology offers methods to assess the global properties of a graph, by associating with it a collection of algebraic objects called *homology groups*. Homology groups of dimension k, $H_k(X)$, provide information about the characteristics of linear combinations (chains) of simple oriented units known as simplices, which can be used to represent a graph through what is known as a simplicial complex. The elements of homology groups are cycles, (i.e, chains with vanishing boundary) and can be computed using methods from linear algebra [1].

An important piece of information when studying brain functional networks is the number of weaker connectivity areas, and to assess which are important features of the network. This is the matter of persistent homology [2], whose purpose is to extract long-lived topological features of the data through the gradual addition of simplices, creating subgraphs at each step until the whole graph is built.

The focus of this work is to extract time-series data from 10 resting state functional magnetic resonance images, construct the functional networks and perform an exploratory analysis of the data using persistent homology, with the intention of investigating the presence of persistent weaker functionally connected areas. This work is organized as follows: In the second chapter an introduction to the relevant theoretical concepts necessary to understand this work is expounded. The third chapter contains all the methods used to pre-process the data and to analyze it, which includes the method from persistent homology used. In the fourth chapter, the results obtained from the analysis are presented. The conclusions and future work are presented in the last chapter.

2. Background

2.1. Fundamentals of Neuroimaging Techniques

In the Magnetic Resonance Imaging (MRI) technique, an oscillating magnetic field is created and the signal emitted by the hydrogen atoms present in the water molecules that exist in the human body is measured. One of the responses to neuronal activity is the supply of oxygenated hemoglobin [3]. In 1980, Thulborn verified that more MRI

signal was present where blood was highly oxygenated and less MRI signal was present where blood was highly deoxygenated. Ogawa, in 1990, speculated that blood-oxygenation-level dependent (BOLD) contrast could enable the measurement of functional changes in brain activity. More active areas resulted in a more intense BOLD signal that could then be detected using a fMRI machine.

In a scanning process of a fMRI machine several slices of the 3D brain are performed using a grid of voxels with a pre-determined thickness, that is, a grid made by 3 dimensional rectangular cuboids with dimensions determined by the thickness of the slice. The slices are acquired several times throughout different points in time and the combination of all slices make the 3D brain. At a given time point, the BOLD signal of each voxel in the grid is measured. After all the slices have been performed, the response of the BOLD signal of each voxel over time is known. Resting state fMRI focuses on measuring low frequency fluctuations in the BOLD signal without the presence of active tasks.

2.2. Algebraic Topology

Given $k+1$ points $v_0, v_1, ..., v_k$ in general position, a *k-simplex* is the smallest convex set containing them, and its dimension is equal to the number of vertices that define it minus one [4]. Every k-simplex has $(k-1)$-faces that are themselves $(k-1)$-simplices.

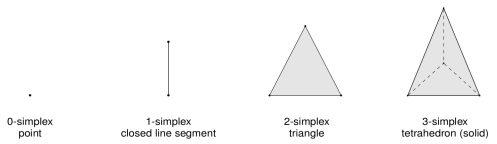


Figure 1: Simplices of the first 4 dimensions.

A finite collection of simplices in some euclidean space X is called a *simplicial complex* if whenever a simplex lies in the collection, so does any of its faces, and whenever two simplices of the collection intersect, they do so in a common face.

The free abelian group generated by the oriented q-simplices of a simplicial complex K is denoted by $C_q(K)$, and an element in this group is called a *q-chain*. A q-chain can be seen as a linear combination of oriented q-simplices of K with integer coefficients, $\lambda_1\sigma_1 + ... + \lambda_s\sigma_s$, keeping in mind that $\lambda(-\sigma) = (-\lambda)\sigma$, where $-\sigma$ corresponds to σ with the orientation reversed.

The boundary of a q-simplex σ_q is denoted by $\partial\sigma_q$, and is the $(q-1)$-chain determined by the sum of its $(q-1)$-dimensional faces, each of the faces taken with the orientation induced by the simplex σ_q. For a specific ordering of its vertices $(v_0, ..., v_q)$, the boundary of the q-simplex σ_q is given by

$$\partial(v_0, ..., v_q) = \sum_{i=0}^{q} (-1)^i (v_0, ..., \hat{v_i}, ..., v_q) \quad (1)$$

where $(v_0, ..., \hat{v_i}, ..., v_q)$ is the ordered $(q-1)$-simplex obtained by deleting the vertex v_i. Changing the orientation of σ_q changes the orientation induced on each of its faces, therefore $\partial\sigma_q + \partial(-\sigma_q)$ is zero. Thus, the boundary operator ∂_q determines an homomorphism

$$\partial_q : C_q(K) \to C_{q-1}(K)$$

where in the special case when $q = 0$, the boundary of a single vertex is defined to be zero and the group $C_{-1}(K)$ is the group $\{0\}$.

A crucial property of the boundary is that the boundary of the boundary is zero for any chain [5]. A *q-cycle* of K is any linear combination of q-simplices with vanishing boundary, therefore, the group of q-cycles $Z_q(K)$ is equal to the kernel of $\partial_q : C_q(K) \to C_{q-1}(K)$. The group $B_q(K)$ refers to the group of bounding q-cycles of K, that is, the group whose elements are any q-cycles that are the boundary of a linear combination of oriented $(q+1)$-simplices. The group $B_q(K)$ is the image of $\partial_{q+1} : C_{q+1}(K) \to C_q(K)$, and is a subgroup of the group of q-cycles, because of the above property.

Homology is an algebraic object that can be associated to any topological space that allows one to detect in some sense the number of *holes* the space has in each dimension [4]. The n^{th} *Homology group* of a simplicial complex K as stated in [1] can be defined in terms of the group of n-cycles $Z_n = ker\ \partial_n = \{z \in C_n : \partial_n(z) = 0\}$, and the group of bounding n-cycles $B_n = im\ \partial_{n+1} = \{b \in C_n : \exists z \in C_{n+1} : b = \partial_{n+1}(z)\}$ as

$$H_n = \frac{Ker\ \partial_n}{Im\ \partial_{n+1}} = \frac{Z_n}{B_n} \quad (2)$$

The *homology class* of an n-cycle z is denoted as $[z]$. Two n-cycles are said to be *homologous or equivalent* if they have the same homology class, that is, if their difference is a bounding n-cycle. The n^{th} *Betti number* of a simplicial complex K, denoted by β_n, is the smallest cardinality of the generating set of infinite cyclic groups in the n^{th} Homology group H_n, and can be written as

$$\beta_n = rank(H_n) = rank(Z_n) - rank(B_n) \quad (3)$$

The first three betti numbers give important information about the topology of the simplical complex (graph). The β_0 represents the number of connected components of the graph, in the case of a connected complex, $\beta_0 = 1$. The number of non

bounding 1-cycles, in other words, the number of 1 dimensional *holes* in a graph is accounted for in β_1, whereas β_2 accounts for the number of 2-dimensional "voids".

To study the importance of a certain topological property in a simplicial complex, one can consider the evolution of a simplicial complex starting from the empty set, and assume that simplices are added to the complex gradually, building a sequence of subcomplexes such that $\emptyset = K^0 \subseteq K^1 \subseteq K^2 \subseteq ... \subseteq K^m = K$, this process is called a *filtration*.

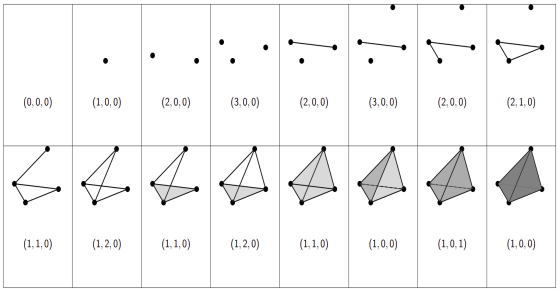


Figure 2: Example of a filtration.

Let h denote a topological attribute and let l denote the step or the index of the filtration in which it appears, then it is said that h is *born* at filtration step l if the homology group of K_l is the first homology group in the filtration to include h. In a similar manner, a topological attribute *dies* at filtration step s if it is present in the homology group of K_{s-1} but not in the homology group of K_s. According to these definitions, *persistence* can be defined as

$$p = s - l \tag{4}$$

Therefore, at each step of the filtration one can compute the betti numbers to gain insight into which topological features persisted for longer filtration intervals (figure 2).

A *Barcode* is a useful way to visualize the persistence of a topological feature, and is a representation in which the rank of the homology groups is present in the vertical axis, as a function of the filtration step (horizontal axis).

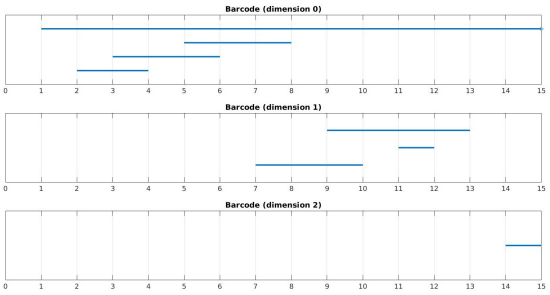


Figure 3: Example of a barcode for filtration in figure 2. On the y axis one can see the betti numbers of the first 3 dimensions and on the x axis one can see the filtration steps.

3. Materials and Methods

3.1. Image Pre-Processing

The experimental data used in this work were collected at Max Planck Institute for Human Brain and Cognitive Sciences, Leipzig, Germany. For each participant, the 3D structural MRI and the resting state fMRI were acquired. Before the scan the participants were instructed to stay awake, focus on a cross and keep their eyes open. For the purpose of this work, only 10 of the 22 images available were used. The anatomical and the resting-state functional data were collected according to the parameters described in [6]. In this work, only the images obtained in the first session were considered. The tools used to analyze the data belonged to the FMRIB Software Library (FSL)), Oxford Center for Functional Magnetic Resonance Imaging of the Brain (FMRIB). The pre-processing steps were performed using the sequence of steps available at https://github.com/cmpetty/mypytools, which include slice time correction, motion correction, brain extraction, registration and normalization and high-frequency noise removal.

3.2. Functional Connectivity Network

To analyze the brain image and build a functional connectivity network it is necessary to create a mapping from the image into the actual network, which starts by identifying a set of functional "nodes", and then attempts to estimate the set of "edges". A node of the network is a region of interest (ROI) and can be seen as a single voxel or a whole brain region. An edge represents a functional connection between two nodes, and a connection can be implied through the correlation between the time-series of any two nodes in the network. In this work, the edges were weighted according to the Pearson product-moment correlation coefficient (PPMCC), which provides a measure of the linear correlation between any two ROIs i and j [7].

For a general functional network, let X be $[x_1, x_2, ...x_N]^T$, where N denotes the number of time points in the time-series, T denotes the number of vertices in the network, and x_k denotes the BOLD signal intensity at time k, then the PPMCC can be formally defined as

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} \cdot C_{jj}}} \tag{5}$$

where C_{ij} is the estimated covariance of x_i and x_j, and the element C_{ii} is the variance of x_i. This coefficient takes values between $+1$ and -1, where $+1$ is a total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation.

Using this similarity measure it is possible to construct a matrix of similarities between the time-

series with the diagonal entries equal to 0, and obtain a functional network whose weighted adjacency matrix or association matrix $A_{i,j} = (a_{ij})_{i,j=1}^{N}$ has elements that indicate the similarity between the time-series of the regions i and j, where the N here represents the total number of ROIs [8].

To build such a network, a parcellation of the image was performed based on a label file with 116 regions of interest, by taking the average time-series of the set of voxels that belonged to each region.

3.3. Persistent Homology Computation

A *n-clique* in a graph G is a subgraph containing n vertices such that every vertex in it is connected to every other vertex. Such a subset of vertices $\{v_1, v_2, ..., v_n\}$ has the property that every nonempty subset of k vertices $\{v_{i_1}, ..., v_{i_k}\}$ spans a k-clique. A clique that is not a subset of a larger clique is a *maximal clique*. The *clique complex* associated to G is defined as the abstract simplicial complex whose n-simplices are the $(n+1)$-cliques in G, and can be computed by finding all cliques in the graph.

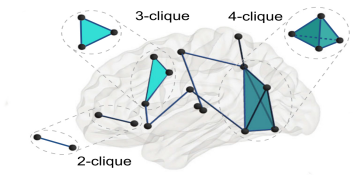
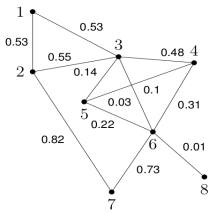


Figure 4: Cliques in the connectivity network (Source: [9]).

In brain connectivity networks, the connections are stronger between some nodes and weaker between others. Therefore, it is convenient to consider a filtration that is based on the connectivity strength between two nodes in the network. In this work, a filtration first proposed in [10] and inspired by the Vietoris-Rips complex was considered. The *Weight Rank Clique Filtration* (WRCF) combines the clique complex construction with a thresholding on weights. Given a weighted network Ω, this filtration starts by ranking all weights of the links from ω_{max} to ω_{min}. The parameter ϵ_t scans the sequence of weights, and by decreasing order of edge ranking, at each step t, one considers the thresholded graph $G(\omega_{ij}, \epsilon_t)$, that is, the subgraph of Ω with links of weight larger than ϵ_t. The clique complex $K(G, \epsilon_t)$ is built for each graph $G(\omega_{ij}, \epsilon_t)$. In this way, the clique complexes are nested according to the growth of t and determine the WRCF. The following figure shows the WRCF for a weighted graph Ω.



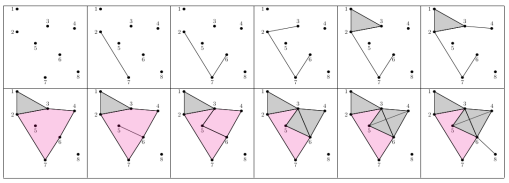


Figure 5: Weight rank clique filtration.

The edges are added by decreasing order of the weight. In the first filtration step, the graph G consists solely of the vertices, and in the next sequence of steps the edges are added. If two edges share the same weight, they are added in the same filtration step. Whenever there are edges connecting all k vertices forming a k-clique, a $(k-1)$-simplex is born. In this example, it is possible to see that when the edges $[1,3]$ and $[1,2]$ are added, a 3-clique is formed giving rise to a $2-$simplex. When the edge $[4,6]$ is added, the first 1-dimensional *hole* is born, which becomes geometrically smaller as the next edges are added.

A weighted network *hole* of dimension 1 with weight w is a 1-cycle of n nodes $i_0, i_1, i_2, ..., i_{n-1}$, such that all edges (i_l, i_{l+1}) (with $i_0 \equiv i_n$) have weights $\geq w$, and all the other possible edges crossing the *hole* have weights strictly weaker than w [10]. Similarly for the case of 2-dimensional weighted *holes*, but instead of edges one considers triangles.

To study the persistence of weighted *holes*, a well documented software designed to compute homology called Javaplex was used [11]. This software allows the user to explicitly define the simplicial complex (or *stream*) through the addition of a simplex or a set of simplices at each filtration step t, and outputs one representative cycle and its lifetime for each *hole* found. In this work, only *holes* of dimension 1 and 2 are being studied, therefore the output from Javaplex is a set of representative cycles determined by the linear combinations of the edges and triangles that enclose the *holes* present, although not necessarily the geometrically optimal ones, as a cycle may be equivalent to multiple cycles. In the last filtration step represented in figure 5, one can see that the cycles formed by the edges $[2,3]+[3,6]+[6,7]+[7,2]$ and $[2,3]+[3,5]+[5,4]+[4,6]+[6,7]+[7,2]$ are equivalent, in the sense that they enclose the same

1-dimensional *hole*, and it might be the case that the first one is geometrically correct and the second one is returned instead. The problem of finding geometrically optimal representative cycles is still an open area of research.

In order to build the clique complex, one needs to compute all the maximal cliques in the subgraph G_t. In this work, all the maximal cliques were found using the `find_cliques()` function available in the Python language software package Networkx, which is an implementation of the algorithm first proposed in [12], as it is the most efficient algorithm known for this problem [8]. Whenever a maximal clique is found at the filtration step t, it is added to the stream if not present yet. For each maximal clique of size k found, one needs to check whether or not all of its faces are already in the stream by finding all the combinations of its elements, which can be done through the computation of the powerset of the set of all elements in the clique.

A Python script that executed the above steps was built with the purpose of writing a Javaplex based text file that contained the explicit definition of each simplex at a given filtration step. Since it is computationally expensive to find all missing faces, as new maximal cliques are being discovered at each step, only the positive entries of the association matrix were considered.

Each weighted *hole* g present in a network can be characterized by its birth index β_g, its persistence p_g and its length λ_g. The birth index of a *hole* is the step index t of its weight w, after all network edges are ordered in descending order. As more and more edges are added to the network, it is possible that an edge with weight $w' < w$, will appear and cross the *hole*, resulting in the *hole's death* γ_g. The persistence p_g corresponds to the difference between the *hole's* death and birth, $p_g = \gamma_g - \beta_g = t' - t$. Lastly, the length λ_g of a *hole* g is equal to the number of simplices composing it. The following two scalar metrics proposed in [10] provide a way to compare the WRCF results from different networks.

$$h_k = \frac{1}{N_{g_k}} \sum_{g_k} \frac{p_{g_k}}{T} \tag{6}$$

$$\tilde{h}_k = \frac{1}{N_{g_k}} \sum_{g_k} \frac{\lambda_{g_g}}{N} \frac{p_{g_k}}{T} \tag{7}$$

where $\{g_k\}$ is the set of generators of the k-th homological group H_k, $N_{g_k} = dim H_k$, N represents the number of nodes in the network and T the maximal filtration rank or maximum filtration value. The first measures the average persistence and is called *network hollowness* h_k, whereas the second takes not only the persistence but also the length of the generators of the k-th homology group into account and is called *chain-length normalized hollowness* $\tilde{h}_k$.

The networks were compared with two null models proposed in [13], to assess if the cycle's lengths, persistences and births were the same as expected by chance. The null models used were weight reshuffling, and both edge and weight reshuffling. The reshuffling of the list of weights was performed using the function `shuffle()` from the Python's library Random, and has the purpose of simply reshuffling the weights in the network, while keeping the overall topology intact. The links were reshuffled using the function `double_edge_swap()` from the Networkx library, with the intent of also reshuffling the topology of the network. The following figure illustrates the null models construction.

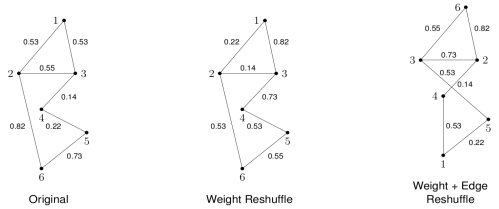


Figure 6: Weight rank clique filtration.

4. Results

To perform a comparison of each network with their randomized versions, each edgelist was truncated at the edge added in the last filtration step before the randomizations were applied. This is necessary since the computation of all the missing faces needs to be performed whenever a new maximal clique is discovered, which takes a long time depending on the number of cliques present in the network. Therefore, the WRCF method was applied to each network until the maximum filtration step was reached, and only the edges that were added until that step were considered to build the randomized networks.

4.1. Persistence, Length and Birth Distributions

The persistences and births collected were normalized by the maximum filtration values, so that the networks of different subjects could be compared. The figure below shows the distributions obtained from the similarity matrix of subject 5.

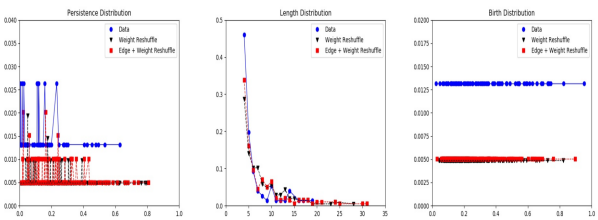


Figure 7: Persistence, length and birth distributions obtained for subject 5. On the y axis one can see the probabilities, and on the x axis the (normalized) persistences, lengths and (normalized) births.

The persistence distributions indicate that most *holes* are born and die in a nearby filtration step. The length distributions show a common tendency for the original networks and their randomizations: there is a higher probability of having 1 dimensional *holes* with a short length. The birth distributions indicate that the births occur uniformly across the filtration and more *holes* are born in early filtration steps. The plots in figure 8 allow the comparison of the distributions obtained for the 10 subjects. Longer persistent intervals and earlier births were found in the randomized versions. The original networks show higher variability of filtration birth times, specially for subject 10. The length distribution comparison shows a higher probability of forming 1-cycles with shorter lengths for all networks.

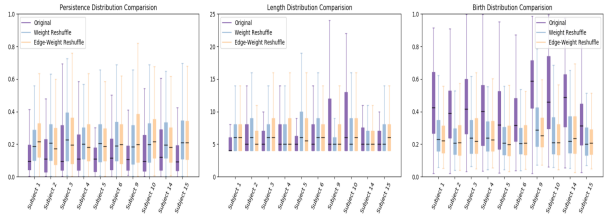


Figure 8: Comparison of the persistence, length and birth distributions.

4.2. Barcodes

The barcodes of the original networks and their randomized versions were collected for all subjects, and similar results were obtained among subjects. The following figure shows the barcodes obtained for the original network of subject 5. The barcodes of the randomized versions are shown in figure 10.

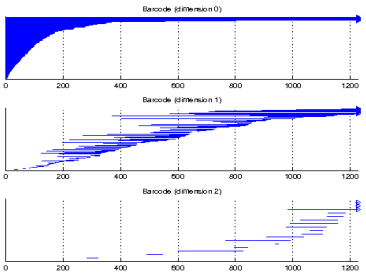


Figure 9: Barcodes of the association matrix of subject 5.

The barcodes of dimension 0, which account for the number of connected components, show that components are merged in earlier steps in the randomized versions for all subjects, and the barcodes of dimension 1 related to 1-cycles show a significantly higher number of *holes* with short persistence born in earlier steps in the randomized versions. These results might reflect the fact that edges that cause the death of these *holes* are added earlier, since both randomization processes change the order by which the edges are added to the complex throughout the filtration.

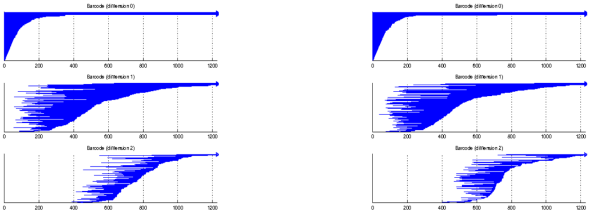


Figure 10: Barcodes of the association matrix of subject 5.

Short persistence 2 dimensional *holes* were found in all networks, and in higher number for the randomized networks, which might result from the fact that since the randomized versions presented a higher number of 1-dimension *holes*, there is a higher tendency for the added edges to give rise to 4-cliques, which are the faces of the 2-dimensional *holes* detected, and then close them as higher dimensional cliques start to be formed, hence the short persistence.

4.3. Network Hollowness

The following table contains the hollowness values obtained for each network and their randomized versions.

Subject	% of edges	h_1	$\tilde{h}_1$	$h_1^{rnd_1}$	$\tilde{h}_1^{rnd_1}$	$h_1^{rnd_2}$	$\tilde{h}_1^{rnd_2}$	h_2	$\tilde{h}_2$
1	31.106	0.137	0.0140	0.224	0.0162	0.249	0.0187	0.109	0.0113
2	37.518	0.156	0.0149	0.231	0.0197	0.219	0.0174	0.0589	0.0109
3	33.516	0.170	0.0195	0.266	0.0192	0.245	0.0206	0.0711	0.00696
4	31.984	0.171	0.0139	0.231	0.0195	0.243	0.0204	0.0870	0.0159
5	36.426	0.136	0.0102	0.237	0.0200	0.220	0.0184	0.0795	0.00824
6	38.997	0.174	0.0151	0.239	0.0219	0.227	0.0177	0.0792	0.0183
9	26.082	0.163	0.0202	0.239	0.0169	0.266	0.0203	0.0191	0.00160
10	31.554	0.176	0.0159	0.237	0.0193	0.254	0.0192	0.0511	0.00712
14	31.730	0.178	0.0143	0.236	0.0162	0.229	0.0155	0.0540	0.00384
15	37.034	0.152	0.0129	0.246	0.0195	0.238	0.0196	0.0971	0.0205

Figure 11: Summary table with the hollowness values for each network.

The table above shows the percentage of edges with positive PPMCC that were included for each subject, and both the hollowness and the normalized hollowness values for dimension 1 and 2 for the original networks. In this work, subjects 7, 8, 11, 12 and 13 were skipped due to reported errors related to the scanning process. The hollowness and normalized hollowness values of dimension 1 for the randomized versions are also shown. The weight reshuffle network is referred to as rnd_1, and rnd_2 refers to the edge plus weight reshuffle network.

Lower hollowness values were obtained for the original networks when compared to their randomized versions, indicating that the randomization process leads to more hollow networks, i.e, networks with a higher number of 1-dimensional "voids". Similar results were obtained for the US Air passenger networks 2000 and 2011, and for the online forum network as one can see in [10], with a small number of 2-dimensional "voids", captured by small values of h_2, and smaller values of h_1 when compared to their randomized versions.

4.4. Representative Cycles

To filter the most persistent 1-cycles from what can be considered noise, the representative 1-cycles returned by Javaplex that persisted for half of the maximum filtration value or more were collected for all subjects and are available in the following table.

Subject	Javaplex Representative 1-cycles
1	[110, 98, 96, 90, 88, 80, 66, 63, 59, 58, 54, 31, 29, 28, 25, 24, 23, 16, 13, 12, 8, 7, 6, 4, 3], [110, 98, 96, 90, 88, 80, 69, 68, 63, 57, 54, 31, 29, 28, 25, 24, 23, 19, 16, 13, 12, 8, 7, 6, 4, 3, 1]
2	[93, 92, 90, 61, 60, 59, 58, 53, 52, 50, 49, 27, 26, 25, 23, 22, 6, 5, 4], [86, 84, 82, 64, 35, 34, 29, 26, 25, 24, 22, 18, 14, 1], [79, 78, 63, 61, 60, 57, 56, 23, 22, 19, 7, 6, 3, 1]
3	[108, 99, 98, 96, 93, 85, 83, 81, 73, 70, 55, 47, 43, 41, 39, 37, 36, 31, 30, 29, 27, 26, 24, 23, 21, 15, 7, 6, 4, 3, 2], [108, 98, 96, 83, 81, 79, 73, 70, 41, 39, 37, 36, 31, 30, 29, 27, 26, 25, 24, 23, 21, 17, 15, 7, 6, 4, 3, 2], [110, 103, 99, 98, 97, 96, 85, 83, 81, 77, 76, 71, 70, 55, 54, 47, 43, 38, 36, 33, 31, 30, 23, 22, 15, 7, 3], [71, 70, 33, 32, 31, 30, 23, 22, 18, 2], [110, 103, 99, 85, 83, 81, 71, 70, 55, 47, 43, 33, 32, 31, 30, 23, 22, 15, 7, 3]
4	[93, 91, 65, 59, 58, 57, 51, 49, 48, 47, 45, 44, 43, 35, 34, 1], [102, 100, 99, 92, 90, 65, 63, 61, 59, 58, 55, 51, 49, 48, 47, 45, 44, 43], [68, 64, 51, 43, 38, 34, 32, 23, 21, 19, 2, 1, 0], [68, 65, 60, 58, 51, 49, 48, 47, 45, 44, 43, 34, 32, 19, 3]
5	[88, 64, 62, 61, 59, 57, 56, 35, 33, 32, 31, 30, 23, 22, 7, 2, 1], [115, 99, 97, 95, 94, 91, 86, 83, 82, 60, 59, 57, 56, 39, 30, 22, 3, 2, 1]
6	[80, 67, 66, 64, 62, 60, 59, 58, 50], [92, 85, 67, 66, 64, 62, 59, 51, 50, 49, 48, 45, 44, 42], [100, 99, 91, 90, 84, 77, 67, 66, 65, 64, 62, 59, 50, 48, 45, 44, 39, 30, 27, 25, 24, 16, 15, 12, 10, 8, 6, 5, 4, 3, 2], [101, 97, 93, 91, 45, 24, 23, 22, 6, 4, 2], [99, 91, 90, 77, 67, 66, 65, 62, 59, 50, 48, 46, 45, 44, 42, 39, 30, 25, 24, 17, 5, 4]
9	[98, 96, 76, 75, 74, 70, 54, 31, 25, 24, 5, 4], [111, 96, 89, 85, 84, 80, 76, 75, 74, 70, 66, 65, 62, 60, 58, 51, 49, 46, 45, 44, 42, 31, 25, 23, 22, 14, 13, 8, 7, 6, 5, 3, 2], [111, 107, 99, 98, 96, 91, 88, 54, 47, 46, 44, 43, 42]
10	[66, 65, 61, 59, 58], [65, 63, 61, 35, 31, 25, 23, 9], [65, 60, 35, 25, 23, 22], [99, 97, 87, 85, 83, 82, 80, 74, 54, 47, 46, 43, 42, 31, 30, 25, 24, 23, 17, 14, 5, 0], [99, 97, 85, 80, 67, 65, 58, 54, 47, 46, 43, 42, 35, 25, 24, 6, 2, 0], [67, 65, 58, 35, 33, 7, 6]
14	[112, 111, 105, 104, 98, 80, 78, 77, 76, 73, 72, 70, 55, 54, 47, 36, 28, 26, 20, 16], [110, 109, 97, 80, 78, 69, 67, 66, 57, 47, 46, 33, 32, 19, 16, 1, 0], [55, 53, 51, 49, 48, 47, 45, 44], [83, 72, 30, 26, 24, 20, 15]
15	[95, 69, 41, 39, 37, 23, 22, 19, 18, 10, 7, 6, 3, 1, 0], [67, 66, 58, 56, 30, 27, 23, 22, 10, 7, 6, 5, 3, 1, 0], [73, 70, 68, 34, 33, 32, 5], [93, 92, 91, 89, 85, 81, 80, 76, 73, 62, 61, 53, 52, 51, 37, 35, 34, 32, 29, 17, 15, 11, 7, 3, 1], [115, 111, 110, 105, 85, 81, 75, 47, 46, 39, 36, 29, 27, 22, 21, 5, 3]

Figure 12: Table with the regions belonging to the 1 dimensional cycles that persisted for half or more than half of the maximum filtration value.

To assess which regions belong to the most persistent 1-cycles found, the following distribution was computed, where on the y axis one can see the fraction of subjects who showed each region in a persistent 1-cycle.

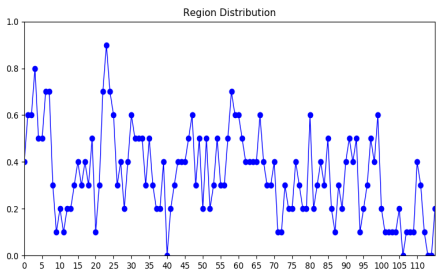


Figure 13: Distribution of regions.

The Frontal_Sup_Medial_R was the ROI that appeared in 9 of the 10 subjects' most persistent representative 1-cycles studied. None of the persistent 1-cycles found contained the regions that correspond to Amygdala_L, Cerebelum_10_L, Vermis_8 and Vermis_9.

5. Conclusions

This work intended to provide a mesoscopic characterization of the resting state networks. To serve this purpose, a comparison of each network with two randomized versions, weight reshuffle and edge plus weight reshuffle was performed. The same comparisons were made in [10] yielding two classes of networks: the ones that differ significantly from the random expectations, with shorter and less persistent cycles that appeared across the entire filtration, and networks that did not differ significantly from the random expectations, with long cycles and late filtration birth times. The results obtained with this dataset indicate that these networks do not belong to either of these two classes, as they appear significantly different from their randomized versions, with short persistence intervals yet with later birth filtration times, which suggests the existence of a larger spectrum of network types.

At least two 1-cycles that persisted for half or more than half of the maximum filtration value were found for all subjects. Further studies should be performed in order to determine if they enclose the same *holes*. Despite the fact that it was not possible to know the exact regions surrounding the 1-dimensional *holes* found, the most persistent representative cycles found still enclose weaker connectivity areas, which are important features of the data, and should inspire future investigations.

The choice of the similarity measure can have an important impact on the topology of the network that is obtained. It would be relevant to build the network considering different similarity measures and to investigate the consistency of these results. Also, the fact that only a percentage of edges with positive PPMCC was considered, starting from edges with highest positive correlation coefficient, might have overlooked the importance of small links to the overall topology of the network. This fact together with the choice of the pre-processing steps can affect the resulting network properties, as is discussed in [14].

Other software packages allow the computation of more than one missing face in the same filtration step through the use of multiple threads (e.g. https://www.jholes.eu/), reducing the computational time needed for this step, but are still very time consuming when one considers large networks. Once better algorithms and data-structures are developed, applying the WRCF method will become less computationally time expensive allowing one to include all network edges, and losses of in-

formation that arise from thresholding the networks will be avoided.

Acknowledgements
I thank professor Alexandre Francisco for providing the conditions for the pre-processing and the analysis of the data, and professor Roger Picken for the availability to discuss the relevant mathematical concepts used in this work.

I acknowledge my family and friends for their love and support.

References

[1] D. Horak, S. Maletić, and M. Raijkovic. Persistent homology of complex networks. *Jornal of Statistical Mechanics: Theory and Experiment*, 2009(3):P03034, Mar 2008.

[2] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(5):511–533, 2002.

[3] S. Huettel, A. Song, and G. McCarthy. *Functional magnetic resonance imaging, volume 1.* Sinauer Associates Sunderland, 2004.

[4] R. Levi. A topological toolbox for neuroscience, Jul 2017. `http://www.sci.kyoto-u.ac.jp/ja/_upimg/kce/dULKCg/files/Kyoto-July-17-Prelim.pdf`.

[5] M. A. Armstrong. *Basic Topology.* Springer, 1983.

[6] K. Gorgolewski, N. Mendes, D. Wilfling, E. Wladimirow, C. Gauthier, T. Bonnen, F. Ruby, R. Trampel, P. Bazin, R. Cozatl, J. Smallwood, and D. Margulies. A high resolution 7-tesla resting-state fmri test-retest dataset with cognitive and physiological measures. *Scientific Data*, 2:140054, Jan 2006. Dataset available at `https://openneuro.org/datasets/ds001168/versions/1.0.1`.

[7] Pearson product-moment correlation coefficient. `https://docs.scipy.org/doc/numpy/reference/generated/numpy.corrcoef.html`. Accessed: 2019-11-12.

[8] B Stolz, H. Harrington, and M. Porter. Persistent homology of time-dependent functional networks constructed from coupled time series. *Chaos*, 27(4):047410, Jan. 2017.

[9] A. Sizemore, C. Giusti, J. Vettel, R Betzel, and D. Bassett. Cliques and cavities in the human connectome. *Journal of Computational Neuroscience*, 44(1):115–145, Feb 2018.

[10] G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino. Topological strata of weighted complex networks. *PLOS ONE*, 8(6):1–8, 06 2013.

[11] Andrew Tausz, Mikael Vejdemo-Johansson, and Henry Adams. JavaPlex: A research software package for persistent (co)homology. In Han Hong and Chee Yap, editors, *Proceedings of ICMS 2014*, Lecture Notes in Computer Science 8592, pages 129–136, 2014. Software available at `http://appliedtopology.github.io/javaplex/`.

[12] C. Born and J. Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, Sep 1973.

[13] T. Opsahl, V. Colizza, P. Panzarasa, and J. Ramasco. Prominence and control: The weighted rich-club effect. *Physical Review Research*, 101(16):168–702, Oct 2008.

[14] A. J. Schwarz and J. McGonigle. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *NeuroImage*, 55(3):1132–1146, 2011.