TÉCNICO LISBOA

Exploratory Topological Data Analysis of Resting-State fMRI Data

Marta Raquel Azevedo Castro

Thesis to obtain the Master of Science Degree in

Mathematics and Applications

Supervisor: Prof. Roger Picken

Examination Committee

Chairperson:

Supervisor:

Member of the Committee:

October 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 networks, with the intent of investigating the presence of persistent weaker functionally connected areas.

Comparisons of the networks with two randomized versions showed that the original networks appear significantly different from their randomized versions, with later birth filtration times and shorter persistence intervals. Regions enclosing weaker connectivity areas were obtained allowing further investigations.

Contents

List of Tables

List of Figures

List of Acronyms

BET   Brain Extraction Tool

BOLD   Blood-Oxygen-Level-Dependent Signal

CSF   Cerebrospinal Fluid

EC   Euler Characteristic

FAST   FMRIB's Automated Segmentation Tool

FEAT   fMRI Expert Analysis Tool

FLIRT   FMRIB's Linear Image Registration Tool

FMRIB   Oxford Center for Functional Magnetic Resonance Imaging of the Brain

FSL   FMRIB Software Library

GM   Gray Matter

MCFLIRT Motion Correction FLIRT

MNI   Montreal Neurologic Institute

MRI   Magnetic Resonance Imaging

MR   Magnetic Resonance

PPMCC   Pearson Product-Moment Correlation Coefficient

ROI   Region of Interest

WM   White Matter

WRCF   Weight Rank Clique Filtration

fMRI   Functional Magnetic Resonance Imaging

rs-fMRI Resting-state fMRI

tb-fMRI Task-based fMRI

Chapter 1

Introduction

In recent years, network theory has been extensively used to approach questions about the global brain organization, and networks have become a common tool in the analysis of neuroimaging data coming from diverse experimental modalities [Bullmore and Sporns, 2009], demonstrating clinical applicabilities [Pandit et al., 2013]. Underlying the standard approaches from network theory are the measures of the strength of the pairwise connections between its microscopic constituent elements. To study functional brain activity one often looks for correlations among the time series of the constituents, which results in a matrix of similarities that allow the representation of the network as a weighted graph, that is then 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 [Horak et al., 2008].

When studying networks at a mesoscopic scale, one is often interested in discerning which features are relevant and which can be considered noise and be safely ignored. An important 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 [Edelsbrunner et al., 2002], 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 exposed. 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.

Chapter 2

Background

In this chapter useful notions of brain imaging are introduced as well as concepts from algebraic topology that are relevant for this work. A mathematical tool to study the persistence of topological features in the brain is also presented.

2.1 Fundamentals of Neuroimaging Techniques

2.1.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) first reported in [Carr, 1953] is a brain imaging technique that relies on the detection of the signal emitted by the hydrogen atoms in the water molecules naturally present in the tissues of the human body. This technique takes advantage of the fact that hydrogen atoms emit a signal in the form of a radio frequency when subjected to an oscillating magnetic field.

The time required for the excited atoms to return to an equilibrium state is the relaxation time and is what determines the contrast between different tissues in the MRI. The relaxation time related to the magnetic field in the z direction is denoted as T1, and the relaxation time denoted as T2 is related to the magnetic field in the x-y plane. Different structures will be given different weights in the MRI according to these two processes of relaxation. For example, in the more fluid environments of the brain such as the cerebrospinal fluid (CSF), the water molecules are less restricted, so these tissues have higher relaxation times when compared with more structured environments like gray matter (GM) and white matter (WM) [Buxton, 2009]. For these structures, in a T1 - weighted image, CSF will appear black, GM will appear dark gray, and WM will appear light gray. In a T2 - weighted image, the contrast between light and dark will measure a different tissue property, and in this case GM will appear light, WM will appear dark and CSF very bright [Huettel et al., 2004].

The study of the structure and the examination of the proprieties of the tissues in the brain can be performed through an MRI. However, structural studies of the brain are limited, in the sense that they cannot reveal short-term physiological changes associated with the active function of the brain. Functional neuroimaging studies are needed in order to create a spatial map of the areas responsible for a given mental process in the brain.

2.1.2 Blood-Oxygen-Level-Dependent Signal

Increased neuronal activity is supported by increased blood flow. Changes in blood flow that accompany increases in neuronal activity are believed to be initiated when active neurons release substances that reach the blood vessels. One of the responses to neuronal activity is the supply of oxygenated hemoglobin [Huettel et al., 2004]. In 1936, Linus Pauling and Charles Coryell conducted an investigation of the molecular structure of hemoglobin in which they discovered that the hemoglobin molecule had magnetic properties that differed depending upon whether or not it was bound to oxygen. 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. In 1990, Ogawa speculated that blood-oxygenation-level dependent (BOLD) contrast, could enable measurement of functional changes in brain activity. It was then verified that BOLD contrast was dependent upon the total amount of deoxygenated hemoglobin present in a brain region, which in turn depended upon the balance between oxygen consumption and oxygen supply. Thus, more active areas resulted in a more intense BOLD signal that could then be detected using a fMRI machine.

2.1.3 Functional Magnetic Resonance Imaging

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.

There are essentially two common types of fMRI experiments, the task-based experiment and the resting state experiment. The task-based fMRI (tb-fMRI) is obtained while the subject is doing some task, for example, squeezing one's hand repeatedly, and has the purpose of measuring the brain's response to that specific task [Huettel et al., 2004]. Another type of fMRI that is commonly performed is the resting state fMRI (rs-fMRI). This scanning process was first described in [Biswal et al., 1995] and is carried out while the subject is at rest, and it focuses on measuring low frequency fluctuations in BOLD signal (< 0.1Hz) without the presence of active tasks [Khanna et al., 2015].

2.1.4 Functional Connectivity

The analysis of functional connectivity using time-series relies on the definition that functional connectivity corresponds to the temporal correlations between different neuronal operations [Friston, 1994]. When different areas of the brain exhibit similar temporal patterns of activity, connectivity among those areas is implied and a functional connectivity network emerges.

The study of the brain's connectivity can be based on anatomical connectivity or functional connectivity. Anatomical connectivity indicates physical patters of neuronal connectivity and tends to be determined by the brain's early development and its heritable traits. Functional connectivity refers to

the influencing patterns of neurophysiological events and can change depending on the environment or experience of the individual's brain [Park et al., 2008]. It is important to stress that the demonstration of a functional connection does not imply the presence of a structural connection, nor does the presence of a demonstrable structural connection imply the presence of a functional connection [Biswal et al., 2010].

2.2   Algebraic Topology

Topology is the branch of mathematics that studies continuous shapes and properties which are unchanged by continuous deformations [Armstrong, 1983]. A sphere can be bent or distorted into an elliptic form, as long as the transformation doesn't involve any cutting or slicing of the sphere. In a more discrete sense, a sphere can also be seen as a deformed tetrahedron, a two-dimensional surface made of triangles glued together along bent edges. When such deformations exist, the shapes are said to be homeomorphic to each other.

The Euler characteristic (EC), published in [Euler, 1752], is defined as $V - E + F$, where V is the number of vertices, E the number of edges, and F the number of faces, and is a quantity that one can associate to any topological object of dimension up to two. The EC of a cube is $8 - 12 + 6 = 2$, which is equal to the EC of a tetrahedron $4 - 6 + 4 = 2$, and both spaces are homeomorphic to the sphere. More generally, if a polyhedron is homeomorphic to a sphere, then its EC is 2. H. Poincaré realized this fact in 1895, simply by noting that any two maps of the same topological space had the same EC. For example, in any mapping of the sphere, adding one vertex will increase the number of edges by one, leaving the EC unchanged. Similarly, adding one edge will lead to an increase in the number of faces by one, which leaves the EC unchanged. This idea allows us to consider convenient mappings of topological spaces knowing that the structure of the spaces will always be preserved.

A *triangulizable* space is a space that can be separated into triangular pieces which can be recognized, and which "fit together properly" [Armstrong, 1983]. As an example one could think of a strip that is chopped up into triangles and then the ends are joined with a half twist. The space obtained is homeomorphic to a Möbius strip, and as a result one has "triangulated" the Möbius strip.
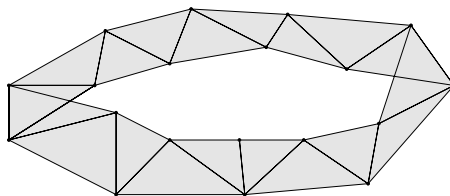
Figure 2.1: Triangulation of the Möbius strip.

The Möbius strip is a two-dimensional surface and so one can create a map of it using triangles as elementary pieces. However, for objects of higher dimension, elementary pieces of higher dimension need to be considered.

2.2.1 Simplicial Complexes

Given $k+1$ points $v_0, v_1, ..., v_k$ in general position, a k-*simplex* is the smallest convex set containing them [Levi, 2017]. In other words, a k-simplex is the convex hull of $k+1$ points in $\mathbb{R}^k$ that don't belong to the same $k-1$ dimensional hyperplane. If $v_0, v_1, ..., v_k$ is such a set of points, then the k-*simplex* they form is the set of points in $\mathbb{R}^k$ of the form

$$x = \lambda_0 v_0 + \lambda_1 v_1 + ... + \lambda_k v_k, \quad \text{where} \quad \sum_{i=0}^{k} \lambda_i = 1 \tag{2.1}$$

in which the vertices are the points $v_0, v_1, ..., v_k$ and the dimension is equal to the number of vertices that define it minus one. The figure below represents the simplices of the first 4 dimensions, the first fundamental building blocks that can be easily drawn.

| 0-simplex | 1-simplex | 2-simplex | 3-simplex |
| point | closed line segment | triangle | tetrahedron (solid) |
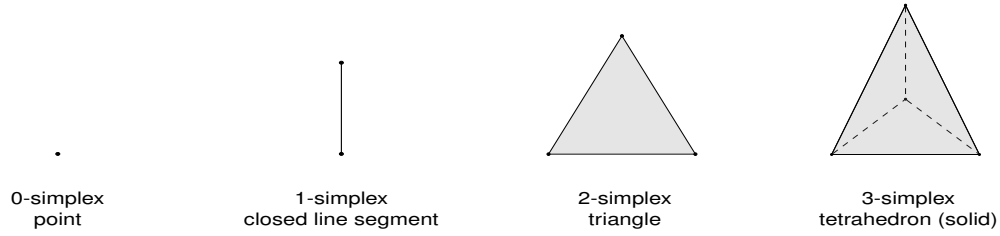
Figure 2.2: Simplices of the first 4 dimensions.

A q-simplex σ_q is a q-*face* of a k-simplex σ_k if every vertex of σ_q is also a vertex of σ_k. Every k-simplex has $(k-1)$-faces that are themselves $(k-1)$-simplices. In the case of a solid tetrahedron, a 0-face is one of its vertices, a 1-face one of the edges and a 2-face one of its triangles. 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. Formally, a simplicial complex K has the property that if the simplex $\sigma \in K$ and $\tau \subseteq \sigma$ is a face, then τ is also a simplex in K.

Two simplices are said to fit together "in a proper way" whenever they intersect in a common face. In the following figure, there is a distinction between simplices that fit well and simplices that don't.

Allowed - fit "in a proper way"   Not Allowed
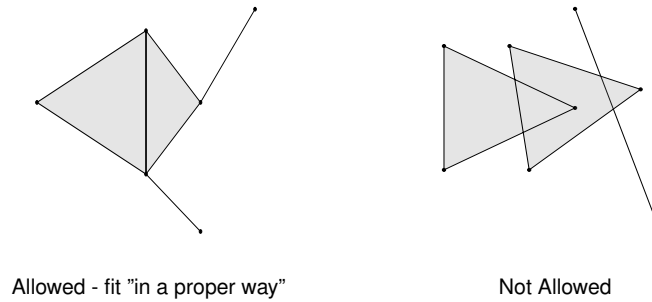
Figure 2.3: Intersection of simplices

Any two simplices of any dimension can be attached along a face of the same dimension, and the

dimension of a simplicial complex is equal to the maximum of the dimensions of the simplices that it contains. The simplicial complex represented on the left in the figure above has dimension two, since the simplex of highest dimension that it contains is a triangle.

2.2.2 Chains, Cycles and Boundaries

When a fixed triangulation K of a topological space X is considered, the edges can be oriented and the orientation is denoted as usual by arrows on edges. An edge that is oriented from the vertex u to v is denoted by the symbol (u, v). Similarly, the symbol (u, v, w) denotes the ordering u, v, w of the vertices of a triangle of K, in this case (u, v, w) denotes a cyclic ordering. A change in the orientation is denoted by a minus sign, so $(u, v) = -(v, u)$ and $(v, u, w) = -(u, v, w)$.

The *boundary* ∂ of an oriented edge (u, v) is defined as $\partial(u, v) = v - u$, which is the difference between the final and the initial vertices. The boundary of an oriented triangle (u, v, w) is sum of its edges $\partial(u, v, w) = (v, w) + (w, u) + (u, v)$, each taken with the orientation induced by the overall orientation of the triangle [Armstrong, 1983].
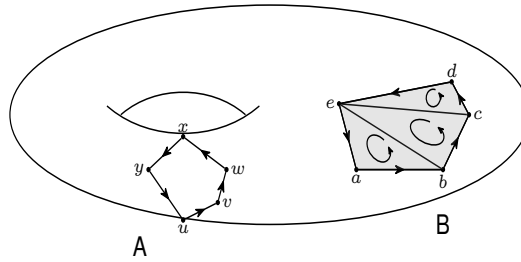
Figure 2.4: Curves on a triangulazed torus.

The curve A represented in figure 2.4 can be written in terms of the sum of its oriented edges

$$A = (u, v) + (v, w) + (w, x) + (x, y) + (y, u)$$

and its boundary can be linearly defined as

$$\partial A = \partial(u, v) + \partial(v, w) + \partial(w, x) + \partial(x, y) + \partial(y, u)$$

$$= v - u + w - v + x - w + y - x + u - y$$

$$= 0$$

i.e all terms cancel out, which is a formal manner of defining a *closed* curve. The curve A is closed and therefore has no boundary. Considering now the oriented curve B, one can see that is also closed and encloses three of the triangles which are part of the triangulated Torus. Taking the orientation of the triangles as represented in the figure, their union can be written as

$$C = (e, a, b) + (e, b, c) + (e, c, d)$$

which allows one to compute its boundary as

$$\partial C = \partial(e, a, b) + \partial(e, b, c) + \partial(e, c, d)$$

$$= (a, b) + (b, e) + (e, a) + (b, c) + (c, e) + (e, b) + (c, d) + (d, e) + (e, c)$$

$$= (a, b) + (b, e) + (e, a) + (b, c) + (c, e) - (b, e) + (c, d) + (d, e) - (c, e)$$

$$= (a, b) + (b, c) + (c, d) + (d, e) + (e, a)$$

$$= B$$

indicating that the curve B bounds a piece of the torus, given that it bounds a collection of simplices of dimension 2.

A 1-*cycle* or a one-dimensional cycle of K, is any arbitrary linear combination of oriented edges of K with integer coefficients $\lambda_1(u_1, v_1) + ... + \lambda_k(u_k, v_k)$, in which the boundary vanishes, recalling that $\lambda(u, v) = -\lambda(v, u)$. The oriented curve A has a vanishing boundary and is an example of a 1-cycle. The 1-cycles together with the operation of addition form an abelian group

$$\sum \lambda_i(u_i, v_i) + \sum \mu_i(u_i, v_i) = \sum (\lambda_i + \mu_i)(u_i, v_i) \tag{2.2}$$

this group of 1-cycles of K is denoted by $Z_1(K)$ [Armstrong, 1983].

A one-dimensional cycle is a *bounding cycle* if there exists a linear combination of oriented triangles whose boundary is the given cycle. The oriented curve B is an example of a bounding cycle, since there are three oriented triangles that when combined have as their boundary B. The boundary of B is zero, and below it will be clear that the group of bounding 1-cycles $B_1(K)$ forms a subgroup of $Z_1(K)$.

The curves A and B are pieces of the triangulated Torus and they are built with low dimensional simplices, but simplicial complexes can contain simplices of an arbitrary dimension. An *oriented simplex* is a simplex together with a specific choice of orientation. A simplex of any dimension can be oriented in two different ways, with the exception of the vertices, which can only be oriented in one way. The free abelian group generated by the oriented q-simplices of 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

$$\partial(v_0, ..., v_q) = \sum_{i=0}^{q} (-1)^i (v_0, ..., \hat{v}_i, ..., v_q) \tag{2.3}$$

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

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

8

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 [Armstrong, 1983]. The group of the q-cycles of K is denoted by $Z_q(K)$ and, since it is the group formed by the oriented q-chains with vanishing boundary, it is equal to the kernel of $\partial_q : C_q(K) \to C_{q-1}(K)$. The group $B_q(K)$ is the group of bounding q-cycles of K and is the image of $\partial_{q+1} : C_{q+1}(K) \to C_q(K)$, and this group is a subgroup of the group of q-cycles, because of the above property.

2.2.3   Homology

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 [Levi, 2017]. An homology group can be associated to a triangulated topological space making use of the concepts of cycles and bounding cycles. The nth *Homology group* of a simplicial complex K as stated in [Horak et al., 2008] 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

$$H_n = \frac{Ker\ \partial_n}{Im\ \partial_{n+1}} = \frac{Z_n}{B_n} \tag{2.4}$$

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 q-cycle. The nth *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 nth Homology group H_n, and can be written as

$$\beta_n = rank(H_n) = rank(Z_n) - rank(B_n) \tag{2.5}$$

The intuition that justifies these definitions and their relation to the number of *holes* of a topological space can be understood through the following examples:

0th **Homology group**

In the simplicial complex of figure 2.5, the group C_0 is generated by the vertices v_0, v_1, v_2, v_3, v_4 and the group C_1 is generated by the oriented 1-simplices a, b, c, d, e.
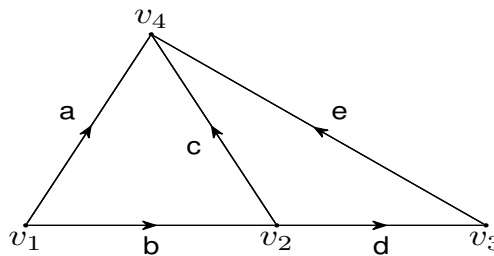

Figure 2.5: Simplicial complex

Since the group Z_0 is formed by the cycles in C_0, and by definition the boundary of a vertex is always zero, then it is generated by $< v_1, v_2, v_3, v_4 >$. The group of bounding 0-cycles B_0 is a subgroup of C_0 and is generated by the boundary of each 1-simplex

$$\partial(a) = v_4 - v_1, \partial(b) = v_2 - v_1, \partial(c) = v_4 - v_2, \partial(d) = v_3 - v_2 \text{ and } \partial(e) = v_4 - v_3.$$

Since B_0 is generated by elements that are themselves linear combinations of the vertices, $B_0 \subseteq Z_0$. And the 0th homology group is

$$H_0(K) = \frac{Z_0(K)}{B_0(K)} \simeq \mathbb{Z}$$

that is, H_0 is an infinite cyclic group generated by one of its vertices, for example $[v_1]$, since all elements of B_0 are being sent to 0, which corresponds to setting all vertices equal to each other. Thus, the 0th Betti number is equal to $rank(H_0)$ which is 1, corresponding to the cardinality of the set with just one of its vertices.

This homology group can be seen as a measure of the number of connected components in a simplicial complex K. If instead of the simplicial complex in figure 2.5, the space simply consisted of two directed edges a and b as in the following figure:
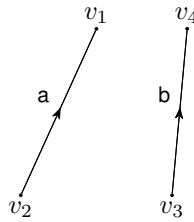
Figure 2.6: Two connected components

Then, B_0 would be generated by $\partial(a) = v_1 - v_2$ and $\partial(b) = v_4 - v_3$. Since Z_0 is generated by the vertices v_1, v_2, v_3 and v_4, the 0th homology group would be

$$H_0(K) = \frac{Z_0(K)}{B_0(K)} \simeq \mathbb{Z} \oplus \mathbb{Z}$$

i.e. an infinite cyclic group isomorphic to $\mathbb{Z}$ generated by each component, noting that an example of a typical element in this group has the form $\lambda_0 v_1 + \lambda_1 v_3$. Thus, the 0th Betti number now is 2, corresponding to the two generators, which is equal to the number of connected components of this simplicial complex.

1st **Homology group**

In the simplicial complex of figure 2.5, a chain α in C_1 is an integer linear combination of its edges, so it can be represented as $\lambda_0 a + \lambda_1 b + \lambda_2 c + \lambda_3 d + \lambda_4 e$. In order to compute the first homology group, it is necessary to determine the 1-cycles of this complex. The boundary of an element α in C_1 is given by:

$$\partial(\lambda_0 a + \lambda_1 b + \lambda_2 c + \lambda_3 d + \lambda_4 e) = \lambda_0 \partial(a) + \lambda_1 \partial(b) + \lambda_2 \partial(c) + \lambda_3 \partial(d) + \lambda_4 \partial(e)$$

$$= \lambda_0 (v_4 - v_1) + \lambda_1 (v_2 - v_1) + \lambda_2 (v_4 - v_2) + \lambda_3 (v_3 - v_2) + \lambda_4 (v_4 - v_3)$$

$$= (-\lambda_0 - \lambda_1)v_1 + (\lambda_1 - \lambda_2 - \lambda_3)v_2 + (\lambda_3 - \lambda_4)v_3 + (\lambda_0 + \lambda_2 + \lambda_4)v_4$$

which can be conveniently represented as a matrix and transformed into row echelon form as follows

$$
\begin{bmatrix}
-1 & -1 & 0 & 0 & 0 \\
0 & 1 & -1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
1 & 0 & 1 & 0 & 1
\end{bmatrix}
\xrightarrow{L_1 \rightleftarrows L_4}
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 \\
0 & 1 & -1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
-1 & -1 & 0 & 0 & 0
\end{bmatrix}
\xrightarrow{L_1 + L_4}
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 \\
0 & 1 & -1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
0 & -1 & 1 & 0 & 1
\end{bmatrix}
$$

$$
\xrightarrow{L_2 + L_4}
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 \\
0 & 1 & -1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & -1 & 1
\end{bmatrix}
\xrightarrow{L_3 + L_4}
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 \\
0 & 1 & -1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\xrightarrow{L_3 + L_2}
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 \\
0 & 1 & -1 & 0 & -1 \\
0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

The equation $\partial(\alpha) = 0$ has now the following solution in the vector form:

$$
\begin{bmatrix}
\lambda_0 \\
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\lambda_4
\end{bmatrix}
= r
\begin{bmatrix}
-1 \\
1 \\
1 \\
0 \\
0
\end{bmatrix}
+ s
\begin{bmatrix}
-1 \\
1 \\
0 \\
1 \\
1
\end{bmatrix}
$$

Both vectors provide a basis for the null space of the matrix that represents the mapping ∂. Since $Ker\ \partial(\alpha) =< -a + b + c, -a + b + d + e >$, the cycles $-a + b + c$ and $-a + b + d + e$ generate the group of 1-cycles. Any other 1-cycle can be written as a linear combination of these two, for example, the remaining 1-cycle $d + e - c$ is the difference between the cycles $-a + b + d + e$ and $-a + b + c$. The simplicial complex doesn't have any bounding 1-cycles which enclose any simplex of dimension 2, thus $H_1(K)$ is simply isomorphic to two infinite cyclic groups $\mathbb{Z}$, one for each of the generators found.

In this example, the 1st Betti number is 2, since it is the cardinality of the set of generators, being the two 1-cycles found. In fact, observing figure 2.5 one could easily conclude that the complex has two 1-dimensional *holes*, that correspond to the two triangles that are glued together and have no interior. The fact that the simplices of highest dimension are edges, allows one to conclude that there aren't any *holes* of dimension higher than one.

Considering now a new simplicial complex in which a 2-simplex is added, it is possible to see how homology changes.
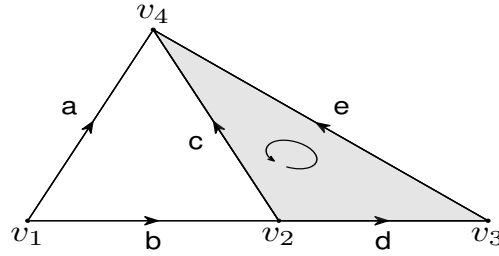
Figure 2.7: Modified simplicial complex

In this simplicial complex, the cycle $d + e - c$ is the boundary of the new 2-simplex that was added, and no longer captures a 1-dimensional *hole* as it captured before. In order to measure 1-dimensional *holes*, the cycle $d + e - c$ has to be zero. Since the difference between the two generators $-a + b + d + e$ and $-a + b + c$ of the group of cycles Z_1 was the cycle $d + e - c$ and $d + e - c$ needs to be zero, the two generators must be equal. Therefore, in order to determine the first homology group, it is necessary to quotient the group of 1-dimensional cycles generated by $< -a + b + d + e, -a + b + c >$ by the subgroup of the bounding 1-dimensional cycles, in this case generated by $< d + e - c >$. Thus, instead of having $H_1(K) = \mathbb{Z} \oplus \mathbb{Z}$ as before, now

$$H_1(K) = \frac{Z_1(K)}{B_1(K)} = \frac{\mathbb{Z} \oplus \mathbb{Z}}{\mathbb{Z}} = \mathbb{Z}$$

In this case, the group C_2 is no longer $\{0\}$ as it was in the previous example, now it is generated by the triangle $T = (v_2 v_3 v_4)$, whose boundary is given by $\partial_2(T) = (v_3 v_4) - (v_2 v_4) + (v_2 v_3)$ which is precisely $e - c + d$, the generator of the group of bounding 1-cycles B_1. This result is in accordance with the definition of nth Homology group given, which can be written in this case for $n = 1$ as

$$H_1(K) = \frac{Ker\ \partial_1}{Im\ \partial_2}$$

The 1st Betti number is equal to 1, which corresponds to the cardinality of the set consisting of only one of its 1-cycles, in this case the cycle $-a + b + c$, which is not the boundary of any 2-simplex. It is also important to note that the 1-cycles $-a + b + d + e$ and $-a + b + c$ capture the same 1-dimensional *hole*, therefore they have the same homology class. And since the difference between the two 1-cycles is the bounding 1-cycle $d + e - c$, they are equivalent cycles.

This same reasoning can be applied to determine the number of *holes* of higher dimension of any triangulation K of a topological space X. In order to understand the importance of a topological feature present in K it is necessary to keep track of the creation and destruction of these homology classes, when simplices are added to the complex or removed from it [Horak et al., 2008].

2.2.4 Filtration

Homology can be used to study the importance of a certain topological property in a simplicial complex. For that purpose, one considers the evolution of the complex starting from the empty set, and it is assumed that simplices are added to the complex gradually. A sequence of subcomplexes constructed

in such a gradual process, that is, a sequence of simplicial complexes such that $\emptyset = K^0 \subseteq K^1 \subseteq K^2 \subseteq ... \subseteq K^m = K$, is called a *filtration*. There are usually two types of filtrations considered: The first one is a total filtration, in which only one simplex is added at each step, establishing a total ordering on the simplices. The second one is a partial filtration, in which case more than one simplex is allowed in the same filtration step [Edelsbrunner et al., 2002]. An example of a total filtration is illustrated in figure 2.8.

2.2.5 Incremental Algorithm

The following algorithm proposed in [Delfinado and Edelsbrunner, 1995] computes the three first Betti numbers ($d = 0, 1, 2$) of a simplicial complex. This algorithm assumes a total filtration of the complex that represents a triangulation contained in the sphere S^3, $K = \{\sigma_1, \sigma_2, ..., \sigma_m\}$, and assumes that $K^i = K^{i-1} \cup \{\sigma_i\}$, for $i = 1, ..., m$.

```
for l := 0 to d do βl := 0 endfor;
for i := 0 to m do
    k := dim σi;
    if σi belongs to a k − cycle of Ki
        then βk := βk + 1
        else βk−1 := βk−1 − 1
    endif;
endfor;
return (β0, β1, β2);
```

Initially the subcomplex is empty, so the three Betti numbers start as being equal to zero. Then, a new simplex σ_i is added at each step in the filtration, which leads to a sequence of m subcomplexes for which the three Betti numbers are computed. Whenever σ_i is a vertex, dim $\sigma_i = 0$ and by definition it belongs to a 0-cycle. An added vertex is always isolated from the rest of the simplices, therefore the number of connected components increases by one and so does β_0. If σ_i is an edge, β_0 decreases by 1 whenever σ_i connects two different components of K^{i-1}, and if σ_i connects two vertices in the same component of K^{i-1}, then β_1 increases by 1. Similarly for the cases in which σ_i is a triangle and a tetrahedron [Basener, 2006]. An example of the execution of this algorithm is illustrated in figure 2.8.

It is important to note that an added simplex σ_i belongs to a k-cycle of K_i if σ_i belongs to a linear combination of k-simplices whose boundary vanishes. By the filter property, no simplex in K_{i-1} has σ_i as a face, thus in this case a k-cycle is necessarily a non-bounding cycle.

A $(k + 1)$-simplex σ_i that belongs to a $(k + 1)$-cycle is denoted as *positive*, otherwise it is denoted as *negative*. Let $pos_k = pos_k^l$ and $neg_k = neg_k^l$ be the number of positive and negative k-simplices in K^l and let $\beta_k = \beta_k^l$ be the kth Betti number of K^l. Then, equation 2.5 can be rewritten as:

$$\beta_k = pos_k - neg_{k+1} \tag{2.6}$$

That is, the Betti number β_k is the number of k-simplices that create k-cycles minus the number of

13

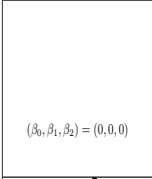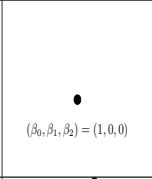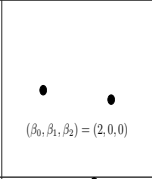$(k+1)$-simplices that destroy k-cycles through the creation of k-boundaries [Edelsbrunner et al., 2002].
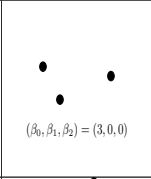

Figure 2.8: Computation of the three first Betti numbers.

2.2.6 Persistence

Persistence is a measure of the lifetime of an attribute of a simplicial complex throughout the filtration steps. Attributes can be ranked by their lifetime in a filtration using the cycle and boundary groups.

Let K^l be the simplicial complex in the lth filtration step, and Z_k^l and B_k^l denote the kth cycle group and the kth boundary group respectively. In order to get the persistent cycles in K^l, it is necessary to quotient its kth cycle group by the kth boundary group of K^{l+p}, p steps later in the filtration. The p-*persistent k^{th} homology group* of K^l can be formally written as

$$H_k^{l,p} = \frac{Z_k^l}{B_k^{l+p} \cap Z_k^l} \tag{2.7}$$

The rank of $H_k^{l,p}$ is the p-*persistent k^{th} Betti number* $\beta_k^{l,p}$ of K^l. As the filtration step p increases by one, the persistence of all non-bounding cycles decreases by one. Negative simplices cancel positive simplices earlier in the filtration as p increases, and short-lived attributes, which correspond to noise of the complex, can be destroyed by increasing p sufficiently [Edelsbrunner et al., 2002].

A different and more intuitive way of defining persistence is by considering birth and death filtration steps. 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, instead of equation 2.7, persistence can be defined as

$$p = l - s \tag{2.8}$$

A topological feature is considered to have *infinite persistence* whenever it persists until the last filtration step [Stolz et al., 2018].

2.2.7 Visualization: Barcodes

Persistent homology allows one to detect the birth and death of each topological feature as the complex gradually evolves. Therefore, it is convenient to represent it in terms of the rank of the homology groups in each filtration step, as a function of some parameter [Horak et al., 2008]. A possible choice of parameter is the time t_i that accounts for the birth of cycles and the time t_j when they disappear. The other possible choice of parameter are the intervals whose endpoints represent the filtered complexes $K_i(t_i)$ and $K_j(t_j)$ at the times when the birth t_i and death t_j of cycles took place. The parameter intervals correspond to the lifetimes of various points in the filtration and they can be represented on the horizontal axis, whereas the homology generators H_k can be represented on the vertical axis, that is, the Betti numbers while a given topological attribute is present. Such a representation is called a *Barcode*, and is a useful tool to analyze the persistence of a topological feature.

Figure 2.9 shows the barcodes for the filtration in figure 2.8. The steps of the filtration are represented on the x-axis, and since it is a total filtration, each step also corresponds to the indices of the simplicial complexes in the filtration. The filtration starts with the empty set, so no topological features are present in the interval $[0,1]$. At step 1 the first vertex is added, and K_1 is a simplicial complex with just one vertex, so the number of connected components increases by one, which is represented by the first bar in the barcode of dimension 0. At step 3 there are three isolated vertices, so three bars are represented in the interval $[3,4]$, and since at step 4 an edge is added, two of the connected components are merged and one of the bars ends at step 4. At step 7 the first 1-cycle appears and only disappears at step 10 when the 2-simplex is added, so in the barcode of dimension 1 the first bar starts at step 7 and ends at step 10. At step 14, the simplicial complex is made up of 4 triangles that are glued together enclosing a 2 dimensional *hole*, therefore the barcode of dimension 2 has a bar that starts at step 14 and ends at step 15 when the 2 dimensional *hole* is filled giving rise to a tetrahedron.
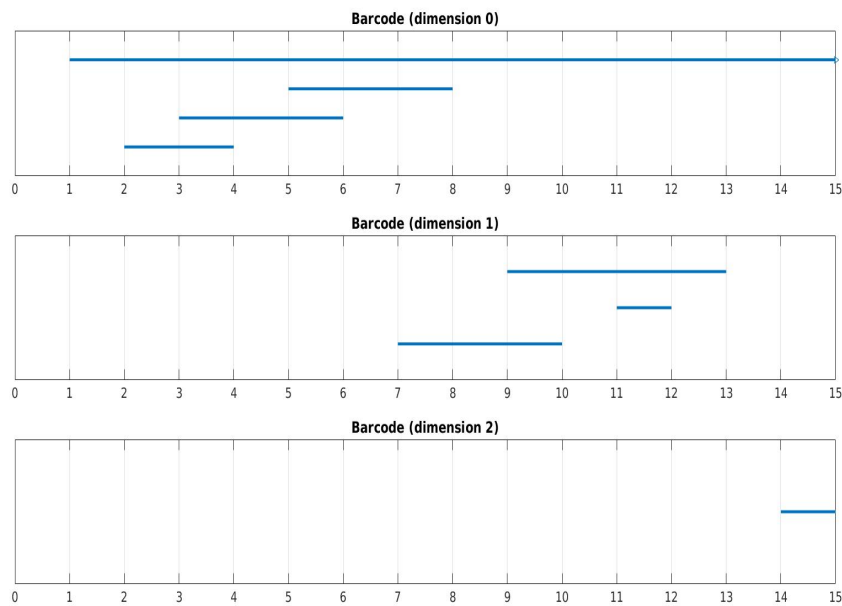
Figure 2.9: Barcodes representing the filtration in figure 2.8.

15

The presence of a homology class that persists over a long range of parameter values is represented by bars that are present during long intervals, whereas short intervals indicate that cycles which are "born" at a given parameter value, "die" at a nearby parameter value. The fact that the filtration in 2.8 is the filtration of a tetrahedron, which consists of only one connected component, is reflected in the barcode of dimension 0 by the long bar that persists throughout all parameter intervals.

Chapter 3

Materials and Methods

In this chapter, the dataset is introduced and the main preprocessing steps performed are briefly explained. It is explained how the functional connectivity network was constructed and what persistent homology computation methods were used to evaluate the topology of the network.

3.1 Image Pre-Processing

All the experimental data used in this work are freely available at [Gorgolewski et al., 2006] and were collected at Max Planck Institute for Human Brain and Cognitive Sciences, Leipzig, Germany using a 7T whole-body MAGNETOM MR scanner (Siemens Healthcare, Erlangen, Germany). For the purpose of this work, only 10 of the 22 images available were used. 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. The anatomical and the resting-state functional data were collected according to the parameters described in [Gorgolewski et al., 2006]. In this work, only the images obtained in the first session were considered.

In order to analyze the fMRI experimental data and extract the functional network, a series of processing steps were performed to remove possible errors due to the data acquisition process. 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. The correction steps and the tools used are described in the following subsections.

3.1.1 Slice Time Correction

The fMRI machine acquires the different brain slices at different points in time guaranteeing that the information of each volume throughout time is recorded. With the purpose of decreasing the complexity of the image analysis, each volume is considered to be taken at the same time point. In this step the FSL function `slicetimer` was used, which outputted the 4D dataset containing the information of the 3D brain structure at each point in time.

3.1.2 Motion Correction

The functional information from the sequence of images acquired during the scanning process is extracted through the application of statistical time-series analysis, which assumes that the location of a given voxel within the brain is the same over time. However, during the scanning process there is usually some degree of subject's head motion within the scanning machine. In order to perform a statistical analysis, this motion needs to be estimated and corrected for. In this step, the adjustment was done using the FSL tool MCFLIRT [Jenkinson et al., 2002]. This tool uses the first time point image acquisitions and performs rotations and translations to them. After all the transformations have been performed, they are compared using a specified cost function. The transformation with the lowest value for the cost function is chosen. In this case, the rotation and translation motion parameters were regressed out of each individual voxel using a linear regression, correcting the linear changes in the intensity of each voxel induced by small head motions.

3.1.3 Brain Extraction

A fMRI dataset often contains non-brain data that need to be removed. To serve this purpose, the BET (Brain Extraction Tool) [Smith, 2002] from the FSL library was used. This algorithm starts by finding the images' maximum and minimum intensities by looking at the intensity histogram. Then, a roughly chosen threshold is calculated that attempts to distinguish between brain matter and background. This threshold is then used to estimate the position of the center of gravity of the brain/head image. For all voxels with intensity greater than t, their intensity is used in a standard weighted sum of positions. Next, the mean "radius" of the brain/head in the image is estimated. All voxels with intensity greater than t are counted, and a radius is found. Lastly, the median intensity of all points within a sphere of the estimated radius and centered on the estimated center of gravity is found.

3.1.4 Registration and Normalization

In the registration step, an alignment and overlay of the fMRI data from each subject with their own separately acquired T1-weighted anatomic images was performed using the FSL tool FLIRT (FMRIB's Automated Segmentation Tool)[Jenkinson et al., 2002, Jenkinson and Smith, 2001]. In the first normalization step, the fMRI data was aligned and warped into a generic anatomic template, in this case the standard space chosen was the Montreal Neurologic Institute (MNI) template `MNI152_T1_2mm_brain` available at the FSL atlases directory. This normalization step serves the purpose of deforming the image into a reference space so that the interpretation of the data across subjects is possible. This normalization step was performed using the FSL tool FEAT (fMRI Expert Analysis Tool) [Woolrich et al., 2001]. Lastly, the T1-weighted anatomical image was segmented using FAST (FMRIB's Automated Segmentation Tool) [Zhang et al., 2001] into two classes, CSP and WM. This algorithm works by attempting to differentiate the two tissue types through the estimation of the intensities of both tissues in a T1-weighted image. The purpose of this last step was to use those tissues masks based on the T1-weighted image and use their estimated intensities to normalize the intensity signal from the functional data using linear regression.

3.1.5 High-frequency Noise Removal

In this step, the frequencies of the signal that come from the voxels were filtered out, and all the components of the signal considered to be noise were removed from the dataset. To achieve this, a band-pass filter was considered, so that only the frequencies of the signal inside a specified range are taken into account. In this case, the passband filter used was available in the `resting_pipeline.py` file. The highpass filter was fixed at 0.001 Hz and the lowpass filter was fixed at 0.08 Hz.

3.2 Functional Connectivity Network

In order 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. Such a mapping usually starts by identifying a set of functional "nodes", and then attempts to estimate the set of connections or "edges" between these nodes [Smith et al., 2011]. 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. The edges can contain weights that quantify the similarity of the associated time series according to a similarity measure. 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 [Pea].

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

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} \cdot C_{jj}}} \tag{3.1}$$

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 times series of the regions i and j, where the N here represents the total number of ROIs [Stolz et al., 2017].

In this work, in order to build such a network, a parcellation of the image was performed based on a label file with 116 regions of interest (see A.1). The average time-series of the set of voxels that belonged to each region was obtained using the FSL function `fslmeants`. The correlations among the regions were obtained using the `corrcoef` from the NumPy library.

In the following figure it is possible to see an example of a connectivity network, in this case a structural network, in which the nodes represent brain regions and the edges represent the structural connections among regions. The strength of the connections is represented by the edges thickness.
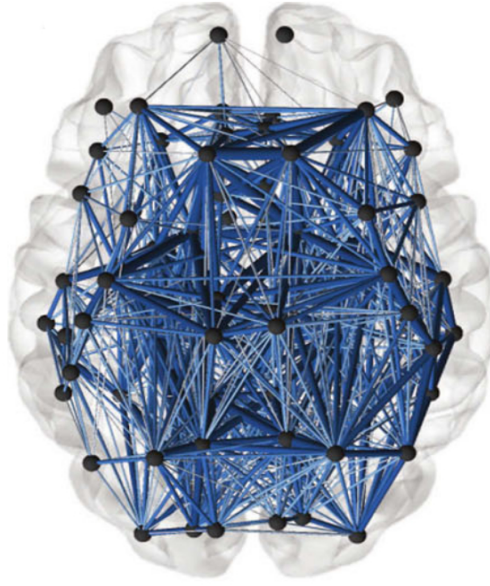
Figure 3.1: Example of a ROI based connectivity Network. Source: [Sizemore et al., 2018]

3.3   Persistent Homology Computation

It is clear that simplicial complexes are an approximation to topological spaces and are capable of capturing their properties. A graph is a topological space that is frequently considered when representing real world networks, and is a common way in which simplicial complexes arise [Levi, 2017]. In graph theory, a *graph* $\mathcal{G} = (V, E)$ is formally defined as a set of nodes or vertices V and a set of edges or connections among nodes E. The nodes of the graph are mathematical abstractions and can refer to any objects, and the edges correspond to the relations among these objects.

3.3.1   Building Simplicial Complexes from Graphs

Let $\mathcal{G}$ be a graph, then a n-*clique* in $\mathcal{G}$ 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 a graph $\mathcal{G} = (V, E)$ is defined as the abstract simplicial complex whose n-simplices are the $(n + 1)$-cliques in $\mathcal{G}$, and is denoted as $C(\mathcal{G})$ [Horak et al., 2008]. From figure 3.1 one can think of the functional brain network as a graph, in which the nodes represent each ROI and an edge a connection between any two ROIs. Therefore, the clique complex can be computed by finding all cliques in the graph. A 3-clique is a subset of three vertices of $\mathcal{G}$ and gives rise to a triangle, and a 4-clique results in a tetrahedron. The following figure illustrates this idea, where in the example of the 4-clique it is also possible to see that taking any subset of its vertices results in a clique of lower dimension, independently of the vertices one chooses.
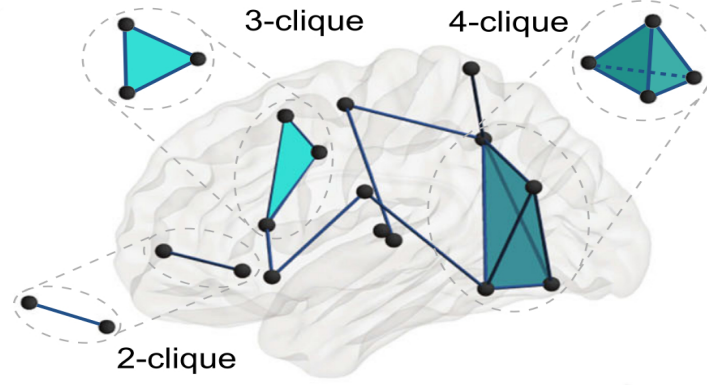
Figure 3.2: Cliques in the connectivity network. Source: [Sizemore et al., 2018]

3.3.2 Building a Filtration

Given a graph $\mathcal{G} = (V, E)$ and the associated clique complex $C(\mathcal{G})$, a filtration for this complex is the sequence $K_0 \subset K_1 \subset ... \subset K_n = K$, such that the ith complex in the filtration is given by

$$K_i = \sum_{j=1}^{i} S_j \qquad (3.2)$$

where S_j is the $j - th$ skeleton of the clique complex, that is, the set of simplices of dimension less than or equal to j [Horak et al., 2008]. Such a filtration is called *Clique filtration* and is illustrated in figure 3.3 for a given graph $\mathcal{G}$. At each step, one builds the simplicial complex according to the dimension of the skeleton of the clique complex. Initially one considers only the edges and vertices, and one gradually fills the cliques by increasing the order of the dimension. First, the 2-simplices are filled for all the 3-cliques present, then in the next step one fills the 3-simplices for all the 4-cliques present, and so on.
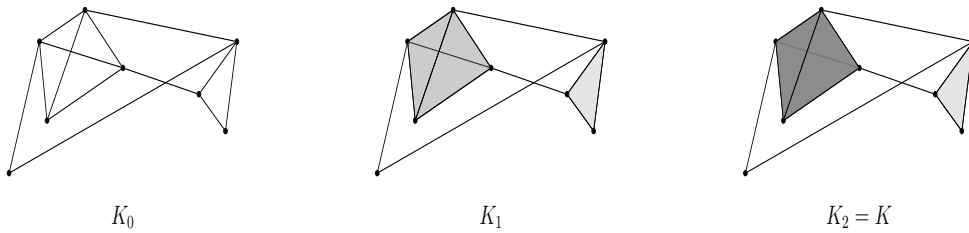
Figure 3.3: Filtration of a clique complex.

In the case in which it is possible to establish a distance between any two vertices in the simplicial complex, the topological space is considered to be a metric space. In such cases, a common method to build a filtration from a point set or a vertex set is by placing a ball B with radius ϵ around each element in the set. For some radius $\epsilon > 0$, one can define a covering of a metric space X as the family $\mathcal{B}_\epsilon = \{B_\epsilon(x)\}_{x \in X}$, and for any subset $V \subseteq X$, one can consider the abstract simplicial complex with

vertex set V, where a family $\{x_0, ..., x_k\}$ spans a k-simplex if and only if $B_\epsilon(x_0) \cap ... \cap B_\epsilon(x_k) \neq \emptyset$. This complex is called the *Čech complex* attached to V and ϵ and is denoted as $C(V, \epsilon)$ [Carlsson, 2009].

The construction of a Čech complex requires the storage of simplices of various dimensions, since it only depends on the radii ϵ, which is computationally expensive. In order to solve this problem the following variant of the Čech construction is considered: Let X denote a metric space with metric d, then the simplicial complex whose vertex set is X such that $\{x_0, x_1, ..., x_k\}$ spans a k-simplex if and only if $d(x_i, x_j) \leq \epsilon$ for all $0 \leq i, j \leq k$, is denoted as $VR(X, \epsilon)$ and is called the *Vietoris-Rips complex* for X attached to the parameter ϵ. And now this simplicial complex can be recovered solely from the edge information, which represents the distance between the pairs of points.

| Covering with radius ϵ_1 | Covering with radius $\epsilon_2 > \epsilon_1$. | Čech Complex | Vietoris-Rips Complex |
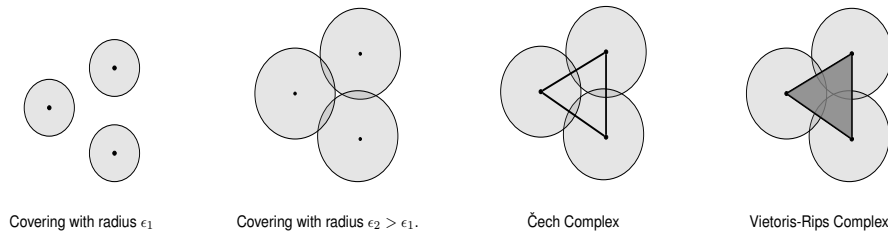
Figure 3.4: Filtrations examples

In figure 3.4 the leftmost two images show two coverings, in the first one no intersections among the balls exist, whereas in the second one the increased radius allows the construction of the Čech and the Vietoris-Rips complexes. Note that the parameter for building the Vietoris-Rips complex is the distance, while in the computation of the Čech complex the parameter ϵ corresponds to the radii of the balls. In the example presented in the image, to build the Čech complex the vertices whose balls of radius ϵ intersect are connected by an edge, since the balls intersect two by two, whereas in the case of the Vietoris-Rips complex the same intersections result in a triangle.

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 [Petri et al., 2013] 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 $\mathcal{G}(\omega_{ij}, \epsilon_t)$, that is, the subgraph of Ω with links of weight larger than ϵ_t. The clique complex $K(\mathcal{G}, \epsilon_t)$ is built for each graph $\mathcal{G}(\omega_{ij}, \epsilon_t)$. 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.
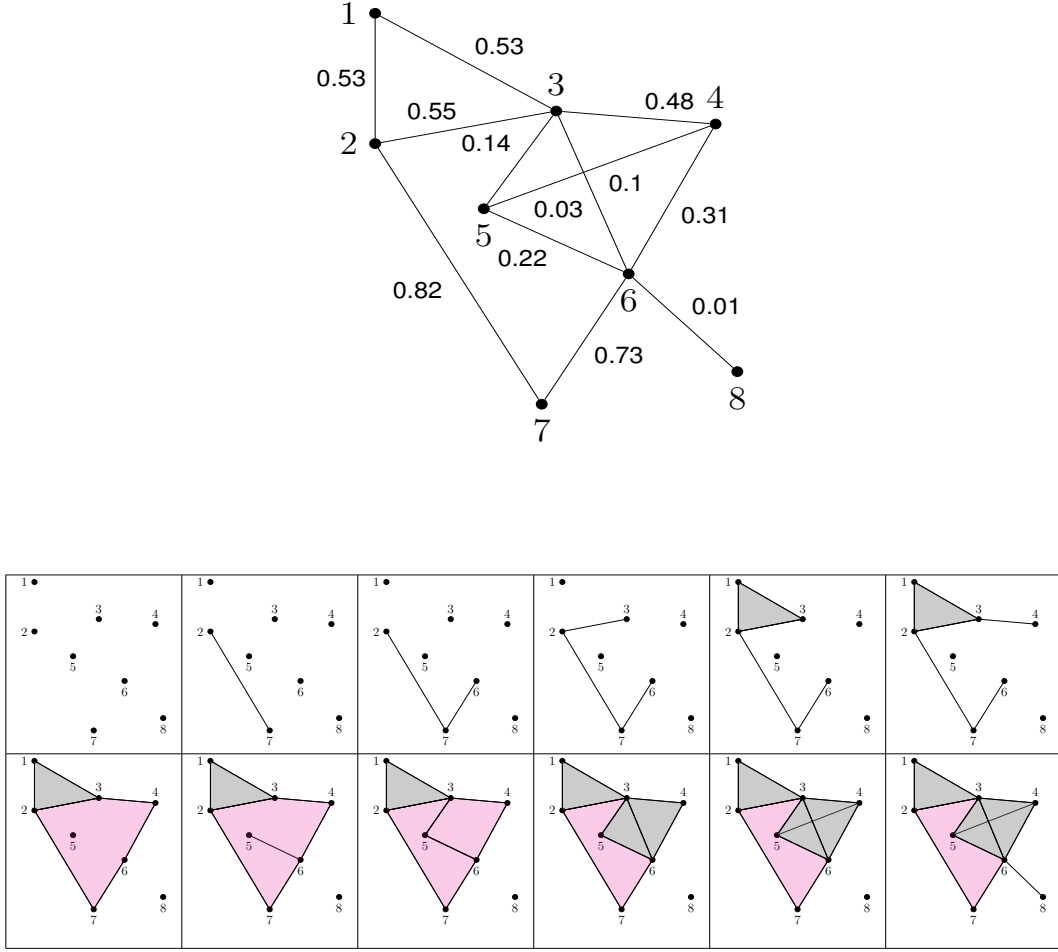
Figure 3.5: Weight rank clique filtration

The edges are added by decreasing order of the weight. In the first filtration step, the graph $\mathcal{G}$ 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.

3.3.3 Topological Measures

The WRCF method preserves the complete weight information as well as the topological structure of the network, allowing one to evaluate the presence of *weighted network holes*. A weighted network *hole* of dimension 1 with weight w is a 1-cycle, as explained in the subsection 2.2.2, 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 $\leq w$, and all the other possible edges crossing the *hole* have weights strictly weaker than w [Petri et al., 2013]. Similarly for the case of 2-dimensional weighted *holes*, but instead of edges one considers triangles.

To study the persistence of a topological feature in a network, each weighted *hole* g present can be

23

characterized by three quantities: 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 $w' < w$, that is, when $t' > t$, 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 edges composing it if it is a 1-dimensional *hole*, and to the number of triangles if it is a 2-dimensional *hole*. The following two scalar metrics proposed in [Petri et al., 2013] do not depend on the number of generators in the network filtration, and provide a way to compare the WRCF results from different networks. Both metrics are defined as:

$$h_k = \frac{1}{N_{g_k}} \sum_{g_k} \frac{p_{g_k}}{T} \tag{3.3}$$

$$\tilde{h}_k = \frac{1}{N_{g_k}} \sum_{g_k} \frac{\lambda_{g_g}}{N} \frac{p_{g_k}}{T} \tag{3.4}$$

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_i, 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* $\tilde{h}_k$.

3.3.4   Detecting Weighted *Holes*

In order to study the persistence of weighted *holes*, a well documented software designed to compute homology called Javaplex was used [Tausz et al., 2014]. 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 3.5, 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. Thus, in the analysis one has to take into account that the ROIs that belong to the representative cycles returned, may not be the geometrically correct ones (i.e. shortest), as this corresponds to finding an optimal basis of homology generators, one that behaves in a geometrically representative way. The problem of finding geometrically optimal representative cycles is still an open area of research.

The WRCF method relies on the gradual addition of weighed edges, therefore one needs to convert the similarity matrix into a weighted edgelist, and then order the edges according to the weights. At each

filtration step t, one adds one or more edges depending on whether or not there are edges that share the same weight. This construction results in a subgraph G_t with edges whose weights are higher than or equal to w_t as required.

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 [Born and Kerbosch, 1973], as it is the most efficient algorithm known for this problem [Stolz et al., 2017]. 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, which resulted in an average of 3317.1 edges.

3.4   Null Models

The networks were compared with two null models proposed in [Opsahl et al., 2008], 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. Both randomizations were designed to break possible existing correlations between the weights and the links, while preserving the degree distribution $P(k)$, (i.e, the probability that a given ROI is connected to exactly k other ROI's) and the weight distribution $P(w)$ of the original networks (i.e, the probability that a given connection has weight w). 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.
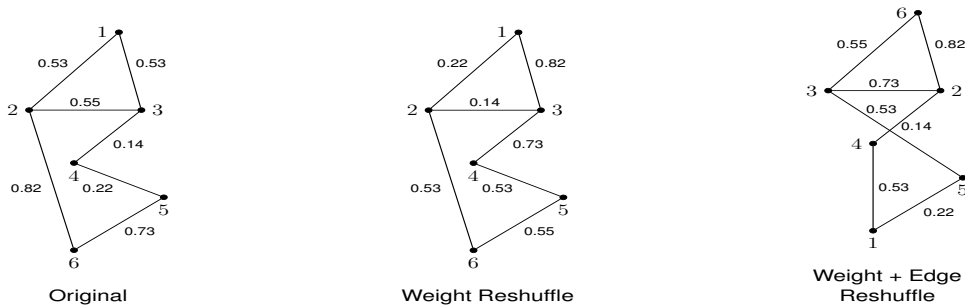

Figure 3.6: Null models construction

Chapter 4

Results

The hollowness values computed are presented in this chapter, together with the barcodes and distributions of persistences, lengths and births of 1 dimensional holes for one of the subjects. A comparison between subjects distributions is also showed. A distribution of the regions that belong to the more persistent 1-cycles is also reported in this chapter.

4.1 Network Hollowness

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. 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	0.409	0.224	2.022	0.249	2.205	0.055	0.010
2	37.518	0.156	0.844	0.231	3.163	0.219	2.852	0.049	0.091
3	33.516	0.170	0.864	0.266	2.836	0.245	3.066	0.047	0.016
4	31.984	0.171	0.581	0.231	2.478	0.243	2.722	0.065	0.042
5	36.426	0.136	0.554	0.237	3.253	0.220	2.873	0.071	0.159
6	38.997	0.174	0.778	0.239	3.716	0.226	3.128	0.073	0.326
9	26.082	0.163	0.606	0.239	1.707	0.266	2.312	0.009	0.001
10	31.554	0.176	0.697	0.237	2.527	0.254	2.524	0.043	0.047
14	31.730	0.178	0.590	0.236	2.294	0.229	2.089	0.032	0.007
15	37.034	0.152	0.675	0.246	3.232	0.238	3.174	0.088	0.329

Table 4.1: 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 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 showed. 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. The normalized hollowness accounting for the 1 dimensional *holes* obtained for the three networks are higher than the non-normalized hollowness for all subjects, which might result from the fact that Javaplex can return much longer representative cycles than would be expected.

4.2   Persistence, Length and Birth Distributions

The distributions of the persistence, length and birth times of 1-dimensional *holes* for each network and randomized versions were obtained for all subjects and are available in B. 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.
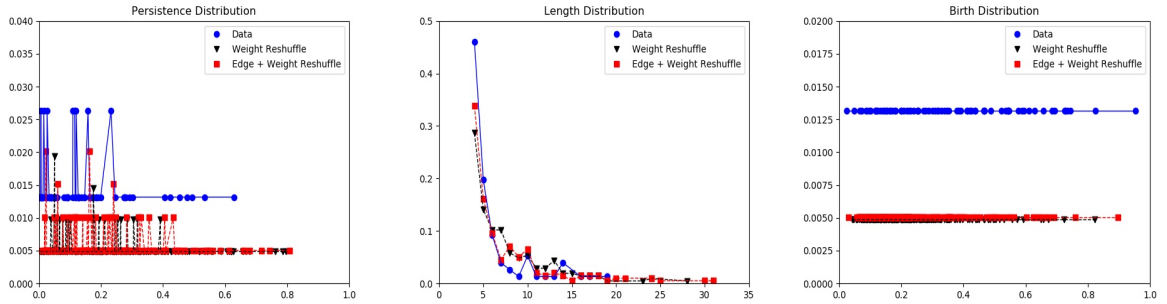

Figure 4.1: Persistence, length and birth distributions obtained for subject 5.

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 4.2 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.
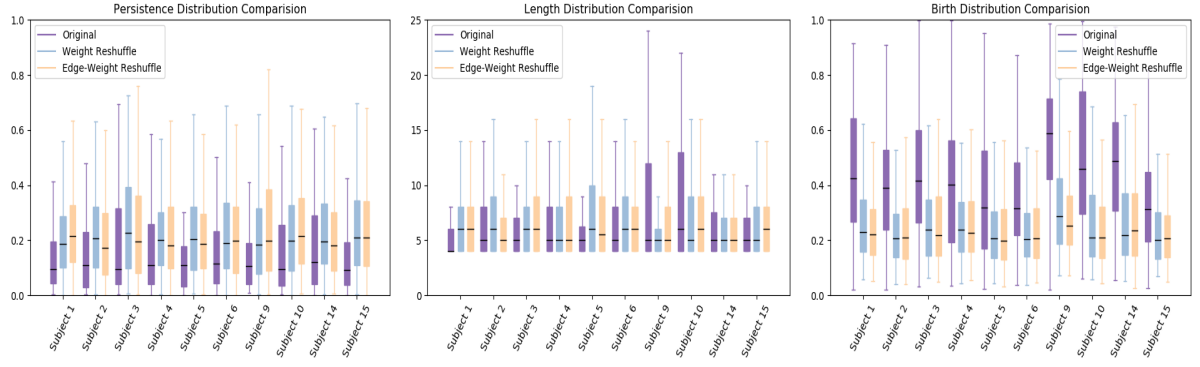
Figure 4.2: Comparison of the persistence, length and birth distributions.

4.3 Barcodes

The barcodes of the original networks and their randomized versions were collected for all subjects and are available in B. The following figure show the barcodes obtained for the original network of subject 5. The barcodes of the randomized versions are shown in figure 4.4.
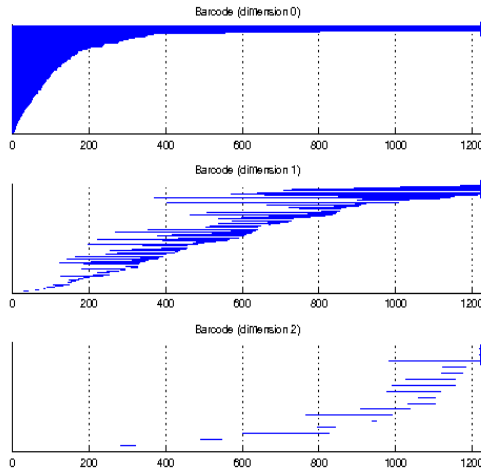
Figure 4.3: Barcodes of the association matrix of subject 5.

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 4.4: Barcodes of the weight reshuffle version (left) and weight plus edge reshuffle version (right).
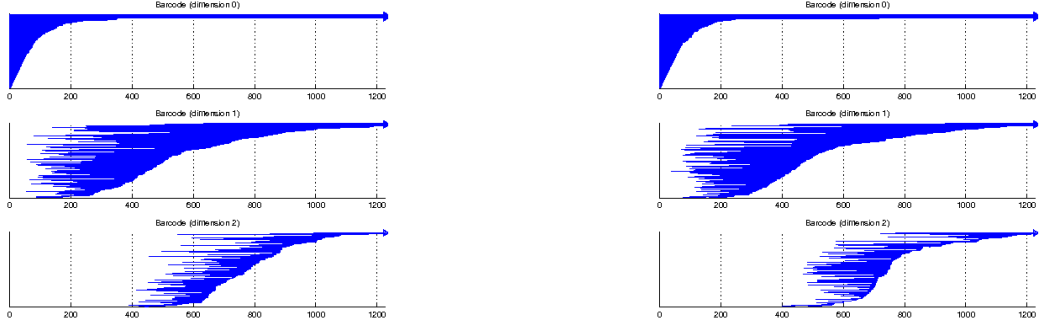
Short persistence 2 dimensional *holes* were found in all networks, and in higher number for the randomized networks, which might result from the fact that since the randomized versions presented a higher number of 1-dimension *holes*, there is a higher tendency for the added edges to give rise to 4-cliques, 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.

4.4   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 A.2. 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.
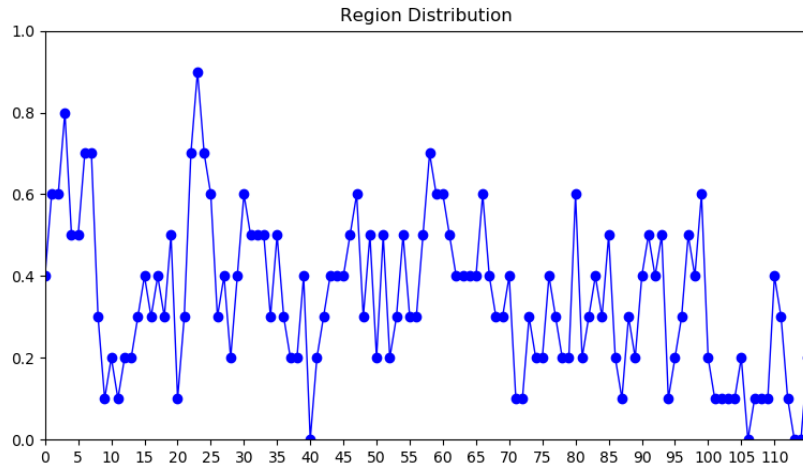

Figure 4.5: Distribution of regions.

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.

Chapter 5

Conclusion

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 [Petri et al., 2013] 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 any 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 that data, and should allow 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 [Schwarz and McGonigle, 2011].

Other software packages (e.g. https://www.jholes.eu/) allow the computation of more than one missing face in the same filtration step through the use of multiple threads, 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 information that arise from thresholding the networks will be avoided.

Acknowledgments

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.

Bibliography

E. T. Bullmore and O. Sporns. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, Mar 2009.

A. S. Pandit, P. Expert, R. Lambiotte, et al. Traumatic brain injury impairs small-world topology. *Neurology*, 80(20):1826–1833, May 2013.

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

H. Edelsbrunner, D. Letscher, and A. Zomorodian. *Discrete & Computational Geometry*, volume 28, pages 511–533. 2002.

H. Y. Carr. *Free precession techniques in nuclear magnetic resonance.* PhD thesis, Harvard, 1953.

R. B. Buxton. *Introduction to functional magnetic resonance imaging: principles and techniques.* Cambridge University Press, 2009.

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

B. Biswal, F. Z. Yetkin, V. M Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34(4):537–41, Oct 1995.

N. Khanna, W. Altmeyer, J. Zhuo, and A. Steven. Functional neuroimaging: Fundamental principles and clinical applications. *The Neuroradiology journal*, 28(2):87–96, Apr 2015.

K. J. Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping*, 2(1-2):56–78, 1994.

C. Park, S. Y. Kim, Y. Kim, and K. Kim. Comparison of the small-world topology between anatomical and functional connectivity in the human brain. *Physica A: Statistical Mechanics and its Applications*, 387 (23):5958–5962, Oct 2008.

B. Biswal, M. Mennes, X. Zuo, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.

M. A. Armstrong. *Basic Topology*. Springer, 1983.

L. Euler. Elementa doctrinae solidorum (eneström 230) novi commentarii academiae scientiarum petropolitanæ. *Opera Omnia series 1*, 26:71–93, 1752.

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`.

C. Delfinado and H. Edelsbrunner. An incremental algorithm for betti numbers of simplicial complexes on the 3-sphere. *Computer Aided Geometric Design*, 2(7):771–784, Nov 1995.

W. F. Basener. *Topology and its applications*, pages 300–311. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006. ISBN 978-0-471-68755-9; 0-471-68755-3.

B. Stolz, T. Emerson, S. Nahkuri, M. Porter, and H. Harrington. Topological Data Analysis of Task-Based fMRI Data from Experiments on Schizophrenia. *arXiv e-prints*, Sep 2018.

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`.

M. Jenkinson, P. Bannister, J. M. Brady, and S. M Smith. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841, Oct 2002.

S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–55, Nov 2002.

M. Jenkinson and S. M. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, 2001.

M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith. Temporal autocorrelation in univariate linear modeling of fmri data. *NeuroImage*, 14(6):1370–1386, Dec 2001.

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imag*, 20(1):45–57, Jan 2001.

S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich. Network modelling methods for fmri. *NeuroImage*, 54(2):875–891, Jan 2011.

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

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.

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.

G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, Apr 2009.

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

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/`.

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

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.

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.

Appendix A

Appendix A

ID	Region	ID	Region	ID	Region
0	Precentral_L	39	ParaHippocampal_R	78	Heschl_L
1	Precentral_R	40	Amygdala_L	79	Heschl_R
2	Frontal_Sup_L	41	Amygdala_R	80	Temporal_Sup_L
3	Frontal_Sup_R	42	Calcarine_L	81	Temporal_Sup_R
4	Frontal_Sup_Orb_L	43	Calcarine_R	82	Temporal_Pole_Sup_L
5	Frontal_Sup_Orb_R	44	Cuneus_L	83	Temporal_Pole_Sup_R
6	Frontal_Mid_L	45	Cuneus_R	84	Temporal_Mid_L
7	Frontal_Mid_R	46	Lingual_L	85	Temporal_Mid_R
8	Frontal_Mid_Orb_L	47	Lingual_R	86	Temporal_Pole_Mid_L
9	Frontal_Mid_Orb_R	48	Occipital_Sup_L	87	Temporal_Pole_Mid_R
10	Frontal_Inf_Oper_L	49	Occipital_Sup_R	88	Temporal_Inf_L
11	Frontal_Inf_Oper_R	50	Occipital_Mid_L	89	Temporal_Inf_R
12	Frontal_Inf_Tri_L	51	Occipital_Mid_R	90	Cerebelum_Crus1_L
13	Frontal_Inf_Tri_R	52	Occipital_Inf_L	91	Cerebelum_Crus1_R
14	Frontal_Inf_Orb_L	53	Occipital_Inf_R	92	Cerebelum_Crus2_L
15	Frontal_Inf_Orb_R	54	Fusiform_L	93	Cerebelum_Crus2_R
16	Rolandic_Oper_L	55	Fusiform_R	94	Cerebelum_3_L
17	Rolandic_Oper_R	56	Postcentral_L	95	Cerebelum_3_R
18	Supp_Motor_Area_L	57	Postcentral_R	96	Cerebelum_4_5_L
19	Supp_Motor_Area_R	58	Parietal_Sup_L	97	Cerebelum_4_5_R
20	Olfactory_L	59	Parietal_Sup_R	98	Cerebelum_6_L
21	Olfactory_R	60	Parietal_Inf_L	99	Cerebelum_6_R
22	Frontal_Sup_Medial_L	61	Parietal_Inf_R	100	Cerebelum_7b_L
23	Frontal_Sup_Medial_R	62	SupraMarginal_L	101	Cerebelum_7b_R
24	Frontal_Med_Orb_L	63	SupraMarginal_R	102	Cerebelum_8_L
25	Frontal_Med_Orb_R	64	Angular_L	103	Cerebelum_8_R
26	Rectus_L	65	Angular_R	104	Cerebelum_9_L
27	Rectus_R	66	Precuneus_L	105	Cerebelum_9_R
28	Insula_L	67	Precuneus_R	106	Cerebelum_10_L
29	Insula_R	68	Paracentral_Lobule_L	107	Cerebelum_10_R
30	Cingulum_Ant_L	69	Paracentral_Lobule_R	108	Vermis_1_2
31	Cingulum_Ant_R	70	Caudate_L	109	Vermis_3
32	Cingulum_Mid_L	71	Caudate_R	110	Vermis_4_5
33	Cingulum_Mid_R	72	Putamen_L	111	Vermis_6
34	Cingulum_Post_L	73	Putamen_R	112	Vermis_7
35	Cingulum_Post_R	74	Pallidum_L	113	Vermis_8
36	Hippocampus_L	75	Pallidum_R	114	Vermis_9
37	Hippocampus_R	76	Thalamus_L	115	Vermis_10
38	ParaHippocampal_L	77	Thalamus_R		

Table A.1: List of regions and IDs used in the parcellation of the image.

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]

Table A.2: Table with the regions belonging to the 1 dimensional cycles that persisted for half or more than half of the maximum filtration value.
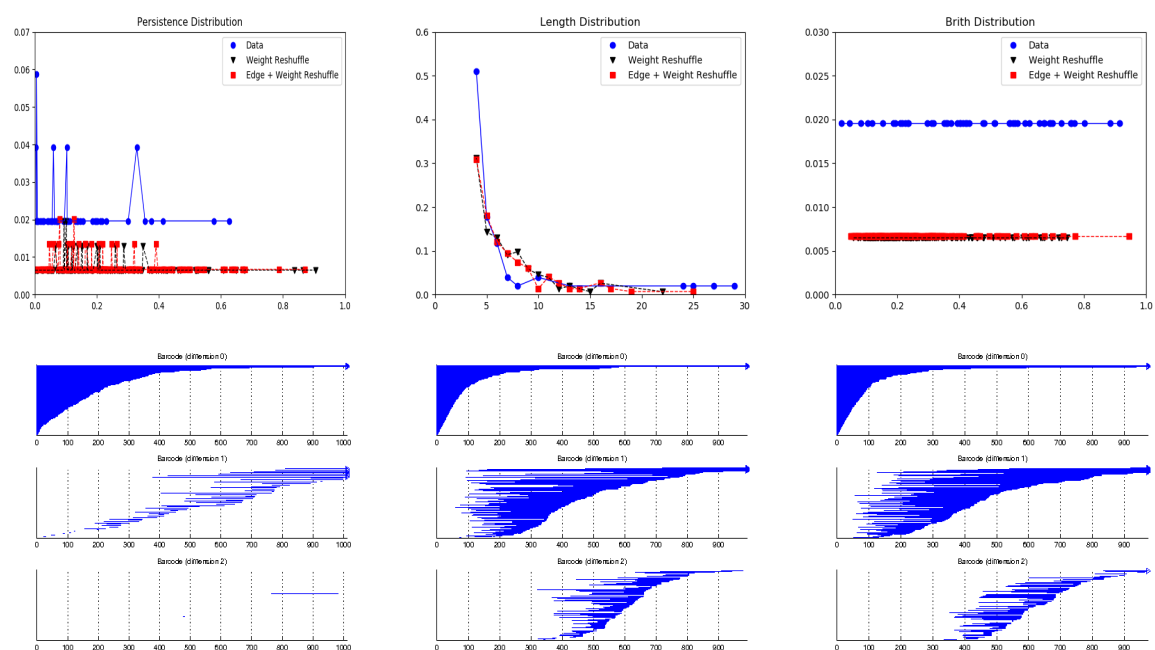
Appendix B

Appendix B

Figure B.1: Results obtained for subject 1. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.
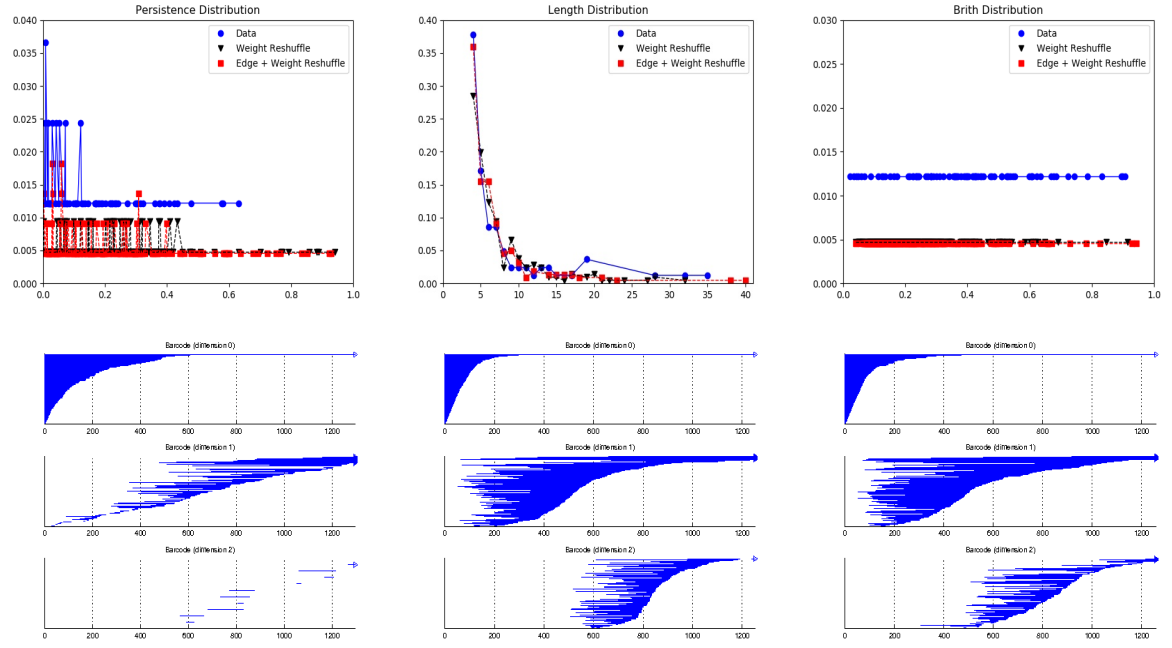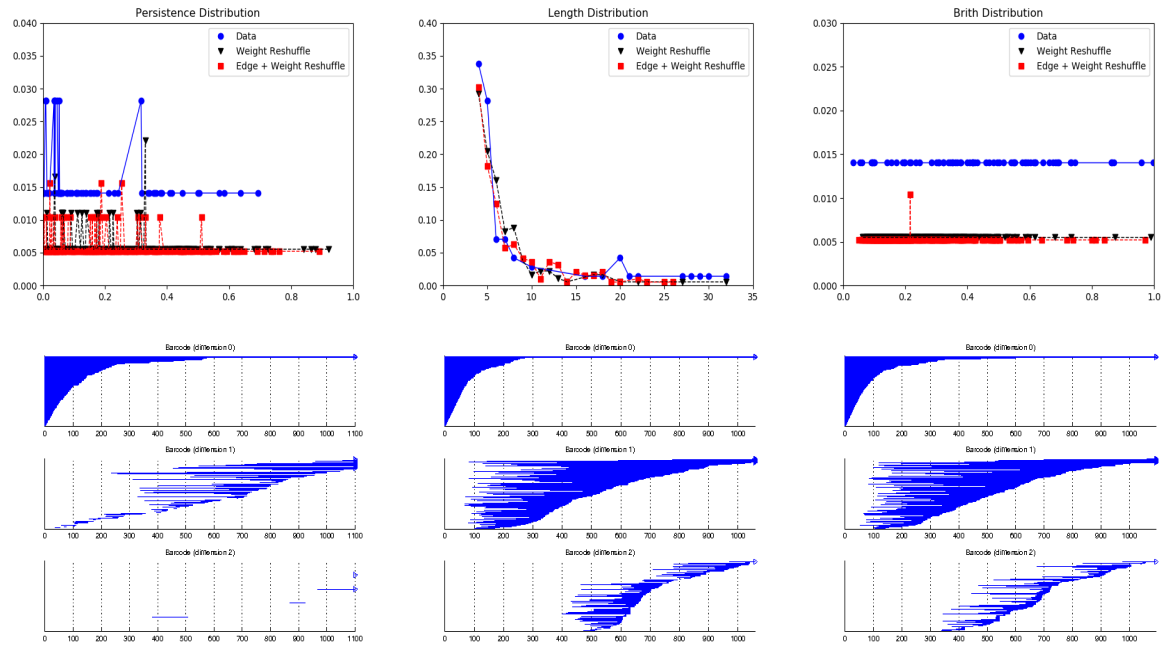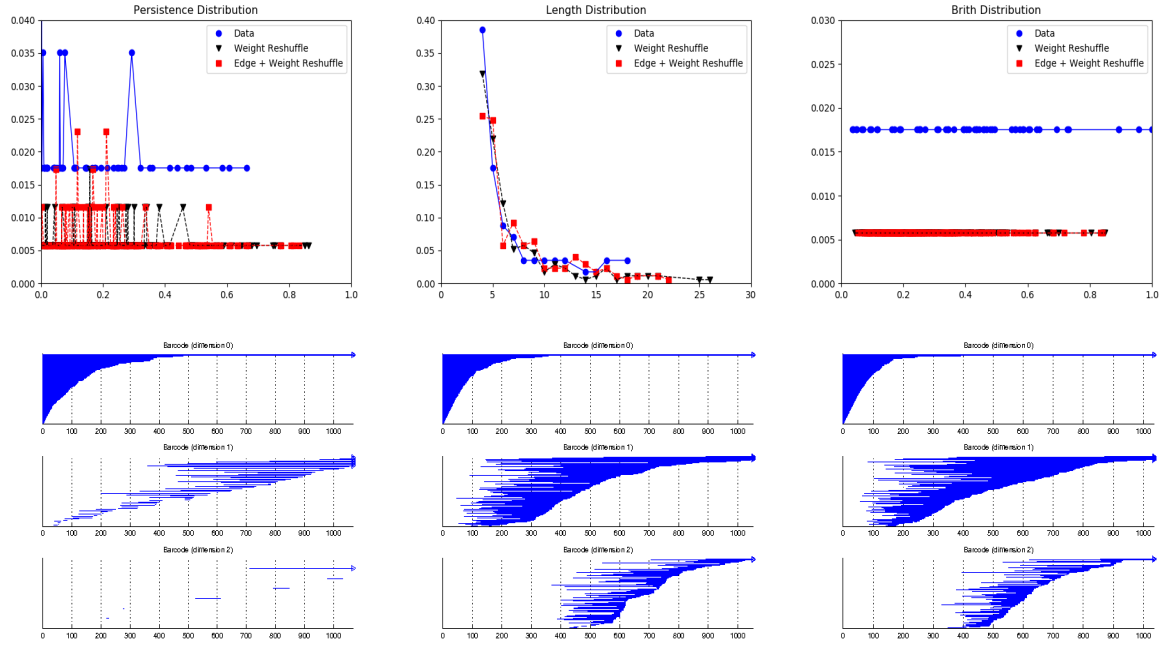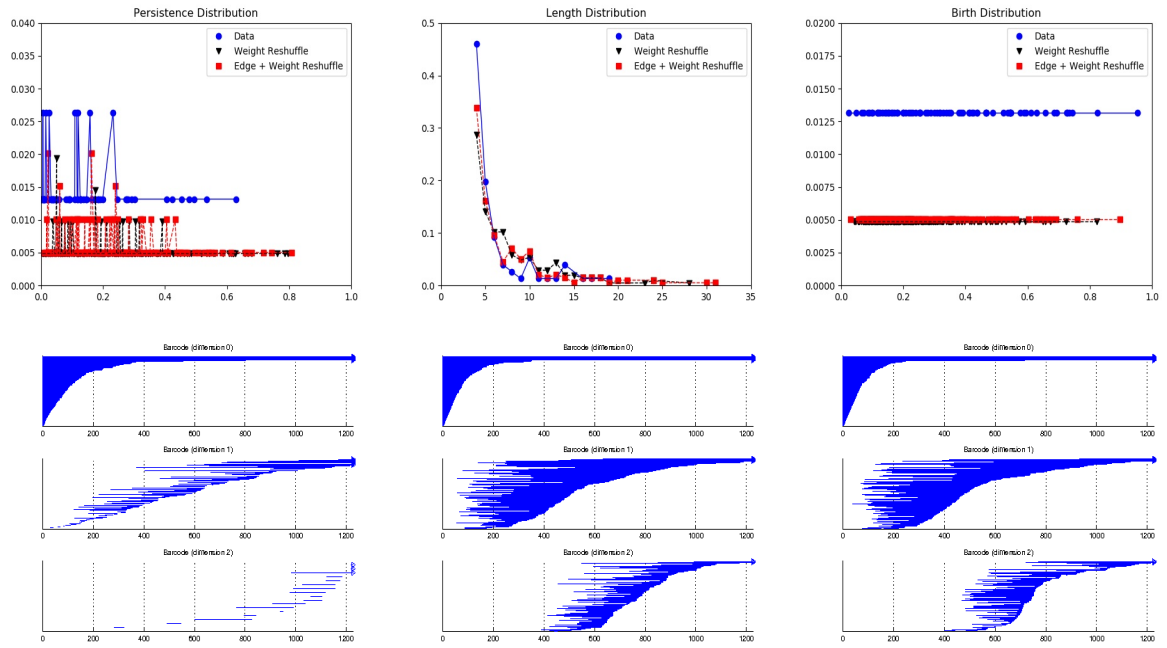
Figure B.2: Results obtained for subject 2. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.

Figure B.3: Results obtained for subject 3. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.
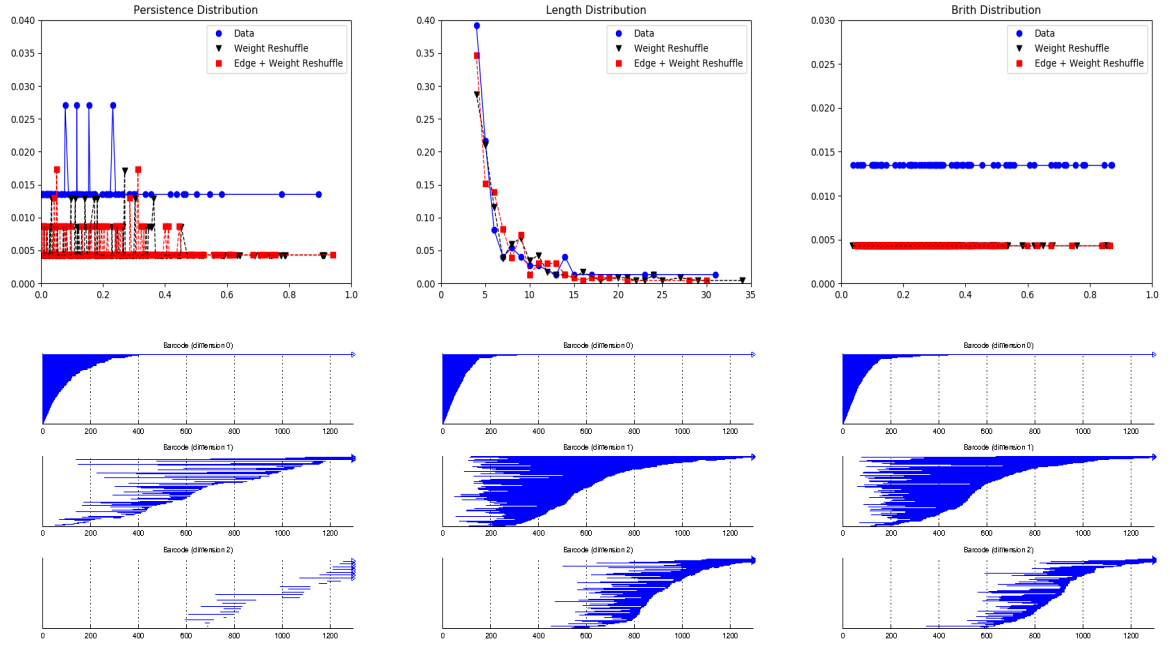
Figure B.4: Results obtained for subject 4. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.

Figure B.5: Results obtained for subject 5. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.
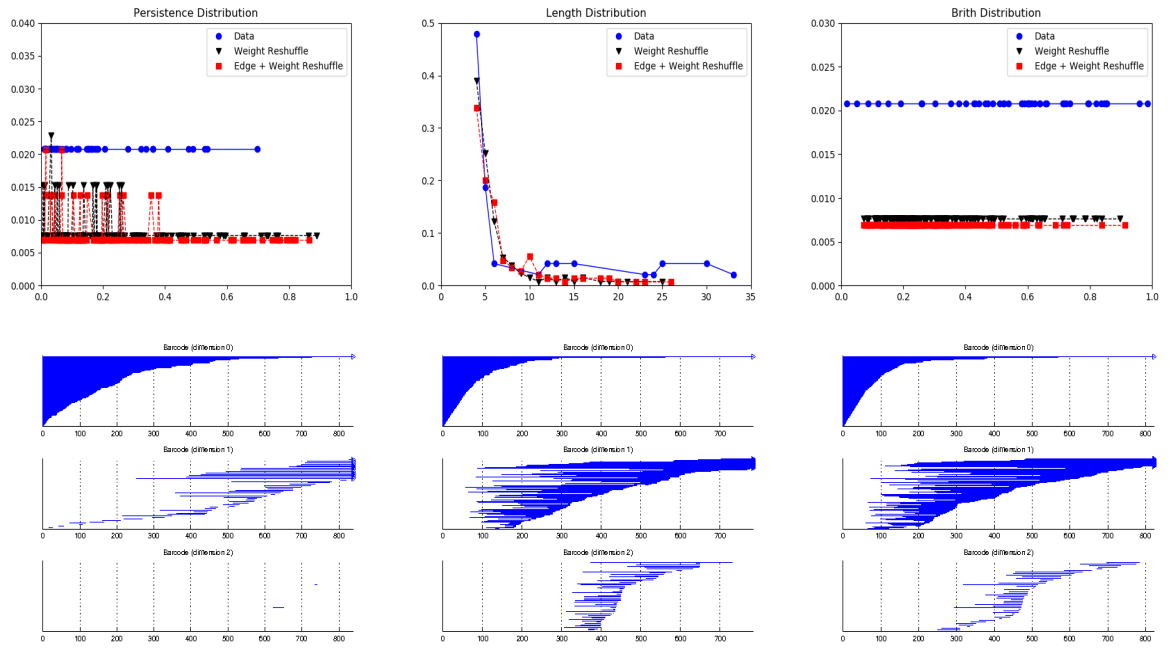
Figure B.6: Results obtained for subject 6. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.
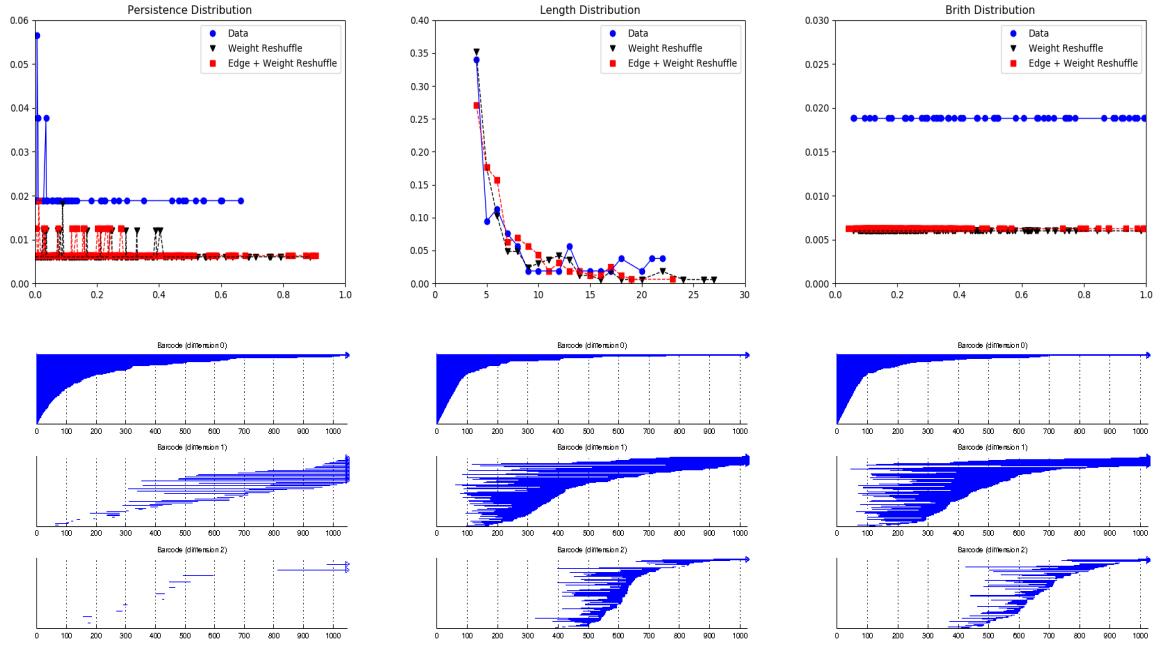

Figure B.7: Results obtained for subject 9. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.

Figure B.8: Results obtained for subject 10. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.
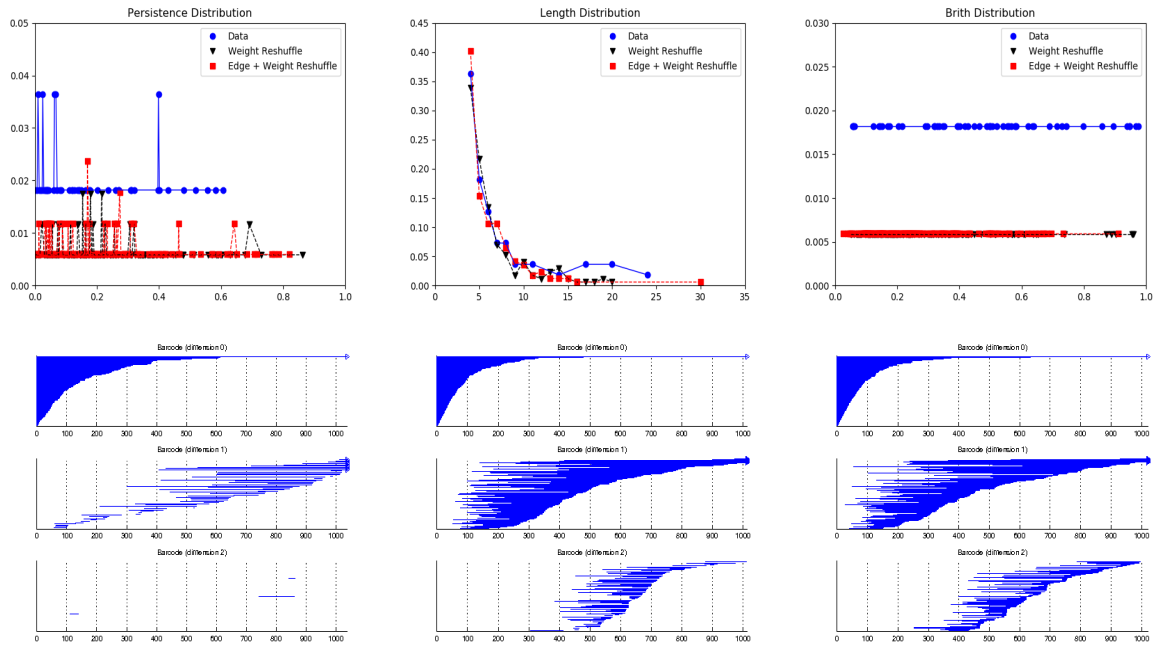

Figure B.9: Results obtained for subject 14. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.
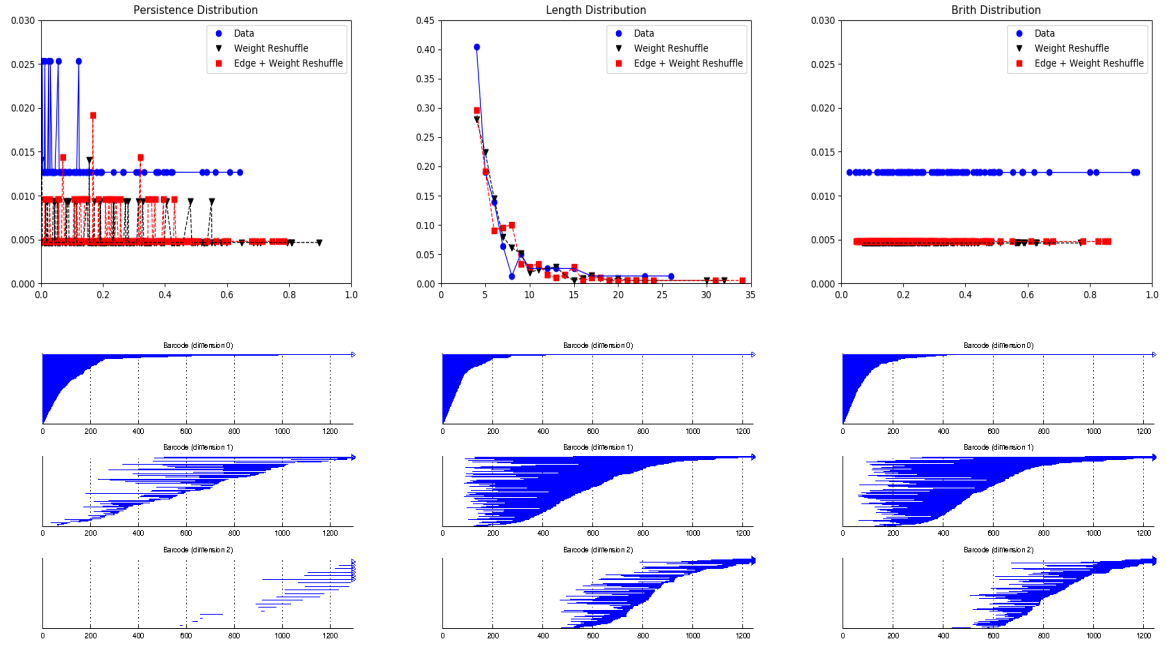
Figure B.10: Results obtained for subject 15. Barcode of the original network on the left, in the middle the barcode of the weight reshuffle network, and on the right the edge plus weight reshuffle network.