POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

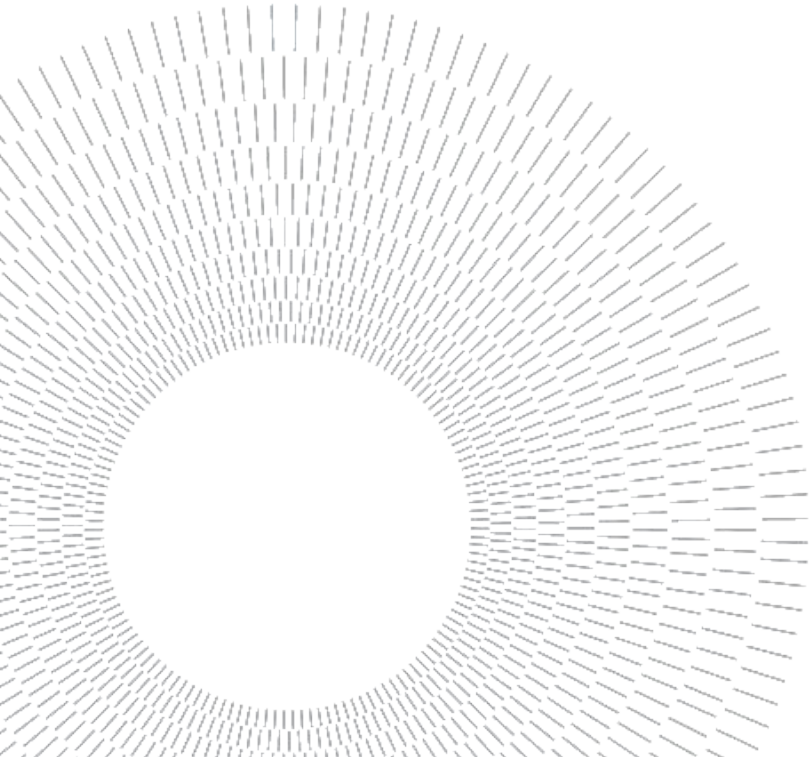
Brain Connectivity assessment of Graph and Sub-Graph Indices: Uncertainty evaluation by Bootstrap and Sensitivity Analysis

TESI DI LAUREA MAGISTRALE IN
BIOMEDICAL ENGINEERING-
INFORMATION BIOENGINEERING (BIF)

Supervisor: Prof. Giuseppe Baselli
Co-Supervisor: Eng. Davide Coluzzi

Author: **Cesare Rocco 952753**

Academic Year 2021/2022



Acknowledgements

I wish to thank Professor Giuseppe Baselli and Davide Coluzzi for allowing me to work on this such interesting topic, for their availability, for their support and for following me from the start to the end of this work. It was really a pleasure and an inspiration working with you.

I also want to express my gratitude to my family, my closest friends and Gabriella for standing beside me and supporting me throughout this whole journey.

My achievements are also yours.

Abstract

Brain networks are formed by pathways which connect cerebral regions. It is possible to investigate them in a non-invasive way thanks to Magnetic Resonance Imaging techniques. In this context, brain networks can be analyzed with respect to graph theory, considering the structural or functional connectivity pathways to analyze the brain's organization and using different graph metrics. However, the MRI techniques are characterized by several limitations. These problems make the statistical investigation of structural or functional connectivity subject to uncertainty. In this context, a sensitivity analysis and the bootstrapping technique were introduced to test the uncertainty of graph indexes.

Three datasets were tested: a functional and a structural connectivity datasets of healthy subjects and also a functional connectivity dataset formed by control and schizophrenic subjects. First, the sensitivity analysis was performed to evaluate the influence of the connectivity weight outliers. Second, the Bootstrapping technique was used to obtain surrogate data of random extraction of connectivity weights, which were normalized according to a probabilistic normalization procedure to obtain more robust results.

The major findings of this work were about the quantification of the variability of the metrics and biases of the distributions. On one hand, the sensitivity analysis confirmed the hypothesis that the great variability of the connectivity weights causes a great variability of the indexes and uncertainties. On the other hand, differences ($p < 0.05$) between healthy and schizophrenic subjects in several graph-based indexes, lacking in the original data, emerged using the bootstrapping procedure.

Then, a software for the analysis of subgraphs (Spider-Net) was used to show the connectivity pathways between specific subsets of brain parcels with the connectogram visualization. The structural connectivity dataset in terms of comparison between two MRI processing techniques and the differences between healthy and schizophrenic subjects in the Default Mode Network (DMN) were qualitatively analyzed.

Abstract in italiano

Le reti neurali sono formate da vie che connettono le regioni cerebrali. È possibile investigare tali reti non maniera non invasiva con tecniche di risonanza magnetica (MRI). Le reti neurali, inoltre, possono essere analizzate secondo la teoria dei grafi, considerando i percorsi di connettività strutturale e funzionale per l'investigazione della topologia cerebrale usando metriche dei grafi. Tuttavia, le tecniche MRI sono caratterizzate da diverse limitazioni. Queste fanno sì che l'analisi statistica di connettività strutturale o funzionale sia affetta da incertezze. In questo contesto, un'analisi di sensitività e il bootstrap sono stati introdotti per testare l'incertezza sugli indici dei grafi.

Tre dataset sono stati considerati: un dataset funzionale e uno strutturale formati solo da soggetti sani e un dataset funzionale formato da soggetti schizofrenici e di controllo. In primo luogo, l'analisi di sensitività è stata usata per analizzare l'influenza degli outliers sui pesi di connettività. Successivamente, il bootstrap è stato usato per ottenere dati surrogati attraverso estrazione random dei pesi di connettività, i quali sono stati anche normalizzati secondo una normalizzazione probabilistica per ottenere risultati più robusti.

I risultati più importanti di questa tesi sono relativi alla quantificazione della variabilità degli indici di grafi e i bias delle distribuzioni di tali indici. Da un lato, l'analisi di sensitività ha confermato l'ipotesi sulla elevata variabilità dei pesi di connettività che causa incertezze e variabilità degli indici. Dall'altro lato, differenze ($p < 0.05$) tra soggetti sani e schizofrenici in molti indici di grafi sono emerse usando il bootstrap .

Inoltre, un software per l'analisi di sottografi, chiamato Spider-Net, è stato usato per mostrare le connessioni tra specifiche parcels attraverso la visualizzazione in connettogrammi. Il dataset strutturale è stato analizzato comparando due tecniche di preprocessing MRI e analizzando quantitativamente le differenze tra soggetti sani e schizofrenici nella Default Mode Network (DMN).

Key Words: Brain Networks; MRI; Graph Theory; Graph Indexes; Structural and Functional Connectivity; Connectograms; Uncertainty; Bootstrap; Sensitivity Analysis; Normalization; Schizophrenia; Spider-Net; Default Mode Network.

Parole chiave: Reti Neurali; MRI; Teoria dei Grafi; Indici dei Grafi; Connettività Strutturale e Funzionale; Connettogramma; Incertezza; Bootstrap; Analisi di Sensitività; Normalizzazione; Schizofrenia; Spider-Net; Default Mode Network.

EXECUTIVE SUMMARY OF THE THESIS:

Brain Connectivity assessment of Graph and Sub-Graph Indexes: Uncertainty evaluation by Bootstrap and Sensitivity Analysis

TESI MAGISTRALE IN BIOMEDICAL ENGINEERING – INFORMATION BIOENGINEERING (BIF)

AUTHOR: CESARE ROCCO 952753

SUPERVISOR: PROF. GIUSEPPE BASELLI

CO-SUPERVISOR: ENG. DAVIDE COLUZZI

ACADEMIC YEAR: 2021-2022

1. Introduction

The brain is the most tangled network known to mankind. Indeed, it comprises billions of neurons which form trillions of synapses between each other. Considering the methods used to investigate brain networks, Magnetic Resonance Imaging (MRI) is the dominant technique mainly because of its safety, spatial resolution and availability throughout the world. On one hand, functional MRI (fMRI) inspects the dynamics of activity in each gray matter (GM) area. The functional activations of the GM areas are based on the Blood Oxygen Level Dependent (BOLD) response. On the other hand, Diffusion Tensor Imaging (DTI) and High Resolution Angular Imaging (HARDI) allow to visualize and examine the organization, the orientation and the trajectories of white matter (WM) tracts. The latter solves the intrinsic limitation of DTI of not detecting crossing configurations characterized by high curvatures, known as the crossing fiber problem [1]. In the context of the analysis of functional brain networks, considering the known dynamic and condition-dependent nature of brain activity, it is obvious that the functional connectivity metrics such as the Pearson correlation coefficient or the wavelet coherence will change over time [1]. This variation of the connectivity metrics is, however, not easy to be explained. The comprehension of brain networks is a complex field of research named connectomics, which addresses the brain networks at all its scales and features [1]. It is possible to map the parcels of the network according to well-known brain atlases from the structural or functional point of view [2]. The brain network analysis is based on the creation of connectivity matrices, which are investigated in their organization and function, through methods from the graph theory. Thus, a graph can be defined as a collection of nodes (brain regions) and links (anatomical or functional connections) between pairs of nodes [2]. The weights of the connectivity matrices can be 0 and 1 (binary) considering the presence/absence of an edge connecting two regions. Conversely, they can represent the strengths of the connections according to the connectivity measure used (such as the number of fibers in the structural case, or the Pearson's correlation in the functional case). In graph theory, it is possible to analyze and compare different networks towards the investigation of their topological features, with binary or weighted indexes such as node degree, node strength, clustering coefficient, efficiency, path length or modularity [2]. These indexes represent relevant properties in brain connectivity since many studies, revealed changes of these metrics in a pathological condition, such as schizophrenia [3], or Alzheimer's disease. Due to the limitations of the MRI techniques [1] the statistical investigation of structural brain connectivity datasets is often subject to uncertainty. In

this context, a sensitivity analysis based on graph perturbation could be useful in understanding and quantifying the uncertainties. In particular, the investigation is based on removing some nodes/edges and investigate how this removal affects the graph results. Considering the great variability of structural or functional connectivity measures from a statistical point of view [1], the Bootstrapping technique can be introduced. This technique is able to use repeated random sampling from a set of measurements, creating surrogate data which allow to obtain more robust estimates of the variability and uncertainty of different graph features. Recently, the investigation of brain networks focused on the visualization of connectivity pathways between specific subsets of interest through connectograms. In this regards, a novel software tool allowing the analysis of subgraphs can be introduced, called SPIDER-Net Tool (SNT) [4]. The aims of this study were:

1. the development of a method for assessing the connectivity considering the intrinsic uncertainty of the measures, employing:
    a. a sensitivity analysis to analyze the variability of the brain connectivity weights;
    b. a Bootstrapping procedure with a probabilistic normalization to obtain more robust brain indexes, testing it on structural and functional healthy control dataset;
2. the validation of the method on a healthy control-schizophrenic study;
3. the extraction and the exploration of connectivity of specific sub-network of interest (such as the Default Mode Network) [1].

2. Materials and Methods

This section first characterizes the main dataset used in this work, then the protocols employed to obtain the results are described.

2.1 Data Acquisition and Study Population

The investigations of this work were done on 3 different datasets: a functional connectivity dataset and a structural connectivity dataset formed by healthy controls, and a functional connectivity dataset of schizophrenic and control subjects. The considered datasets are derived from ethically approved protocols, informed consent and anonymization included. The first dataset used was acquired from resting state fMRI scans in a group of 10 healthy subjects, 5 males and 5 females, with age between 30 and 43 years (mean age + SD = 36 ± 6.32). In order to construct the connectivity matrices, AAL atlas was used to divide the brain in 90 parcels which were considered as nodes in the network. A 90x90 correlation matrix of Pearson's correlation coefficients between all possible connections of node pairs was, thus, computed. The functional connectivity matrices, initially fully-connected, were density-thresholded at 50% with SNT.

The structural dataset consists of 17 healthy control (HCs) subjects (7 males and 10 females; mean age ± SD: 52.5 ± 8.3 years) acquired with both DTI and HARDI techniques. The brain was divided in parcels, 165 in total, according to the Destrieux atlas. The structural connectivity matrices were obtained by computing the edges as the number of the reconstructed fibers normalized by the sum of the nodes volume. Further information and details on data acquisition are reported in [4].

The functional connectivity dataset of Schizophrenic and Healthy subjects in formed by fifteen healthy volunteers (mean age 33.3 years, SD = 9.2 years, 14 male) and 12 subjects with chronic schizophrenia (mean age 32.8 years, SD = 9.2 years, 10 male) diagnosed according to standard operational criteria in the Diagnostic and Statistical Manual of Mental Disorders IV (American Psychiatric Association, 2000). Nodes were labelled according to the AAL segmentation defining a 74×74 connectivity matrix for each subject. The functional connectivity matrices were adjusted to reset negative connections and put the diagonal weights to zero. Further information and details on data acquisition are reported in [5].

2.2 Protocol

Bootstrap was applied with the aim of creating resampled data to evaluate the topology of the functional network. After the creation of the surrogate data, the network indexes were computed through the Brain Connectivity Toolbox (BCT) [2] and the CI were computed, with the Standard formulas used in this context to compute both the boundaries, such that:

$$upper, lower\ boundary = mean \pm 1.96 \cdot \frac{std(BootstrapResamples)}{\sqrt{number\ of\ iterations}} \quad (1)$$

Bootstrap method was analyzed with respect to the expected results in terms of the reduced variability of the index distributions and alignment to the original data. Specifically, the bias of the mean of the distributions of the values of the different graph metrics is computed such that:

$$bias = abs\left(\frac{mean(InitialData) - mean(BootstrapResamples)}{mean(InitialData)} \cdot 100\right) \quad (2)$$

In order to analyze the huge variability of the connectivity weights resulting from the limitations of the MRI techniques, a sensitivity analysis was performed to investigate the variability of the graph indexes. In particular, the contribution of strongest/weakest connections was explored. The variability of the indices was examined by resetting one at a time the connectivity weights outside the 10^{th} – 90^{th} percentile range and computing the value of the graph metrics at each resetting. Weighted Clustering Coefficient, Node Strength and Weighted Efficiency were considered for this analysis. The aim of this analysis was to verify that the outliers weights are the ones that cause the variability of the values of the graph indexes. Furthermore, to reduce the variability of the connectivity weights, the data were normalized according to a probabilistic normalization procedure, whose flowchart is represented in Figure 1.

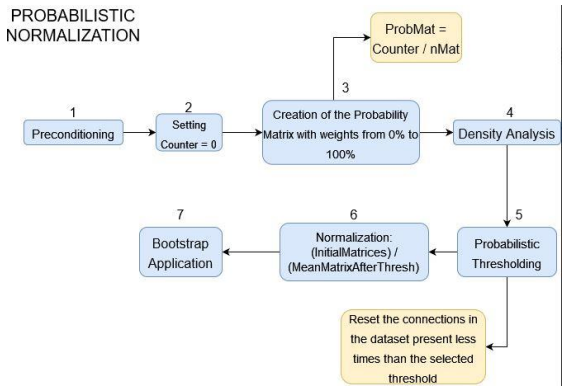


Figure 1 *Flowchart of the Probabilistic Normalization procedure*

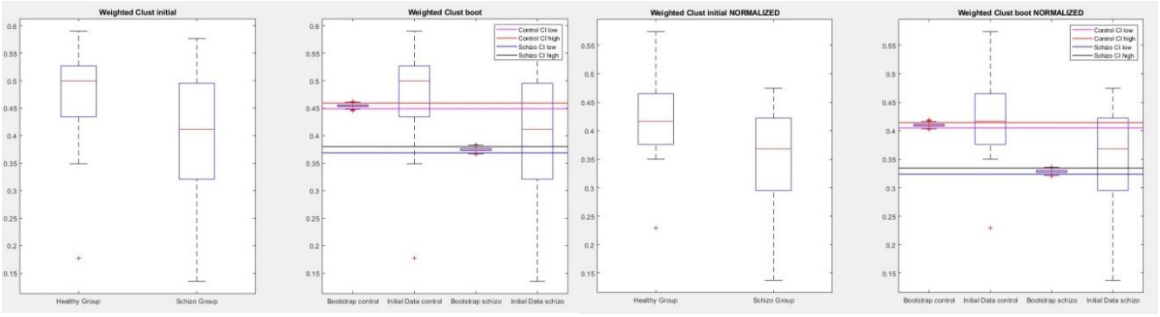
In particular, after the preconditioning phase (1), a counter is initialized (2), which counts how many times each i^{th},j^{th} element in the matrices is present. Obtained this counter matrix, the next step concerns the creation of a probability matrix (3) by dividing each element of the counter matrix by the total number of matrices present in the dataset, thus, obtaining a matrix of probabilities characterized by i^{th},j^{th} elements ranging from 0 to 1 (0% to 100%). Then, a Density Analysis (4) is performed to investigate the variation of the density of the dataset at the resetting of the rarest connections. A threshold that did not vary the density of the dataset was selected and applied according to the probability of each connection to be present in the dataset (5). Afterwards, the mean matrix of the so obtained dataset is evaluated. The normalization (6) is then achieved by dividing the initial matrices by the Mean Matrix obtained at the previous step. Unpaired t-tests for the graph indexes within healthy and schizophrenic groups were performed. They were computed from: the original data; the normalized data; the surrogate data obtained from the standard bootstrapping method and the surrogate data obtained from the normalized bootstrapping method. In the end, the DMN-subnetwork connectograms of healthy and schizophrenic subjects were extracted in SNT and analyzed.

3. Results

An example of the Weighted Clustering Coefficient is reported in the Figures 1 and 2 for illustrative purposes. In particular, Figure 1 (a) and (c), represent the distribution of the index extracted from the initial data, whereas Figure 1 (b) and (d) the distributions after the bootstrapping procedure with the boundaries of the confidence intervals represented with different colors. It is possible to notice the restriction of the distributions assumed by the values of the index after the bootstrap with both techniques. Moreover, a bias of the distribution can be seen from this figure, particularly in the Standard bootstrapping method (Figure 1 (b)), while it is reduced with the normalization (Figure 1 (d)).

Table 1 and 2 represent the numerical results of the distribution of the indices of the two populations after the bootstrapping procedure applied in the Standard and Normalization cases, respectively. The only statistically significant index (p-value BeforeBoot < 0.05) on the initial data before the application of the bootstrap is the Modularity. After the normalization of the data, the statistically significant indexes (p-value BeforeBoot < 0.05) are the Weighted Path Length, Weighted Efficiency and Strength. After the application of the bootstrap, all indexes in both Standard and Normalized cases become statistically significant (p-value AfterBoot < 0.05). The biases of the distributions considering before and after the bootstrap are higher in the pathological group for all indexes in both the Standard and Normalized cases, except from the Modularity in the Normalized case. This result is in line with what evaluated also in the functional and structural connectivity datasets. Indeed, distributions of all graph-indexes from the healthy subjects are shrinked after the bootstrap, and the biases reduced with the application of the normalization (from 17% to 8% in the functional connectivity dataset and from 13.3% to 11.5% in the structural connectivity dataset for the Weighted Clustering Coefficient). Figure 2 shows the results of the Sensitivity Analysis for two sample subjects from both control and pathological groups. In particular, the blue circles represent the values the index corresponding to the removal of the connections having outlier weights. On the other hand, the red dots represent the values of the index removing the connections leading to the greatest changes. This analysis confirmed that the outliers of the connectivity matrices, thus the variability of the weights, influence the variability of the values of the indexes. The

sensitivity analysis results highlighted that more weights in the schizophrenic population will cause a change in the values of the considered indexes, as expected from the different biases found through the bootstrap. Finally, Figure 3 shows the connectograms extracted with SNT of the DMN of the sixth control and tenth schizophrenic subjects aimed at identifying disruptions and a reduced activation of this subnetwork. It can be noticed that the schizophrenic case presents sparser connectograms and a reduced activity of the DMN. Indeed, the density of the DMN in the control population was 97.7% ± 3.8%, while it was 95.4% ± 4.6% in the schizophrenic population. Furthermore, the schizophrenic population is characterized by a weaker activation of the DMN as can be seen by Figure 3 middle panel, where the prevalence of yellow (weak) connectivity pathways is a sign of a network characterized by a low activation. Conversely, in the control connectogram, it is evident the prevalence of red (strong) connectivity pathways. Furthermore, the weights in the healthy control group (HC) are stronger than the pathological one (SCH), as highlighted by the strength (HC = 20.0 ± 4.1; SCH = 16.5 ± 5.1). After the removal of the weakest connections, the density results were 88.6% for the control group and 79.3% in the pathological one.



(a) (b) (c) (d)

Figure 2 *Distributions of the* ***Weighted Clustering Coefficient*** *for the Standard (figures (a) and (b)) and Normalized (figures (c) and (d)) data. The values of the confidence intervals evaluated after the bootstrap (figure (b) and (d) for standard and normalized procedures, respectively) are shown with different colors, such that the red and the magenta horizontal line represent the upper and lower value (respectively) of the confidence interval (CI) in the control case, while the black and the blue horizontal lines represent the upper and lower values (respectively) of the confidence interval in the schizophrenic group.*

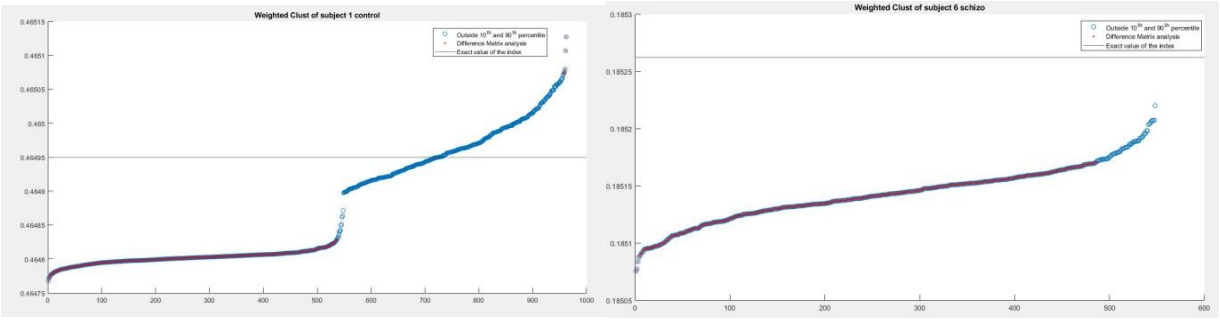


Figure 3 *Scatterplots of the Sensitivity analysis for the* ***Weighted Clustering Coefficient*** *for the first subject of the control group and the sixth subject of the schizophrenic one.*

Graph-Based Indexes	**Binary**			**Weighted**		
Standard	Control	Schizophrenic	p-value	Control	Schizophrenic	p-value
Clustering Coefficient	*Mean*: 0.486 ± 0.002 [0.486; 0.487] *Bias*: 0.409%	*Mean*: 0.418 ± 0.003 [0.417; 0.418] *Bias*: 1.182%	*BeforeBoot: 0.162* ***AfterBoot < 0.05***	*Mean*: 0.454 ± 0.003 [0.454; 0.455] *Bias*: 2.783%	*Mean*: 0.375 ± 0.003 [0.374; 0.375] *Bias*: 5.542%	*BeforeBoot: 0.156* ***AfterBoot < 0.05***
Degree/Strength	*Mean*: 71.750 ± 0.121 [71.741; 71.760] *Bias*: 0.002%	*Mean*: 69.211 ± 0.201 [69.201; 69.232] *Bias*: 0.009%	*BeforeBoot: 0.132* ***AfterBoot < 0.05***	*Mean*: 35.391 ± 0.154 [35.381; 35.402] *Bias*: 0.002%	*Mean*: 30.262 ± 0.185 [30.251; 30.270] *Bias*: 0.017%	*BeforeBoot: 0.147* ***AfterBoot < 0.05***
Path Length	*Mean*: 1.017 ± 0.002 [1.017; 1.018] *Bias*: 0.002%	*Mean*: 1.052 ± 0.003 [1.051; 1.052] *Bias*: 0.008%	*BeforeBoot: 0.132* ***AfterBoot < 0.05***	*Mean*: 0.208 ± 0.012 [0.207; 0.209] *Bias*: 31.85%	*Mean*: 0.108 ± 0.008 [0.107; 0.108] *Bias*: 49.03%	*BeforeBoot: 0.158* ***AfterBoot < 0.05***
Efficiency	Mean: 0.991 ± 8.265 * 10^{-4} [0.991; 0.992] Bias: 8.216 * 10^{-4} %	Mean: 0.974 ± 0.001 [0.974; 0.975] Bias: 0.004%	*BeforeBoot:* 0.132 ***AfterBoot < 0.05***	Mean: 0.525 ± 0.002 [0.524; 0.525] Bias: 4.578%	Mean: 0.484 ± 0.002 [0.483; 0.484] Bias: 8.497%	*BeforeBoot*: 0.158 ***AfterBoot < 0.05***
Modularity	Na	Na	Na	Mean: 1.647 ± 0.231 [1.633; 1.662]	Mean: 1.775 ± 0.312 [1.756; 1.794]	***BeforeBoot < 0.05*** ***AfterBoot < 0.05***

		Bias: 0.844%	Bias: 11.94%

Table 1 *Numerical results of the distribution of the indices of the two populations after the bootstrapping procedure applied in the* ***Standard case*** *(no normalization applied). Mean Values and Standard Deviation, Confidence Intervals within square brackets, p-values and biases percentages are shown. Na: Index not available for binary/weighted cases.*

Graph-Based Indexes	**Weighted**		
Normalized	Control	Schizophrenic	p-value
Clustering Coefficient	*Mean*: 0.410 ± 0.002 [0.409; 0.410] *Bias*: 2.148%	*Mean*: 0.329 ± 0.003 [0.328; 0.330] *Bias*: 4.697%	*BeforeBoot: 0.054* ***AfterBoot< 0.05***
Node Strength	*Mean*: 31.41 ± 0.156 [31.40; 31.42] *Bias*: 0.019%	*Mean*: 25.99 ± 0.177 [25.98; 26.00] *Bias*: 0.023%	***BeforeBoot < 0.05*** ***AfterBoot < 0.05***
Path Length	*Mean*: 0.272 ± 0.013 [0.271; 0.273] *Bias*: 16.06%	*Mean*: 0.133 ± 0.010 [0.133; 0.134] *Bias*: 39.34%	***BeforeBoot < 0.05*** ***AfterBoot < 0.05***
Efficiency	*Mean*: 0.460 ± 0.002 *Standard CI:* [0.459; 0.460] *Bias*: 3.064%	*Mean*: 0.410 ± 0.002 *Standard CI:* [0.409; 0.410] *Bias*: 6.389%	***BeforeBoot < 0.05*** ***AfterBoot < 0.05***
Modularity	*Mean:* 1.652 ± 0.263 [1.635; 1.668] *Bias*: 10.38%	*Mean:* 1.730 ± 0.352 [1.708; 1.752] *Bias*: 6.164%	*BeforeBoot: 0.103* ***AfterBoot < 0.05***

Table 2 *Numerical results of the distribution of the weighted indices of the two populations after the bootstrapping procedure applied in the* ***Normalized case.*** *Mean Values and Standard Deviation, Confidence Intervals within square brackets, p-values and biases percentages are shown.*

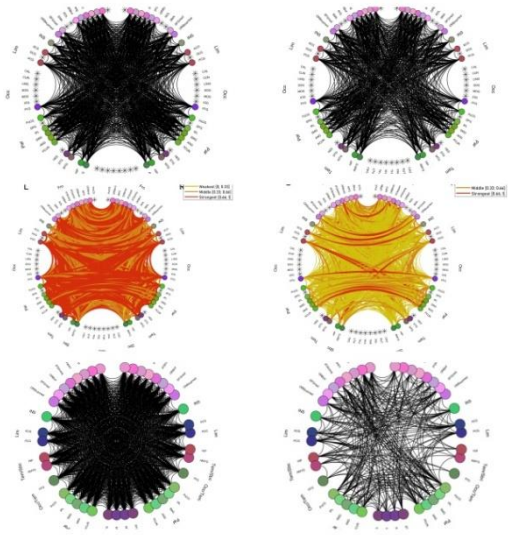


Figure 4 *Connectograms obtained by extracting the DMN with SNT from the sixth subjects in the control (left) and schizophrenic (right) datasets. In particular, the top panel shows the DMN subgraphs without thresholding, the middle panel shows the subgraphs with the connectivity weights shown, such that that red: 0.66 < weight < 1, orange: 0.33 < weight < 0.66, yellow: 0 < weight < 0.33. The bottom panel shows the connectograms of the DMN subnetworks whose weakest connections (weights < 0.2) were reset to evidence the disparity in terms of disconnections between the pathological and the control population.*

4. Discussion

In accordance with literature [3], it is concluded that the schizophrenic population is characterized by disrupted brain networks, which can be investigated in terms of graph indexes by decreased Clustering Coefficients, Efficiencies, Strength, Degree and higher Path Lengths and Modularity. These results were confirmed with and without normalization in Tables 1 and 2 with exception of the Weighted Path Length, greater in the control population. Then, statistically significant differences ($p < 0.05$) were found in the Modularity index (without normalization) and in Strength, Path Length and Efficiency (with normalization). Thus, the normalization with respect to the group allowed to find new differences. This first result highlights the importance of applying it for the problem of the great variability of the connectivity measures [1]. Regarding the bootstrap analysis, it is first worth noting that, in Figure 1, a bias of the distributions emerged. For this reason, a Sensitivity Analysis was performed to verify that the abovementioned problem of the variability of the weights also influences the bias. The results highlighted that: most of the weights that affect the

indexes most are the outlier weights, confirming the hypothesis that a normalization is needed; the previous result is greater for schizophrenics, thus expecting more influence in this case. The bootstrap application confirmed the latter, for all indexes as shown in Table 1 and 2. Considering the CI, both bootstrap methodologies granted to obtain appropriate boundaries in the ideal purpose of producing a normality range. The bootstrap brought advantages considering the p-values. As can be seen by the numerical values shown in the Table.1 and Table.2, it can be stated that the bootstrap is beneficial in the statistical investigation of the distribution of the graph metrics. Before the application of the bootstrap, instead, all indexes were not statistically significant except from the Modularity. These advantages are related to the possibility of having more robust values for all indexes, with respect to the original population, and to have aligned distribution, with respect to Standard Bootstrap, that can be affected by a bias. Furthermore, these results are in line with what obtained from healthy controls being promising for the investigation of normality ranges. Regarding the connectograms results shown in Figure 3 top panel, it can be seen that the disruptions and reduced activity of the DMN is not straightforwardly identifiable in the schizophrenic subject, as can also be seen by the numerical results of the mean densities of the two populations. In Figure 3 middle panel, it is possible to see that the schizophrenic DMN is characterized by a much weaker activation than the control one. In this context, the weakest connections (weights less than 0.2) were removed, and disruptions of some DMN regions are identifiable, as shown by Figure 3 bottom panel. Further investigations on the functional role of these regions can be performed in future analyses.

5. Conclusion

Considering the method proposed in this work to analyze graph indexes, it is possible to state that it represents a robust technique to investigate brain networks from the graph theory point of view, even considering the well-known limitations affecting the reliability of data processed from MRI techniques [1]. Indeed, several preprocessing features can be treated in different ways, bringing several differences in the reconstruction of fibers, such as stopping criteria, subject-motion or background artifacts removal methods. Further developments are related to validating bootstrap and sensitivity analysis as a tool to evaluate uncertainty also for other widely used metrics. In the context of the normalization of the data, a logarithmic normalization could be a development in the validation of the results on a single case study subject. In the context of the second-level analysis treated in this work, schizophrenia is a mental disorder characterized by altered perception, irregular emotion regulation, hallucinations and weakened working memory, while the DMN is a network deeply involved in social behavior, control of the emotional state of the individual [3]. SNT connectograms allowed to visualize the disconnections and the weaker activation of the DMN in a straightforward way, thus, validating the usefulness of this tool in connectomics. Furthermore, the analysis of the DMN graph indexes allowed to evidence the typical disruptions and disconnections from this point of view, which deserves further investigation. Future developments for this topic might be related to evidence specifically which brain regions and brain networks present more altered connectivity pathways.

6. Bibliography

[1] A. Fornito, A. Zalesky, and E. T. Bullmore, *Fundamentals of Brain Network Analysis*. 2016.

[2] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[3] Y. Liu *et al.*, "Disrupted small-world networks in schizophrenia," *Brain*, vol. 131, no. 4, pp. 945–961, 2008

[4] D. Coluzzi *et al.*, "Development and Testing of SPIDER-NET: An Interactive Tool for Brain Connectogram Visualization, Sub-Network Exploration and Graph Metrics Quantification," *Front. Neurosci.*, vol. 16, no. March, 2022

[5] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: Identifying differences in brain networks," *Neuroimage*, vol. 53, no. 4, pp. 1197–1207, 2010

List of Figures

List of Tables

Contents

Introduction

1.1 Aims of the thesis

The investigation of the networks of the brain has been an open issue for decades in the field of neurosciences; however, nowadays, the introduction of advanced magnetic resonance imaging techniques made possible to examine and quantify structural and functional connections based, respectively, on the fibers connecting brain areas and the correlation of their functional activity. The two approaches are indicated as Structural Connectivity (SC) and Functional Connectivity (FC), and lead to two complementary descriptions of the networks connecting brain areas. In this context, graph theory gives the possibility to represent, compare and investigate the human's brain anatomical and functional systems in a consistent and relatively uncomplicated way, such that it is strongly used to analyze data coming from neuroimaging techniques.

This thesis begins with the description of the basic concepts of brain connectivity, such as the definition of the different types from the different modalities, and the graph theory in the context of brain networks investigation. Next, the focus is put on the mapping of the brain in parcellation atlases, which represents a fundamental step for a flexible and straightforward analysis of the huge datasets coming from neuroimaging data. To complete the first introductive part, the physical principles undergoing magnetic resonance imaging are described, together with the advanced techniques such as Diffusion Weighted Imaging (DWI) and HARDI (High Angular Resolution Diffusion Imaging) used for SC. Moreover, the physical principles, advantages and limitations of fMRI used for FC are dealt with at the end of the introductive part.

Following this framework, attention is put on the definition of the brain connectivity matrix, which is, in graph theory, the fundamental point in the description of the network metrics, which were used thoroughly in this work. Thereafter, a novel software for brain network and sub-network mapping and visualization is introduced, called Spider-Net (Software Package Ideal for Deriving Enhanced Representations of NETworks) [4]. This software is the result of a collaboration between Politecnico di Milano, Milano (IT) and CADiTeR, MRI

Laboratory of Fondazione Don Carlo Gnocchi, Milano (IT). The aim was to develop a software capable of exploring the brain networks in an undemanding way, which is fundamental in the clinical research context.

The investigation of changes in structural and functional connectivity based on graph indexes is a powerful tool in the clinical context for the analysis of brain organization, mostly from a statistical point of view, to compare groups of subjects characterized by the different conditions or an individual vs. the normality ranges. However, there are some issues to fulfill this purpose, such as the sensitivity to noise of the neuroimaging techniques and some limitations in the preprocessing techniques. In this context, the focus in this work is put on the analysis of the uncertainty through the Bootstrap technique which allows to estimate the statistical features of a population characterized by the same condition from a limited number of measurement samples. This procedure is based on a random sampling of the initial dataset to create "surrogate" data which can better explain the statistical properties of the population under analysis. Indeed, the bootstrap procedure was applied in this work to allow a comparison between the different network metrics extracted from different brain connectivity datasets taking into account a certain degree of uncertainty.

Moreover, the Spider-Net software was used to investigate the advantages and disadvantages in terms of the detection of brain connectivity pathways with two different advanced magnetic resonance imaging techniques, such as (Diffusion Tensor Imaging (DTI) and HARDI (High Angular Resolution Diffusion Imaging).

Eventually, both these approaches were used in the investigation of a dataset involving healthy control and schizophrenic subjects. In particular, bootstrapping was applied to analyze the differences in terms of network metrics in the two populations. Then, the Spider-Net software was used to inspect and visualize the brain connectivity dissimilarities in the whole-brain connectivity matrix and in sub-networks of interest, such as the Default Mode Network (DMN). This is a functional network known for being active when a person is not focused on the outside world and the brain is at wakeful rest, besides, the network is also activated for emotion control and recollecting past memories. From different studies, it was clear that this circuit is altered in schizophrenia, pathology characterized by disturbs of behavior, hallucinations, and cognitive deficits [6], [3], [7], [8]; thus justifying the aim of this work in investigating the variability of the schizophrenic subjects connectivity pathways.

1.2 Brain Networks

The brain is the most tangled network known to mankind. Indeed, it comprises billions of neurons which form trillions of synapses between each other. The collective actions of individual nerve cells linked organized in networks of connections makes behavior, thought, memory and consciousness possible. In this context, clinical disorders, like dementia or schizophrenia, are characterized by an alteration in the connections' paths. It is, thus, understandable that neurosciences have been focused on the topic of brain network connectivity in the last decades. However, neuroscientists cannot fully comprehend the brain functions without a multiple scale approach. This is based on classifying the networks which organize neurons in functional brain regions and associate brain and body in a thorough organism. Thus, the comprehension of network interactions across various levels of organization is fundamental for a more complete understanding of the brain as an integrated system. This utmost wide and complex field of research is called connectomics and addresses the brain connectome at all its scales and features. Within this framework, the introduction of the concept SC and FC, addressing connection at the scale of brain areas to describe the respective connectivity matrices and represent all possible pairwise connections between the parcellation of areas in the brain, is a huge simplification but also an essential scale of brain areas classified by their anatomical position and function in well-established atlases is in the order of 100 parcellations, each representing a connectivity matrix entry, or graph node. This concept thus relies on an extremely simplified representation of the cellular wiring and organization of the brain, yet supported by the current knowledge of well localized brain areas with specific functions. In particular, *Fig.1 (left panels)* shows how a gray matter (GM) parcellation atlas *(top)* is projected onto a subject's brain *(bottom)* after a co-registration.

Passing from the definition of graph nodes to the weight of graph edges, different features are considered in SC compared to FC.

In particular, SC considers the links established between GM areas by the fibers travelling in the brain white matter (WM). The procedure for edge mapping is shown in *Fig.1 (right panels).* WM tractography *(right, bottom panel)* consists in extracting all possible virtual fibers (stream-lines) given the diffusion weighted images (DWI) of the brain in many directions. The resulting pattern of stream-lines is very complex and imbricated, nonetheless, it permits the extraction of neural fiber tracts (alias, fascicles or bundles) based on anatomical sorting rules. In the case of SC matrix extraction, the rule for each edge is to select all stream-lines touching at their ends the two connected areas *(right, top panel)*. Weights in the SC

matrix (*Fig.1, central bottom panel*) are finally derived based on quantitative and/or qualitative properties of the connecting streamlines.

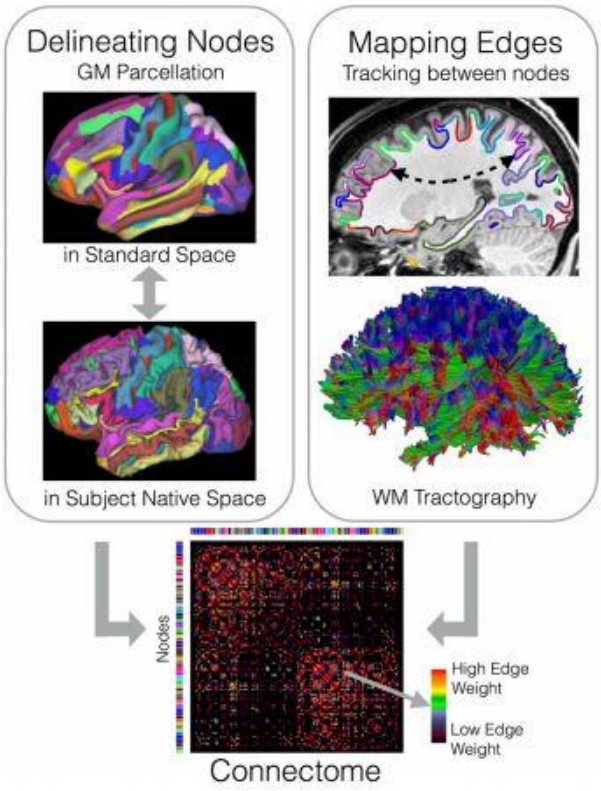


Figure 1 *Creation of the connectome. On the left, the definition of the nodes of the matrix representing the brain parcellations is depicted. On the right, the mapping of the edges of the matrix representing anatomical connections extracted through white matter tractography is shown* [9].

Passing to the FC matrix and graph, the definition of GM areas as network nodes it exactly the same shown in *Fig.1 (left panels)* for SC. Concerning the weights of graph edges, the functional interactions between areas can be obtained by the correlation of the spontaneous dynamics of activity in brain areas. The cortical activity changes in time can be accurately mapped by the functional magnetic resonance imaging (fMRI), which indirectly derives it by the neurovascular coupling, implying higher oxygenation in phases of higher activity. Since the underlying hemodynamic response function has a duration of about 10 s, fMRI can follow only slow fluctuations. Noninvasive electrophysiological signals such as EEG (electroenchephalography) and MEG (magnetoencephalography) conversely provide faster dynamic features and alternative ways to study FC [9].

1.2.1 Graph theory for the analysis of brain networks

Complex networks have topological properties that show a more or less elaborate organization, neither random nor regular. These systems range from societies to molecular relationships between organisms, and their complexity stands on the fact that they are represented by millions and millions of elements interacting with each other. Different methods were developed to deal with this type of data, such that it became mathematically appealing to investigate the properties of the complex networks in terms of their organization. In this context, a modern field of network analysis was borne, based on representing complex networks and investigating their organization and function: graph theory [1].

The first creation of a graph used to understand a real-world system is related to the mathematician Leonard Euler, who lived in a city surrounded by seven bridges linking the four main lands of the urban area. The problem at the time was about the possibility to walk around the town with a pathway which crossed each of the seven bridge only once. Euler solved this issue by representing the four land masses as nodes and the seven bridges as edges as represented in *Fig.2*. From this graph, it was shown that it was impossible to find a route around the city which crossed each bridge only once [10].

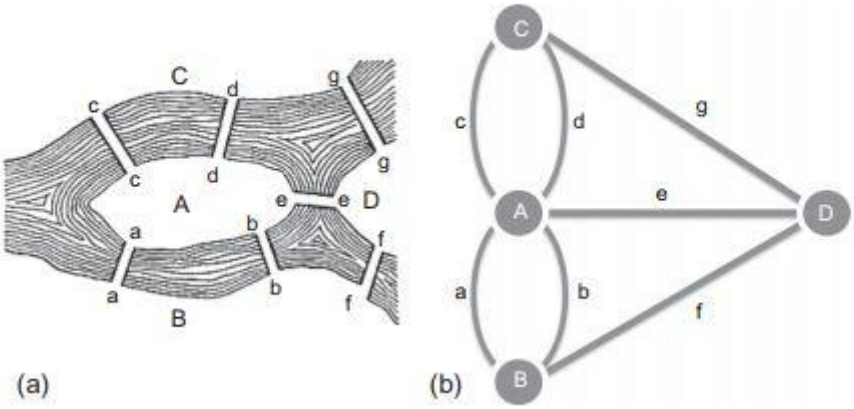


Figure 2 *The origins of graph theory: (a) Simplified geographical map of the Prussian city where Euler lived, characterized by four landmasses marked from A to D and linked by seven bridges marked from a to f. The problem was related to the possibility of finding a pathway which allowed to cross each bridge only once. (b) Graphical representation of the issue described by Euler's binary graph where nodes represent the landmasses and edges representing bridges* [10].

In general, a graph is a mathematical representation of a real-world complex system and can be defined as a collection of nodes (vertices) and links (edges) between pairs of nodes [2]. Thus, nodes and edges are the fundamental elements of brain networks and their definition is cardinal for a graph theoretical approach of complex networks [11]. Focusing on brain networks, what are nodes and edges in the nervous system? The answer to this question is dependent on which scale was used to represent the brain. Moreover, the definition of nodes and edges in also influenced by the chosen measurement technique. Indeed, for example, a graph constructed from electrophysiology can be analyzed in the same way as a graph constructed from functional magnetic resonance imaging. Nonetheless, the nodes and edges derived from these two techniques do not refer to the same biological processes.

Generally speaking, connectomics distinguishes between three spatial scales to define brain connectivity: microscopic, mesoscopic and macroscopic [1]. The brain connectivity at the microscopic scale is characterized by neurons interconnected by synapses which are represented as nodes of the graph, while axonal projections and synapses are the edges of the graph. The brain connectivity at the macroscopic scale analyses white matter tracts (edges) interconnecting confined cortical areas (nodes) across the whole brain. The mesoscopic scale acts as a bridge between macroscopic and microscopic. In particular, at this scale, the investigations address local networks among and within a limited region of brain areas. The integration of the various scale is one of the most ambitious goals for neuroscience for many years to come. The analyses considered in the present study are limited to the macroscale of brain parcellation into distinct areas.

1.2.2 Macroscale connectomics

Nowadays, the macroscale techniques are the most appropriate for mapping the human connectome with cognitive and behavioral associations [12]. Indeed, the interpretation of macroscale measurements is the easiest way to detect lesions and to conduct brain imaging studies at the scale of structures (GM areas, WM tracts, etc.). Furthermore, the imaging techniques for the measurement of the macroscale connectomics, MRI (Magnetic Resonance Imaging), EEG (Electroencephalography) and MEG (Magnetoencephalography), are non-invasive for in vivo studying of the human brain networks with respect to the same aim at the microscale. In particular, the MEG measuring system in shown in *Fig.3(a)* together with the frequency bands (*Fig.3(b)*) used to identify the different aggregation of neurons, forming the networks (*Fig.3(c)*). The benefits of these methods are based on their

safety from the clinical point of view and the possibility of analyzing the whole-brain connectivity across different pathological or healthy subjects or across different periods of time [1]. These advantages allow to investigate the evolution and the changes at the macroscopic scale. Nevertheless, these techniques can only distinguish elements on the scale of millimeters and centimeters which corresponds to the aggregation of large populations of neurons, axons and synapses. This assortment obviously decreases the precision with which it is possible to delineate nodes and edges of the brain networks.

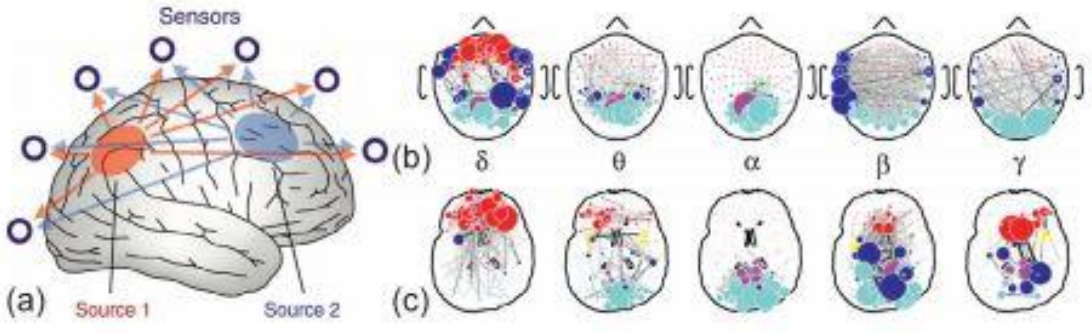
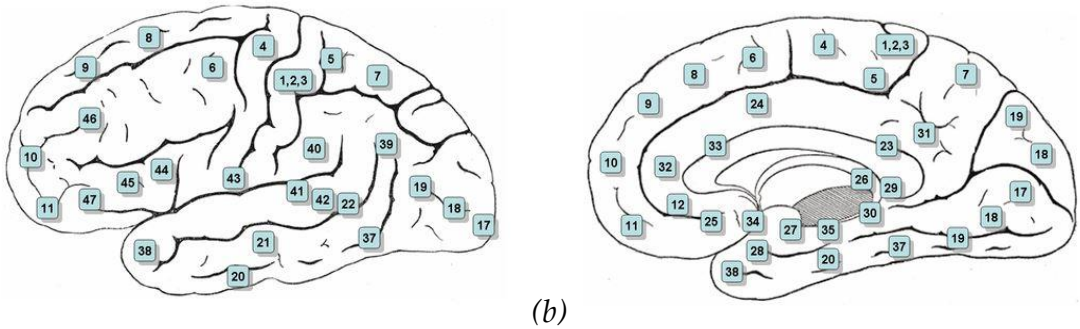


Figure 3 *Macroscale functional connectivity networks measured with MEG: (a) representation of how the sensors are used outside the head of the patient to detect the electrical and magnetic activities of neurons. (b) MEG frequency bands. In particular, the colors are related to the different anatomical divisions, mainly related to the lobes. (c) Functional connectivity networks represented with respect to the anatomical parcellations used to estimate the signals coming from the different sources* [1].

Considering the methods used for macro-connectomics, magnetic resonance imaging (MRI) is dominant mainly because of its safety, spatial resolution and availability throughout the world. Moreover, MRI allows to investigate structural and functional features at the macroscale with diffusion-weighted magnetic resonance imaging (DWI) and functional MRI (fMRI), respectively. The former allows to visualize and examine the organization, the orientation and the trajectories of white matter tracts. The latter inspects the dynamics of activity in each GM area, thus, in resting state condition (i.e. with no stimulus) it can provide the correlation between the spontaneous activations/inhibitions between each pair of areas.

The mapping of a connectome at the macroscale is based on the problem of the definition of nodes [9]. Historically, in humans, the first attempt in classifying the cerebral cortex into discrete parcels or areas was represented by the Brodmann areas, defined by the homonymous neuroscientist in 1909 [13], based on the cytoarchitectural organization of neurons, where the lateral and medial view's numerical organizations are shown in *Fig.4*.



(a) *(b)*

Figure 4 *Lateral (a) and mesial (b) view and numerical categorization of the Brodmann's areas*

However, the Brodmann's parcellation scheme introduces some limitations related to not taking into account the variability of the cytoarchitectonic boundaries across individuals. In this context, information additional to cytoarchitecture was introduced thanks to the evolution and the increasing precision available nowadays with magnetic resonance imaging, which is behind the development of actual atlases (see next Section 1.2.3.). Still these atlases refer their improved parcellations to the numeration of areas introduced by Brodmann.

1.2.3 Brain Atlas Segmentation

During the last decades, it was felt the need of flexible and suitable techniques aimed at investigating and dividing appropriately big data related to brain acquisitions. Within this framework, it is possible to define the network's nodes based on the macroscopic landmarks visible with MRI, such as sulci or gyri [14] [15]. Nonetheless, the parcels obtained with this kind of approach can be characterized by variability in size which can alter the analysis of the connectome. Moreover, the anatomical atlases may not establish a match between these kinds of landmarks and functional boundaries [16]. In this context, functional maps are a possible solution to these issues. The functional atlases are, usually, derived from multi-modal data coming from resting-state fMRI, task-based fMRI or diffusion weighted MRI to acquire a whole-brain parcellation scheme characterized by considering different features of the brain's organization. For example, Glasser and colleagues, in their study, fuse information coming from myelin content, cortical folding, rs-fMRI and task-fMRI to obtain an atlas which is properly able to specify the cortical properties [17]. The main atlases in the literature and those related to the present work are described in the following.

An example is provided by the Talaraich atlas, [18], which is a volume-filling, hierarchical, anatomical labelling scheme where anatomical structures are defined as a collection of voxels, each voxel described by x-y-z coordinates. Standardized 3D coordinates describe with precision the locations of the anatomical labels within the atlas. The Talaraich atlas labels are defined with a subdivision of a structure at level N into its substructures of level N-1. In particular, the brain is subdivided into its volumetric structures, cerebrum, cerebellum and brainstem, at level 1 as shown in *Fig.5*. Besides, the cerebrum is subdivided into lobes and sublobar regions at level 2 and so on for every structure until level 5 for this atlas. This type of segmentation, however, cannot be performed with complete accuracy since a surrogate atlas with characteristics similar to the subject of interest would be needed to have a more correct labelling. Most clinicians use the Talaraich atlas to investigate the localization of the brain regions which are activated or deactivated in different conditions, thus, in functional connectivity studies [19]. Nonetheless, assigning a set of coordinates to anatomical labels can be inaccurate due to many ambiguities when the objective is to define a set of coordinates between different brain areas.

Hierarchy level	Structure label							
Hemisphere (level 1)	Cerebrum (R/L)							
Lobe (level 2)	Lobes				Sublobar			
Gyrus (level 3)	Gyri		Subgyral		Nuclei		Extranuclear	
Tissue type (level 4)	GM	WM	WM	CSF	GM	WM	WM	CSF
Cytoarchitecture (level 5)	BA	Tract	Tract	Space	Subnuclear	Tract	Tract	Space

Figure 5 *The anatomical structure-naming used for organizing the anatomical regions based on volume occupancy. In particular, the volumes of interest are organized into five hierarchical levels: Hemisphere, Lobe, Gyrus, Tissue type and Cytoarchitecture. BA: Brodmann Area; WM: White Matter; GM: Gray Matter; CSF: CerebroSpinal Fluid* [18].

This kind of procedure was improved by the definition of an automatic labelling of the activation of the different brain regions to classify from a hierarchical point of view the different Brodmann's areas. Indeed, a more precise classification requires the use of automatic algorithms. The purpose of the study produced by Tzourio-Mazoyer and colleagues [19] is to present an automated anatomical labelling of activations detected with fMRI which does not aim at resolving the inter-individual anatomical differences, which are about 9 to 18 mm depending on the brain regions considered. Indeed, the inter-individual variability does not provide the absolute anatomical localization that can only be obtained with a reference to the particular anatomical map. Preferably, the work suppresses the

ambiguity related to the relationship between a set of coordinates and the corresponding anatomical labelling. This, indeed, represents a crucial point, since the strength of a brain atlas is to provide a common system to be used in functional connectivity studies reporting the localization of the activation of the different brain regions. This automatic anatomical labelling (AAL) atlas is based on the main sulci used as landmarks for the 3D definition of 45 anatomical regions of interest for each hemisphere. The procedure involves, firstly, a software which divides the brain according to the sulci, then the regions of interest were manually drawn with the same software every 2 mm on the axial slices; eventually, each of the total 90 regions are described with a label [19].

In the development of brain atlases, two different perspectives are considered.

Firstly, the approach involves the selection and identification of a group of individuals with the same characteristics, such that the atlas built in this way is specific for the particular group. On one hand, this method achieves great accuracy, on the other hand, it has problems from a practical point of view in the assessment of different properties across populations.

The second method is based on the development of a more general atlas which is are less accurate within a single population but applicable across different groups. In the creation of the structural Desikan atlas [14], each hemisphere is divided in 34 regions. The labelling procedure is based on a sulcal approach by a manual tracing from one sulcus to another to define the different labels. This kind of procedure was performed with respect to different sources of information such as neuroanatomical conventions or modifications of previous studies. These were used to define the region-of-interest on the images, that are then transposed onto the inflated cortical surface of the reconstructed brain. The inflated surface allows the representation of the cortical surface with respect to sulci and gyri. The localization of sulco-gyral structures of the human cerebral cortex is important for the description of structural/morphological data. This type of labeling is potentially useful for delineating regions-of-interest on the anatomical images acquired on the particular patients.

However, this kind of labelling is time consuming due to the complexity of the anatomical brain structures and it also requires important experience by clinicians. To solve this issue, automatic techniques were improved substantially to produce a rapid and precise reconstruction of the human brain. In this context, the study provided by Destrieux et al. describes a complete parcellation of the cortical surface using standard nomenclature and criteria, available in the FreeSurface package software [15].

The procedure involves first a computer assisted hand sulco-gyral parcellation subdividing the brain into 84 labels per hemisphere. In this way, the cortex is divided into sulcal and gyral structures depending on the values of the local mean curvature and the mean convexity from the reconstructed output of the FreeSurfer software; in particular, structures with average convexity value below a given threshold are considered as sulci, while vertices above or equal to this threshold are considered as gyri. Secondly, after the classification of the cortical surface in gyri or sulci, the limits between adjacent gyri and sulci are drawn manually using tools included in the FreeSurface software. Eventually, each cortical structure is associated with a label chosen from the database name, with particular large structures divided in further sub-parcellations, such as the cingulate gyrus which is divided into middle-anterior, middle-posterior, posterior-dorsal, posterior-ventral parts. After this whole procedure, each structure of the cortical surface is assigned to an anatomical label [15].

Finally, it is possible to state that the problem of the definition of the connectome at the mascroscale and the construction of brain atlases is critical. Indeed, if clinicians cannot be sure that the nodes are matched with respect to appropriate parcels of the brain, it is difficult to determine if the results coming from the investigation on the connectome are meaningful or ill-posed. Often, it is useful to repeat the same analysis across different kinds of parcellations to check for the repeatability and steadiness of the results.

1.2.4 Integration/Segregation Paradigm and Definition of Functional Networks

In the investigation of brain networks, it is important to highlight the fact that the brain is characterized by functional dynamical changes to support different tasks execution. In particular, the brain reconfiguration is related to the specific task to be performed and relies on either independent specialized subsystems (segregation) or cooperation between subsystems (integration), as shown in Fig.6 exhibiting the concept of integration and segregation of nodes on a networks. [20].

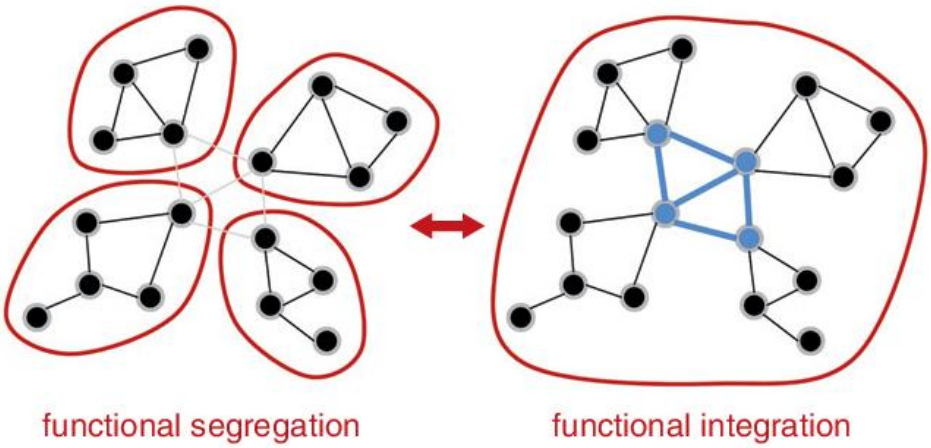


Figure 6 *Segregation (left) and integration (right) paradigms of nodes interconnected to form brain networks*

The brain activity at rest is supposed to balance segregation and integration demands, which need to be explicitly defined and quantified [20]. In this context, the knowledge of some structural network topological features, such as node clustering or path length, provides information about the functional segregation and integration of network interactions [21]. Moreover, combining the knowledge coming from these two topological features, it is possible to define the small-worldness, a dimensionless parameter to be integrated with the anatomy of the brain to have some kind of information about the activation of functional networks. It was studied that the small-world topology of brain networks is efficiently combined with physical distances, defined as *economical model* embedded in physical space [22], [23]. In particular, the edges of a brain network characterized by high clustering tend to be at short anatomical distance (low path length), while the edges at further distance (high path length) tend to be low clustered [24]. In this context, it is possible to define the brain as an economical small-world network, such that the term "economical" is related to the concept of "value for money" [25]. In particular, nodes characterized by high clustering are anatomically close to each other, therefore minimizing the wiring cost (path length). However, this minimization of the wiring cost, although maximizing the small-worldness of the network, makes the network to lose its integrative capacity which is increased in case of an increased path length, thus, reducing the small-worldness of the brain network [24]. The economical idea is related to the concept that brain networks find a compromise between the minimization of the wiring cost (low path length, high clustering coefficient, therefore increased small-worldness) and the maximization of the integration between nodes (high path length, low clustering coefficient, therefore, decreased small-worldness). Thus, the

brain networks can be generated by economical models for different ranges of the two parameters controlling the wiring cost and the topological integration [24].

In the context of the characterization of the functional brain networks which are activated during different tasks, it was first reported by Mesulam [26] the division in 5 core functional brain networks: a spatial attention network; a language network anchored in Wernicke's and Broca's areas; a memory network anchored in the hippocampal-entorhinal complex and inferior parietal cortex; a face-object recognition network and a working memory-executive functional network [21]. Thus, anatomical areas and functional circuits are activated during tasks performed to complete complex cognitive functions such as speech, language, visual processing and sensorimotor actions. However, the integration between these systems represents still an open issue in formulating how these circuits cooperate. In particular, the issue is related to how consciousness is formed in the individual and studies in literature are focused on this topic such as the one developed by Casali and colleagues based on determining an index of consciousness independent from sensory processing or behavior [27]; or the one developed by Rosanova and colleagues based on recovering effective connectivity pathways to detect and track recovery of consciousness in brain-injured patients who are unable to exchange information with the external environment [28].

The brain regions forming these networks have been detected from fMRI activations during specific tasks. Separately from the different tasks, the functional organization of the brain at rest can show specific patterns of activation whose difference with respect to the patterns activated during tasks allows to understand cognitive performance. The functional brain networks are in close correspondence in the analysis of resting and task-related connectivity patterns, suggesting that these networks are coupled at rest and during cognition. This allowed to put the focus on brain networks which are activated during rest or social cognitive and deactivated during cognitively demanding tasks, an example of which represented by the Default Mode Network (DMN) [21]. The DMN was introduced in 2001 by the neurologist ME Raichle by referring to the state in which the brain is not activated, thus, not involved in any task requiring some kind of skill [29]. Structurally, the DMN is formed by three brain's zones, defined as the ventral medial prefrontal cortex, the dorsal medial prefrontal cortex and the posterior cingulate cortex, together with the entorhinal cortex and the lateral parietal cortex [30]. In particular, the ventral medial prefrontal cortex receives the information coming from the senses and the body, thus, it is involved in social behavior and control of the emotional state of the individual. The dorsal medial prefrontal cortex, instead, is associated with self-reference and judgement, while the other

regions involved in this network are generally referred to the recollection of past memories and the experiences involved during wakefulness and sleep.

1.3 Magnetic Resonance Imaging (MRI)

For the largest part of history, brain anatomy and functions were considered as unresolved, hidden beneath the cranium and suitable only for post-mortem studies [31]. The evolution of new methodologies and tools uncovered organizational and operational features of the brain. In particular, anatomical approaches unveiled the pattern of connectivity between brain cells and regions, while physiological techniques yielded functional schemes, such as sensory and motor responses maps. Nowadays, brain activity can be studied using different methods, each characterized by its own arrangement considering spatial and temporal resolution and invasiveness.

Magnetic resonance is a measurement technique used to investigate atoms and molecules based on the relationship between a created magnetic field and a particle characterized by spin and charge [1]. The Magnetic Resonance Imaging (MRI) focuses on the nuclei of particles, with a phenomenon known as Nuclear Magnetic Resonance (NMR), experimentally determined by Purcell in 1946 [32]. As the name states, NMR implies the resonance phenomenon of the nuclei and the external magnetic field. The NMR requires a magnet for a very high and homogeneous main field B_0 (up to 3 Tesla, in current clinical MRIs). The body is inserted into a bore surrounded by a powerful coil (superconductive for high fields such as 1.5 or 3 T) which generated B_0 [33]. Under the magnetic field, the body nuclear spins are in one of the two possible energy states: low energy state ("spin-up", parallel to B_0), or high energy state ("spin-up", antiparallel to B_0). The atomic nuclei all include protons and neutrons and are characterized by a net positive charge [34]. Some atomic nuclei, for example the hydrogen $_1H$ or the phosphorus $_{31}P$, show the "spin" property based on the number of protons which can be represented as a spinning of the nucleus around its own axis at a constant rate. The rotation of the nucleus positive charge (atomic number multiplied by the proton charge: $N \cdot e^+$) creates a dipole field μ. Normally, all dipoles are randomly directed and the net magnetization M is zero (*Fig.7 (a)*). However, when the nuclei are inside the strong magnetic field B_0, the prevalence of the spin-up status creates a net magnetization M directed as B_0 (*Fig.7 (c)*). Nonetheless, their angular momentum J, due to the spinning, impedes the alignment due to the gyroscopic

effect (*Fig.7 (b)*). Eventually, they get a precession motion around the axis of B_0 at the Larmor frequency [33]:

$$\omega = \gamma \cdot B_0 \tag{1}$$

where γ is defined as the gyromagnetic ratio, calculated as:

$$\gamma = \frac{\mu}{J} \tag{2}$$

thus the ratio of the magnetic dipole and the angular momentum.

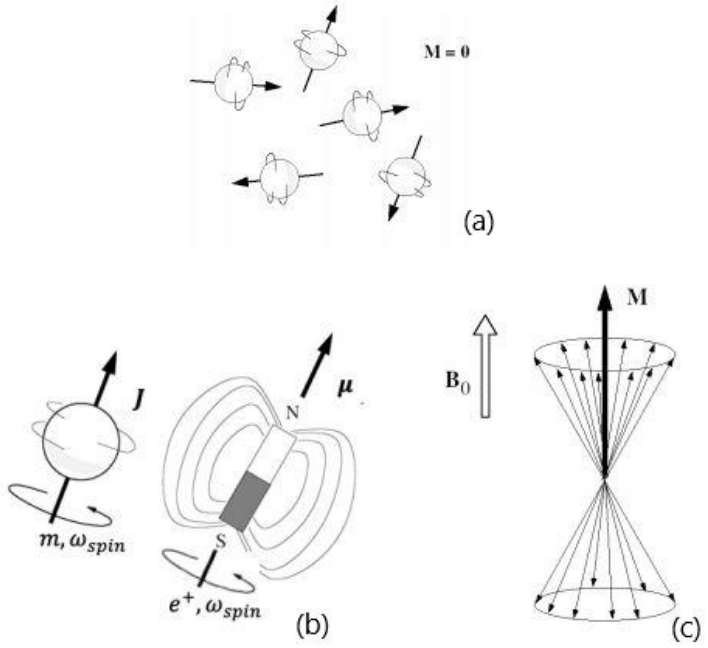


Figure 7 *(a): initial null net magnetization due to the random orientation of the dipoles. (b): angular momentum J of the spins and representation of the magnetic dipole μ. (c): when the dipoles are inserted into a strong magnetic field B_0, they align their axes producing the magnetization vector M* [33]

A limited number of values for the spin are available in nature, such that the spin number , I, is quantized to certain discrete values. There are only three groups of values for the spin number I: zero (indicating no spin), integer and half-integer. The hydrogen nucleus ($_1$H), consisting of a single proton, is the most used one in MRI thanks to its half integer spin and due to the fact that its response to an external magnetic field is one of the largest found in nature. Moreover, body tissues are mostly formed by water and fat which both contain $_1$H [33]. I.e., they have a high proton density, where "proton" stands for the $_1$H nucleus.

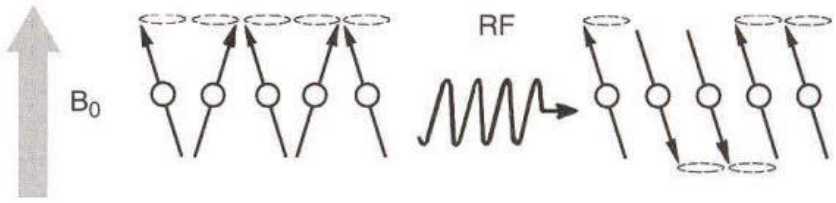
Nuclei that are characterized by a spin number different from zero can be excited inside the strong magnetic field B_0 thanks to the application of a second field B_1 perpendicular to B_0, activated at short radiofrequency (RF) pulses. For a large collection of protons inside a certain volume of tissue, the effect is absorption and emission, happening during the RF pulse and depending on the state of the proton. Nonetheless, since there are usually more protons in the lower energy state with respect to the higher one, the main effect is the absorption of energy by the tissue. The orientation difference between B_0 and B_1 allows the coupling between the RF pulse and the magnetization vector M_0, thus, the energy is conveyed to the protons, as shown in *Fig.8(a)*. The macroscopic effect is that of moving the magnetization vector by a flip angle α proportional to the RF pulse duration τ and amplitude B_1 [35]:

$$\alpha = \gamma \cdot \tau \cdot B_1 \tag{3}$$

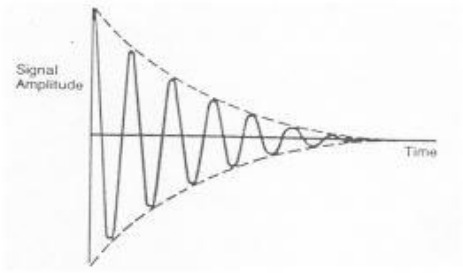
This produces the development of a transverse magnetization component:

$$M_{xy} = M_0 \cdot \sin \alpha \tag{4}$$

which is the only measurable effect since it is a dipole rotating at RF, thus, transmitting out a RF signal consisting in the observed NMR, also called Free Induction Decay (FID), exhibited in *Fig.8 (b)*.



(a)



(b)

Figure 8 *(a) Coupling between the RF pulse and the magnetization vector which moves the vector proportionally with respect to the flip angle α. (b) Representation of the FID signal* [35].

Magnetic resonance imaging is not only caused by the dissimilarities in proton density, but also by differences in nuclear processes also known as relaxation. Indeed, in the behavior of spins, two relaxation mechanisms are noticed: the longitudinal relaxation with a time constant T_1 explained by the spin-to-lattice interactions, which absorb the spin-down excess of energy and converts it into heat; and the transverse relaxation with time constant T_2 (T_2^*) explained by the spin-to-spin interactions.

1.3.1 Diffusion Weighted Imaging (DWI)

Diffusion weighted imaging (DWI) has been employed since the 1990s to investigate the possibility of early stroke and other neurological disorders [36]. One of the most important advantages of this methodology is that it uses the intrinsic contrast between tissues without the introduction of any exogenous contrast material. In this context, within the last decade, DWI was continuously improved to become a routine clinical application in ischemia and is also the methodology used in research in other neurological diseases such as multiple sclerosis, dyslexia and schizophrenia [37]. Moreover, there are examples of DWI studies on the liver, the kidneys and the lymph nodes [38], [39]. Considering the physical background of DWI, it refers to the random (Brownian) motion of water molecules in a fluid that results from the thermal energy carried by these molecules [40].

The amount of diffusion depends on the diffusion coefficient, D. In particular, in a homogenous mean, the diffusion coefficient is the same in every direction, or isotropic. Conversely, in some biological tissues (e.g. neural fibers), the diffusion coefficient is different, or anisotropic. The diffusion coefficient is expressed in terms of $\frac{mm^2}{s}$ and describes the mean displacement, d, with respect to the motion within a time, τ, and it refers to the standard deviation of the position, such that [41]:

$$d = \sqrt{6 \cdot \tau \cdot D} \tag{5}$$

The apparent diffusion coefficient (ADC) of water molecules in tissues measured by DWI is much lower than that in free water due to hindering and restriction created by the cellular structures and macromolecules. Importantly, the ADC is related to the average microstructure within a voxel and, in anisotropic tissues, also to the explored direction. So, DTI, which will be discussed later in the paragraph 1.3.2, measures water molecule diffusion in different directions in every

pixel of a magnetic resonance image, and it is derived by performing DWI in more than six non-collinear directions.

As stated in the beginning of paragraph 1.3.1, the NMR phenomenon combined with magnetic field gradients grants to build images characterized by different types of contrasts. Namely, DWI requires to insert in a sequence with a long echo-time (TE), hence T_2-weighted, a bipolar gradient made by a dephasing and a rephasing gradient pulse separated by a long (20-30 ms) diffusion time $\tau = \Delta$. This permit to evaluate the amount of diffusion motion of the protons (mainly, in water molecules) along the direction of the bipolar gradient [42]. The most frequent technique used for DWI is by means of a single shot readout of an entire slice called echo-planar-imaging (EPI). The need to provide a consistent diffusion time Δ is satisfied by the EPI readout applied over a spin echo with a long echo time ($TE > \Delta$), which causes a T_2-weight to be combined with the diffusion one [36]. Diffusion weighing is introduced by a bipolar gradient made of two opposite gradient pulses: the former dephases all spins proportionally to their position along the gradient direction, while the latter, delivered after the diffusion time Δ, rephases them.

However, the random diffusion displacements during the diffusion time Δ cause the moved protons to experience a rephasing different from the initial dephasing. The resulting dispersion of spin phases produce an attenuation of the magnetic resonance signal, indicating the level of ADC.

With the application of the diffusion gradients in a single spatial direction, only the motion due to diffusion parallel to this direction would be detected. On one hand, this investigation is sufficient to acquire the diffusion properties in an isotropic medium. On the other hand, muscles or nerves are highly anisotropic tissues and the diffusion will be lower in the direction orthogonal to fibers compared to the parallel one, resulting in a decreased apparent diffusion coefficient [37]. This issue detected can be solved with the diffusion tensor which represents the diffusion no longer with a single scalar, but with a matrix (described later in paragraph 1.6.3).

The degree of signal attenuation can be calculated as:

$$\mathrm{SI} = SI_0 \cdot e^{-b \cdot ADC} \tag{6}$$

where, SI_0 is the signal intensity of the T_2-weighted image with no diffusion gradient (i.e., with b = 0) applied and b is the degree of diffusion weighting (b-value):

$$b = \gamma^2 \cdot \delta^2 \cdot G^2 \, (\Delta - \frac{\delta}{3})$$

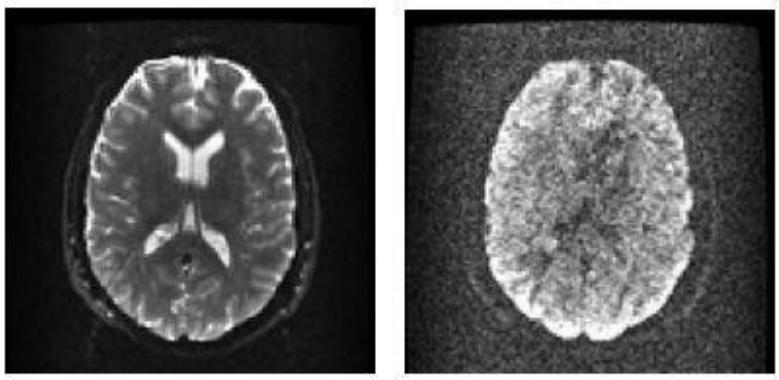
where γ is the gyromagnetic ratio, while δ and G are the duration and size of the gradient pulses.

This value, in particular, tunes the diffusion attenuation and it is measured in $\frac{s}{mm^2}$. In the context of imaging it is fundamental to set the b-value properly [43], since different values of b are considered to have the most accurate diffusion coefficient able to correctly characterize different kinds of tissues with sufficient SNR, as shown in *Fig.9*. Describing the role of the b-value in acquiring images, it is important to realize that the higher its value, the more the water molecules are sensitive to their molecular displacement, thus, increasing the signal attenuation, , as shown in *Fig.9*, but also increasing the scanning time due to the fact that in the clinical practice the only way to increase the b-value is to raise the diffusion time. The most used values of b range from 600 to 1500 $s \cdot mm^{-2}$, where the lower values in this interval are typically used for the imaging of premature babies, while the higher values are commonly used for stroke patients. [44].

The ADC can be evaluated by solving the previous equation with respect to this term, such that [44]:

$$ADC = \frac{ln\left(\frac{Si}{S_0}\right)}{b} \tag{7}$$

In the clinical practice, the final image, characterized by different ADC for each pixel, is referred to an ADC map [36]. In anisotropic diffusion the molecular mobility is not equal in all directions producing distinct ADCs in the different directions.



$b = 0$ s/mm^2 $b = 700$ s/mm^2

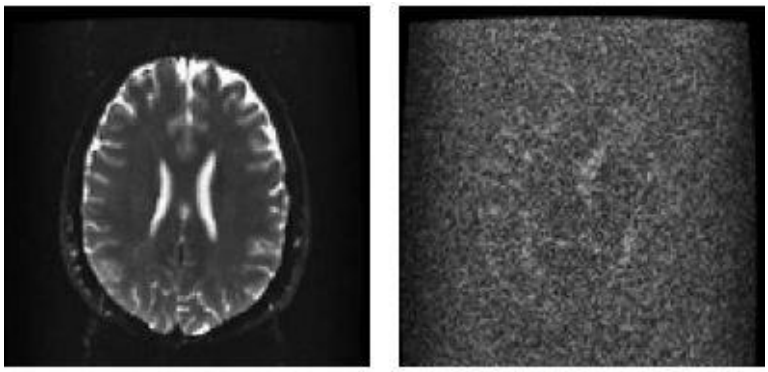


Figure 9 *Diffusion Weighted Images for different b-values. It is important to appropriately tune the b-value to avoid, on one hand, a very low signal attenuation when b is too low or a reduced signal-to-noise ratio when b is too high*[45] .

1.3.2 Diffusion Tensor Imaging (DTI)

As discussed in the previous section (1.3.1), the variations in the diffusion-weighted signal intensity related to anisotropy of white matter tracts complicates the interpretation of the DW images in the clinical practice, except if ADC maps are computed [44]. The anisotropic motion of water molecules was found to be much faster along the white matter pathways than the perpendicular direction with respect to them [46]. The problem is solved by composing DWIs in several direction and describing anisotropy by a proper model. The simplest possible model maintains the Gaussian nature of free diffusion but with an anisotropic structure described by the principal axis (eigenvectors) of a 3D Gaussian. The variance of displacements is hence no more described by a scalar but by a covariance tensor which is a 3X3 symmetric matrix

Diffusion Tensor Imaging revolutionized the field of white matter mapping. From its presentation in the 1990s by the neuroscientist Basser [46], it was provided a straightforward technique to describe the complex neuroanatomical information coming from anisotropic white matter tracts.

The diffusion tensor is a 3x3 symmetric matrix of vectors representing a mathematical model of the pattern of diffusion anisotropy of white matter pathways. The tensor, D, is dependent on the signal intensities, S and S_0 (defined before in paragraph 1.6.2), such that [46]:

$$\frac{S}{S_0} = e^{-(\sum_{i=x,y,z}\sum_{j=x,y,z} b_{i,j}\cdot D_{i,j})} \tag{11}$$

where,

$$b_{i,j} = \gamma^2 \cdot gd_i \cdot gd_j \cdot \delta^2 \cdot \left(\Delta - \frac{\delta}{3}\right) \tag{12}$$

Six or more measurements of S are needed to estimate D. Redundancy of measures in respect to the 6 unknowns is usually dealt with by least square fitting methods.

Eventually, the diffusion tensor is represented as:

$$D = \begin{bmatrix} Dxx & Dxy & Dxz \\ Dxy & Dyy & Dyz \\ Dxz & Dyz & Dzz \end{bmatrix}$$

The most intuitive way to represent the information described by the diffusion tensor is to see it from a geometrical point of view as the angular variation of the ADC values in the shape of a 3D ellipsoid, as shown in *Fig.10* [44].

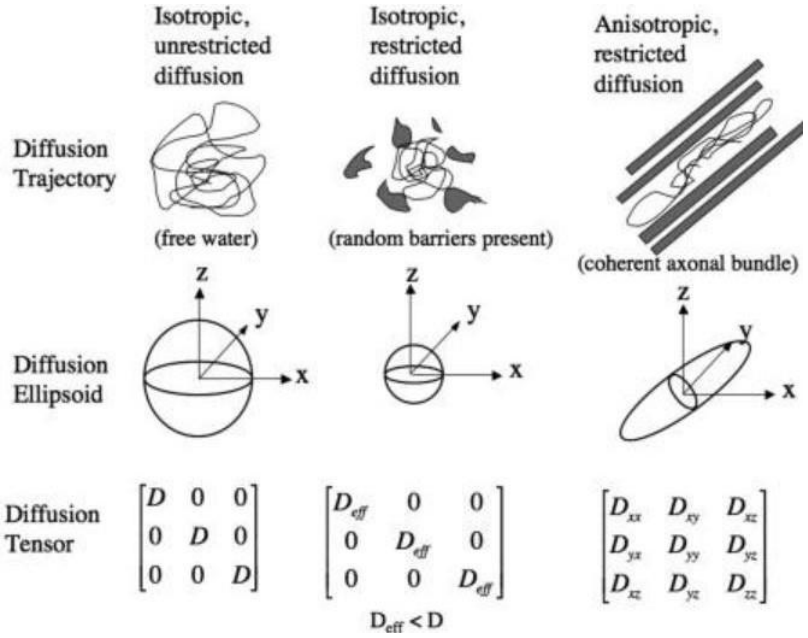


Figure 10 *Geometrical information provided by the diffusion tensor. In the image, the diffusion ellipsoids and tensors are shown for isotropic unrestricted diffusion, isotropic restricted diffusion, and anisotropic restricted diffusion* [44].

In this context, the diffusion ellipsoid is described by six variables which represent the ADC of water molecules in each direction. For isotropic diffusion (*Fig.10* on the left middle row), the ellipsoid is a sphere since the ADC is the same in every direction, on the other hand, the anisotropy is modelled with an elongated ellipsoid which assumes this kind of shape due to the fact that there is an increased diffusion distance along the main axes of the ellipsoid (*Fig.10* on the left middle row). In all the different types of diffusion, the elements of the tensor above the diagonal are always equal to the ones below. Thus, there are only 6 independent parameters of the tensor which need to be identified with a minimum of 6 non-collinear diffusion-encoding acquisitions.

1.3.2.1 DTI Metrics

Three main parameters can be extracted from the diffusion tensor, known as the mean diffusivity, the diffusion anisotropy and the main diffusion orientation. The mean diffusivity, D_{mean}, represents the diffusion coefficient averaged over the spatial directions and it is calculated as a third of the sum of the three diagonal elements of the tensor (trace of a matrix), such that [47]:

$$D_{mean} = \frac{Dxx + Dyy + Dzz}{3} \tag{13}$$

The diffusion anisotropy, instead, or fractional anisotropy (FA), indicates how much the diffusion phenomenon of interest deviates from being isotropic. In order to calculate this parameter from the diffusion tensor, the tensor is represented in a frame rotated along the ellipsoid principal axes (i.e. the tension eigenvectors) where the tensor is reduces to a diagonal matrix, such that:

$$D = \begin{bmatrix} \lambda 1 & 0 & 0 \\ 0 & \lambda 2 & 0 \\ 0 & 0 & \lambda 3 \end{bmatrix}$$

with $\lambda 1$, $\lambda 2$, $\lambda 3$ being the eigenvalues corresponding to the eigenvectors v1, v2, v3. After this mathematical manipulation, the FA is evaluated as [47]:

$$FA = \left(\sqrt{\frac{3}{2}}\right) \cdot \frac{\sqrt{(\lambda 1 - D_{mean})^2 + (\lambda 2 - D_{mean})^2 + (\lambda 3 - D_{mean})^2}}{\sqrt{\lambda 1^2 + \lambda 2^2 + \lambda 3^2}} \tag{14}$$

Furthermore, it is also possible to define the Relative Anisotropy (RA), such that [44]:

$$RA = \left(\frac{1}{\sqrt{3}}\right) \cdot \frac{\sqrt{(\lambda 1 - D_{mean})^2 + (\lambda 2 - D_{mean})^2 + (\lambda 3 - D_{mean})^2}}{D_{mean}} \tag{15}$$

The Fractional Anisotropy allows to describe diffusion anisotropy differences by building gray-scale maps characterized by limits between zero and one. Indeed, in white matter, the FA is high, near one in the most organized areas; the increased FA value specifies the fast diffusivity along the fibers and the slow diffusivity perpendicular to them [48]. In gray matter and cerebrospinal fluid, instead, the FA is near zero since diffusivity is similar in all directions. Due to its ease of interpretation and calculation, FA and RA have become through the years the most extensively used parameter of anisotropy in the clinical practice.

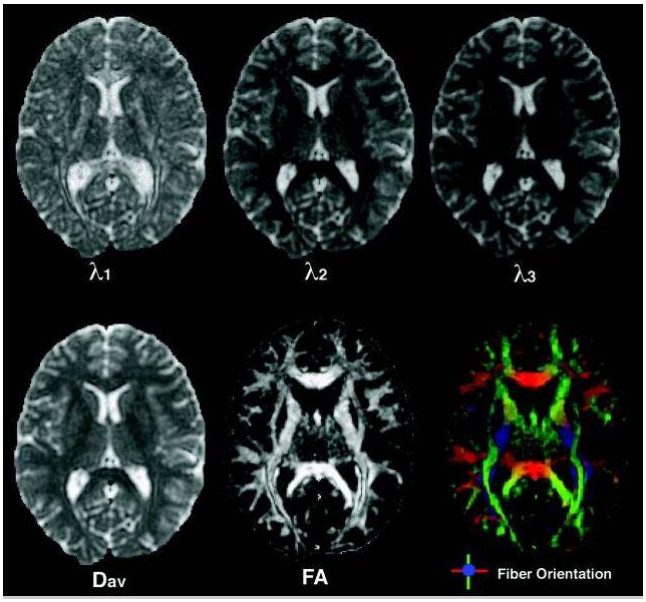


Figure 11 *Diffusion Tensor maps. In the first row, the first, second and third eigenvalues (in decreasing order) are shown with the same intensity scaling. In the second row, the left image represents the map derived from the averaged diffusivity, which corresponds to the mean of the 3 eigenvalues. The middle image in the bottom row shows the FA map, while the right one describes*

the map which was colored to show the orientation of the primary eigenvector with left-to-right fibers colored as green, anterior-to-posterior as red, and inferior-to-superior as blue [44].

Each eigenvector v1, v2, v3 obtained by the tensor describes the directions and lengths of the three diffusion ellipsoid axes in decreasing order of magnitude. In this context, the largest eigenvector is the primary direction of water diffusion, also called "axial diffusivity". This eigenvector is fundamental for the fiber tractographic algorithmic rules since it corresponds to the orientation of the axonal fiber pathways [40]. Indeed, the primary diffusion vector is also called the "longitudinal diffusivity" since it indicates the rate of diffusion along axonal fiber bundles. The second and the third eigenvectors (v2, v3) are perpendicular to the primary eigenvector; the associated eigenvalues ($\lambda 2$, $\lambda 3$) describe the strength of diffusion in the transversal plane. Within this frame of reference, it is possible to define the "radial diffusivity" as the mean between these two eigenvalues [49].

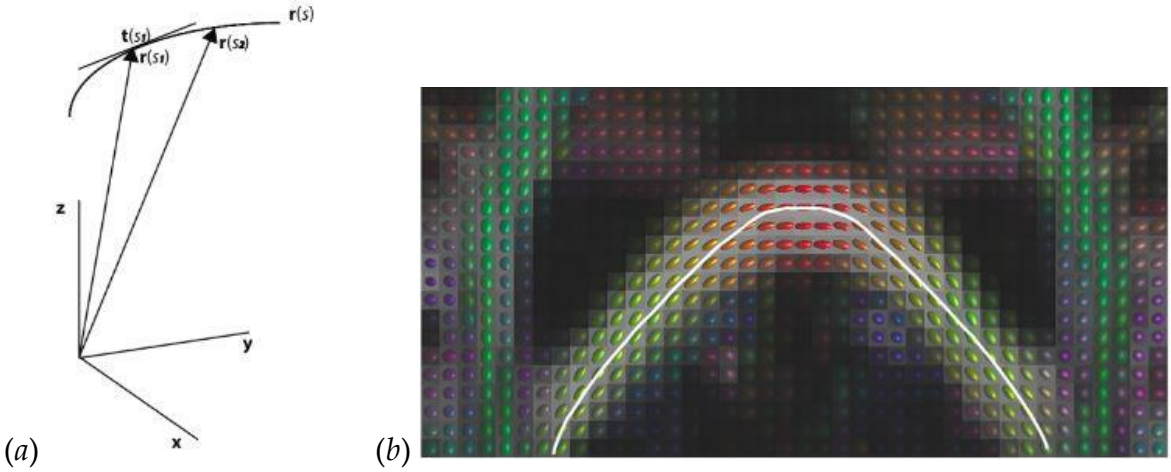
A different typology of the representation of the eigenvectors is by the mapping of colors. In this context, the largest eigenvector v1 is associated with x, y and z components which can be shown as grayscale maps (*Fig.11* top row). These maps are usually multiplied by the FA map (*Fig.11* middle image in the bottom row) to distinguish between low and high-anisotropy regions where the first are characterized by absence of dominant fibers, so enhancing the high anisotropy regions which allow to form images based on more information from an anatomical point of view. The x, y, and z components forming the image are put in correspondence to the RGB (Red-Green-Blue) color scheme and the components are unified to form a single color-coded map (*Fig.11* right image in the bottom row). The RGB color scheme which is usually used to describe the orientation of the eigenvectors is chosen such that: blue is superior-inferior, red is left-right and green is anterior-posterior directions of fibers [49].

1.3.3 Fiber Tracking in Diffusion Tensor Imaging

The mapping of the structural connectivity of the human brain has been a crucial scientific purpose for decades [50]. Nowadays, the sole safe, noninvasive technique to fulfill this goal is diffusion MRI tractography, which uses knowledge coming from the displacement of water molecules in the brain. Conversely, in the past, invasive tract tracing was the most used technique to map connectivity in the brain. However, in these kinds of experiments, after the injection of the radiotracer dye, the animal under interest is euthanized to dissect the brain and to map the different kinds of neural elements. Hence, the ethical limits are clear in performing these experiments in animals, while they are not applicable on humans. In

particular, this method can be used to describe the trajectory of white matter fibers or the connectivity among different regions of the brain. The usefulness of DTI tractography is related to the simultaneous non-invasiveness feature and the possibility of localization and estimation of neural pathways applied to diagnostic procedures, both in the clinical and surgical practice. Indeed, this method is used not only by neuroscientists to acquire information about brain functioning, but also by surgeons for surgery planning and by neurologists to assess the healthiness of the main WM tracts [50]. It is, thus, clear that the accuracy in classifying the presence or absence of connections and in delineating the fiber pathways is essential for correct diagnosis and proper surgical results.

Tractography is a technique which integrates the information coming from fiber orientations into a pathway which connect brain regions. This pathway is called "streamline" which can be defined as a curve whose tangent is always parallel to the vector field and they are reconstructed by starting from a seed point and following the local vector information, an example of which is shown in *Fig.12(b)* [51].



(*a*) (*b*)

Figure 12 *Streamline tractography. (a) Mathematical representation of the streamline tractography such that the location is parametrized with a vector r and the length of the streamline is known as s. The tangent to the streamline is called t(s) and it is the estimate of local fiber orientation. (b) The white streamline follows the orientation of the diffusion with least obstacles* [51].

From the mathematical point of view, the location of the streamline, *r*, is expressed in function of the distance along the streamline from the start, knows as the arc length, *s*. In particular, the tangent to the streamline at arc length is considered to be the estimation of the fiber orientation. (*Fig.12(a)*). Considering the diffusion tensor model, described in the previous paragraph, the tangent at the arc length

can be defined as the first eigenvector of the diffusion tensor as: $t(s) = v1(r(s))$, where $r(s)$ is the location that is distance s along the streamline [51]. At this point, it is possible to define the evolution of a streamline, as:

$$\frac{dr(s)}{ds} = const \cdot v1 \tag{16}$$

with $r(0) = r_0$ being the seed point. In practice, tractographic algorithms proceed numerically, propagating a streamline from one voxel to the next by following the direction of the main eigenvector $v1$.

The defined trajectories will pursue the largest eigenvector of the diffusion tensor from pixel to pixel, with end-stopping criteria based on the maximum turning angle of the streamline or the minimum FA within a pixel or both rules could be used simultaneously. The choice of the proper end-stopping criterion is fundamental to avoid false positives, thus, to dodge the possibility to detect connections which in reality are missing. This largely applied deterministic propagation can be substituted by stochastic approaches, which are particularly useful when anisotropy has more complex models as in HARDI protocols (see next). In summary, the DTI ellipsoid or the more complex HARDI patterns are seen as the probability of fiber propagation in any direction. The problem is hardly solved by analytical ways and Monte Carlo algorithms are used. I.e., fibers are propagated by random extraction according to the probability of each direction in each voxel, repeating the random extraction a large number of times to get statistically representative patterns.

1.3.4 DTI limitations

The initial condition of the differential equation described in the previous paragraph, is defined a priori and allows to describe the seed point of the streamline. The characterization of the initial condition is the key step in the DTI-based tractography procedure; in order to do this, it is important have knowledge about the anatomy of the white matter of the chosen region-of-interest (ROI) [48]. This, however, can be an issue if considering modest fiber systems where knowing the pathways of the fibers might be difficult or in patients where the anatomy of the white matter is altered for example by a lesion.

The selection of the appropriate ROI to start the fiber tractography algorithm is not the only problem considering these types of procedures. Indeed, the DTI model, as it is defined, is different from the real anatomical situation. It is assumed

that the diffusion in white matter is considered as Gaussian which is obviously not true in real conditions due to the presence of restrictions in the anatomical pathways. Moreover, only an average fiber population is modelled for every voxel which represents a fundamental limitation when the voxels are characterized by multiple crossing fibers. When this situation is met, the tensor model is not valid anymore, since these voxels can no longer be represented by the Gaussian distribution. Indeed, voxels in which at least two different fiber directions can be identified represent about 66% to 90% of the total voxels representing the white matter. It is, thus, critical to go beyond DTI and consider higher order models and new reconstruction techniques able to describe non-Gaussian voxels. This necessity led to the development of the HARDI (High Resolution Angular Imaging) technique (more details in Section 1.3.5).

Another fundamental limitation related to the DTI include the inability of distinguishing the neural direction of axonal pathways. Considering SC, we cannot have directed graphs indicating the weights of fibers from area A to area B and vice versa. Indeed, the polarity of a neural pathway does not affect diffusion, thus, it cannot be detected by this technique [52]. This represents still an open issue and its solving could provide important information about the structural changes in white matter pathways, which can predict functional changes in a pathological subject depending on polarity. It should be remarked that this limitation is inherent to DWI measures and is shared by DTI and the more sophisticated HARDI models. In fact, it's the physics of diffusion itself to be symmetric on both orientation of an axis. In perspective, this problem could be solved by fusing information about the streamline endings and a priori anatomical knowledge. But this is by far not accessible nowadays due to the limited resolution of DWI (in the order of millimeters). Indeed, even detecting the direction of a pathway (efferent or afferent), within a voxel containing a bundle of parallel axons, there will be axonal pathways running in both directions [52].

Furthermore, one more problem related to DTI to be analyzed is because the voxel signal is the sum of all the tissue signals within the voxel. This causes a mixture of signals at the interface of two tissues, also named Partial Volume Effect (PVE), which causes loss of contrast within edges, provoking even the impossibility to detect small lesions near these edges [53]. This effect is not negligible since causes issues in the ROI detection considering DTI and errors in the volume measurements in MRI. Besides, this effect is more acute when the difference between the signals coming from two different tissues is greater or when the boundaries within the tissue interface bends with a shallow angle with respect to the edge of the voxel. One of the main source of PVE in DTI can be found in cerebrospinal fluid (CSF) contamination of gray matter on the surface of the

cortical ribbon [53]. Indeed, the CSF gray matter is characterized by high contrast in MRI images and the cortical ribbon is denoted with considerable undulations, producing shallow surfaces. The PVE causes an overestimation of the Mean Diffusivity and underestimation of Fractional Anisotropy, representing a critical issue in the investigation of the brain structures. A major problem is that PVE are more pronounced in subjects with altered brain integrity. A possibility to solve the PVE is to reduce the size of imaging voxels to avoid the signal's contamination between different tissues; however, this causes a floss in the SNR [53].

Eventually, it is also necessary to highlight the impact of the pre-processing steps on the connectivity metrics to be derived from the diffusion signal. Indeed, there are multiple software packages which are able to analyze the signal coming from diffusion MRI data. However, it is important to select the appropriate one, since an erroneous choice would lead to differences in data quality, and the potential power of the investigation of to detect connectivity differences between subjects [54]. In the context of erroneously interpreting connectivity measures, it is important to state that the tractography does not provide a quantitative information of connectivity strength. Indeed, in many studies, it is chosen to not rely on the information about weights and perform investigations on binary diffusion tensors, as analyzed in Section 2.1.1 [55]. In particular, these are obtained from the binarization of the weighted network, by setting edges which have at least one streamline as one, and edges characterized by the absence of streamlines as zero. This processing is usually performed because tractography algorithms are not able to quantify in an anatomical context neither the number of axonal projections nor the connection strength.

1.3.5 HARDI (High Angular Resolution Diffusion Imaging)

As stated in the previous paragraph, improvements can be gained by solving the limitations of DTI related to the problem of the crossing fibers in order to have more precise and rich robust tractographic reconstructions [56]. The issue here is how to overcome the Gaussian assumption. One possible answer to this matter could be to consider a different number of Gaussians. Nonetheless, since a single Gaussian distribution has six parameters, obtained from six diffusion-weighted images, parametrized by the diffusion tensor; the higher the number of Gaussians, the more the unknowns to estimate, thus, the more the diffusion measurements which are needed. Another possible answer to the problem related to the Gaussian assumption is to consider model-free techniques. This gives birth to two possible

families of methods: model-free and mixture model techniques. These have in common to take into account more data and more acquisition and optimization schemes necessary for reconstruction of images. Mixture model methods make biophysical assumptions on the fiber populations within an imaging voxel and describe how the water molecules diffuse in this assumed environment [57]. On the other hand, model free methodologies do not make physical assumptions about tissue properties and represent the diffusion signal with respect to diffusion features. In this context, in the HARDI (High Angular Resolution Diffusion Images) acquisition schemes, the diffusion signal is measured along many (order of 100) gradient directions uniformly distributed on the sphere, with a single b-value or even repeated with two or more b-values, called "shells". Nowadays, there are three acquisition strategies used in 3D advanced diffusion imaging which are diversified with respect to the number of directions, N, and the number of b-values. The aim of single-shell techniques is to build an angular function whose maxima are aligned with respect to the fiber orientation structure. In this context, $45 < N < 200$ in terms of the number of measurements, while the b-value needs to be chosen usually between 2000 and 4000 $\frac{s}{mm^2}$ and the acquisition time is between 5 and 20 minutes. Consequently, the advantage of single-shell HARDI is related to the taking of a single sample of a sphere in the q-space, which reduces the acquisition time despite the high angular resolution results. Furthermore, the signal-to-noise ratio is improved by the choice of the appropriate b-value. Nowadays, most HARDI acquisition protocols involve a b-value equal to 1500 $\frac{s}{mm^2}$, 60 directions and 2 *mm* isotropic resolution for a good signal and angular contrasts.

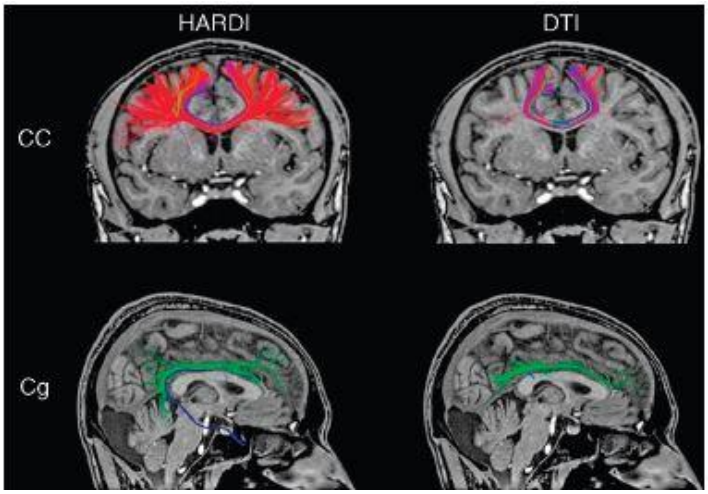


Figure 13 *HARDI fiber tracking versus DTI. It is shown how HARDI is superior with respect to DTI such that there are obtained more accurate reconstructions of complex fiber crossing configurations characterized by high curvatures such as the corpus callosum (above CC) and the cingulum (below Cg)* [57].

As DTI, HARDI is also characterized by some disadvantages. Indeed, although, on one hand, it allows to describe crossing fibers if enough measurements are performed, as shown in *Fig.13* exhibiting the superiority of HARDI in the detection of the streamlines of the Corpus Callosum and the Cingulum. On the other hand, it is important to investigate which kind of crossing voxels can be analyzed by this technique. As a matter of fact, the method is not able to distinguish between crossing, branching or curving configurations of fibers producing ambiguity in the interpretation of data. This represents a problem in the tractography schemes, and new methods have started to investigate beyond the voxels to not come across these ambiguities. Nevertheless, these studies are still in the preliminary phase and, thus, the problem of the choice of the tracking method in general remains an open question. Furthermore, it is necessary to consider the number of false-positive connections with HARDI tractography algorithms. These connections are commonly filtered in the post-processing phase by selecting fibers of a particular length or considering a specific cortical region which, however, is not able to control all false-positives/negatives.

Another limitation is related to the scanning time required with high direction numbers and in multi-shell methods [57]. Moreover, a small difference in seeding or stopping criterion can produce a critical change on fiber tracking output, as for example a small change in the FA threshold which can switch from a situation in which there are no fibers passing through the area of interest to thousands of crossing fibers in the same zone. In this context, improved visualization techniques are developed to understand these kinds of uncertainties in the acquisition of images and the appropriate selection of the tracking parameters.

Thus, tractography outcomes can be seen with unbelief due to the presence of these limitations. As of today, nevertheless, clinicians are convinced in the usefulness of HARDI tractography applications for structural connectivity mapping purposes, connectivity-based parcellation and functional/structural connectivity analyses [57].

1.3.5 Functional MRI (fMRI)

The functional organization of the brain can be defined by the amount of information being processed [58]. The interconnections between brain regions is provided by synchronized activity, thus, brain networks consist in spatially and functionally connected regions which process information. To explain the contrast mechanism used by fMRI it is necessary to introduce brain metabolism [59].

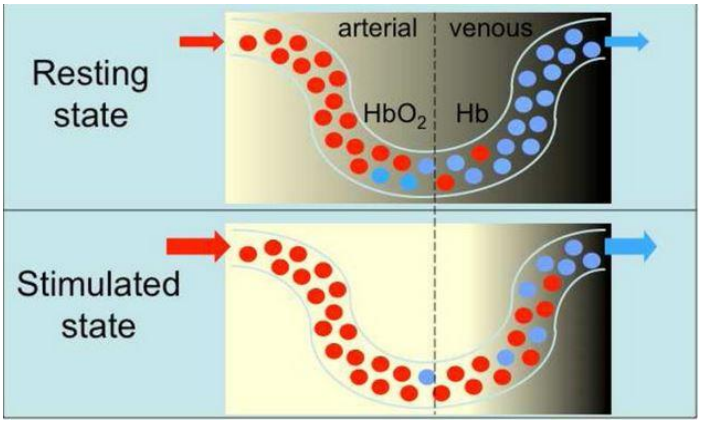


Figure 14 *Representation of brain tissue capillary during rest (above figure) and activation (below figure). In particular, red circles are red blood cells which are fully oxygenated (HbO_2), while blue circles represent red blood cells which are fully deoxygenated (Hb). The magnetic resonance signal is increased in the activated state due to the fact that the blood flow is increased and this causes the deoxyhemoglobin to be reduced with the oxyhemoglobin; this swap causes the BOLD signal to increase* [59].

All the activities of neural signaling in the brain need energy in the form of adenosine triphosphate (ATP). When a brain region is activated by a specific task the supplementary neural firing result in an increased energy requirement, producing an increased cerebral metabolic rate of oxygen in that particular brain region, as shown in *Fig.14*. The increased metabolic rate of oxygen brings about a dilatation in the adjacent blood vessels, causing an increase in blood flow. The intensification of blood flow to match the deficit in the metabolic oxygen demand produces, however, more oxygen than is needed. This creates a decrease in the deoxygenated hemoglobin (deoxy-Hb) in the local micro-vascular bed [59]. A decreased concentration of the paramagnetic deoxy-Hb decreases the magnetic field inhomogeneity ΔB_0 thus enhancing the T2*-weighted signal according to the formula:

$$\frac{1}{T2^*} = \frac{1}{T2} + \gamma \Delta B_0$$

In this way a Blood Oxygen Level Dependent (BOLD) contrast is obtained by means of the dynamic analysis of T2*-weighted contrast. The typical MRI sequence applied is echo-planar-imaging (EPI), which, being a gradient echo sequence, has the needed characteristics of being very fast in scanning a brain slice (order of 100 ms or less) and to provide T2*-weighted contrast. Currently, a stack of slices covering the whole brain volume (about 30 slices) can be acquired at a repetition time TR of about 2–3 s. Hence, the BOLD signal can be analyzed voxel wise at this sampling rate, comparing periods of activation and periods of rest.

The basic fMRI applications were developed inducing activation by the performance of different kinds of tasks to induce distinct changes in the neural activity, such as visual or auditory stimuli, and the objective is to build activation maps which compare the specific brain region activity before and after its activation, as exhibited in *Fig.15*.

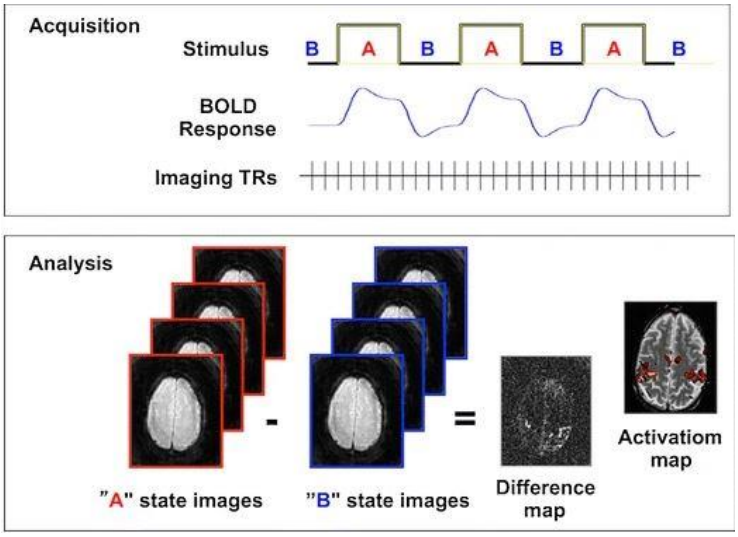


Figure 15 *Acquisition scheme and analysis of fMRI maps. The acquisition (top) is based on repeating the stimulus after a certain repetition time (TR) at which the images are acquired, while the analysis (bottom) is performed by considering the difference of the resting-state and the stimulated images which allow to create the activation map* [59].

It is worth remarking that the experimental design shown in *Fig.15* searches for correlation of voxels or entire areas with the external stimulus related to the task. Conversely, FC analyzes the BOLD signal dynamics searching for correlations between brain voxels or areas. For this reason, the typical experimental design for FC is considering a few minutes (about 5–10) of resting-state (rsMRI or rsfMRI) in the absence of any external stimulus, with the subject asked to stay awake with no particular thought in his/her mind.

Functional MRI imaging has inaccurate time resolution in the order of seconds but precise spatial resolution in the order of millimeters. Indeed, this technique was used to construct functional brain networks related to different type of tasks operating at low frequencies, lower than 0.5 Hz [6]. To construct functional brain networks, different measures of coordination between the ongoing BOLD signal can be defined, being the simplest one the Pearson linear correlation coefficient. This measure, indeed, computes the linear correlation of the BOLD signal between pairs of voxels or pairs of areas (in this case getting the signal averaged on each area) and thus a link forming the network is built if the correlation exceeds an arbitrary threshold [60]. Different FC measures exist, apart from the Pearson linear

correlation coefficient, such as the mutual information, coherence and wavelet coherence which take into account different aspects of the statistical dependence between BOLD signals (such as information domain, frequency-domain and wavelet-domain) [61].

Nowadays, there are several studies describing the effects of different diseases and disorders considering fMRI brain networks. For example, Liu et al. study in schizophrenia showed that graph metrics such as clustering coefficient and small-worldness are related with an inverse proportionality to the duration of the illness [3]. There are several findings of functional connectivity disturbances in schizophrenia, reflected in the alteration of the values of the graph metrics and disconnections in the brain networks. Another example of a study investigating schizophrenic subjects was developed by Zalesky and colleagues who analyzed a dataset formed by 15 control subjects and 12 pathological ones. In particular, they used a statistical tool to identify disruptions in the pathological population characterized by fronto-temporal and occito-temporal disconnections [5].

Another study conducted by Supekar and colleagues was related to Alzheimer's Disease (AD) functional connectivity networks, and it showed that clustering coefficient was largely reduced in AD patients, and, thus, it could be used to distinguish between healthy and unhealthy patients [62].

The analysis of functional networks can also be performed with other techniques rather than the fMRI, known as Electroencephalography (EEG) and Magnetoencephalography (MEG). It is necessary to highlight that they carry information about the brain's electromagnetic activity over an extensive range of frequencies, from 1 to 100 Hz, with appropriate time resolution in the order of milliseconds and coarse spatial resolution in the order of centimeters. Thus, MEG and EEG were widely employed to analyze connectivity in parallel or in place of fMRI to study organizational and topological changes of brain networks. Indeed, considering an example of study using MEG data, Stam and colleagues analyzed AD patients who were characterized by reduced functional connectivity strength, decreased clustering coefficient and increased path length. The innovation of this study is related to the multimodality, indeed the authors investigated data coming from both MEG and fMRI techniques [63].

1.3.7 fMRI limitations

To conclude this paragraph, it is important to mention some limitations related to the fMRI and the functional networks analysis. Indeed, in the interpretation of the results coming from this technique, stationarity is often assumed. However, considering the known dynamic and condition-dependent nature of brain activity, it is obvious that the functional connectivity metrics such as the Pearson correlation coefficient will change over time [64]. This variability allows to define the paradigm of dynamic functional connectivity analysis. These changes over time are related to task demands, learning and large state transitions such as sleep or anesthesia [64]. Moreover, it is relevant to consider the dynamic functional connectivity, since this type of connectivity also varies within the same subject and even between time windows within the same session. In this context, the variation of the functional connectivity metrics, such as the correlation coefficient or the wavelet coherence, is not easy to be explained. Indeed, functional connectivity acquisitions are characterized by low signal-to-noise ratio (SNR) and non-neural noise related to cardiac and respiratory processes and hardware instability. Besides, what complicates the interpretation of the results coming from functional connectivity studies is the fact that the networks can overlap, such that the time series coming from two nodes can have correlations with different networks. Thus the functional connectivity between two regions involved in a particular network can change if the time series are not separated. [64]. In this context, some strategies were developed to interpret these variations of the time series derived from the BOLD signal fluctuations. The most used strategy for translating these functional connectivity dynamics is the sliding window approach. In particular, a fixed-length time window is selected and the time-series signal inside that window is used to calculate the functional connectivity metric. Then, the window is shifted time by time. This technique allows to quantify the variation in time of the chosen metric given a sufficient number of data points where to shift the window.

Methods

2.1 From the Connectivity Matrix to Graph Theory

As stated earlier, graph theory allows to describe mathematically the complex networks. A undirected (or directed) network $G = (N,L)$ is defined as two sets N and L with N being any system different from the empty system, while L is a set of unordered (or ordered) pairs of elements of N. In this context, the elements of N are the nodes of the graph G, while the elements of L are the edges [65]. A node is defined by its order i in the set N. In an undirected graph, each edge is defined by a pair of nodes i and j, known as $l_{ij,}$ where the order is not important. In a directed graph, the order of the two nodes is, conversely, fundamental, indeed, l_{ij} is the edge which goes from i to j, and l_{ij} is different from l_{ji}. In graph theory, the usual way to picture a network is by representing the nodes as dots and edges by links connecting pairs of dots. An example of a directed and an undirected graph is drawn in *Fig.16*.

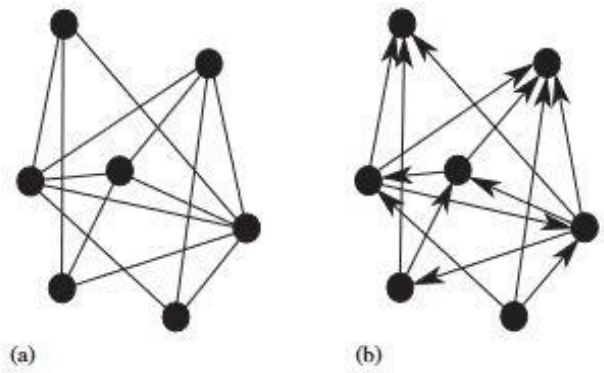


Figure 16 *Undirected (a); Directed (b) graphs with 7 nodes and 14 links. In the directed graph (b), adjacent nodes are connected by arrows, indicating the direction of each link* [65].

An important notion related to graph theory is the reachability of two distinct nodes. In particular, a *walk* from node i to node j can be defined as a sequence of nodes and edges which begins with node i and node j, and the *length* of the walk is represented as the number of edges in the walk. In this context, it is important to define *trails* and *paths*, where the former ones are walks with no edge repeated, while the latter ones are walks where no node is touched more than once. Considering paths, the minimal length of a walk between two nodes is known as

the *shortest path*; moreover, a graph is denominated as *connected* if, for every combination of different nodes *i* and *j*, it does exist a path from *i* to *j*, differently, the graph is said to be *disconnected* [65].

Considering the background of graph theory, the fundamental point in the description of the network measures is the appropriate definition of the connectivity matrix, which is the cardinal representation of network connectivity [66]. Considering undirected graphs, the adjacency matrix is squared and symmetric *N x N*, where the dimension of the matrix corresponds to the number of nodes in the network:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

The diagonal elements of the adjacency matrix ($a_{11}...a_{nn}$) are defined as the connectivity strength of each node with itself. Nonetheless, in the context of the analysis of brain networks, the diagonal is usually neglected, imposing the diagonal to zero. On the other hand, the off-diagonal elements describe the connectivity strength between a pair of nodes, with values depending on the metrics used for connectivity estimation. These values can be used to describe the type (excitatory or inhibitory) and strength of connectivity between each pair of nodes [1].

The off-diagonal elements can be divided into a lower triangle, defined by all values beneath the matrix diagonal, and an upper triangle, all elements above the diagonal. The upper and lower triangles refer to the earlier defined concept of directed and undirected networks. In particular, if the matrix is characterized by different values in the upper and lower triangles, the matrix is asymmetric and the graph is directed. Conversely, if the upper and lower triangle of the connectivity matrix are identical, the matrix is symmetric and the graph is undirected [1].

Considering the a_{ij} values in the connectivity matrix, in binary, or unweighted, graphs, these assume only 0 and 1 considering the presence or absence of an edge connecting two nodes (*Fig.17(b)*). On the other hand, in weighted networks, the adjacency matrix values represent the strengths of the edge connecting two nodes w_{ij}, where an example of weighted networks is shown in *Fig.17(a)(c)*. As previously stated in the Section related to the DTI limitations (1.3.4), the information about connectivity strength is sometimes discarded and the connectivity matrices binarized according to the number of detected streamlines being below (0) or above (1) a given threshold. Clearly, this process might lose important

information, but, on the contrary, will not consider weight changes due to errors intrinsic to the tractographic methods, thus providing more robust outcomes.

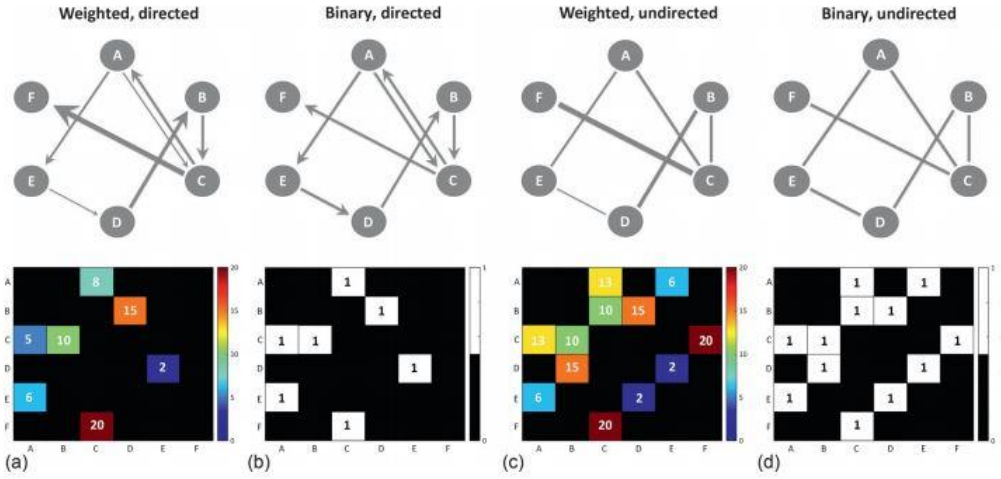


Figure 17 *Equivalence between graphs and matrices. (a) A weighted, directed graph is shown in the upper panel with connection strengths represented as variations in the edges' thickness, the corresponding connectivity matrix is shown in the lower panel where the different colors of the* a_{ij} *elements describe the different weights. (b) A binarized, directed graph is shown in the upper panel with edges characterized by the same thickness due to the binarization, also the corresponding connectivity matrix in the lower panel is formed by either 1s or 0s indicating the presence or absence of a connection, respectively. (c) A weighted, undirected network is shown in the upper panel, where the absence of directionality is described by the lack of arrows, the corresponding connectivity matrix is symmetric such that* $a_{ij} = a_{ji}$*. (d) A binarized, undirected network is shown with its corresponding connectivity matrix* [1].

2.1.1 Thresholding Methods and the Importance of Weak Connections

The main limitation of computing binary networks by setting a threshold is that weak connections are missed, possibly losing important information coming from the investigation of the adjacency matrix. Indeed, the analysis of the connectivity matrix can provide comprehension about the structure of the connectome by putting the spotlight on which connections exist and where they are positioned. It is expected that the analysis of weighted connectomes instead of binary ones could solve the problem of the influence of weak connections. However, the same procedure of removing weak connections is applied also by the thresholding of weighted graphs, also known as "pruning" [67]. To use this technique, it is

important to select the appropriate method, such that, removing all the connections up to a given connection strength, by cutting out the connections up to a given density or deleting the connections according to the values of different graph metrics. However, without a standardization of the thresholding approach, the same study on the same data could produce different results due to divergences in thresholding methodologies.

Recently, many studies investigated the potential usefulness of weak connections in recognizing different behaviors and conditions of different diseased or control subjects. Indeed, often, in studies, the exploration of human brain networks has focused on the strong connectivity patterns of the particular brain regions, while the role of weaker connections has not been widely considered. To analyze this topic in further depth, Santarnecchi and colleagues investigated the different connection weights to explain individual differences in Intelligent Quotients (IQs) [68]. In particular, it was concluded that, inspecting the functional connectivity networks of 98 individuals of different age, the differences in IQs were mostly explained by long-distance and weak connections, with only a limited contribution by strong brain connectivity. In another study by Bassett and colleagues, the complexity of the human's brain is investigated in schizophrenic patients using a multi-level analysis of low frequency resting-state fMRI data [6]. It was found that weak connections could be used as clinical biomarkers correlated with attention, memory and schizophrenic symptoms, as shown in *Fig.18*. Before investigating the connectomes, it is usually performed some kind of processing and filtering of the connections, for example before binarization. A common procedure is to threshold the matrix to lower the presence of low-weight/spurious connections of the network, which, however, can generate some issues in the interpretation of the connectivity pathways due to the removal also of useful information.

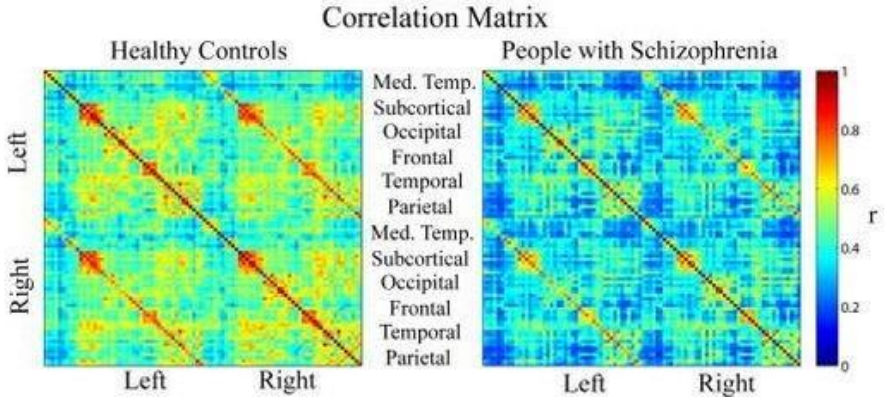


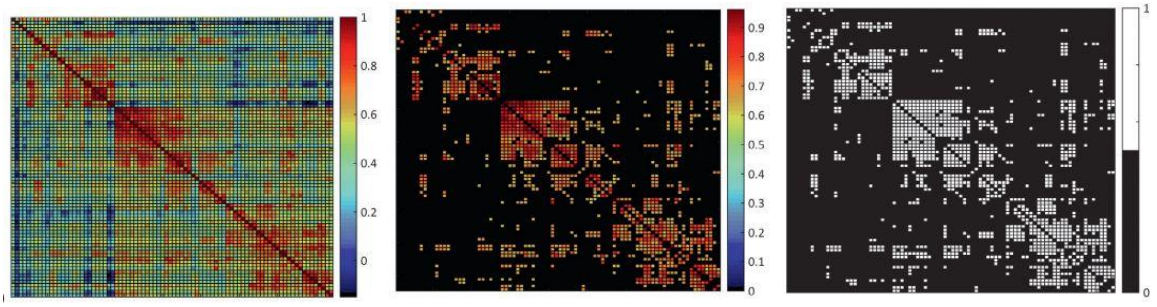
Figure 18 *Diagnostics of healthy and schizophrenic patients thanks to the characterization of weak connections. In the study conducted by Bassett and colleagues* [6], *the connectivity matrices were computed according to the AAL atlas. The strength of the connections is represented in the*

different $a_{i,j}$ elements of the functional connectivity matrix and it is scaled with colors such that the bluer the weight, the weaker is the connection. It is visible that the connection weights are significantly decreased in patients, indicating that the magnitude of functional connectivity was lower in the schizophrenic population.

Absolute thresholding applies a single cut-off threshold, τ, to individuate which connections are to be set to zero, such that:

$$a_{ij} = \begin{cases} a_{ij}\ if\ a_{ij} > \tau \\ 0\ otherwise \end{cases}$$

This operation aims at the cancellation of spurious connections due to noise. Nonetheless, it is difficult to determine which specific threshold should be used, so the proper choice of τ. This approach is also known as "absolute thresholding" [69]. Examples of the thresholding and binarization procedure performed over a functional connectivity matrix are exhibited in *Fig.19 (a)(b)(c)*.



(a) (*b*) (*c*)

Figure 19 *Thresholding and binarization of a functional connectivity matrix. In all figures it is possible to notice the zeros on the main diagonal, while, in figure (a), being a fully connected network, every non-diagonal element is different from zero. However, some kind of thresholding of the connectivity matrix is necessary to remove spurious connections characterized by being noisy. In figure (b) it is shown the same matrix of figure (a) thresholded. In figure (c) the network is binarized* [1].

This is an elementary and powerful approach to investigate networks but, however, characterized by some issues. Indeed, the selection of the absolute threshold changes the number of edges to be kept across datasets and also across populations giving problems in the analysis of the topological features, such as graph metrics.

To overcome the above limitations, an approach named as "proportional or density-based thresholding" was proposed, which adapts the threshold of each subject to maintain an equal number of connections, thus equalizing the density of the considered datasets [69]. In the procedure of proportional thresholding, however, having the same number of connections (same density) across networks

means to include also low correlations which increase the possibility of random and noisy fluctuations in the networks.

2.1.2 Network Metrics

In graph theory, it is possible to analyze and compare different networks towards the investigation of their topological features at different spatial levels.

Local and Global topological properties:

The node degree, k_i, for the i_{th} node, n_i, is defined as the number of nodes connected to node, n_i, and can be calculated from the binary adjacency matrix, such that [1]:

$$k_i = \sum_{j=1}^{N} a_{ij} \tag{18}$$

If the graph is directed, the degree of the node is characterized by two components: the number of outgoing links $k_i^{out} = \sum_j a_{ij}$, also known as the out-degree of the network, and the number of ingoing links $k_i^{in} = \sum_j a_{ij}$, also known as the in-degree of the network [66]. Thus, the total degree is defined as:

$$k_i = k_i^{out} + k_i^{in} \tag{19}$$

In weighted networks, the analogous parameter with respect to degree is the node connection strength, s_i, such that [1]:

$$s_i = \sum_{j=1}^{N} w_{ij} \tag{20}$$

Often, strength values are normalized with respect to the number of weights, such that:

$$s_i' = \frac{1}{N-1} \cdot \sum_{i \neq j} w_{ij} \tag{21}$$

Moreover, the edge weights can be signed, such that a positive signed connection indicates an excitatory connection between nodes, while a negative one represents antagonism. Within this framework, it is possible to compute the positive strength, s_i^+, and the negative strength, s_i^-, of the ith node, by summing the negative edge weights, w_{ij}^-, and positive edge weights, w_{ij}^+, such that:

$$s_i^+ = \sum_{i \neq j} w_{ij}^+ \text{ and } s_i^- = \sum_{i \neq j} w_{ij}^-$$

(22)

These signed parameters can be also normalized by dividing the sum for N – 1. From this perspective, Rubinov and Sporns in their study [70] proposed a unified normalized strength measure which considers the signed and normalized parameters of the strength, such that:

$$s_i^* = s_i'^+ - \left(\frac{s_i^-}{s_i^+ + s_i^-} \right) \cdot s_i'^-$$

(23)

This parameter makes sure that positively weighted edges make stronger contribution with respect to negative ones. Indeed, in clinical studies, it is common practice to emphasize the contribution of positive edge weights thinking them as hubs of brain networks which integrate different elements.

Another measure of network connectivity is defined as the graph density, ρ, and it describes the ratio of the number of edges in the graph and the possible maximum number of edges. In particular, this parameter is used to detect potential changes in the networks which can be related to thresholding, for example. The total number of possible connections is defined as $N \cdot (N-1)/2$, where N is the number of nodes in the network and $N \cdot (N-1)$ is the number of off-diagonal elements in the connectivity matrix. Thus, the connection density is calculated as [65]:

$$\rho = \frac{k_i}{N \cdot (N-1)}$$

(24)

Being a ratio, the density varies between zero and one, where $\rho = 0$ corresponds to the absence of connections, while $\rho = 1$ indicates that the graph is fully connected.

A further network metric to be described is the path length; indeed, shortest paths are fundamental in representing carrying and communication of information. In graphs, a geodesic path length, d_{ij}, is the smallest number of edges required to connect node i to node j. The maximum value of d_{ij} is the diameter of the network. The parameter which is usually used to define the separation between two nodes

is called the *average shortest path length*, known as *characteristic path length*, defined as the mean path length over all pairs of nodes. The equation of this metric can be written as [65]:

$$L = \frac{1}{N \cdot (N-1)} \cdot \sum_{i,j \in N,\ i \neq j} d_{ij} \tag{25}$$

However, this representation of L is characterized by the fact that it diverges if there are disconnections in the graphs, indeed, if two nodes are not connected, the path length is supposed to be infinite. To solve this issue, it is possible to either limit the sum in the formula only to pairs of connections belonging to connected components, or to consider the mean of geodesic paths, defining the *efficiency E*, such that [65]:

$$E = \frac{1}{N \cdot (N-1)} \cdot \sum_{i,j \in N,\ i \neq j} \frac{1}{d_{ij}} \tag{26}$$

Indeed, with this quantity, the problem of the divergence is solved due to the fact that any pair of nodes belonging to disconnected components produces a contribution equal to zero to the sum. It is also possible to define the path length for weighted networks, d_{ij}^{w}, thanks to a Dijkstra's algorithm [71]. This is an iterative algorithm which provides the shortest path from one particular starting node to all other nodes in the network. For example, applying this concept to a real case, representing the nodes of the graph as cities, and edge paths as driving distances between pairs of cities connected by a direct road, the algorithm is used to identify the shortest path between one city and all the others. In the context of paths, the process of determining the low-cost paths starting at node n_i starts with connected nodes, and is repeated for path lengths, with the increasing of the number of iterations until all the (N-1) low-cost paths are obtained. This allows to determine the mean geodesic weighted-path length for node i, such that:

$$d_i'^{w} = \frac{\sum_{j=1,\ i \neq j}^{N} d_{ij}^{w}}{N-1} \tag{27}$$

Furthermore, another important graph metric is the *clustering coefficient*. To define it, however, it is first necessary to describe network motifs. In particular, a motif can be defined as a subgraph with a peculiar configuration, the purpose of motifs investigation is to represent which of these recur in a network with a specific repetition. Considering for example a triplet of nodes, in an undirected graph,

these can be connected in only two possible ways, indeed, they form either an open or a closed triangle. In a directed network, instead, there are 13 potential ways of connecting these nodes [65]. In particular, all the potential patterns of connections are shown in *Fig.20*, such that each of the outlines represent a possible motif of the networks.

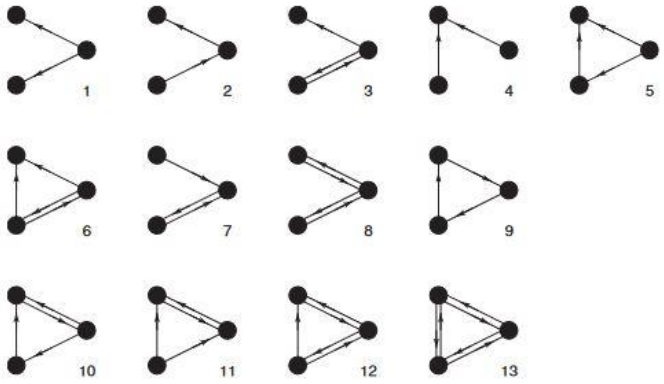


Figure 20 *All the possible 13 motifs formed by three nodes' subgraphs* [65].

Conventionally, a motif is represented as a node-connected subgraph formed by M nodes connected by at least M-1 edges; for every motif there is a particular number of possible wiring compositions, known as *classes*. In this context, clustering, c_i, is a measure of the presence of motifs of triangular connections for the node, n_i. The computation of the clustering coefficient, c_i, of node i in a binary, undirected network is done by counting the number of edges in the graphs, which correspond to the i's neighbors connected to each other. Mathematically, this is expressed as the probability of having $a_{jm} = 1$, between two nodes j and m. This value is then normalized between zero and one by the total numbers of i's neighbors, defined as $k_i \cdot (k_i - 1)/2$. Thus, the equation can be written as [65]:

$$c_i = \frac{\sum_{j,m} a_{ij} \cdot a_{jm} \cdot a_{mi}}{k_i \cdot (k_i - 1)} \tag{28}$$

Considering now graphs, the binary clustering coefficient of the network, C, is obtained by averaging the c_i of all the nodes, such that:

$$C = \frac{1}{N} \cdot \sum_{i \in N} c_i \tag{29}$$

Where the division by the total number of nodes N is done to have the value comprised between zero and one. In the context of brain networks, a value of the clustering coefficient near one indicates efficient communication and complex task

processing. In light of weighted graphs, Onnela and colleagues in their study [72] introduced the weighted clustering coefficient, $c_{i,O}$, which changes the sum of binary triangular elements with the scaled edge weights, thus, normalized by the maximum edge weight in the network. The equation which allows to calculate this parameter can be expressed as:

$$c_{i,O} = \frac{1}{k_i \cdot (k_i - 1)} \cdot \sum_{j,m=1}^{N} \left[w'_{ij} \cdot w'_{jm} \cdot w'_{im} \right]^{\frac{1}{3}} \tag{30}$$

where,

$$w'_{ij} = \frac{w_{ij}}{\max(w_{ij})} \tag{31}$$

As can be seen by the equations, w'_{ij}, or w_{ij}, tends to one in the triangles, the weighted clustering coefficient tends to the binary clustering coefficient, making sure that the network is converted from weighted to binary it will produce the same results.

Network Configurations and Small-Worldness property

Graphs can have different configurations and at the changing of the local and global topological properties defined earlier in this Section, it is possible to characterize networks with distinct models.

The small-world feature of graphs, as an example, was observed in different realistic networks, in particular technological ones. Indeed, its idea was observed from social networks which are characterized by clustering since there is a great possibility that two friends of a single person are also friends with themselves [1]. Moreover, from the work of the psychologists Travers and Milgram [73], it was shown that messages in social networks travel long distances in a few number of steps, so, according to graph theory, characterized by a relatively low average path length. In this context, Watts and Strogatz in their seminal work in 1998 [74] have defined a new class of networks characterized by being neither random nor regular, exhibited in *Fig.21* in the middle, indeed featured with low path length, which a sign of randomness, shown in *Fig.21* on the right, and high clustering coefficient, which is a feature of regularity, shown in *Fig.21* on the left.

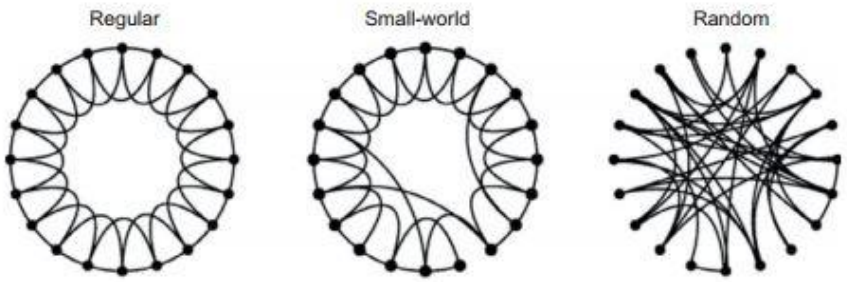


Figure 21 *Regular, Small-World and Random Networks. In the work of Watts and Stogatz in 1998* [74]*, it was shown that in between completely random and regular networks, there is a class of networks. small-world, which can be generated by randomly rewiring an arbitrary proportion of edges in a regular network* [1].

From a mathematical point of view, it is necessary to be more precise in the quantification of the small-worldness of a network than by simply defining a network with "low" average path length and "high" clustering coefficient. In particular, it is critical to quantify when the clustering coefficient can be considered as high and when path length can be thought as low. Humphries and Gurney, in their work, introduced the small-worldness parameter which considers the path length and clustering coefficient obtained from random networks with the same dimension with respect to the network under analysis, and define the ratio between the considered and the random measures, such that it is possible to compute the normalized clustering coefficient and normalized path length as [1]:

$$\gamma = \frac{c_g}{c_N} \text{ and } \lambda = \frac{l_g}{l_N} \tag{32}$$

where c_g and c_N are the average clustering coefficient of the network of interest and the one computed on an ensemble of random ones, respectively The same applies to l_g and l_N. At this point the small-worldness is obtained by [75]:

$$\sigma = \frac{\gamma}{\lambda} \tag{33}$$

In a network characterized by small-world properties, it is expected that $\gamma \sim 1$ and $\lambda > 1$, thus describing a network with a comparable average path length and a greater clustering than random, respectively. In this context, the scalar σ is greater than one, becoming an indicator of this feature.

In clinical research, the small-worldness index is used to distinguish between different disorders of brain networks linked with either a more random

organization or a more organized topology. In particular, disorders in which σ is decreased are associated with randomization of networks, instead, diseases in which the index is increased are characterized by regularity. However, it is necessary to be cautious about drawing conclusions related to brain organization from this kind of parameter. Indeed, being a global measure, it can hide different variations occurring at the individual nodes level. For this reason, the examination on global measures needs to be integrated by an analysis of properties at a lower level to recognize different properties of the brain network under investigation.

Intermediate scale analysis of networks

Analyzing networks at a scale intermediate between local node analysis and the whole graph, the nodes can aggregate into subgroups which are defined as modules. In these, a strong connectivity is observed with respect to the other parts of the network. Furthermore, these modules are organized in a hierarchical way, thus, containing modules within modules at different levels of resolution [1]. Usually, networks characterized by modularity are defined by small-world properties, thus, as stated in the previous paragraph, high clustering coefficient and low path length. To further analyze network properties, it can be seen that nodes within a module, characterized by strong connectivity, share the same properties. On the other hand, weak connections are a sign of nodes which belong to different modules. This division of nodes into modules, characterized by strong connectivity within and weak connectivity between them, is a class of complications related to data clustering. The main focus, in particular, is to recede a large set of observations into a smaller subset of clusters, by detecting hubs characterized by the same features. Within this framework, the solution to this kind of problems can be divided into two main categories: agglomerative and divisive. The former starts from individual nodes and agglomerate these into larger clusters; the latter begins with all nodes related to a single cluster and divide it into subsets of observations. Nonetheless, there is no way in determining which clustering solution is better than the other, thus, the quality of the hierarchical partitioning. In this context, Newman and Girvan in their work [76]. defined two possible criteria for accurate division. Firstly, a partition agglomerates nodes into cohesive modules, such that nodes belonging to the same module will be characterized by high connectivity. Secondly, the connectivity within the same module will be defined by a higher connectivity with respect to networks in which the edges are placed at random. To describe these criteria from a mathematical point of view, it is possible to define a parameter of partition quality, also known as the modularity index. At first, it is necessary to count the number of edges between nodes belonging to the same module. In particular, for a binary, undirected network, the modularity index is given by [1]:

$$M = \frac{1}{2} \cdot \sum_{ij} a_{ij} \cdot \delta(m_i, m_j)$$

(34)

where $\delta(m_i, m_j)$ is termed the indicator which has the property of being equal to one if nodes *i* and *j* are associated with the same module and zero otherwise. The indicator makes sure that only edges between nodes within the same module are counted.

Moreover, it is also possible to analyze the mesoscale connectivity of networks with other organizations, such as the rich-club structures, where the graph is characterized by highly connected and dense nodes; thus, a structure defined by ranking the nodes in descending order of degree. Another example is provided by core-periphery organization, where the core nodes, defined as a critical information processing subgraph, occupy a central position in the graph and are highly interconnected with each other, while the peripheral nodes are moderately connected to the core ones and sparsely connected with each other [1].

To conclude this part, a summary table of the proposed graph metrics is proposed in *Table.1.*

Index	Mathematical Expression	Definition
Degree	$k = \frac{1}{N} \sum_{j=1}^{N} a_{ij}$	Average of all degrees (number of connections that link one node to the rest of the network)
Strength	$s = \frac{1}{N} \sum_{j=1}^{N} w_{ij}$	Analog parameter of the node degree in weighted networks
Density	$\rho = \frac{k_i}{N \cdot (N-1)}$	Measure of sparsity of the matrix; it is the number of actual connections with respect to the maximum number of connections
Characteristic Path Length	$L = \frac{1}{N \cdot (N-1)} \cdot \sum_{i,j \in N,\ i \neq j} d_{ij}$	Measure of integration, expressing the average shortest path between nodes pair. d_{ij} is the distance computed according to the connection weights

Global Efficiency	$E = \frac{1}{N \cdot (N-1)} \cdot \sum_{i,j \in N,\ i \neq j} \frac{1}{d_{ij}}$	Measure of high efficiently the information travel through the whole network. It is the average inverse of the characteristic path length
Global Clustering Coefficient	$C = \frac{1}{N} \cdot \sum_{i \in N} \frac{\sum_{j,m} a_{ij} \cdot a_{jm} \cdot a_{mi}}{k_i \cdot (k_i - 1)}$	Counts the number of edges in the graphs, which correspond to the *i*'s neighbors connected to each other to form a triangle
Weighted Clustering Coefficient	$c_{i,O} = \frac{1}{k_i \cdot (k_i - 1)} \cdot \sum_{j,m=1}^{N} \left[w'_{ij} \cdot w'_{jm} \cdot w'_{im}\right]^{\frac{1}{3}}$	Generalization of clustering coefficient in weighted case, where the number of triangles is replaced with the geometric mean of its weights
Small-Worldness	$\sigma = \frac{\gamma}{\lambda}$	Ratio between normalized clustering coefficient and normalized path length
Modularity	$M = \frac{1}{2} \cdot \sum_{ij} a_{ij} \cdot \delta(m_i, m_j)$	Quantifies to what extent the intra-/inter-community link densities are anomalous in comparison to chance.

Table 1 *Graph metrics with mathematical expressions and definitions*

2.2 Connectograms for graph and sub-graph representation and analysis

As illustrated in Chapter.1, brain connectomics allows to map the human brain with respect to its neural units and their connections formed by white matter pathways, which represent an investigation from the structural connectivity (SC) point of view, or conversely analyze the degree functional connectivity (FC) detected in resting state fMRI experiments. Importantly, the network nodes are fixed once an atlas has been chosen for the parcellation of the brain cortex.

In this context, on one hand the adjacency matrices are a powerful tool to represent brain networks, on the other hand, it is difficult to visualize the connectivity patterns, which often hide important pieces of information, due to the large size of these matrices. It is, thus, fundamental to derive methods to interpret visually the connectivity patterns, in particular for explorative investigations of case studies or for the interpretation of the differences in a second level analysis between groups, discriminating healthy and pathological individuals.

In order to bridge the gap between quantitative analyses and intuitive graphical representation of brain networks of the results, connectograms were proposed. Indeed, they are circular graphs in which the different nodes of the networks are written along the perimeter of the circle according to the particular atlas chosen to investigate the network, while edges are represented as arcs connecting the nodes [4]. Connectograms are widely produced using Circos (http://circos.ca/software/) [77] which is a software designed for creating illustrations about different kinds of data in a very flexible way.

However, the addressed SC and FC networks, with more than 100 nodes, may show thousands of links, thus producing a representation which is difficult to interpret due to the large density of the links, as can be seen for example in *Fig.22*. Within this framework, it is important to extract sub-networks in the investigation of brain connectivity patterns which produces the possibility of having an easier analysis of data and also to inspect particular connections between nodes belonging to specific brain regions.

Beyond visualization and qualitative purposes, it is also necessary to quantitatively investigate the features of brain networks in terms of connectivity measures, such as node degree, small-worldness, modularity and clustering coefficient, defined in Section 2.1 of this work.

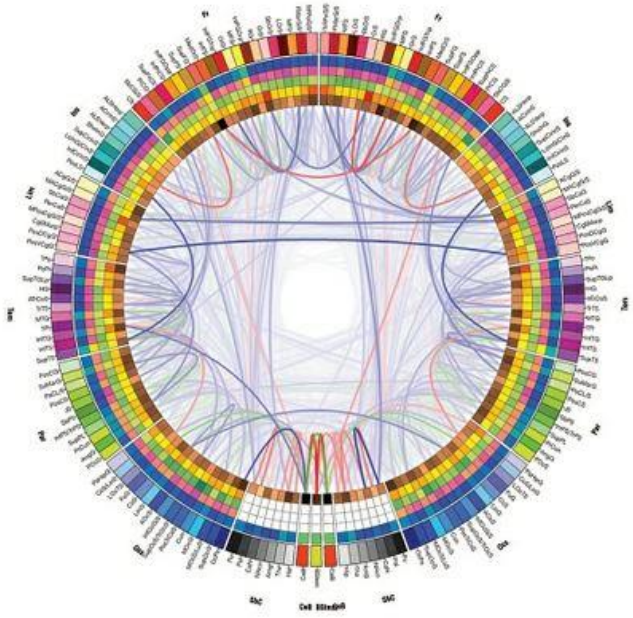


Figure 22 *Example of a connectogram extracted with the software Circos* [77].

Thus, the investigation of the architecture of brain networks in parallel with the calculation of these indices allows to both qualitatively and quantitatively analyze different brain diseases and conditions. However, gold standards with respect to standardized procedures for the network construction, the values of the connectivity measures and the topology of networks are still missing in literature and represent an open issue. Hence, further accessible tools for assessing the topology and the architecture of networks are necessary to be. In this context, generating Spider-Net connectograms could be a good general strategy to test the robustness of the processing pipeline, including the connectivity metrics formation, further conditioning (e.g., thresholding or binarization), and global or local graph indices [4].

Spider-Net (SNT) is a software developed as a collaborative project between Politecnico di Milano, Milano (IT) and CADiTeR, MRI Laboratory of Fondazione Don Carlo Gnocchi, Milano (IT). The idea for software development started with the motive of improving visualization of the brain connectivity and extend current functionalities provided by Circos. This software provides superior visualization properties by allowing the user to select regions of interest in brain and subsequently extract the corresponding subgraph. This is achieved thanks to a flexible and user-friendly Graphical User Interface (GUI) which allows a fast network exploration and the possibility of the creation of subgraph based on user selections (*Fig.23*).

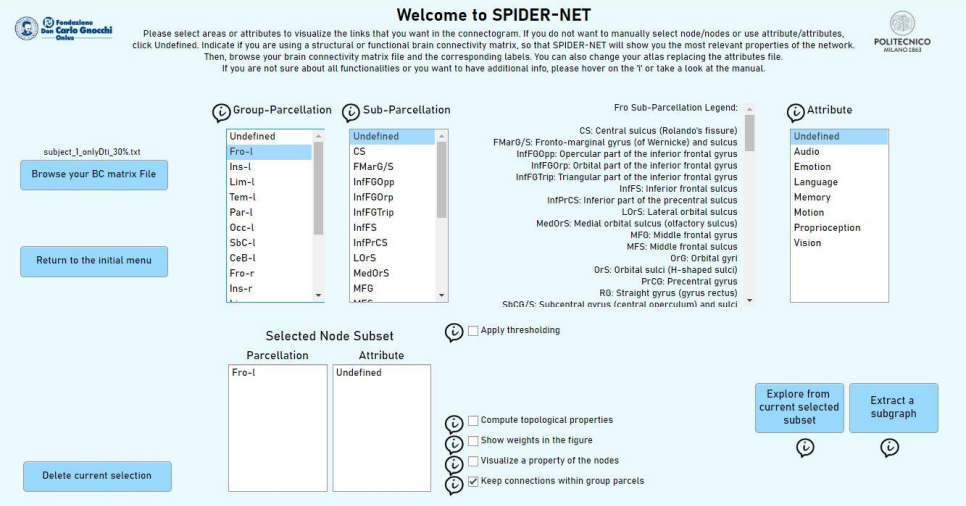


Figure 23 *Graphical User Interface (GUI) of Spider-Net.*

The software, furthermore, allows to automatically threshold, compute graph metrics and define the characteristics of the output connectogram. The interface of Spider-Net requires 3 input files: an Atlas file, a Label file and a Connectivity Matrix file. Before uploading these, it is necessary to choose if the investigation will be performed on a structural or a functional connectivity matrix in SNT Homepage (*Fig.24*).

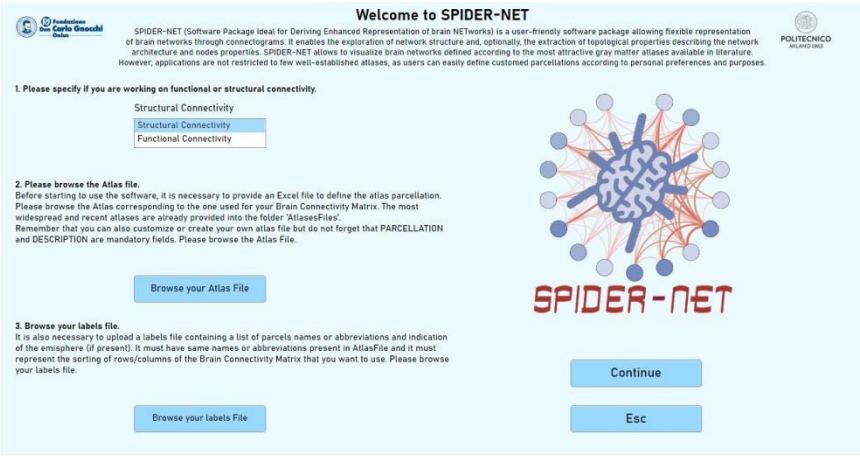


Figure 24 *Homepage Spider-Net.*

The first file is an XSL/XSLX Excel file which describes the atlas used to describe the connections between nodes. The list of the Atlas parcels is shown in a column of this Excel file, which are then represented as nodes in the connectogram generated by Spider-Net, with their sorting in the *.xsl* file determining the position of the parcels in the connectogram. For example, the first parcel is represented on the top of the circle. In the Excel file, it is also possible to group parcels with respect to, for example, brain lobes or resting state networks. There is also shown a column called "Attribute" which comprises optional attributes related with each

parcel, such as functional characteristics of each node. These two features allow to select entire groups of parcels (Group-Parcellation or Sub-Parcellation shown in *Fig.25*). In this context, it also possible to customize or create new atlases according to the aim of the investigations by creating or modifying *.xsl* files.

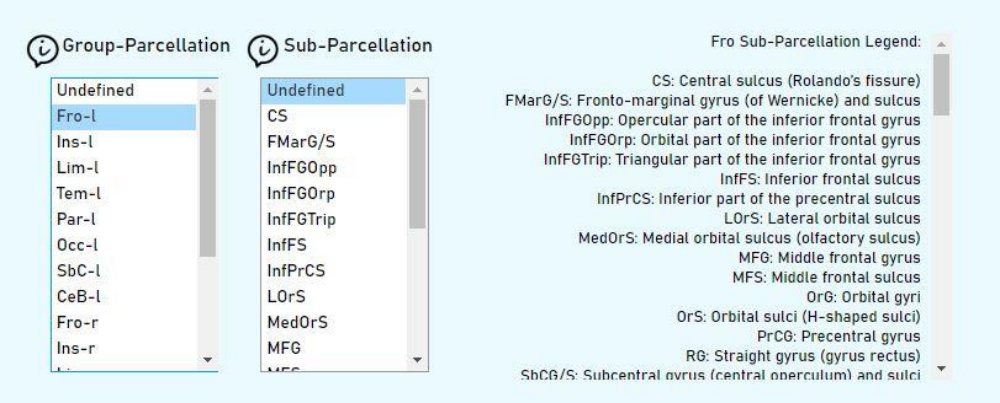


Figure 25 *Selection of Group-Parcellations and Sub-Parcellations with their definition on the right.*

The second file needed as input to create connectograms is the Label file, which can be either a *.xsl* file or a *.txt* file. It is simply a list of the parcels name where the sorting of them needs to be the same as the order of rows and columns in the Connectivity Matrix and their name needs to match the ones in the first column of the Atlas file.

Eventually, the Connectivity matrix file is a *.txt* file representing the connectivity weights in the different rows and columns. It needs to be square and symmetric with the diagonal weights conventionally set to zero. Once the inputs have been uploaded, the selection of either single parcels, groups of parcels or different attributes is available in the Graphical User Interface, exhibited in *Fig.23*.

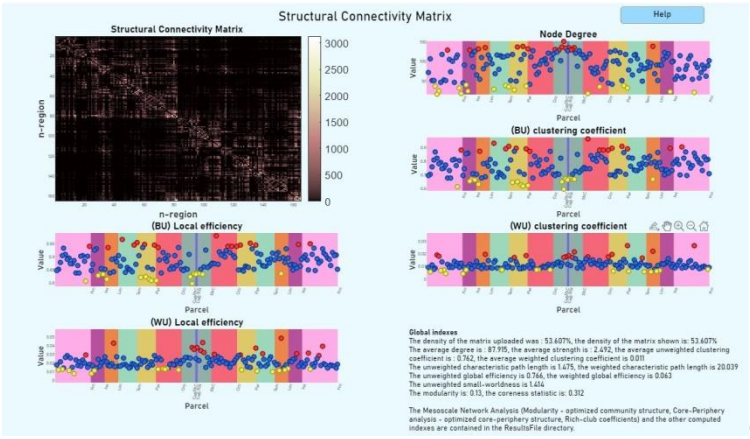
Two of the main outputs of Spider-Net are: the computation of graph topological properties and visualization of connectivity.

2.2.1 Computation of topological properties

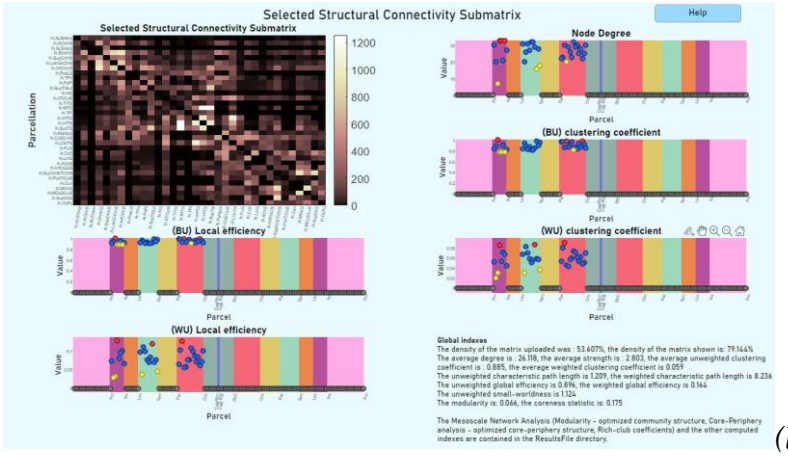
The calculation of graph properties is performed according to the implementation of the graph-based indexes in the software of the Brain Connectivity Toolbox (BCT) and the main ones are:

1. Density;
2. Average node degree;
3. Average node strength;
4. Unweighted Clustering Coefficient;
5. Weighted Clustering Coefficient;
6. Path Length;
7. Global Efficiency;
8. Modularity;

The computed parameters (if requested) are obtained at each plot of the connectogram, comparing the original unmanipulated connectivity matrix and the selected subset of nodes with its particular characteristics. Thus, beyond the plot of the connectogram, the output provided by Spider-Net is represented also by local graph metrics, depicted by scatter plots, where on the x-axis there are the names of the parcels written according to the particular atlas, while the y-axis shows the values of the metrics. Under these plots, there are also written the values of the global indexes (*Fig.26 a-b*). All these results are saved in the output folder after the software terminates its particular task.



(a)



(b)

Figure 26 *First Outputs of Spider-Net analysis. In figure (a) the initial structural connectivity matrix with the heat-map describing the values of the weights, the scatterplots of the main graph metrics concerning the columns of the adjacency matrix and the values of the global indices are shown on the bottom right. In figure (b) the same indications obtained from a sub-network of the initial connectivity matrix are shown. Indeed, the selected structural connectivity matrix characterized by a smaller size with respect to the main one and the unselected nodes highlighted in gray are shown.*

2.2.2 Connectograms Extractions

Considering the visualization of the connectogram, two possibilities are offered to the user by the software, namely the modalities called: "Explore from current selected subset" and "Extract a subgraph". In the first modality, the user is allowed to select one parcel or sub-parcel of the brain and the software draws in the connectogram the edges from this source to the selected targets, if selected, otherwise all links between the starting parcel and all parcels are shown (*Fig.27*). This option is useful for the investigation of possible alterations due to brain lesions or to control the presence of potential errors in the connectivity matrix generation. For example, in the context of structural connectivity studies, the explorative mode can be used to check for the presence of connections which should exist by general anatomical knowledge regarding the presence of connections between a seed node and all the others.

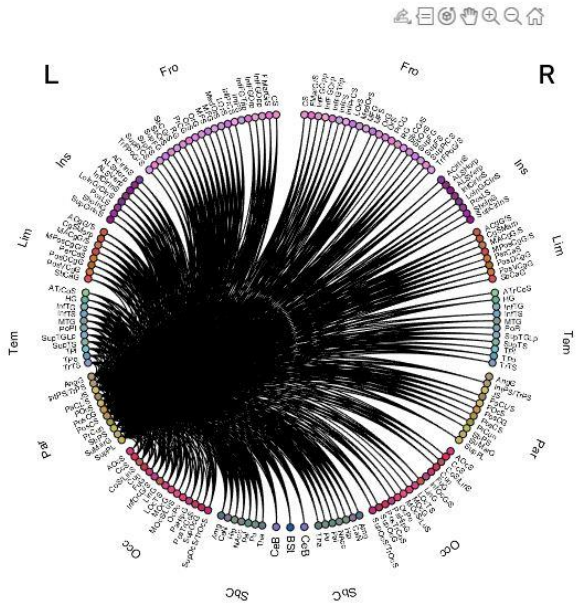


Figure 27 *Example of use of the "Explore Mode" selecting the Left Parietal Lobe as seed.*

The visualization of the connectogram allows also to extract a subgraph of all common links within the selected parcels or sub-parcels (*Fig.28*). The extraction mode is particularly valid considering the inspection of specific brain circuits or an analysis of single well-known connections by literature, to have a quality control of the acquisition/preprocessing phases [4].

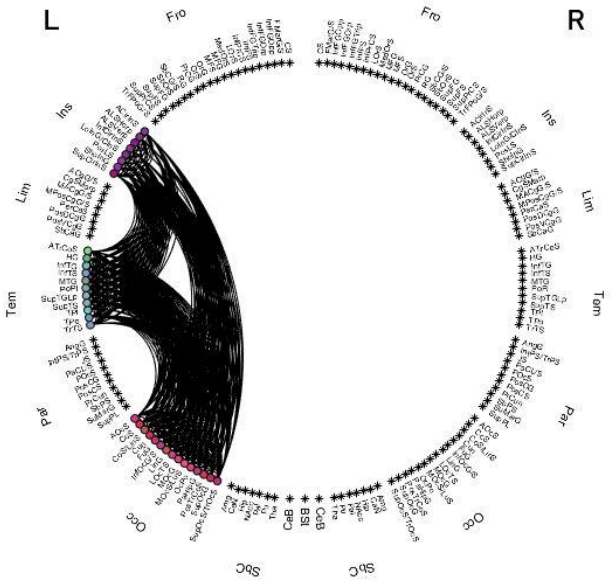


Figure 28 *Example of "Extraction of a Subgraph" mode. In particular, the subnetwork is formed by the Left Insular, Left Temporal and Left Occipital lobes.*

2.2.3 Other settings

In the end, the figure and all related files are automatically saved in a folder created at each execution of the software. Nonetheless, interactive changes to the output connectogram are possible within the software GUI. For example, single nodes can be hidden/shown with respect to the selected labels.

In the context of the modification of the connectograms, the software provides the chance to perform density-based thresholding in the main software interface, such that, firstly, the goal density is selected and then the connectivity matrix which approximates at best the density chosen by the user is extracted by removing the weak connections. (*Fig.29(a)*). It is also possible to display the value of weights of the connectivity matrix in the connectogram, such that the links connecting nodes are colored according to the value of the weight in the connectivity matrix (*Fig. 29*

(b)). The possible colors are yellow, orange, red and blue, corresponding respectively to weak, middle, strong and negative connections. Eventually, in every modality (Explore or Extract of a Subgraph), it is also possible to select to keep or not the connections within the same parcel in order to highlight the presence or absence of connections between different regions or to investigate the connections within the same parcel (*Fig. 29 c-d*).

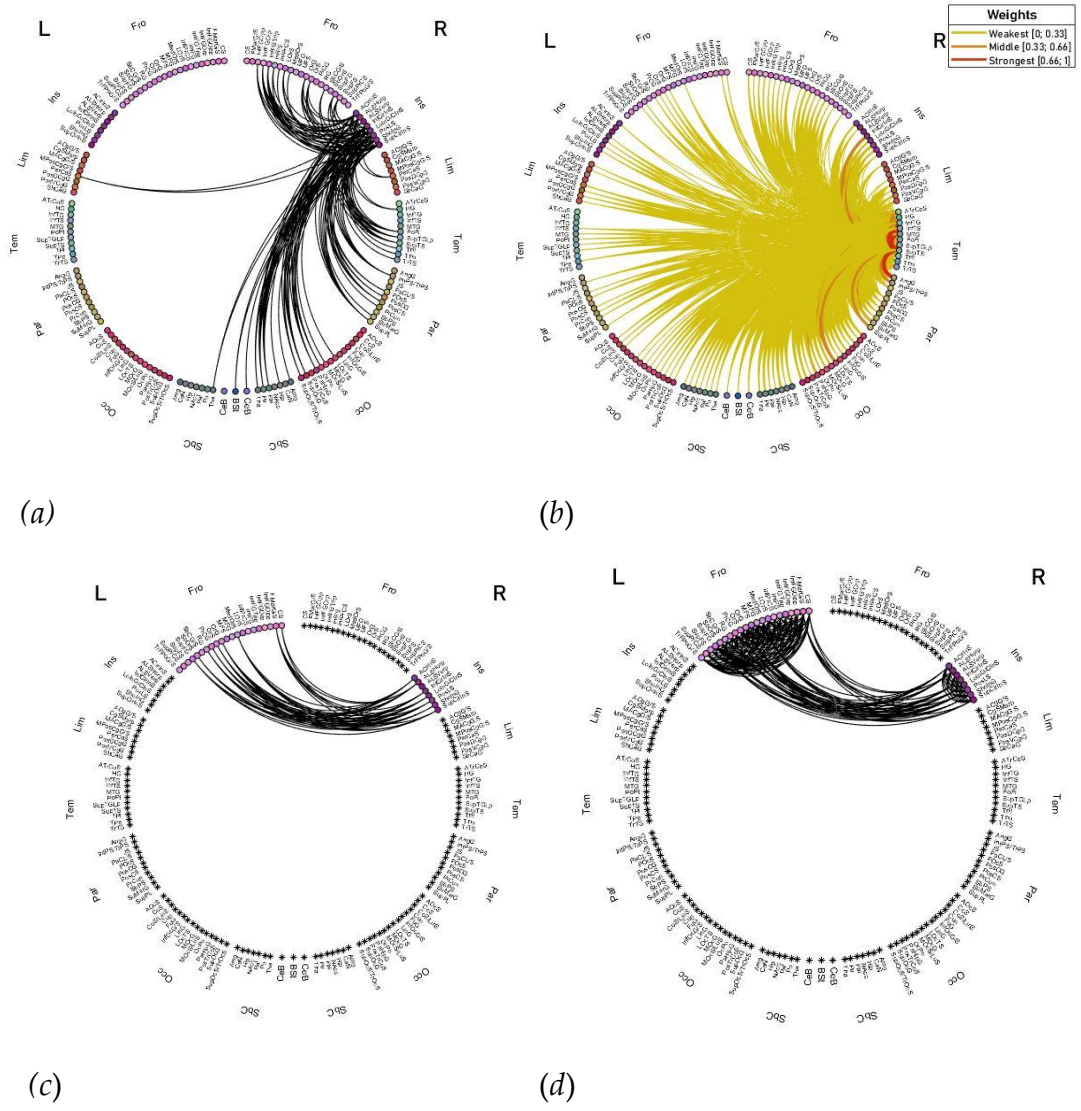


Figure 29 *Different kinds of connectograms extractable from Spider-Net. In figure (a) the connectogram referred to the explorative mode with the Right Insular lobe with 10% thresholding to keep the strongest connections is shown. In figure (b) the connectogram related to the "Show Weights" mode in the explorative modality with respect to the Right Temporal lobe with the colors of links representing the strength of the connection is shown. In figures (c) and (d) two connectograms in the extraction of a subgraph modality from Left Frontal to Right Insular lobes are shown. In figure (c) the intra-parcel connections are not shown, whereas in figure (d) the intra-lobe connections are shown.*

2.3 Bootstrapping

The investigation of changes in structural and functional connectivity is a powerful tool to analyze brain organization, mostly from a statistical point of view, with respect to a group of subjects characterized by the same condition. At the individual level, statistical tests have the uncertainties related to the acquisition of MRI images. In particular, the main issues for structural data are related to the assumption of a Gaussian distribution, which is questionable together with the other limitations explained before in Section 1.3.4 [78]. To solve these problems, different probabilistic tractography algorithms were developed to estimate the pathways variability by distinct probability density functions which describe the direction of the fibers, such as the one provided by the study of Lazar and colleagues in 2003 [79] or the one developed by Parker in 2005 [80]. However, these algorithms may be ineffective when the variance of the model does not describe the variance in the data since the initial dataset size is too small for estimating, in the proper way, the statistical properties of the considered distribution. Also considering functional connectivity data investigations, several limitations affect these data as stated before in Section 1.3.7, such as the assumption of stationarity of the data.

In this context, the bootstrap method allows to estimate the statistical features of a population from a limited number of measurement samples, without any assumptions about the distribution assumed by the initial data [81]. Indeed, bootstrap is based on randomly sampling the initial data, thus that statistical features can be investigated in the resampled data. This approach represents an alternative to traditional hypothesis testing, since it does not require to have a test statistic satisfying certain assumptions largely dependent on the experimental design and to know the properties of the data. Conversely, it is possible to compute a bootstrap test statistic from bootstrap samples, simply observing the sampling distribution. The central assumption is that the original sample accurately represents the actual population. As a result, the main advantage of the method is that the uncertainty variability of the estimator can be quantified, characterizing the dispersion and other errors in the null hypothesis [82], [81], [83].

In the context of investigating network data, the quantification of the uncertainties intrinsic to the data is essential for their scientific usefulness. In the study provided by Green and colleagues, [84], the bootstrap was applied on random graphs. It was seen that the resampling of the data was able to approximate the distributions of motif densities, such as, the number of times fixed subgraph appear in the network, remembering the definition of motif in Section 2.1.1 [84]. In this work, the bootstrap was applied with the same purposes, such as to quantify

the uncertainties about network metrics through the resampling of the initial data. Another example is given by the study provided by Gel and colleagues, where Bootstrap was applied for the quantification of estimation uncertainties in graph degree distributions [83].

Indeed, when applying the bootstrap over a distribution of data characterized by a certain variability and certain size, the expected result are about a better definition of the variability of the distributions centered with respect to the initial mean of the same distribution at the increasing of the number of iterations used to create resampled data [85], [86]. Considering the number of bootstrap samples used, the resulting statistics represent a random sample with replacements from the initial distribution characterize by a smaller size. It is worth remarking to consider that obtaining thousands of bootstrap observations from the initial data is not the same as collecting new data. Indeed, the approach is based on an ensemble of simulated data (known as "surrogates") and the usefulness of bootstrapping is related to the quantification of statistical quantities such as the standard error, a possible bias and confidence intervals of a particular sample of data.

Considering a real application of the bootstrapping in the context of brain connectivity analysis, in the study provided by Wei and colleagues [87] resting state functional connectivity records were investigated. In particular, in the clinical practice, the correlation analysis of functional connectivity data is performed to obtain feature selections. Nonetheless, due to the limitations of the fMRI techniques, explained in Section 1.3.7, the correlation analysis can predict uncertain features. To reduce this uncertainty, the study proposed a bootstrap approach applied to the functional connectivity data to create surrogate graphs. The bootstrap used to create surrogate graphs allowed to reduce the uncertainty in the feature selection scheme applied to four different prediction models to forecast cognitive skills. However, a limitations of this approach is related to computational costs since the bootstrapping methods are based on multiple resampling of the original dataset, creating surrogate graphs, thus it can be computationally expensive, especially for large datasets [87].

Considering the validity and the reliability of the method in different contexts and applications, bootstrapping was proposed and adapted for the evaluation of the uncertainty in the brain connectivity of different populations (see Section 3.3.1).

2.4 Sensitivity Analysis

A mathematical model study can be associated with different sources of uncertainty. The sensitivity analysis investigates the influence of these uncertainties by identifying, quantifying, and analyzing the responses of output variables of the model, impacted by the uncertainties of the input variable [88]. In general, a sensitivity analysis in an observational study assesses the robustness of significant findings to unmeasured confounding. A common procedure used for the analysis of uncertainty of data is through the introduction of perturbation in the data to understand how it changes the results. In particular, this method has been employed in different types of graphs, such as in the study provided by Ouyang and colleagues [89] whose purpose was to investigate the importance of edges in real-world networks. Indeed, their importance was quantified by removing some edges and quantifying how this removal affected the connectivity data, thus perturbing the initial data and drawing conclusions on the thus obtained data. In particular, an importance measure was proposed, named nearest-neighbor connectivity-based edge importance, and used to quantify the importance of a single edge or a set of edges after their removal from the network [89]. Another study involving the graph perturbation was developed by Wang and colleagues [90], who investigate the robustness of binary graphs describing artificial neural network classifiers after the simulation of the loss of edges from the network. This removal is indeed thought to lessen the capability of the network to predict and classify, thus, justifying the analysis of how varies the robustness of data after the erasing of some edges [90]. Another study provided by Mishkovski and colleagues [91] examined the vulnerability of networks after a certain number of nodes had been removed from the network. The study proposed a vulnerability index to analyze the robustness of synthetic and real-world networks, finalizing that the Watts-Strogatz model of small-world network [74] is the most robust one after the removal of different nodes [91]. A similar approach to these studies was implemented in this work such that some connections, causing the great variability in the distributions and thought to be the ones causing uncertainties in the data, were removed and the influence on different network metrics was investigated (see more detail in Section 3.3.3).

Protocol

This chapter of the thesis focuses on the tools and the technologies used in the present work, beyond that, the specification of the datasets considered.

In particular, a novel software called Spider-Net used for visualizing and studying the brain connectivity is introduced. First, Spider-Net was tested on a functional connectivity dataset and on a structural connectivity dataset acquired with both DTI and HARDI techniques. Both functional and structural datasets were composed of only healthy controls. This software was employed for the visualization and the study of whole-brain and regional connectivity through the construction of "connectograms". Afterwards, a second-level analysis on a functional connectivity dataset of patients with schizophrenia and healthy controls was performed. All protocols for data acquisition, the processing techniques employed and the experimentation performed are described. Specifically, the study of uncertainty through sensitivity analysis and the Bootstrapping methods conceived for the comparison of the two groups are reported.

All calculations of graph metrics and plots formation in the present work were carried out in *MATLAB ver. R2019a* [92]. Calculation of networks metric is based on Brain Connectivity Toolbox (BCT) which is a feature available in Matlab which allows to calculate the graph metrics in both the binary and weighted cases (http://www.brainconnectivity-toolbox.net).

3.1 Study Populations and Data Acquisitions

The investigations of this work were done on 3 different datasets, a functional connectivity dataset and a structural connectivity dataset formed by healthy controls, and a functional connectivity dataset of schizophrenic and control subjects. The considered datasets are derived from ethically approved protocols, informed consent and anonymization included.

3.1.1 Healthy Functional Connectivity Dataset

The first dataset used was acquired from resting state fMRI scans in a group of 10 healthy subjects, 5 males and 5 females, with age between 30 and 43 years (mean age + SD = 36 ± 6.32). The images acquisition was performed with a 1.5T Siemens MRI scanner in IRCCS Fondazione Don Gnocchi in Milan. Data were acquired from 5 minutes resting-state and consist in BOLD EPI images characterized by a TR equal to 2500 milliseconds, TE equal to 20 milliseconds and resolution equal to 3.125 x 3.125 x 2.5 mm^3. Images were then processed within the FMRIB (FSL) software library. Afterwards, to construct the connectivity matrices, AAL atlas [19] was used to divide the brain in 90 parcels which were considered as nodes in the network. The mean time courses were extracted from each ROI and time-point (for each 3D volume). Finally, a 90x90 correlation matrix of Pearson's correlation coefficients between all possible connections of node pairs was computed.

3.1.2 Healthy Structural Connectivity Dataset

The dataset consists of 17 healthy control (HCs) subjects (7 males and 10 females; mean age ± SD: 52.5 ± 8.3 years).

All the participants data were acquired with a 1.5 T MRI scanner. In particular, the acquisition protocol included first a high-resolution 3D T1-weighted Magnetization Prepared Rapid Gradient-Echo (MPRAGE) image, (repetition time (TR)/echo time (TE) = 1,900/3.37 ms, Field of View (FoV) = 192 × 256 mm2, resolution = 1 × 1 × 1 mm3, 176 axial slices); then, a diffusion-weighted echo planar images (EPI) image along 64 directions (b- value 1,500 s/mm2, TR/TE 7,800/109 ms, matrix size = 102 × 102 × 46, resolution = 2.5 × 2.5 × 2.5 mm3) and 3 b0 images. Eventually, a dual-echo turbo spin echo proton density PD/T2-weighted image (TR = 4,540 ms, TE = 28/112 ms, matrix size = 320 × 320 × 60, resolution = 0.75 × 0.75 × 2 mm3) [4].

Afterwards, the obtained volumes were divided in parcels, and labeled into 75 cortical parcels for each hemisphere (150 in total) according to the Destrieux atlas [15]. Moreover, some more segmentations were added according to the study provided by Fischl and colleagues [93] who labelled also the left and right thalamus, caudate, putamen, pallidum, nucleus accumbens, amygdala and hippocampus for a total of 165 parcels. Diffusion-weighted images were preprocessed using the FMRIB's Software Library (FSL) to correct geometric

artifacts, head movements and eddy current distortion, according to the protocol defined in the study provided by Pellizzari and colleagues [94]. Then, the diffusion tensor was estimated for each voxel using the FSL DTIFIT toolbox [95], [96], [97].

In addition, diffusion weighted data were processed to estimate the fiber orientation according to the HARDI spherical deconvolution processing [98]. The cortical and subcortical parcels, obtained from the 3D T1-weighted images, were registered to the respective diffusion-weighted space using the FSL flirt toolbox [99]. Successively, for each subject, WM tracts connecting each pair of parcels were reconstructed with TrackVis software.

3.1.3 Functional Connectivity Dataset of Schizophrenic and Healthy Subjects

Fifteen healthy volunteers (mean age 33.3 years, SD = 9.2 years, 14 male) and 12 people with chronic schizophrenia (mean age 32.8 years, SD = 9.2 years, 10 male) diagnosed according to standard operational criteria in the Diagnostic and Statistical Manual of Mental Disorders IV (American Psychiatric Association, 2000) [5]. Acquisition A 1.5 Tesla GE Signa scanner (General Electric, Milwaukee, WI) located at the BUPA Lea Hospital, Cambridge, UK, was used to acquire T2*-weighted echo-planar images depicting blood oxygenation level-dependent contrast as participants laid quietly in the scanner with eyes closed. Imaging parameters were as follow: repetition time: 2 s, echo time: 40 ms, flip angle: 70 degrees, voxel size: 3.05×3.05×7 mm, slice gap: 0.7 mm, flip angle: 70 degrees, number of volumes: 512. Each subject's functional volumes were realigned using a rigid-body transformation to correct for geometric displacements associated with head movements and rotations [100]. Temporal motion correction was then performed by regressing the current and lagged first and second order displacements against the time series of the realigned images. Nodes were labelled according to the AAL segmentation in parcels. In addition, all nodes comprising the cerebellum were excluded as well as any nodes for which the node-averaged time series could not be accurately estimated. As such, a node-averaged time series was estimated in a total of 74 nodes spanning the cortex and subcortex. A 74×74 connectivity matrix was then populated for each subject, where the correlation in the preprocessed times series between the i^{th} and the j^{th} node was stored in element *(i, j)* [5].

3.2 Preconditioning of the data

This sections briefly highlights the manipulation of the different datasets performed before applying the protocols described in the next section (3.3).

In particular, the first functional connectivity matrices described in Section 3.1.1, initially fully-connected, were density-thresholded at 50% with the Spider-Net software. The thresholded matrices were then saved and analyzed through the bootstrap approach.

Considering the structural connectivity dataset, the 17 165x165 matrices acquired with both techniques (DTI and HARDI) related to the structural Destrieux atlas were first divided with respect to the corresponding volume matrices, in order to take into account the differences of the brain volumes. Practically, The connectivity matrices were obtained by computing the edges as the number of the reconstructed fibers normalized by the sum of the nodes volumes [101].

In the context of the functional connectivity dataset formed by control and schizophrenic subjects, the 27 matrices of both the control and the schizophrenic groups were implemented in Matlab and adjusted to reset negative connections and put the diagonal weights to zero.

3.3 Experimentation and Evaluation

This section focuses on the protocols which were employed to explore and analyze the data, assess and interpret the results. Specifically, the description of the Bootstrap approach, the calculation of graph metrics and confidence intervals, the Sensitivity Analysis, the Probabilistic Normalization, the Spider-Net Explorations and the Statistical Tests are described.

3.3.1 Bootstrapping Application

Bootstrap was applied to evaluate the topology of the networks. Indeed, the graph features extracted from data with particular regards to their weighted versions can be highly conditioned by various sources of noise and errors, as reported in Sections 1.3.4 and 1.3.7.

For this reason, a bootstrapping procedure was implemented, as it is summarized in *Fig.30*.

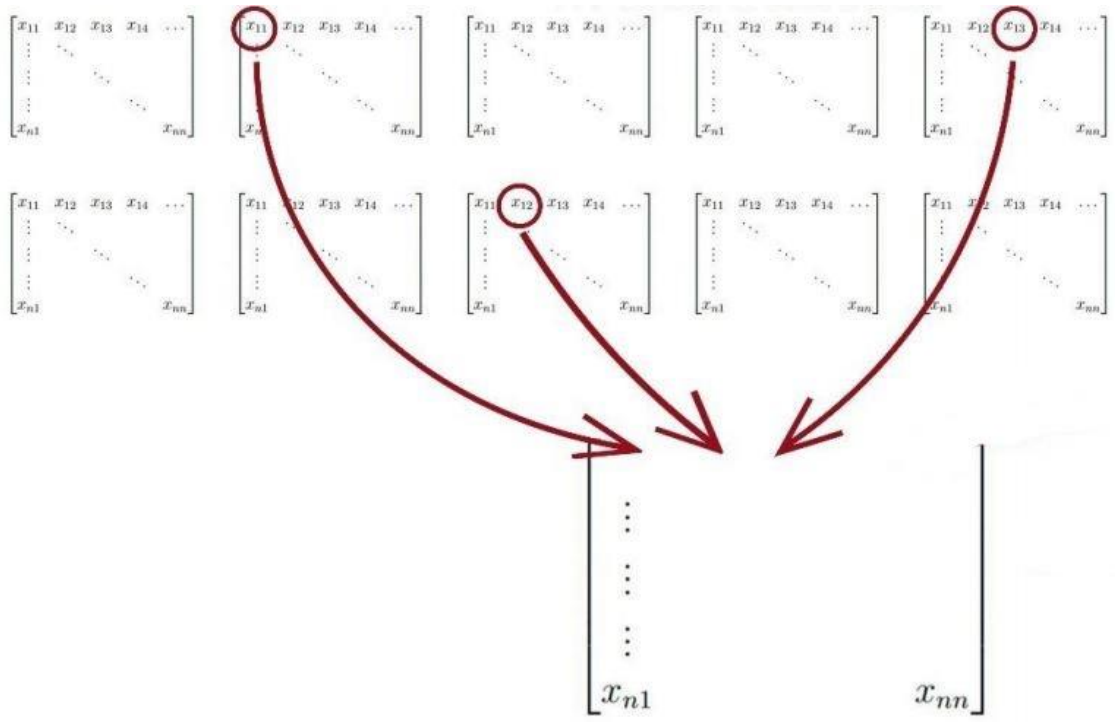


Figure 30 *Bootstrapping procedure based on the creation of surrogate connectivity matrices*

Specifically, surrogate data are created where the i^{th},j^{th} element of the k^{th} surrogate matrix defined at the k^{th} iteration is chosen randomly with respect to the same i^{th},j^{th} element from one of the matrices forming the dataset. This procedure at this step of the work was repeated with 100, 200, 500, 1000 and 5000 iterations resulting in 100, 200, 500, 1000, 5000 surrogate matrices. In this way, a more reliable analysis due to the creation of a surrogate dataset composed by a higher number of matrices than the initial one is provided and it is also possible to evaluate how many iterations are needed to obtain stable distributions of the bootstrapped indices. In particular, the 100, 200, 500, 1000 and 5000 random extractions to

compute the surrogate data were considered with the functional and structural connectivity datasets formed by healthy subjects, while for the functional connectivity dataset describing control and schizophrenic subjects, the step was performed at 1000 iterations.

3.3.2 Graph Metrics and Confidence Intervals (CI) Computations

The purpose of the Bootstrapping approach was to obtain a confidence interval to ideally find normality ranges of the different network measures. In particular, after the creation of the surrogate data, these confidence intervals were computed with respect to different graph metrics chosen, which, for the functional connectivity dataset formed by healthy subjects were:

1- Node Degree;
2- Node Strength;
3- Binary Clustering Coefficient;
4- Weighted Clustering Coefficient.

Instead, the structural connectivity dataset formed by healthy subjects was analyzed with respect to:

1- Node Strength
2- Weighted Clustering Coefficient
3- Weighted Path Length

Finally, the functional connectivity dataset formed by control and schizophrenic subjects was investigated with respect to:

1- Weighted Clustering Coefficient;
2- Binary Clustering Coefficient;
3- Node Strength;
4- Node Degree;
5- Weighted Path Length;
6- Binary Path Length
7- Weighted Efficiency;
8- Binary Efficiency
9- Modularity;

These indices were computed thanks to the Brain Connectivity Toolbox (BCT) [2] formulas which were implemented in Matlab at each step of the Bootstrapping

procedure. Before the calculation of the confidence intervals, the boxplots relative to each index at each step were plotted and analyzed. The boxplot, in particular, is a graphical representation used to describe the distribution of the values assumed by the index in each of the surrogate matrices. It is represented vertically by a rectangle delimited by first and the third quantiles, defined as the 25th and the 75th quantiles and divided by the second quantile, also known as the mean of the distribution. In this way, it was possible to notice the changes in the distributions of the indexes with the number of extractions of the bootstrapping procedure.

The Confidence Intervals (CI), eventually, were evaluated in two different ways: the first one is the Percentile method which consists in the straightforward construction of the 95% confidence interval from the resampled distribution. In particular, for example, for 1000 bootstrap surrogates, it is possible to use the 25th and the 975th value of the ranked values in increasing order assumed by the index of interest as the boundaries of the 95% confidence interval because, in this way, the interval comprises the central 95% of the distribution, also known as the percentile interval. Translating this example with 5000 resamples, the 95% confidence interval is comprised from the 125th (lower boundary) to the 4875th element (upper boundary) of the increasing ranked vector describing the distribution of the values assumed by the index of interest. The second method used to calculate the confidence interval is defined by the Standard formulas used in this context to compute both the boundaries, such that:

$$lower\ boundary\ = mean - 1.96 \cdot \frac{std(BootstrapResamples)}{\sqrt{number\ of\ iterations}} \tag{35}$$

$$upper\ boundary = mean + 1.96 \cdot \frac{std(BootstrapResamples)}{\sqrt{number\ of\ iterations}} \tag{36}$$

Fig.31 sums the steps involved in the procedure described above, such that the first step is based on Preconditioning the matrices forming the different datasets (if necessary). Then, the second step is based on the creation of the surrogate bootstrapped data. Afterwards, the third step involves the computation of the chosen graph metrics. Next, the confidence intervals are computed with respect to the two methods defined before in this paragraph. Eventually, the bias of the mean of the distributions of the values of the different graph metrics is computed such that:

$$bias = abs\left(\frac{mean(InitialData) - mean(BootstrapResamples)}{mean(InitialData)} \cdot 100\right) \tag{37}$$

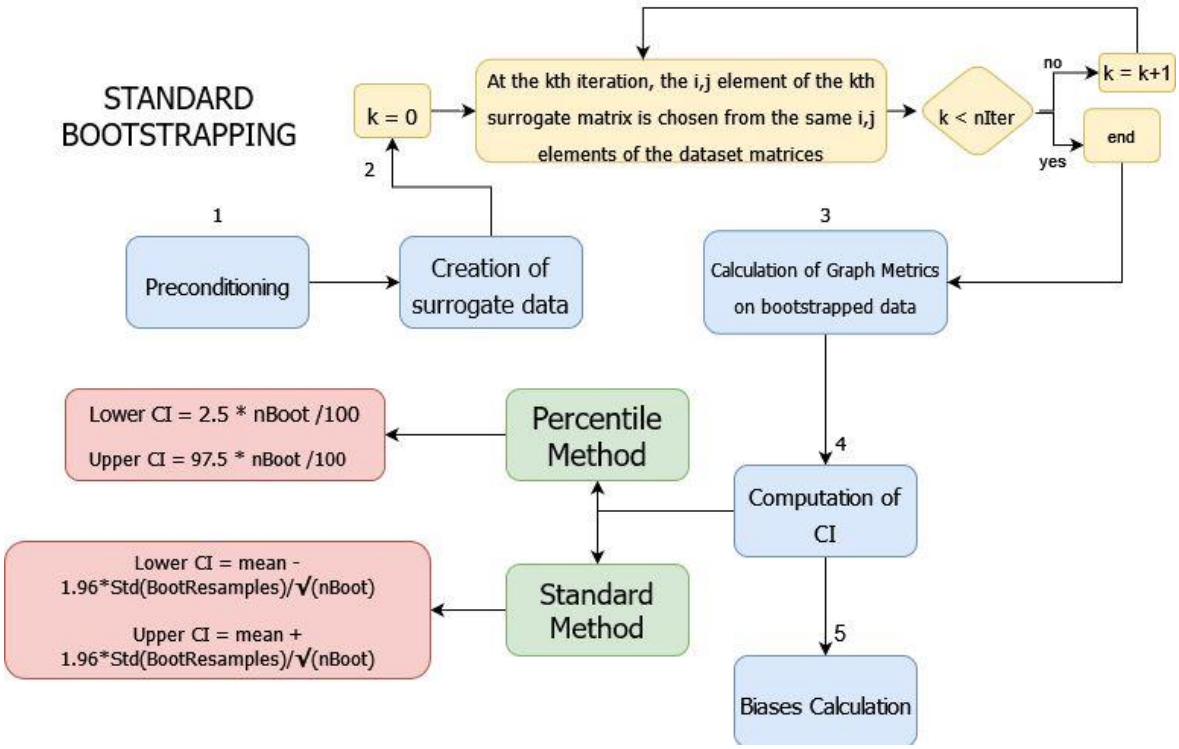


Figure 31 *Flowchart of the Standard Bootstrapping: (1) Preconditioning, (2) Creation of surrogate data, where "nIter" is the number of resamples and "i,j" refers to the row and column, respectively, of the matrix, (3) Graph Metric calculation), (4) CI and (5) biases computation procedure.*

3.3.3 Sensitivity Analysis

This first investigation on the distributions of the values assumed by the indexes after the application of the bootstrap showed, however, an unexpected behavior. Indeed, as can be seen in *Fig.32,* a non-negligible bias of the distributions of the values of the Weighted Clustering Coefficient before ("*Initial Data*") and after ("*Bootstrap*") the application of the bootstrapping. Indeed, the distribution after the bootstrap is not centered with respect to the initial value. This characteristic is visible in the distributions of most of the considered indexes. Hence, to analyze in further depth this issue, a sensitivity analysis was performed.

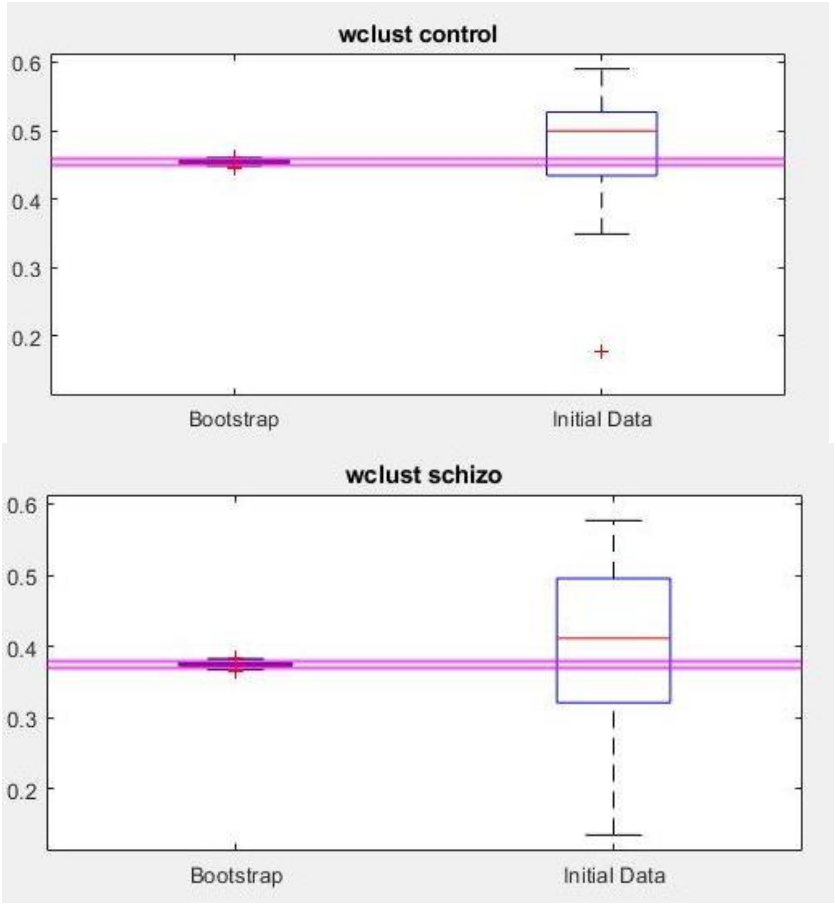


Figure 32 *Boxplots of the Weighted Clustering Coefficient's distributions evaluated on the control (upper figure) and schizophrenic (lower figure) population of the functional connectivity dataset before (right) and after (left) the bootstrap at 1000 iterations. A bias between the two distributions considering before and after the application of the bootstrap is visible.*

The sensitivity analysis of this work, in particular, focused on the connectivity weights, whose variability thought to be the ones causing the biases of the distributions of the graph metrics that occur after the application of the Bootstrapping procedure. Indeed, considering the structural connectivity analysis, the data processing which allow to pass from the raw data to the connectivity matrices is often complex, computationally intensive and requires expert quality control [102]. Moreover, the weights are often so variable that connectivity thresholds are applied to remove spurious connections, a practice considered as necessity in the analysis of structural connectivity datasets [103]. It is possible to examine the huge variability of the connectivity weights before the normalization of the weights between 0 and 1 in *Fig.33*, which is mainly formed by outliers, represented by red crosses spanning over 4 orders of magnitude.

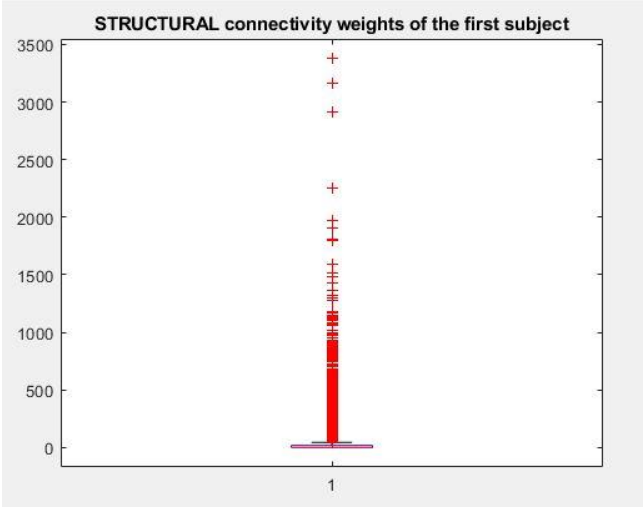


Figure 33 *Boxplot showing the distributions of the connectivity weights of the first subject of the structural connectivity dataset. The variability of the connectivity weights is shown by their variation over 4 magnitude orders..*

The variability of the connectivity weights thought to cause the biases of the distributions of the graph metrics was also observed in both the functional connectivity datasets analyzed in this work.

In this context, the sensitivity analysis was performed on the functional connectivity dataset composed by healthy and schizophrenic subjects and was based on inspecting which are the weights i^{th},j^{th} in the connectivity matrices that deviate most from the other connections in the same matrix. In particular, the 10^{th} and the 90^{th} percentile of the weights for each matrix were evaluated. An example of the distribution of the weights of the first control subject and the sixth schizophrenic one is shown in *Fig.34*.

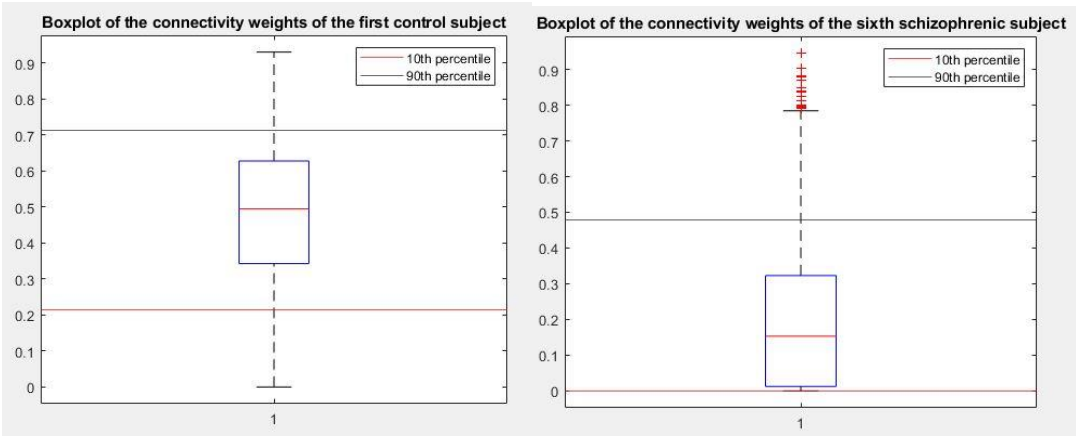


Figure 34 *Boxplots of the connectivity weights distribution of the first control subject (left figure) and the sixth schizophrenic subject (right figure). The red horizontal line represents the value of the 10^{th} percentile, whereas the black one the 90^{th} percentile.*

In order to investigate the variability of the values of the indexes causing non negligible biases, the graph metrics were computed removing the connection whose weight was out the 10^{th} – 90^{th} percentile range. Hence, the contribution of strongest/weakest connections was explored.

In particular, the position i^{th}, j^{th} of each weight which was lower than the 10^{th} percentile or greater than the 90^{th} one was saved. Next, the variability of the indices was examined by resetting one at a time the positions outside the 10^{th} – 90^{th} percentile range and computing the value of the graph metrics at each resetting. Weighted Clustering Coefficient, Node Strength and Weighted Efficiency were considered for this analysis.

Then, a difference matrix for each of the considered matrix was created. In particular, the i^{th},j^{th} element of this matrix is computed by resetting every weight of the matrix one at a time, calculating the graph metrics at each resetting and quantifying the difference between the so obtained value and the initial one before any resetting. Successively, the positions of the highest differences in the matrix are saved. The comparison between the two sets of positions (the ones outside the 10^{th}-90^{th} percentile range and the ones obtained with the difference matrix analysis) is done by overlapping them. For each of the resetting of the overlapped positions, the corresponding value of the graph metrics is computed.

The resulting values were analyzed through a scatterplot to show the overlapping of the weights that influence most the indexes and the weights that deviate most from the distributions. Hence, variability of the connectivity weights in relation to great changes in the distributions of the indexes were, thus, verified to assess the emerging biases (See Section 4.3.1).

3.3.4 Probabilistic Normalization

One of the main purposes of this work was to investigate the changes in the biases of the different graph metrics by applying a normalization of the data before the application of the bootstrap. This was done because normalizing the weights with respect to group weights distributions, the variability responsible of causing the biases is reduced. The chosen normalization was a probabilistic one and the flowchart of the technique is shown in *Fig.35.*

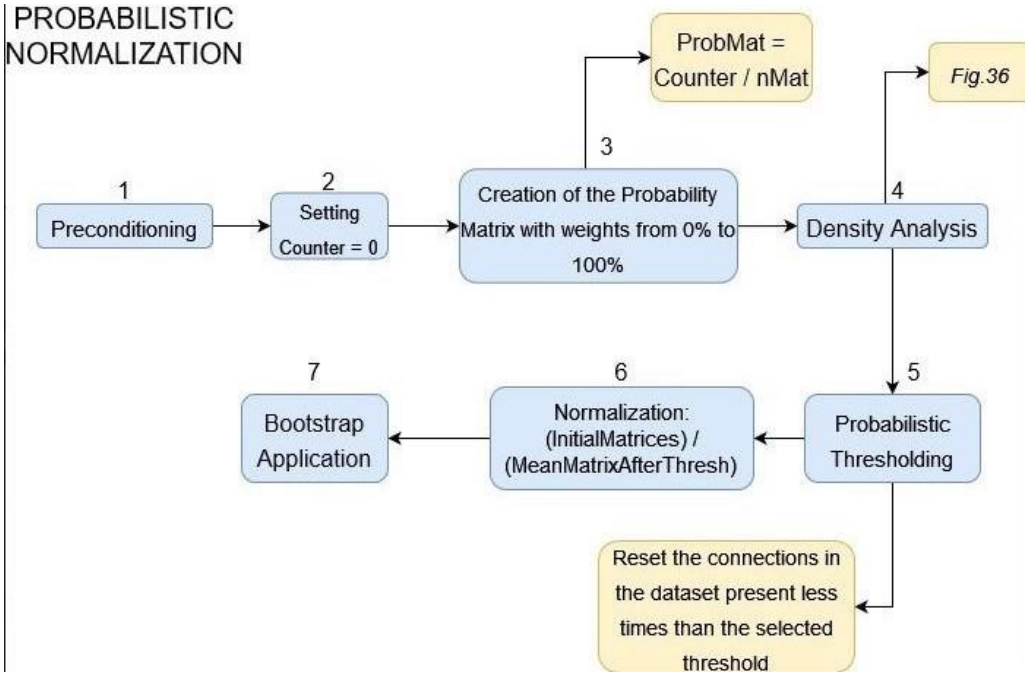


Figure 35 *Flowchart of the Probabilistic Normalization procedure: (1) Preconditioning of the data, (2) Setting of the Counter of every element from 0 the maximum number of matrices forming the dataset counting how many times each connection is present in the data, (3) Creation of the Probability Matrix by dividing the Counter with the number of matrices of the dataset to obtain the probability of having each connection in the data, (4) Density Analysis to investigate the variation of the density of the dataset at the resetting of the connections present less times than the different values of the probability matrix through a scatterplot. An example of the scatterplot representing the changing of the density of the functional connectivity dataset formed by healthy people changing density is presented in Fig.36. (5) Probabilistic Thresholding; (6) Normalization of the Data by dividing the initial matrices with respect to the Mean Matrix of the dataset obtained after the probabilistic thresholding; (7) Application of the bootstrap.*

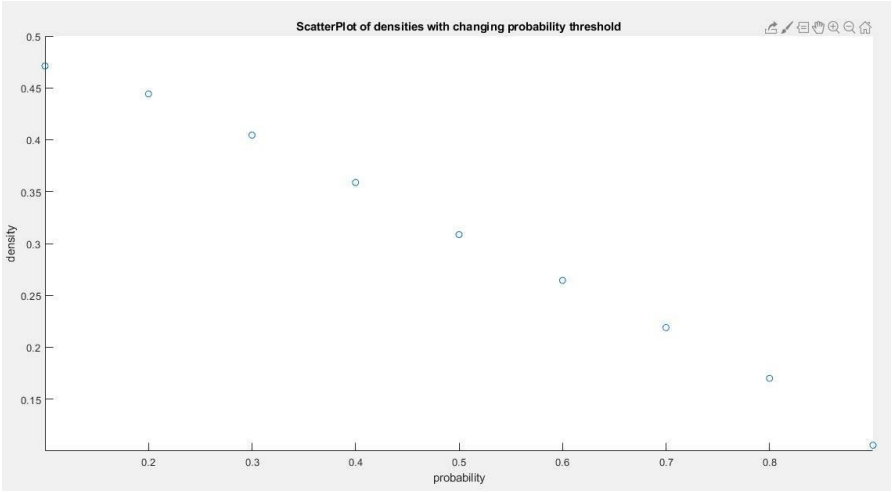


Figure 36 *Example of a Scatterplot of the density of the functional connectivity dataset formed by healthy subjects with changing probability threshold.*

In particular, after the preconditioning phase, a counter is initialized, which counts how many times each i^{th},j^{th} element in the matrices is present. Obtained this counter matrix, the next step concerns the creation of a probability matrix by dividing each element of the counter matrix by the total number of matrices present in the dataset, thus, obtaining a matrix characterized by i^{th},j^{th} elements ranging from 0 to 1 (0% to 100%). These values express the percentage of times that each connection is present in the dataset. Afterwards, a density analysis is performed. In particular, a scatterplot is created by putting on the x-axis the probability values from 0% to 100% with a step of 10%, and on the y-axis the mean density of the dataset that results from the elimination of the connections which are present in less percentage with respect to each percentage step (as shown in *Fig.36*). This is done to find an arbitrary threshold value which allows to eliminate the rarest connections of the dataset without changing too much the density of the dataset. Next, a probabilistic thresholding on the dataset is performed, consisting in resetting the rarest i^{th},j^{th} connections according to the selected threshold. Afterwards, the mean matrix of the so obtained dataset is evaluated. The normalization is then completed with the division of the initial matrices by the Mean Matrix obtained at the previous step. This normalization is thought to reduce the biases of the mean of the graph indices considering before and after the bootstrap. Indeed, dividing the most variable weights in the connectivity matrices, which are probably caused by some kind of errors, by the mean value of them is thought to reduce their variability, thus reducing the values of the biases.

The normalization step is finalized by setting the *Inf* and *NaN* values of the normalized dataset to zero with the BCT function *weight_conversion(dataset, 'autofix')* [2]. Achieved the probabilistic normalization, the bootstrapping procedure is applied in the same way as described in section 3.3.1 of this work, together with the exploration of the distributions of the graph metrics, the computation of the confidence intervals and the calculation of the biases.

3.3.5 Spider-Net Explorations

This Section focuses on how the SNT software was used in the investigation of the structural connectivity dataset formed by healthy subjects and the functional connectivity dataset formed by control and schizophrenic subjects.

Considering the structural connectivity dataset formed by 17 healthy subjects acquired with both DTI and HARDI, the differences throughout the data acquired with the two different techniques and the different connections detected by the

two methods were analyzed. In particular, the purpose was to find the connections detected by HARDI and not by DTI and the ones detected by DTI but not by HARDI. The main expected results are that HARDI should be able to reveal a far greater number of connections with respect to the DTI technique. Indeed, HARDI acquisition schemes are able to solve the limitations of the DTI related to the problem of the crossing fibers in order to have more robust tractography measurements [56].

Specifically, to have coherent data to investigate, normalization and thresholding procedures were performed. In particular, the matrices were first normalized with respect to the greatest value of the total number of streamlines between the data acquired with DTI and HARDI for each subject. For example, considering the first of the 17 subjects, calculating the number of streamlines with respect to the DTI and HARDI methods, the highest value is used to normalize each weight in the corresponding matrix. This was done to equalize the data since the total number of streamlines is greatly higher with HARDI due to its ability in reconstructing fibers. Next, the datasets were absolutely thresholded at 30%, 20% and 10% in order to draw conclusions regarding the weight of the connections in the two datasets and investigate which method was able to detect the most important ones, indeed, the higher the thresholding, the more the weaker connections are eliminated. Next, the matrices composing the datasets were binarized at the same thresholding levels, which implies to set a 1 all non-null weights produced by the thresholding. Following the binarization, to find clearly which connections were detected by HARDI and not by DTI and viceversa, a simple subtraction was performed, such that [56]:

$$Resulting\_Matrix = HARDI\_Matrix - DTI\_Matrix$$

In this way, in the Resulting Matrix, the weights detected by both HARDI and DTI are set to zero, the weights detected by HARDI but not by DTI are set to 1 and the weights detected by DTI but not by HARDI are set to -1. The difference matrices, evaluated for each of the 17 healthy subjects, were then split into two datasets considering the 1s and the -1s in order to have a better visualization of the connections detected by one technique in Spider-Net. In particular, the software was used to obtain the connectogram in the *"Explore from the current Subset"* modality, referred to the Left and Right Temporal Lobe for each of the 17 difference matrices obtained with the normalization with respect to the number of streamlines and thresholded at 30%, 20% and 10%. The temporal lobe was selected since DTI often loses several connections within this region due to the crossing fiber issue between the corpus callosum and the cortico-spinal tract [104]. Thus, it was worth investigating if HARDI was able to correctly reconstruct these missing streamlines.

Considering the functional connectivity dataset formed by 15 control and 12 schizophrenic subjects, it was investigated in terms of visually representing the connectograms describing the connectivity pathways regarding the Default Mode Network (DMN). The DMN, as described in section 1.2.4, is a network deeply involved in social behavior, control of the emotional state of the individual and recollection of past memories. On the other hand, schizophrenia is a mental disorder characterized by altered perception, irregular emotion regulation, hallucinations and weakened working memory. Thus, the investigation of the DMN for this kind of pathology became a focus of research in the clinical context [30], [7]. In particular, the investigation of the dis-connectivity patterns present in the subnetwork formed by the regions involved in the DMN and its weaker activation can provide a deeper analysis on the severity and importance of the pathology [29].

In order to perform this kind of analysis, Spider-Net was used to represent the connectivity pathways underlying the network of interest.

Specifically, after adding the possibility of relating an attribute to each parcel of the AAL atlas used as input of SNT (as shown in Section 3.3.5), the regions involved in the DMN were highlighted. The choice of parcels was performed by investigating in the literature, considering different studies and labelling which parcels belonging to the AAL scheme are also associated to the Default Mode Network [30], [29], [105], [106], [107], [108], [109].

In this way, it was possible to visually represent all the connectivity edges linked to the DMN (within or outwards) in the connectogram visualization.

3.3.6 Statistical Tests

A t-test is defined as an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related. . Mathematically, the t-test takes a sample from each of the two sets and establishes the problem statement. It assumes a null hypothesis that the two means are equal. Using the formulas, values are calculated and compared against the standard values. The assumed null hypothesis is accepted or rejected accordingly [110].

This kind of investigation was performed in the comparison between healthy and schizophrenic subjects. The statistical tests for the different graph indexes were performed through to the "*ttest2*" function implemented in Matlab, which returns a test decision for the null hypothesis that the data in input vectors comes from

independent random samples from normal distributions with unknown and unequal variances, using the two-sample *t*-test. As output this function produced a logical value of either 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise; together with the p-value which is a number describing how likely it is that data would have occurred under the null hypothesis of the statistical test.

In the context of a second-level (group) analysis, the purpose was to investigate the statistical differences of the two populations in the distributions assumed by the values of the graph metrics considering before and after the application of the Bootstrap. As regards as bootstrap hypothesis testing, standard errors and confidence intervals were calculated.

Results & Discussion

This chapter summarizes the results and discusses the investigation performed over the 3 datasets described in Section 3.1. In particular, the results after the application of the Standard and Probabilistic Normalization bootstrap methods on the functional connectivity dataset formed by 10 healthy subjects are discussed. Next, the same techniques are applied on the structural connectivity dataset composed by 17 subjects acquired with DTI. For both cases, surrogate data results with the reduction of the emerging bias are reported and discussed. Moreover, the analysis of the structural connectivity dataset also compares the DTI and HARDI techniques, which are differently affected by possible errors and source of uncertainties.

The hypothesis reported for the first two datasets about the connectivity weights and biases are then investigated in further depth through a Sensitivity Analysis on the functional connectivity dataset formed by healthy and schizophrenic subjects. Afterwards, the Standard and Probabilistic Normalization bootstrap methods results on this dataset are reported. Thus, the group (alias, second-level) analysis was performed to evaluate statistically significant differences between the two groups and to study the uncertainty which can be related to fMRI processing and connectivity measures extraction. Finally, connectivity patterns related to the DMN were investigated through Spider-Net Tool (SNT) to explore the involvement of the network in schizophrenia.

4.1 Healthy Functional Connectivity Dataset Results

This section highlights the results on the functional connectivity dataset formed by 10 healthy subjects with both Standard and Probabilistic Normalization bootstrap methods.

4.1.1 Standard Bootstrapping Results

The numerical results of the mean values of the distributions of the considered graph metrics before and after the application of the bootstrap, together with their standard deviation are represented in *Table.2.*

Graph Indexes ***Standard Bootstrap***	Mean Values and Standard Deviation		Confidence Intervals		Bias Percentage
	Initial Data	Standard Bootstrap	Percentile Method	Standard Method	
Node Degree	44.475 ± 0.069	44.482 ± 0.388	[43.700; 45.233]	[44.471; 44.492]	0.016 %
Node Strength	16.733 ± 5.443	16.733 ± 0.158	[16.419; 17.039]	[16.728; 16.737]	0.000%
Binary Clustering Coefficient	0.295 ± 0.104	0.244 ± 0.003	[0.238; 0.250]	[0.243; 0.244]	17.288%
Weighted Clustering Coefficient	0.273 ± 0.103	0.225 ± 0.003	[0.220; 0.231]	[0.224; 0.225]	17.582%

Table 2 *Numerical results of the considered graph metrics distributions before and after the Bootstrapping procedure, Confidence Intervals and the Bias of the Mean considering Before and After Bootstrapping. The boundaries of the CI are computed as described in Section 3.3.4 and reported within square brackets. The bias percentage is calculated as exhibited in Section 3.3.4.*

The results obtained from each index with the respective distributions are reported below in box-plots are reported below in box-plots to visualize the effect of the bootstrap technique. Namely, the following *Fig.37, Fig.38, Fig.39, Fig.40* show the distributions of the index values as the number of resamples (iterations) increases. The horizontal black lines drawn on the leftmost panel in all figures represent the calculated confidence interval with the percentile method.

First, the Node Degree, *Fig.37*, refers to a count of links with 50% thresholding, as stated in the preconditioning section 3.2. The results are in line with what expected (as explained in section 2.3). However, a slight increase in the distribution dispersion was found, on the order of 1 or 2 links compared to an average of 44.5 links. This small change is not significant and progressively defined as the number of surrogate extractions is risen up to 5000 (from right to left). Importantly, a virtually null bias of the average value is seen with a % value at 5000 extractions equal to 0.016 (*Table.2*).

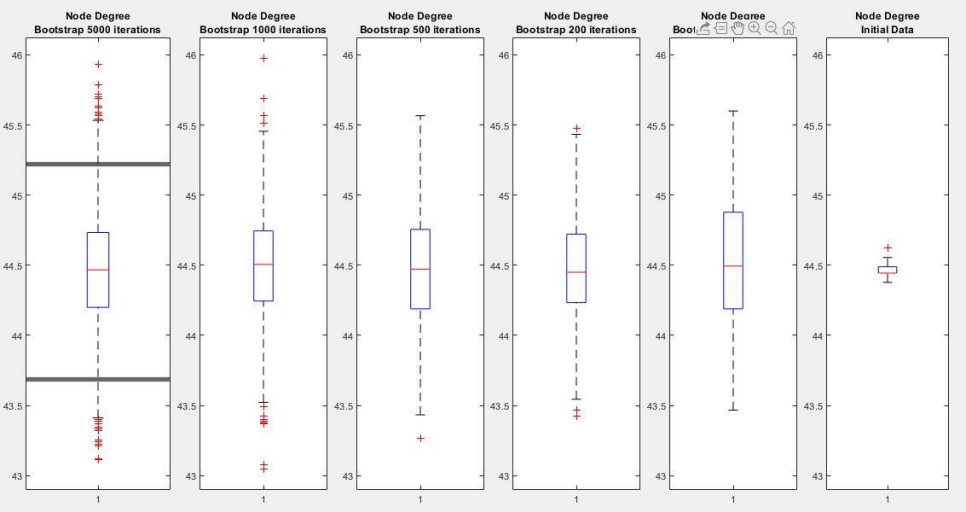


Figure 37 *Distributions of the* ***Node Degree*** *considering, from right to left, the initial 10 functional connectivity matrices, 100, 200, 500, 1000 and 5000 bootstrap iterations.*

In the context of the Node Strength, which is defined as the sum of all edge weights connected to the nodes (see Section 2.1), it is possible to state that the expected results from the bootstrap application were better confirmed. Indeed, the distributions shrink even with a limited number of extractions and keeps steady as the number is increased up to 5000. Hence, a low variability of the distribution of the Node Strength is obtained, as reported in *Fig.38*. Moreover, the numerical results shown in *Table.2* are optimal due to the absence of any bias.

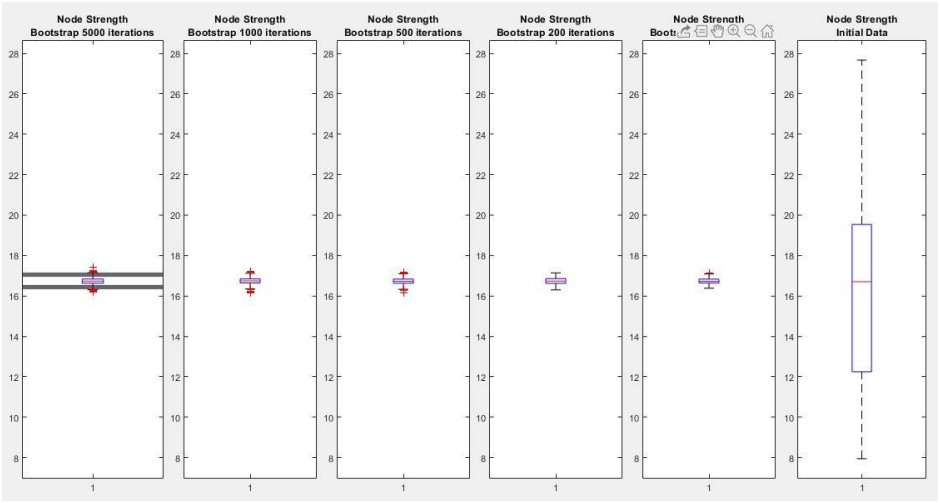


Figure 38 *Distributions of the* ***Node Strength*** *considering, from right to left, the initial 10 functional connectivity matrices, 100, 200, 500, 1000 and 5000 bootstrap iterations.*

The expected results about the shrinking of the distribution were also confirmed for the values assumed by the Binary Clustering Coefficient, defined as the fraction of node's neighbors which are neighbors to each other. In this case, as can be seen by the numerical results shown in *Table.2* and in the distributions reported

in *Fig.39*, it exhibits a greater bias (17.288%) with respect to the previously analyzed indexes. Despite this unexpected bias, the variability of the distribution was significantly reduced as expected.

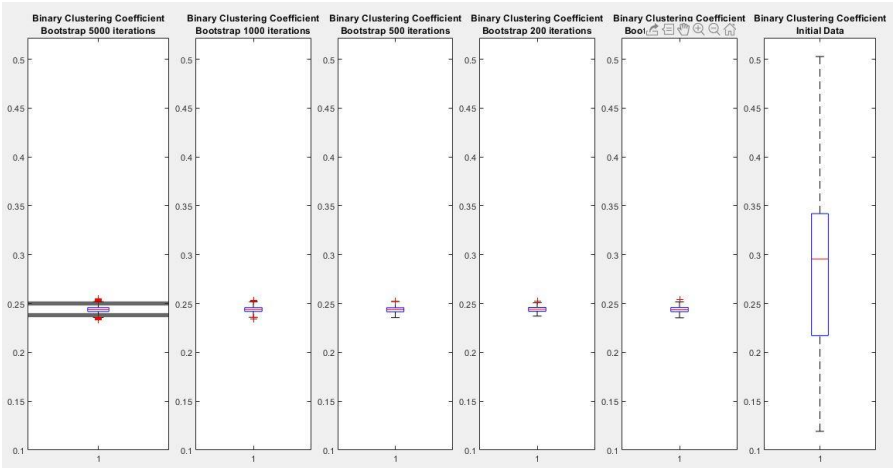


Figure 39 *Distributions of the* ***Binary Clustering Coefficient*** *considering, from right to left, the initial 10 functional connectivity matrices, 100, 200, 500, 1000 and 5000 bootstrap iterations.*

In view of the Weighted Clustering Coefficient, defined as the average "intensity" (geometric mean) of all triangles associated with a node, the results are in line with those obtained with the Binary Clustering Coefficient. Indeed, the distribution of the assumed values, shown in *Fig.40*, shrinks with respect to the increasing of the number of iterations involved in the Bootstrapping procedure. However, as justified by the numerical results in *Table.2*, a bias is present (17.582%) similar to the binary index.

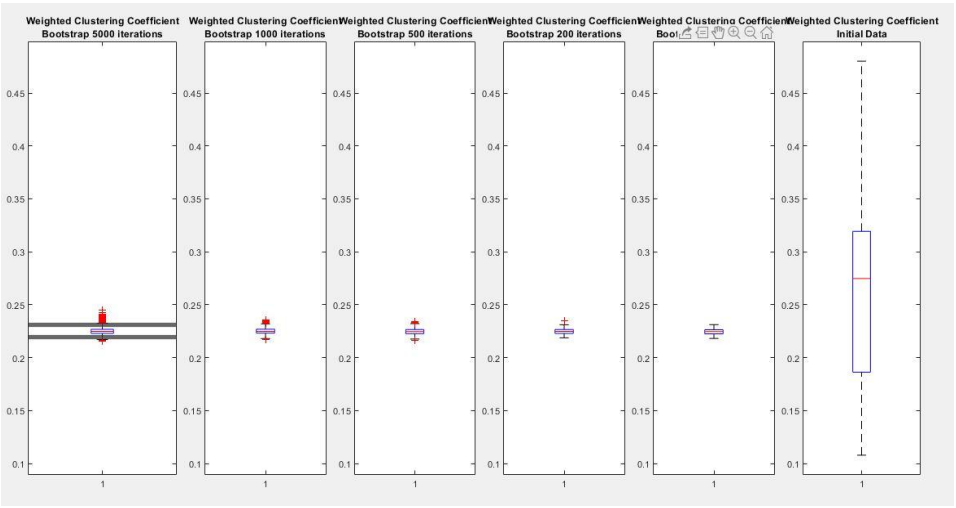


Figure 40 *Distributions of the* ***Weighted Clustering Coefficient*** *considering, from right to left, the initial 10 functional connectivity matrices, 100, 200, 500, 1000 and 5000 bootstrap iterations.*

In general, the numerical results reported in *Table.2* concerning the confidence interval boundaries calculated with respect to the Percentile Method and the Standard Formula for all the considered indexes are also shown. It is possible to notice how the standard method produces narrower intervals for all the graph metrics, highlighting the possibility of reaching a standard value for the indices.

4.1.2 Probabilistic Bootstrapping Results

The first step of the probabilistic normalization procedure was the density analysis, which was performed to deal with the bias issue reducing the variability of the indexes distribution, as described in section 3.3.4. The result of this analysis is represented in *Fig.41*. It is worth noting that the maximum density, the one at the initial step is of about 0.5 (the leftmost point in *Fig.41*), and then this density decreases as the percentage of times that an element is removed with respect to its presence in the dataset increases. The chosen value of the probability threshold (shown on the x-axis) was 0.3, such that the i^{th},j^{th} elements which are present less than 3 times in the dataset composed by 10 subjects were reset. This value produces a reduction of the mean density of the healthy functional connectivity dataset only from about 0.5 to 0.42 by the removal of the rarest connections among all subjects.

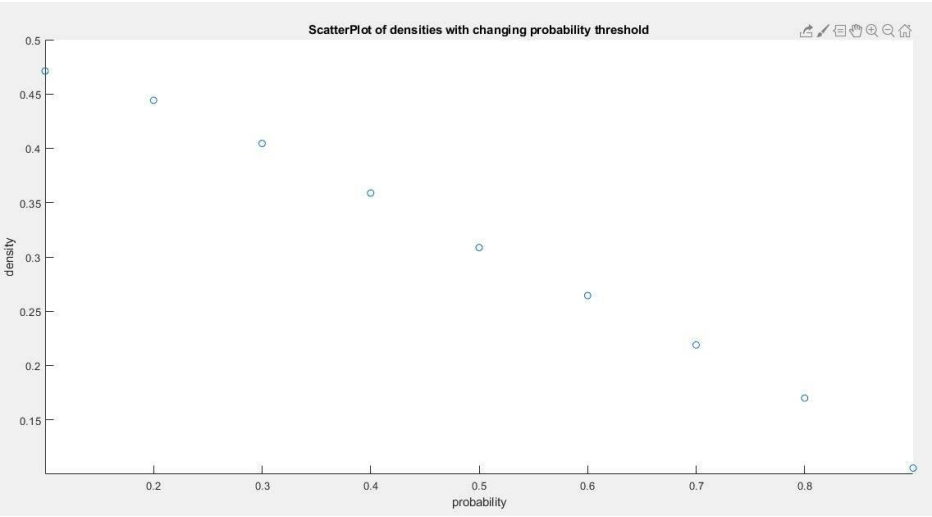


Figure 41 *Scatterplot of the density of the dataset (y-axis) with changing probability threshold (x-axis).*

The results are described and shown as in section 4.1.1. The table with all indexes is first reported, then the distributions for each investigated index are discussed (Node Degree, Node Strength, Binary and Weighted Clustering Coefficient).

Graph Indexes ***Probabilistic Normalization Bootstrap***	Mean Values and Standard Deviation		Confidence Intervals		Bias Percentage
	Initial Data	ProbNorm Bootstrap	Percentile Method	Standard Method	
Node Degree	39.433 ± 1.758	39.008 ± 0.323	[38.334; 39.644]	[38.999; 39.017]	1.089%
Node Strength	12.998 ± 2.526	13.118 ± 0.153	[12.815; 13.418]	[13.115; 13.123]	0.914%
Binary Clustering Coefficient	0.237 ± 0.040	0.195 ± 0.003	[0.189; 0.201]	[0.194; 0.195]	17.722%
Weighted Clustering Coefficient	0.226 ± 0.044	0.206 ± 0.006	[0.196; 0.220]	[0.205; 0.207]	8.765%

Table 3 *Numerical results of the considered graph metrics distributions before and after the Probabilistic Normalization Bootstrap, Confidence Intervals and the Bias of the Mean considering Before and After the Bootstrapping. The boundaries of the CI are computed as described in Section 3.3.4 and reported within square brackets. The bias percentage is calculated as exhibited in Section 3.3.4.*

First, considering the Node Degree, in *Fig.42 (top)* it is possible to notice that the results are more accurate with the probabilistic normalization in terms of the reduction of the variability of the distributions. Indeed, contrary to what shown in *Fig.37*, where the width of the boxplot slightly increased with respect to an increasing number of iterations; in this case, the boxplot shrinks as the number of iterations grows. With respect to the numerical results shown in *Table.3*, compared with the ones shown in *Table.2*, it is noticed that the Node Degree with the probabilistic normalization assumes lower values due to the resetting of the rarest connection. Indeed, since the degree is defined as the number of links connected to a node, a lower degree will be obtained. Besides, the bias remains negligible, as shown by its value of about 1%. The Node Strength, also shown in *Fig.42(bottom)* presents a negligible bias (0.914%) and a clearly visible shrinkage almost unchanged with the number of surrogate extractions.

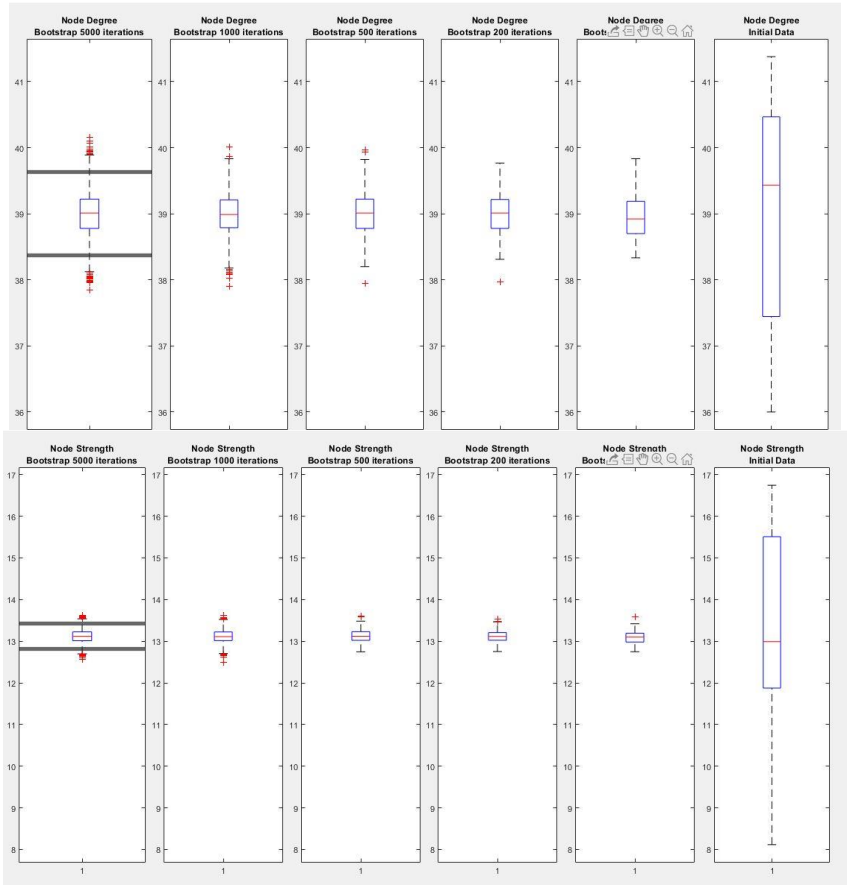


Figure 42 *Distributions of the* ***Node Degree (top panel)*** *and the* ***Node Strength (bottom panel)*** *obtained with the Probabilistic Bootstrapping technique considering, from right to left, the initial 10 functional connectivity matrices, 100, 200, 500, 1000 and 5000 bootstrap iterations.*

Also concerning the Binary and the Weighted Clustering Coefficients with the Probabilistic Normalization (shown in top and bottom panels, respectively, in *Fig.43*), the results regarding the shrinking of the distributions behaved as expected with the increasing of the number of surrogate samples. However, the bias for the Binary Clustering Coefficient is comparable to that in the standard bootstrapping case (17.288% in the Standard Bootstrapping, and 17.722% in the Probabilistic case). On the other hand, the bias of Weighted Clustering Coefficient reduces from 17.582% in the Standard case to 8.765% in the Probabilistic one.

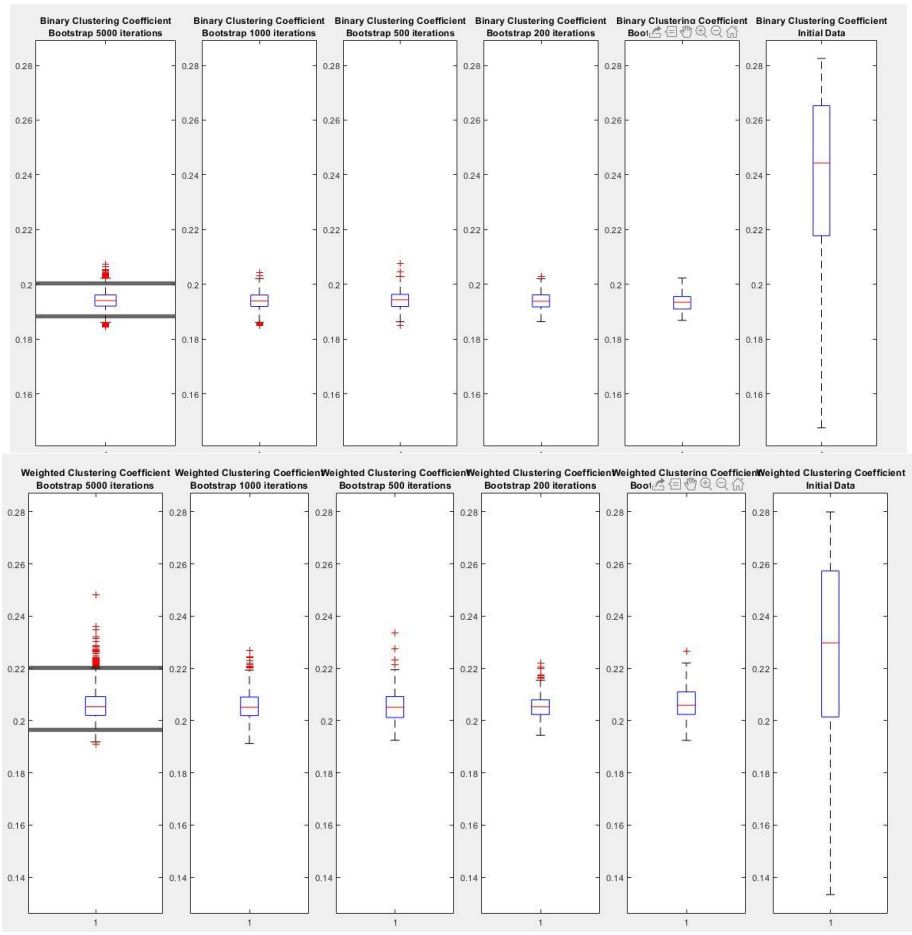


Figure 43 *Distributions of the* ***Binary Clustering Coefficient (top panel)*** *and the* ***Weighted Clustering Coefficient (bottom panel)*** *obtained with the Probabilistic Bootstrapping technique considering, from right to left, the initial 10 functional connectivity matrices, 100, 200, 500, 1000 and 5000 bootstrap iterations.*

4.1.3 Discussion about Bootstrapping Results over healthy functional connectivity dataset

As a whole the bootstrapping allowed to quantify the dispersion and obtain more robust values of the indexes under investigation, both with the Standard procedure and the Probabilistic Normalization. This was confirmed by the reduction of the distributions dispersion, with a slight exception of the Node Degree in the Standard Bootstrapping, characterized by a marginal broadening of its distribution of about one or two links. Moreover, it was possible to notice that a

large increasing in the number of iterations used to create surrogate data does not coherently change the variability of the distribution assumed by the values of the indexes. Hence, we speculate that this might be due to the fact the number of iterations needed in the case of 10 subjects is lower than those selected. However, we acknowledge that there are considerable and well-known discussions among researchers about the optimal number of surrogate data (iterations) to generate [111], [112], [113] and that further and specific experimentation must be performed in future to optimize it.

Analyzing the graph metrics, it is possible to state that the index whose result is more accurate in terms of respecting the desired outcome was the Node Strength in both the Probabilistic and the Standard cases. On the other hand, the worst performing index in this perspective is the Binary Clustering Coefficient, given its non-negligible bias, not reduced by the Probabilistic Normalization approach.

When dealing with connectivity matrices, a common procedure is to absolutely threshold the matrix to lower the presence of low-weight/spurious connections of the network [6], [114]. However, absolute thresholding can generate some issues in the interpretation of the connectivity pathways due to the removal also of useful information [6]. To overcome the above limitations, an approach named as "density-based thresholding" was proposed, which adapts the threshold of each subject to maintain an equal number of connections, thus equalizing the density of the considered datasets [69]. In the procedure of density-based thresholding, however, having the same number of connections (same density) across networks means to include also low correlations which increase the possibility of random and noisy fluctuations in the networks. Besides, equalizing the densities of a control and a pathological subject could bring the loss of information (for example a lesion) since a reduced number of connections might be related to the pathology itself. Thus, the correct thresholding technique to be used in the context of brain connectivity analysis is still an open issue. In this work a thresholding method based on the probability that a connection is present in a population is proposed. Analyzing the proposed results, it seems to be a useful technique to both preserve information and reduce the problem of the bias in the investigated functional connectivity dataset formed by healthy subjects.

With regards to the results shown in the previous sections, it is possible to see an improvement in the distributions assumed by the indexes with the Probabilistic Normalization performed to eliminate the rarest connections of the dataset and to normalize the connectivity weights of the matrices to reduce their variability. Precisely, this procedure allows to reduce the widening of the distribution of the Node Degree and to lessen the bias referred to the Weighted Clustering Coefficient. In terms of Confidence Intervals, with both the Standard and

Probabilistic methodologies, the Bootstrap procedure granted to obtain appropriate boundaries with both the Percentile method and with the Standard Formula, where this last produced narrow boundaries for the Confidence Intervals.

4.2 Healthy Structural Connectivity Dataset Results

This Section exhibits the Standard and Probabilistic Normalization Bootstrapping results of the protocol described in Chapter 3 on the structural connectivity dataset formed by 17 control subjects acquired with the DTI technique.

This investigation was performed to analyze if the Bootstrap allowed to obtain the desired results in terms of variability of the indexes reduction, calculation of the confidence intervals and computation of the biases in a different scenario in terms of uncertainty of the data. The aim of the protocol was to have a more robust estimation of the data through surrogate data, that is independent from the different noise sources in the acquisition of functional and structural connectivity data.

The dataset regarding structural connectivity matrices was not thresholded to maintain all the available information and the procedures are the same as the ones described previously in this work in Sections 3.3.1, 3.3.2. The considered segregation indices were the Node Strength, the Weighted Clustering Coefficient, as before, plus the Path Length. The results are shown in the same way as Section 4.1 of this work, using 1000 iterations for the bootstrapping procedure instead of 5000 for the graph metrics. This was done for computational reasons regarding the calculation of the Path Length with the Brain Connectivity Toolbox (BCT) [2], which requires two steps: first, the definition of the distance matrix and then the calculation of the Path Length. In particular, the distance matrix contains the lengths of the shortest paths between all pairs of nodes as defined by the Dijkstra's algorithm [71] defined earlier in this work, in section 2.1. An entry, i^{th},j^{th} element of the matrix, represents the length of the shortest path from node i to node j. With this matrix the Path Length is, then, calculated. It is, thus, possible to notice the computational problem in calculating this index for 5000 surrogate samples which takes lot of time to run the Matlab code. Moreover, for our investigation purposes, the number of iterations (1000) suggested by the previous investigation of the functional connectivity dataset are sufficient.

4.2.1 Standard Bootstrapping Results

The results shown in this section exhibit the numerical values of the boundaries of the confidence intervals obtained with the Standard Formula and the Percentile Method. Furthermore, the bias between the means with the Standard Bootstrapping procedure without any manipulation of the initial dataset (no thresholding and no normalization) is shown. All the numerical results are summarized in *Table.4*.

Graph Indexes ***Standard Bootstrap***	Mean Values and Standard Deviation		Confidence Intervals		Bias Percentage
	Initial Data	Standard Bootstrap	Percentile Method	Standard Method	
Node Strength	3.008 ± 0.319	3.009 ± 0.021	[2.968; 3.053]	[3.008; 3.010]	0.033%
Weighted Clustering Coefficient	0.015 ± 0.002	0.013 ± $1.264 * 10^{-4}$	[0.011; 0.013]	[0.012; 0.013]	13.333%
Path Length	$8.376 \cdot 10^{-4} \pm 1.140 \cdot 10^{-4}$	$7.918 \cdot 10^{-4} \pm 1.748 \cdot 10^{-5}$	$[7.571 \cdot 10^{-4}; 8.257 \cdot 10^{-4}]$	$[7.907 \cdot 10^{-4}; 7.928 \cdot 10^{-4}]$	5.784%

Table 4 *Numerical results of the considered graph metrics distributions before and after the Standard Bootstrapping, applied on the Structural Connectivity dataset. Confidence Intervals and the Bias of the Mean considering Before and After the Bootstrapping are shown. The boundaries of the CI are computed as described in Section 3.3.4 and reported within square brackets. The bias percentage is calculated as exhibited in Section 3.3.4.*

Examining the numerical results of the Node Strength, it is possible to state that the index is characterized by a negligible bias, whereas the Weighted Clustering Coefficient is characterized by a greater bias (13.333%) than the other investigated metrics. Considering the Path Length, the numerical value of the bias of the mean shows that it is small (5.784%). Moreover, the numerical results concerning the

confidence interval boundaries. It is possible to notice that the Standard Method produces narrower intervals for all the graph metrics, than the Percentile Method.

Considering the distributions of the values assumed by the considered graph indexes, all of them are characterized by the desired outcomes in terms of the reduction of the variability of the distributions after the creation of surrogate data, as exhibited in *Fig.44*.

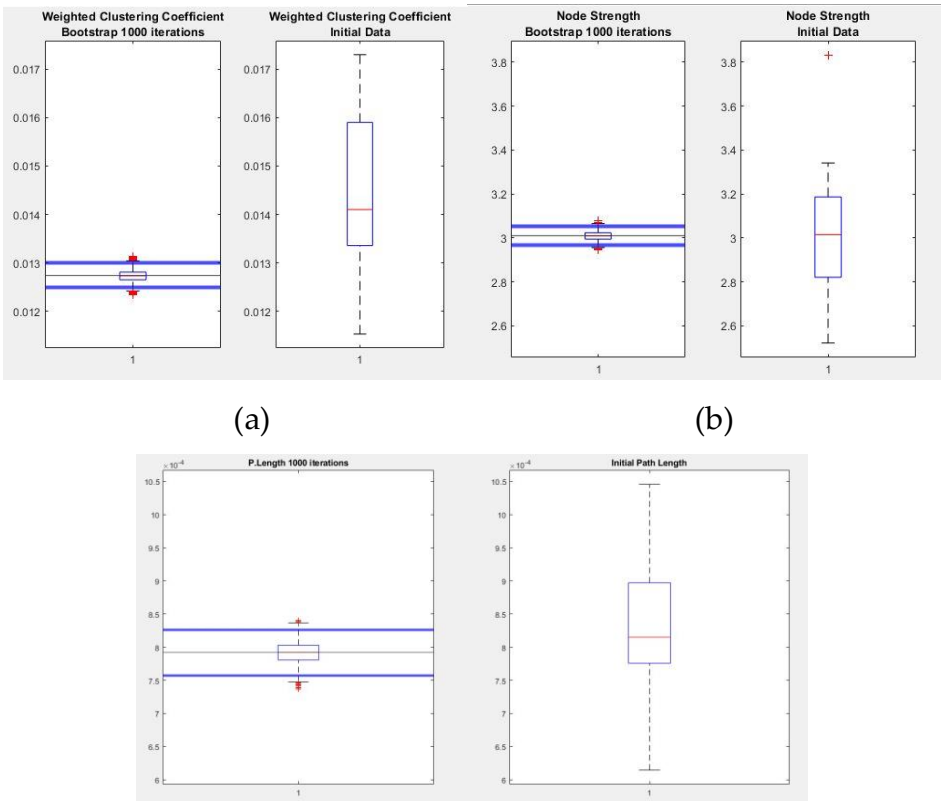


Figure 44 *Distributions of the* ***(a) Weighted Clustering Coefficient (top panel, left figure), (b) Node Strength (top panel, right figure) and (c) Path Length (bottom panel)*** *obtained with the Standard Bootstrapping technique regarding DTI structural connectivity data. The blue horizontal lines in the left graph displaying the boundaries of the Confidence Intervals obtained with the Percentile Method and the black horizontal line representing the mean value of the distribution of the index.*

4.2.2 Probabilistic Normalization Bootstrapping Results

The *Fig.45* shows the scatterplot created by putting on the x-axis the values that the probability matrix values from 0% to 100%, with a step of 10% between each one, and on the y-axis the mean density of the structural connectivity dataset that

results from the elimination of the connections which are present in less percentage with respect to each step. This is done, again, to find a threshold which allows to eliminate the rarest connections of the dataset without changing too much its density. This time, the threshold was chosen with respect to a lower elimination of the connections to analyze the performances of the Bootstrapping without manipulating too much the data. In particular, only the connections present less than 20% of the times with respect to the total 17 were removed.

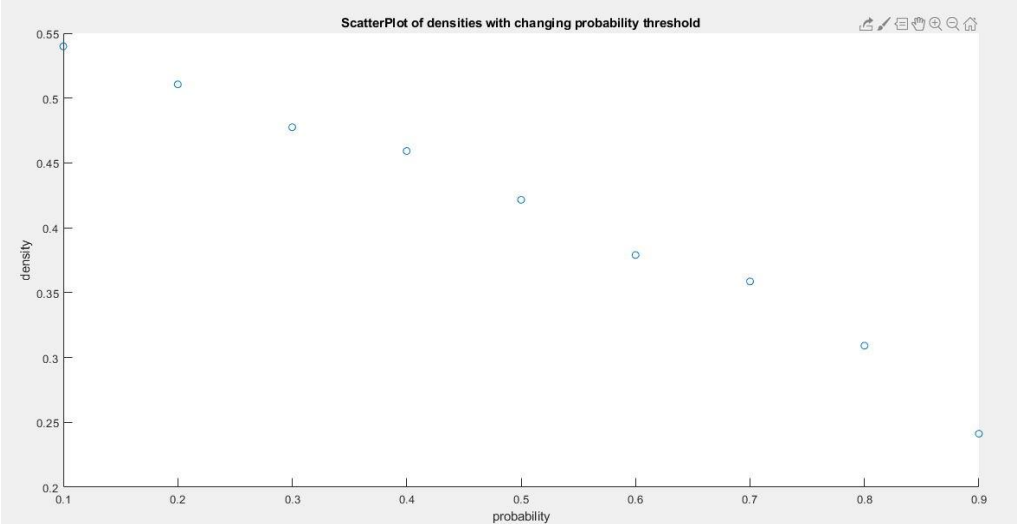


Figure 45 *Scatterplot of the density of the structural connectivity dataset with changing probability threshold.*

As can be seen in *Fig.45*, indeed, the density of the structural dataset passes from the initial 55% to 50% after the resetting of the rarest connections, while in the previous functional connectivity case, the decrease of the density due to the setting to zero of the least present connections was greater (from 50% to 42%).

The results about the table and the graph metrics boxplots, biases and confidence intervals are shown in the same way as in the previous section.

Considering the numerical results of the Node Strength, shown in *Table.5*, the bias is, again negligible also with the Probabilistic Normalization. Looking at the Weighted Clustering Coefficient, the bias slightly reduces with respect to the standard case, passing from 13.333% (*Table.4*) to 11.538% (*Table.5*). Conversely, in the case of the Path Length, the bias slightly increases (6.667%), with respect to the Standard Bootstrapping value (5.784%).

Graph Indexes ***Probabilistic Normalization Bootstrap***	Mean Values and Standard Deviation		Confidence Intervals		Bias Percentage
	Initial Data	ProbNorm Bootstrap	Percentile Method	Standard Method	
Node Strength	7.441 ± 1.089	7.439 ± 0.084	[7.276; 7.608]	[7.437; 7.441]	0.027%
Weighted Clustering Coefficient	0.052 ± 0.007	0.046 ± $4.724 * 10^{-4}$	[0.045; 0.047]	[7.437; 7.441]	11.538%
Path Length	0.015 ± 0.002	0.014 ± $4.055 \cdot 10^{-4}$	[0.013; 0.015]	[0.013; 0.014]	6.667%

Table 5 *Numerical results of the Probabilistic Normalization Bootstrapping procedure on the Structural Connectivity dataset of the considered graph metrics of the Mean Values Before and After Bootstrapping procedure together with their Standard Deviations, CI boundaries computed with the Percentile and the Standard Method within square brackets, and the Bias of the distributions.*

As in the Standard Bootstrapping case and with the previous functional connectivity dataset, the Node Strength is characterized, also with the Probabilistic Normalization performed before the Bootstrapping in terms of the negligible bias of the mean regarding before and after Bootstrapping, as exhibited numerically in *Table.5.*

Also in the Probabilistic Normalization case, the expected behavior of the distributions dispersion was found, as highlighted by the standard deviation and the Confidence Intervals (CI).

4.2.3 Comparison between Structural and Functional results and discussion

It is, thus, possible to state that the bootstrap on structural connectivity dataset, as for the functional one, allowed to reduce the variability of the values assumed by all the considered indexes. This result was highlighted by the Confidence Intervals values (computed with both the Percentile and the Standard Methods). Comparing the advantages of the Probabilistic Normalization in both the datasets,

in the functional one the improvements were clearer in the functional dataset, mainly in terms of the reduction of the bias of the Weighted Clustering Coefficient. In the structural case, the developments were reduced, probably due to the consideration of more rare connections and to the noisy and spurious connections characteristic of this technique. Indeed, before, in the functional connectivity case, the bias regarding the Weighted Clustering Coefficient dropped considerably, passing from 17.582% in the Standard Case to 8.765% with the Probabilistic Normalization. In the Structural Connectivity dataset, instead, the bias still reduced, but very slightly, passing from 13.333% in the Standard case to 11.538%.

In general, the Bootstrapping procedure allowed, again, to reduce the variability of the distribution of the values assumed by the indexes of interest (Node Strength, Weighted Clustering Coefficient and Path Length) also with the Structural Connectivity dataset. However, the probabilistic thresholding was able to remove the rarest connections, a practice which positively influenced the bootstrap results in terms of alignment of the distributions. From the comparison between the probabilistic normalizations performed on the functional and the structural connectivity datasets, the protocol resulted more robust and, thus, less characterized by biases when a more selective thresholding was performed. This, however, brings inevitably the loss of the information provided by weak connections, which was analyzed in literature to be important [68], [6].

In this context it is possible to state that every connection, even having low weight is important to evaluate the whole topology. In particular, the importance of the weak connections was analyzed at different levels of absolute thresholdings in the next section (4.2.4) by comparing DTI and HARDI.

On the other hand, the contribution of the outlier weights was considered in the Sensitivity Analysis to explain the great variability of the graph indexes, whose results are shown in Section 4.3.1.

4.2.4 HARDI/DTI COMPARISON

To further investigate the information coming from the structural connectivity matrices, the same 17 healthy subjects considered before, and acquired with the DTI technique, were explored with respect to the HARDI technique. This comparison had the aim of analyzing the benefits of the HARDI technique, described before in section 1.3.5, mainly in the solving of the crossing fiber

problem and to have more robust tractography data. It is worth remarking that the DTI and the HARDI tractographic reconstructions were generated by the same DWI scans performed on a high number of directions (64) [4]. This permitted both the simplified DTI modelling and the more complex HARDI analysis by spherical deconvolution. In this way, DTI provided a single streamline direction per voxel, while HARDI could disentangle the two directions in voxels with crossing fibers. This allows to obtain more accurate reconstructions of complex crossing fibers configurations characterized by high curvatures, as can be seen previously in *Fig.13*. The differences throughout the data acquired with the two different techniques and the different connections detected by the two methods were analyzed. In particular, the purpose was to find the connections detected by HARDI and not by DTI and the ones detected by DTI but not by HARDI using SNT. In order to accomplish this, some operations (normalization for number of streamlines, thresholding etc.) and subtraction between the two matrices were performed as reported in Section 3.4.5. SNT was used to obtain the connectograms referred to the Left and Right Temporal Lobe for each of the 17 Resulting Matrices. The temporal lobe was selected since DTI often loses several connections within this region due to the crossing fiber issue between the corpus callosum and the cortico-spinal tract [104]. In *Fig.46, Fig.47, Fig.48* the connectograms extracted from the Subject 5 of 17 matrices describing the connections detected by HARDI and not by DTI (on the left), and the ones detected by DTI and not by HARDI (on the right) starting from the thresholded structural connectivity matrices at 30% (*Fig.46*), 20% (*Fig.47*) and 10% (*Fig.48*), considering the Left and Right Temporal Lobes are shown. Furthermore, the density values of the selected subset of nodes are shown in *Table.6* (30% thresholding), *Table.7* (20% thresholding), *Table.8* (10% thresholding).

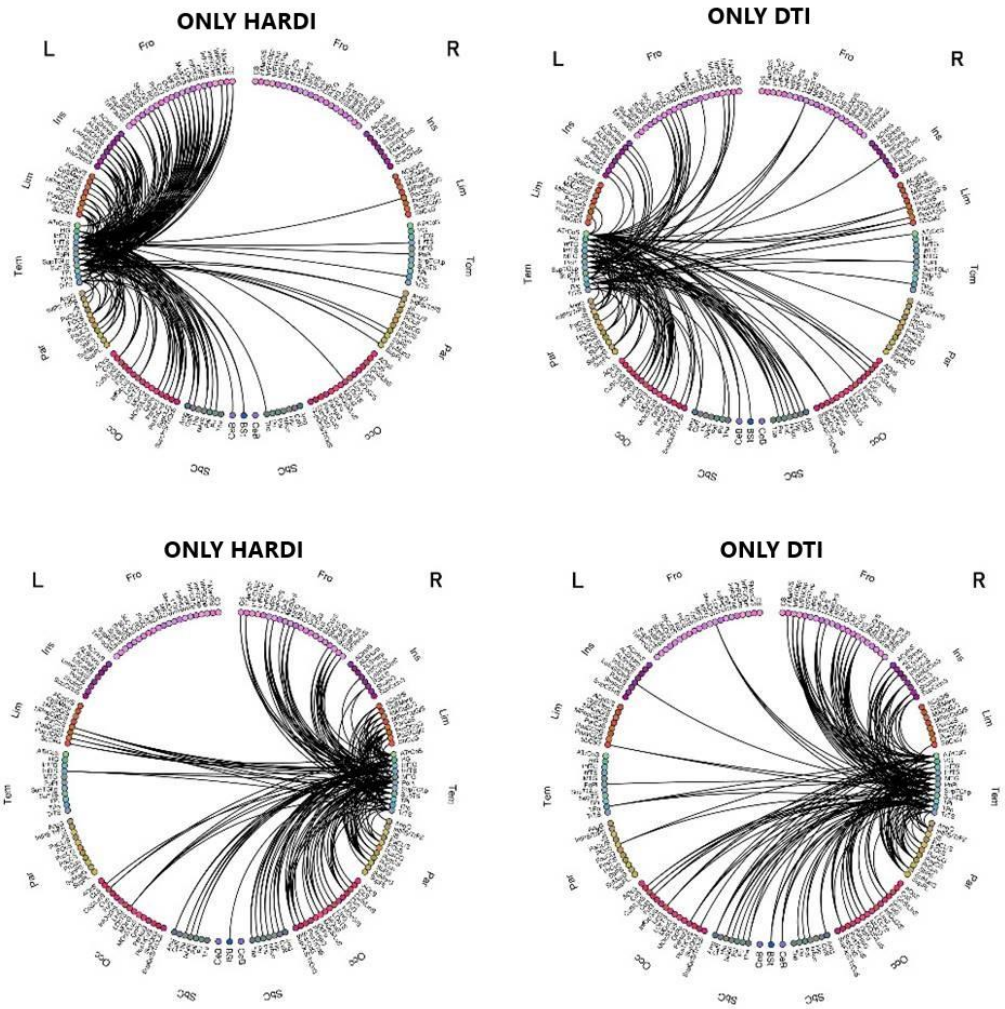


Figure 46 *Connectograms of the Subject number 5 showing the Left (top figures) and Right (bottom figures) connections detected only by HARDI (Left figures) and only by DTI (Right Figures) considering the thresholding at* ***30%*** *of both the structural connectivity datasets.*

30% Thresholding	***Only HARDI***	***Only DTI***
Density	*Left Temporal:* 1.38% *Right Temporal:* 1.28%	*Left Temporal:* 0.93% *Right Temporal:* 1.18%

Table 6 *Numerical values of the Density of the corresponding connectograms of the Subject number 5 shown in the previous figure (Fig.46).*

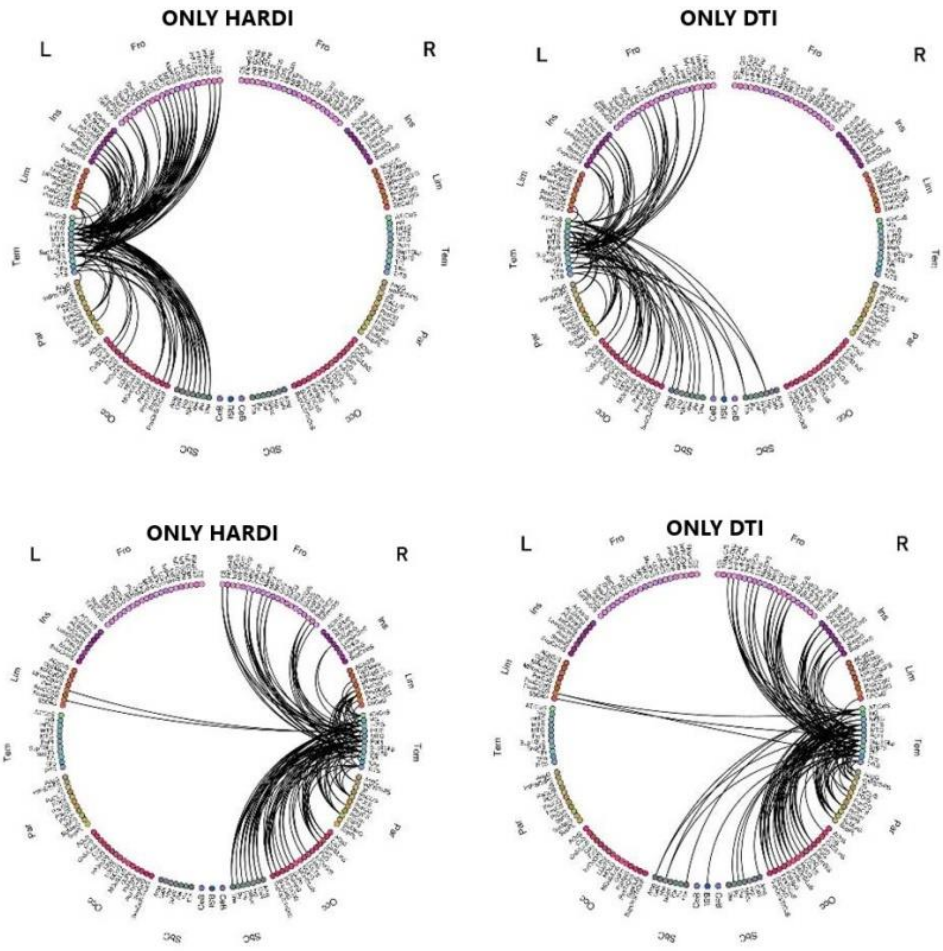


Figure 47 *Connectograms of the Subject number 5 showing the Left (top figures) and Right (bottom figures) connections detected only by HARDI (Left figures) and only by DTI (Right Figures) considering the thresholding at* ***20%*** *of both the structural connectivity datasets.*

20% Thresholding	***Only HARDI***	***Only DTI***
Density	*Left Temporal:* 0.87% *Right Temporal:* 0.87%	*Left Temporal:* 0.59% *Right Temporal:* 0.89%

Table 7 *Numerical values of the Density of the corresponding connectograms of the Subject number 5 shown in the previous figure (Fig.47).*

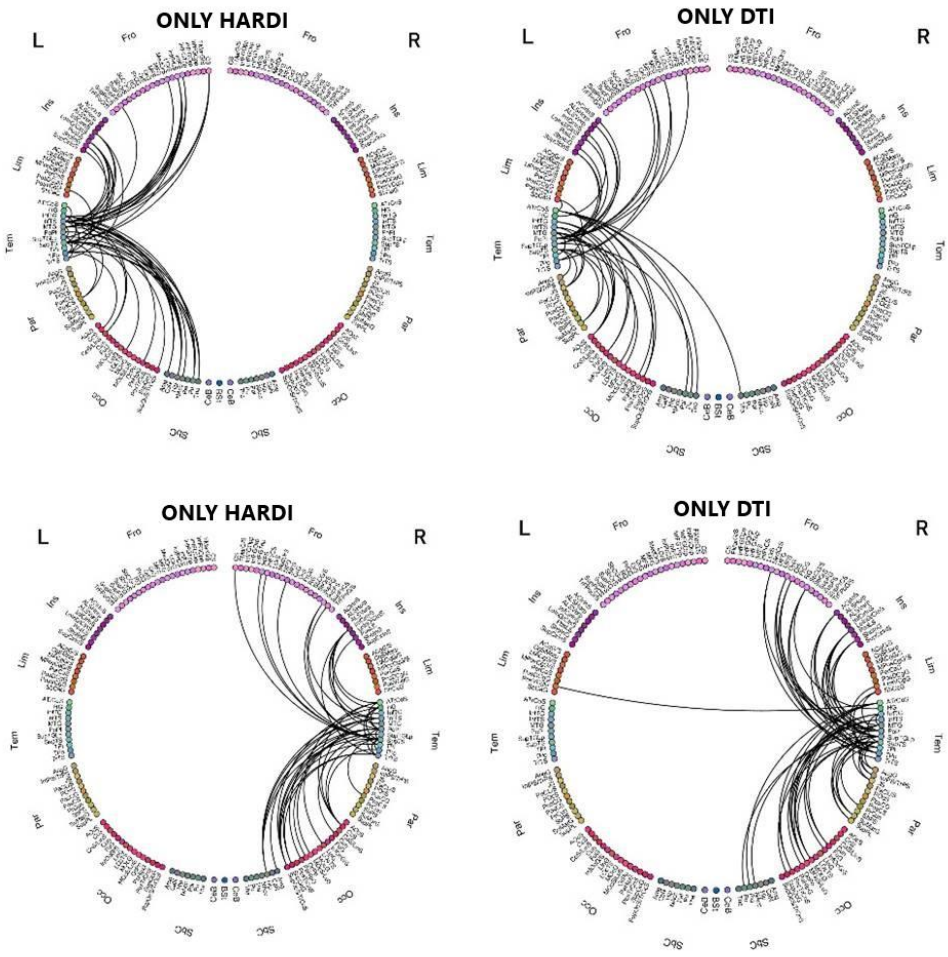


Figure 48 *Connectograms of the Subject number 5 showing the Left (top figures) and Right (bottom figures) connections detected only by HARDI (Left figures) and only by DTI (Right Figures) considering the thresholding at* ***10%*** *of both the structural connectivity datasets.*

10% Thresholding	***Only HARDI***	***Only DTI***
Density	*Left Temporal:* 0.37% *Right Temporal:* 0.32%	*Left Temporal:* 0.28% *Right Temporal:* 0.42%

Table 8 *Numerical values of the Density of the corresponding connectograms of the Subject number 5 shown in the previous figure (Fig.48).*

The previous figures and tables about the subject 5 highlighted the possibility of HARDI in detecting more connections regarding the left and right temporal lobe. Going in deeper into these kinds of analysis for all the 17 subjects acquired with both techniques, the results are controversial. Indeed, the results are opposed with respect to the desired ones, since the density of the temporal lobe subgraph was higher in the DTI case with respect to the HARDI one. This can be seen in *Table.9*,

showing the mean densities of the connectograms of the Temporal Lobe for all the 17 subjects processed with DTI and HARDI. In particular, all the values of the densities are greater in the DTI acquisitions.

30% Thresholding	***Only HARDI***	***Only DTI***
Mean Density	*Left Temporal:* 1.13% *Right Temporal:* 0.96%	*Left Temporal:* 1.14% *Right Temporal:* 1.17%
20% Thresholding	***Only HARDI***	***Only DTI***
Mean Density	*Left Temporal:* 0.74% *Right Temporal:* 0.58%	*Left Temporal:* 0.82% *Right Temporal:* 0.91%
10% Thresholding	***Only HARDI***	***Only DTI***
Mean Density	*Left Temporal:* 0.33% *Right Temporal:* 0.21%	*Left Temporal:* 0.41% *Right Temporal:* 0.48%

Table 9 *Table showing the Left and Right mean numerical results of the densities of the connectograms obtained only by HARDI and only by DTI for all subjects considering the thresholding at 30%, 20% and 10% of both the structural connectivity datasets.*

4.2.5 Discussion about the HARDI/DTI comparison results

The controversial results of the DTI/HARDI comparison were analyzed in further depth by referring to a study with the same aims conducted by Prckovska and colleagues [115]. In their study, graph-based measures were analyzed on the data acquired from 22 healthy subjects comparing the connectivity matrices obtained with DTI, HARDI and Diffusion Spectrum Imaging (DSI, not considered in the present study). The investigated graph matrices were: Global Efficiency, Node Strength, Density and Path Length. Highlighting the differences in these metrics comparing HARDI and DTI, it was found that HARDI showed higher Global Efficiency, Node Strength (evidencing an overall higher connectivity), Density (indicating a more sparse connectome in DTI than HARDI), while the Path Length was lower in HARDI with respect to DTI, as shown in *Fig.49(a), 50(a), 51(a), 52(a),*. In the figures, the results coming from DTI are represented in red, while those related to HARDI in green.

The same indices were calculated in this work on the population of 17 165x165 structural connectivity matrices. However, the results were in contrast with respect to the study provided by Prckovska and colleagues [115], as is represented in *Fig.49(b), 50(b), 51(b), 52(b)*, where the boxplot on the Left shows the results coming from DTI and on the Right the ones coming from HARDI. In particular, in the structural dataset considered in this work, Density, Global Efficiency and Node Strength were higher in the DTI acquisitions while the Path Length was higher in HARDI.

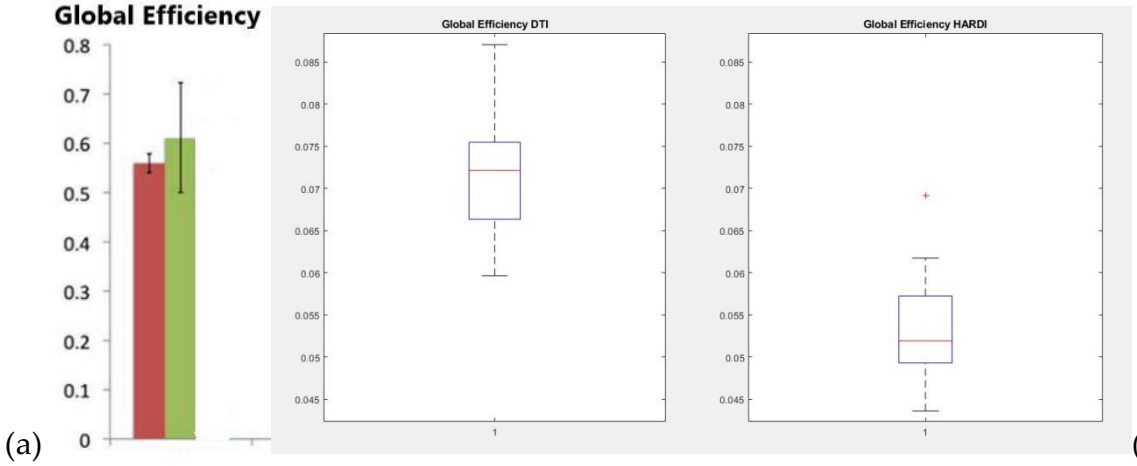


Figure 49 *(a),(b) Global Efficiency calculated on DTI and HARDI datasets. (a) Distributions of the Global Efficiency in the DTI (red) and HARDI(green) cases evaluated in the study provided by Prckosvka and colleagues* [115] *(b) Boxplots of the Global Efficiency in the DTI (left) and HARDI (right) cases evaluated on the 17 healthy structural connectivity matrices.*

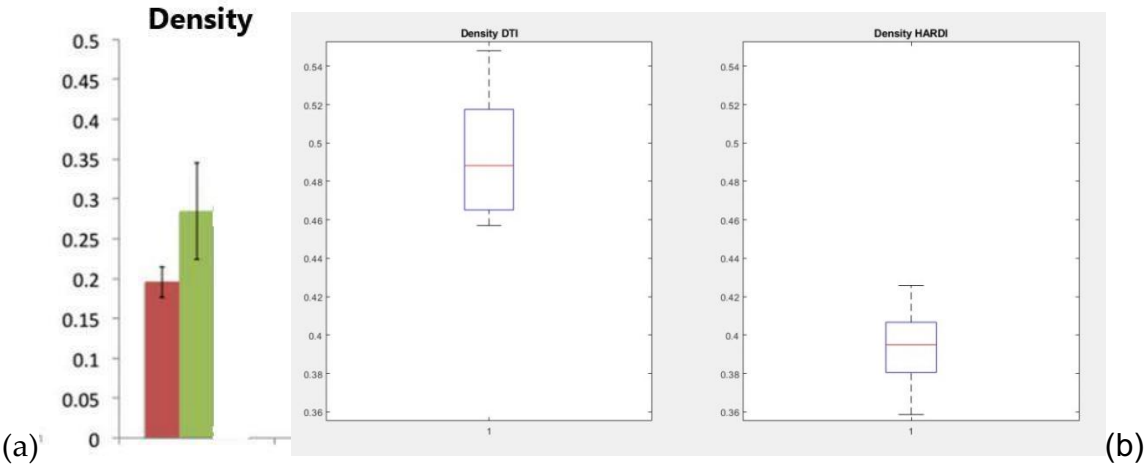
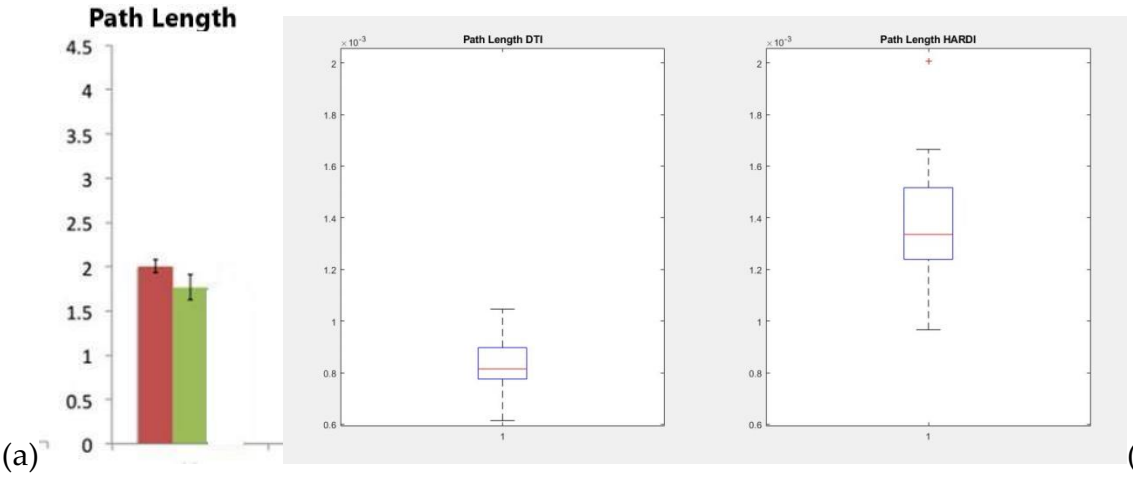
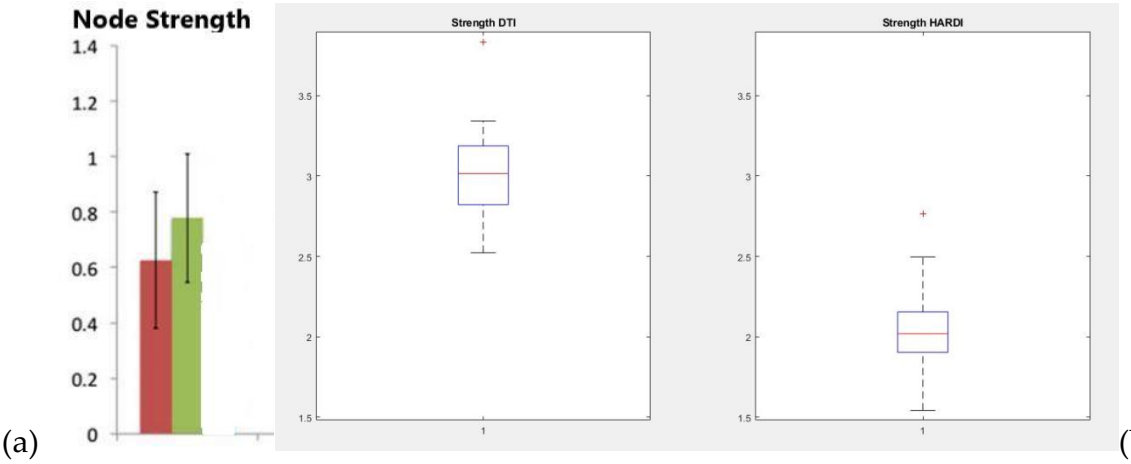


Figure 50 *(a),(b) Density calculated on DTI and HARDI datasets. (a) Distributions of the Density in the DTI (red) and HARDI(green) cases evaluated in the study provided by Prckosvka and colleagues* [115]*(b) Boxplots of the Density in the DTI (left) and HARDI (right) cases evaluated on the 17 healthy structural connectivity matrices.*



(a) (b)

Figure 51 *(a),(b) Path Length calculated on DTI and HARDI datasets. (a) Distributions of the Path Length in the DTI (red) and HARDI(green) cases evaluated in the study provided by Prckosvka* [115]*(b) Boxplots of the Path Length in the DTI (left) and HARDI (right) cases evaluated on the 17 healthy structural connectivity matrices.*



(a) (b)

Figure 52 *(a),(b) Node Strength calculated on DTI and HARDI datasets. (a) Distributions of the Node Strength in the DTI (red) and HARDI(green) cases evaluated in the study provided by Prckosvka and colleagues* [115]*(b) Boxplots of the Node Strength in the DTI (left) and HARDI (right) cases evaluated on the 17 healthy structural connectivity matrices.*

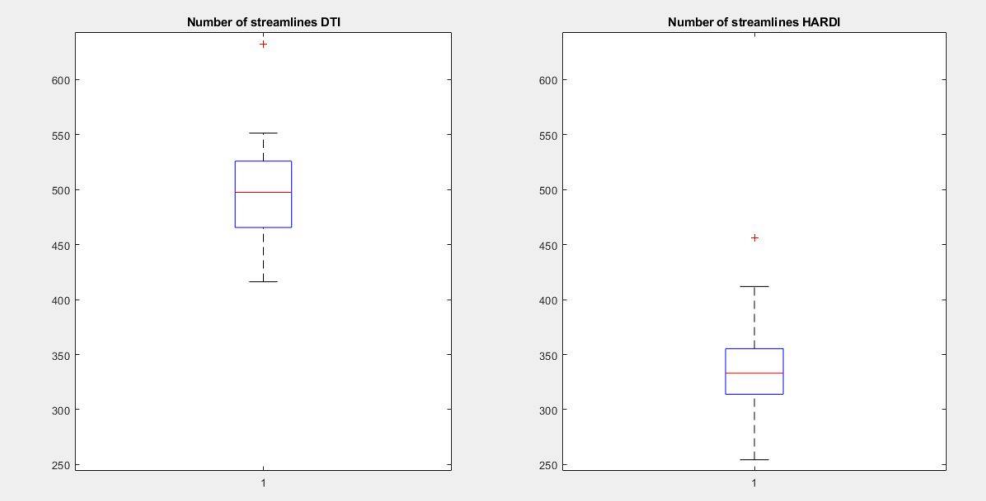


Figure 53 *Boxplots of the number of streamlines evaluated on the 17 structural connectivity matrices in the DTI (left) and HARDI (right) cases.*

Hence, our results were unexpected, highlighting the importance of having a coherent preprocessing pipeline in DTI and HARDI. Indeed, several preprocessing features can be treated in different ways, bringing several differences in the reconstruction of fibers, such as stopping criteria, subject-motion or background artifacts removal methods. [54], [116], [117], [118], [119], [98], [120].

In particular, HARDI connectivity matrices should be characterized mainly by a higher density and also a higher number of streamlines with respect to DTI, as can be seen in the example provided in *Fig.13* showing a far greater density and number of connections in the HARDI case than the DTI one. The number of streamlines is evaluated by summing all the connectivity weights in the matrices composing the two datasets regarding the same healthy subjects and, again, the results were not expected, due to the fact that the DTI dataset had greater number of streamlines with respect to the HARDI one, as shown in *Fig.53*. Further explorations are needed to verify the influence of the preprocessing pipelines in DTI and HARDI data

In conclusion, it is possible to state that the results are ambiguous. Indeed, the comparison of some subjects acquired with DTI and HARDI confirms the conclusions of the literature studies exhibited in Section 1.3.5, thus solving the crossing fiber problem. In particular the temporal lobe investigation allowed to individuate the connections with HARDI even if DTI often loses some connections in this region due to the crossing fiber problem between the corpus callosum and the corticospinal tract [104]. However, the results present some ambiguities. Indeed, in some other subjects, the HARDI advantages related to the detection of more connections by solving the crossing fiber problem are not evident. This is exhibited by the different results in terms of the graph metrics

investigated in the study provided by Prckosvka and colleagues [115] and the results acquired with a different subject of the dataset. Further explorations are needed to verify the influence of the preprocessing pipelines to reconstruct DTI and HARDI data.

4.3 Schizophrenic-Control Dataset Analysis

The final part of this work is related to the analysis of a dataset provided by A. Zalesky and colleagues [5], composed by fifteen healthy volunteers and 12 subjects with chronic schizophrenia. Schizophrenia is a critical mental disease characterized by behavior disorders, hallucinations and cognitive deficits; it is, thus, clear that this kind of sickness causes altered brain connectivity pathways which were captured thanks to the development of MRI [7].

Representing the functional connections between brain regions in terms of graph theory to analyze the topology of the network, the first step involved in the experimental protocol on this dataset was about the investigation of the results of the sensitivity analysis, as explained in Section 3.3.3.

4.3.1 Results & Discussion about the Sensitivity Analysis

The following scatterplots of the weighted indices described in section 3.3.3 (Weighted Clustering Coefficient, Node Strength, and Weighted Efficiency) are referred to two illustrative example cases from the control group and the schizophrenic one. The figures show the overlapping of the outlier weights with the weights influencing most of the graph indexes.

First, it is possible to notice that for all the indexes considered a high amount of the most influential connectivity weights are the outlier weights, confirming a great variability in the index distributions.

Then, the analysis of the weights outside the 10^{th} and 90^{th} percentile range showed that the control subject was characterized by a higher number of connections outside this range with respect to the schizophrenic one. Despite this, considering the same number of differences obtained from the Difference Matrix, as described in Section 3.3.3; more positions overlapped in the schizophrenic case than the

control one for the considered weighted indices. This can be, indeed, seen from the next figures (*Fig.54, Fig.55, Fig.56*). In particular, these scatterplots allow to appreciate how many overlaps occur for the considered graph metrics, considering the values obtained by resetting each of the weights outside the percentile range of interest and the positions saved from the greatest values of the Difference Matrix.

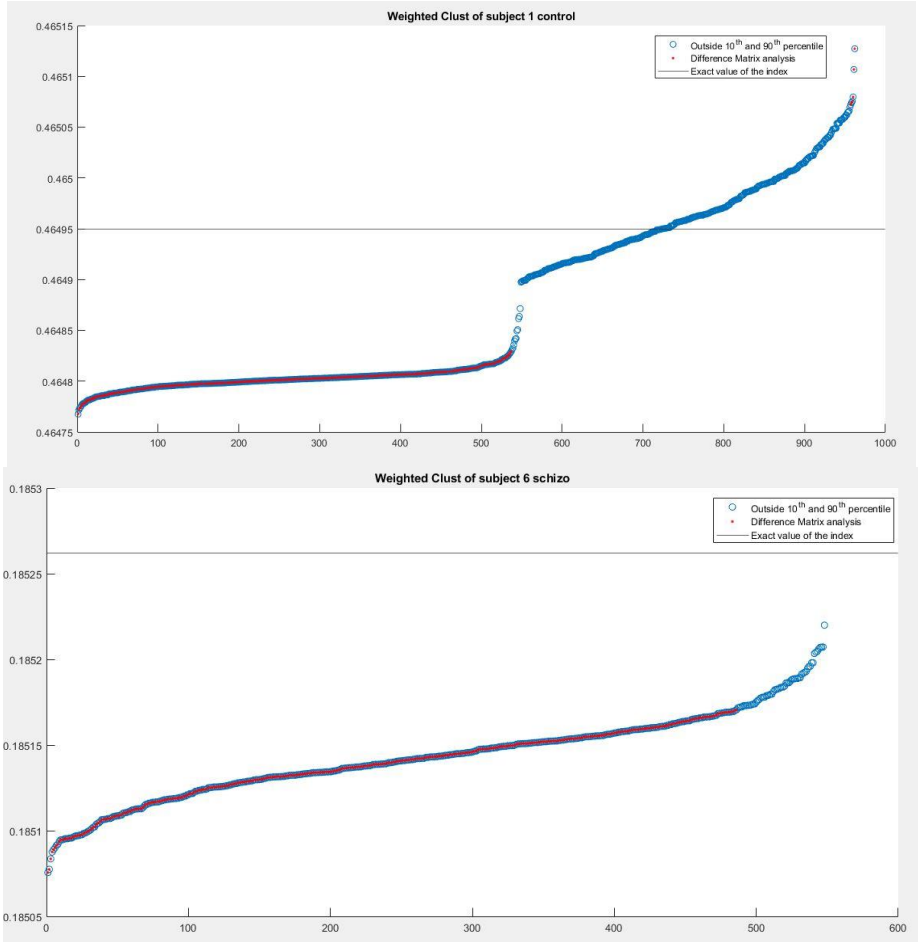


Figure 54 *Scatterplots of the Sensitivity analysis for the* ***Weighted Clustering Coefficient*** *for the first subject of the control group (top) and the sixth subject of the schizophrenic one (bottom). The red and the blue dots represent the value of the index after the removal of the connections with most deviating weights and greatest change, respectively. The black horizontal line representing the value of the metric computed for the specific matrix before any resetting of connections*

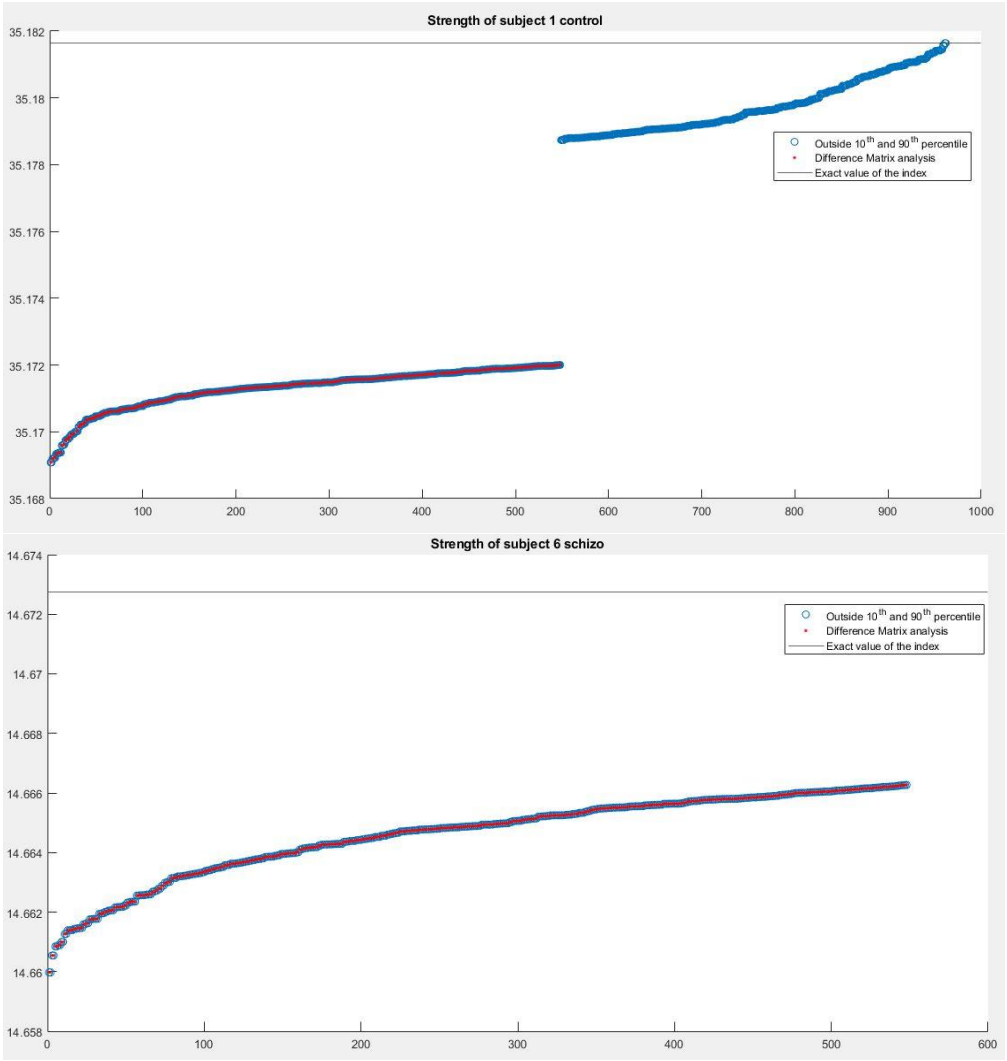


Figure 55 *Scatterplots of the Sensitivity analysis for the* ***Node Strength*** *for the first subject of the control group (top) and the sixth subject of the schizophrenic one (bottom). The red and the blue dots represent the value of the index after the removal of the connections with most deviating weights and greatest change, respectively. The black horizontal line representing the value of the metric computed for the specific matrix before any resetting of connections*

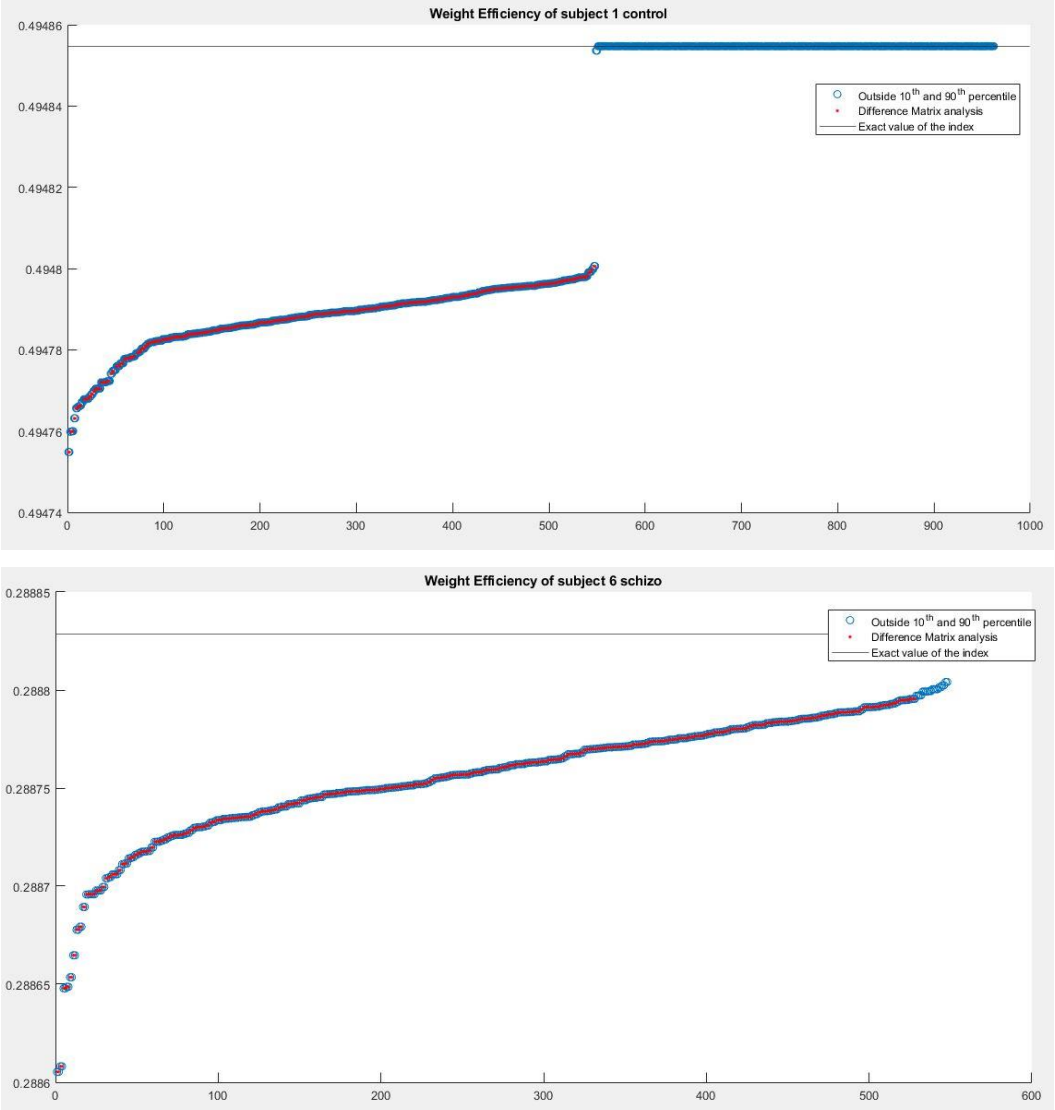


Figure 56 *Scatterplots of the Sensitivity analysis for the* ***Weighted Efficiency*** *for the first subject of the control group (top) and the sixth subject of the schizophrenic one (bottom). The red and the blue dots represent the value of the index after the removal of the connections with most deviating weights and greatest change, respectively. The black horizontal line representing the value of the metric computed for the specific matrix before any resetting of connections*

This analysis allowed to state that the outliers of the connectivity matrices, thus the variability of the weights, influence the variability of the values of the indexes. The sensitivity analysis results highlighted that more weights in the schizophrenic population will cause a change in the values of the considered indexes. It can be, thus, expected that the biases of the mean considering before and after the bootstrapping procedure would be greater in the schizophrenic population with respect to the control one, as confirmed by the results of the Standard and the Probabilistic Normalized bootstrapping techniques examined in the Section 4.3.3.

4.3.2 Graph Metrics Results & Discussion

The proposed indices (Weighted and Binary Clustering Coefficient, Strength, Degree, Weighted and Binary Path Length, Weighted and Binary Efficiency and Modularity) were calculated initially for both the populations and the distributions of the values assumed by these metrics are shown in the following figures (from *Fig.57* to *Fig.63*).

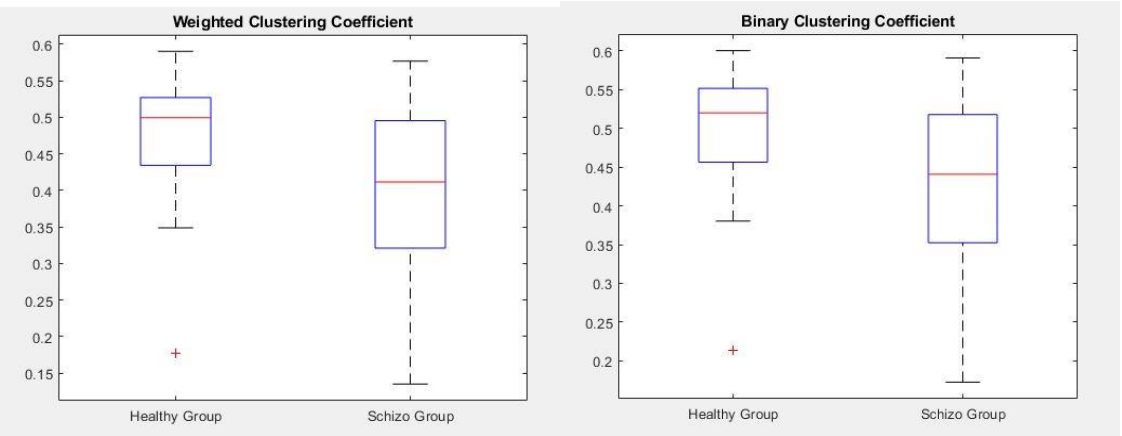


Figure 57 *Distributions of the* ***Weighted*** *(subplot on the left) and* ***Binary Clustering Coefficient*** *(subplot on the right) of the control (left) and the schizophrenic (right) groups.*

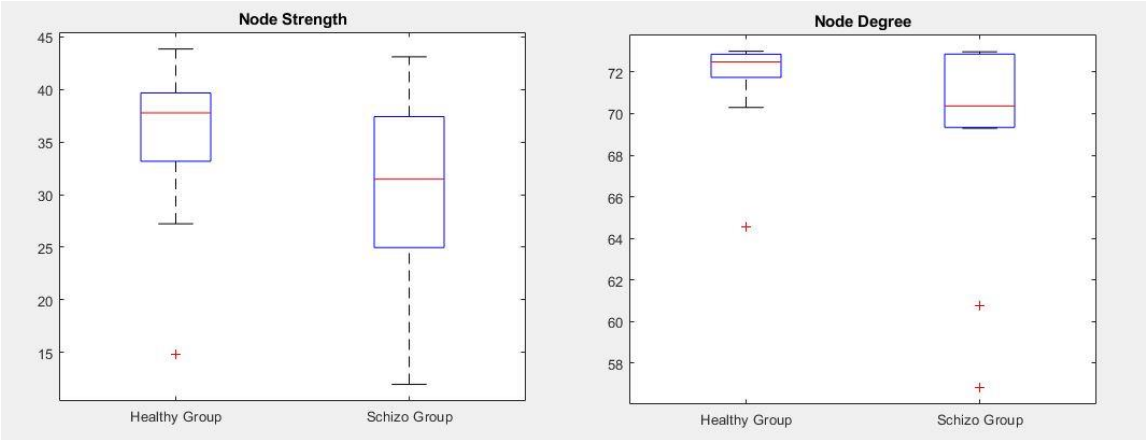


Figure 58 *Distributions of the* ***Node Strength*** *(subplot on the left) and* ***Degree*** *(subplot on the right) of the control (left) and the schizophrenic (right) groups.*

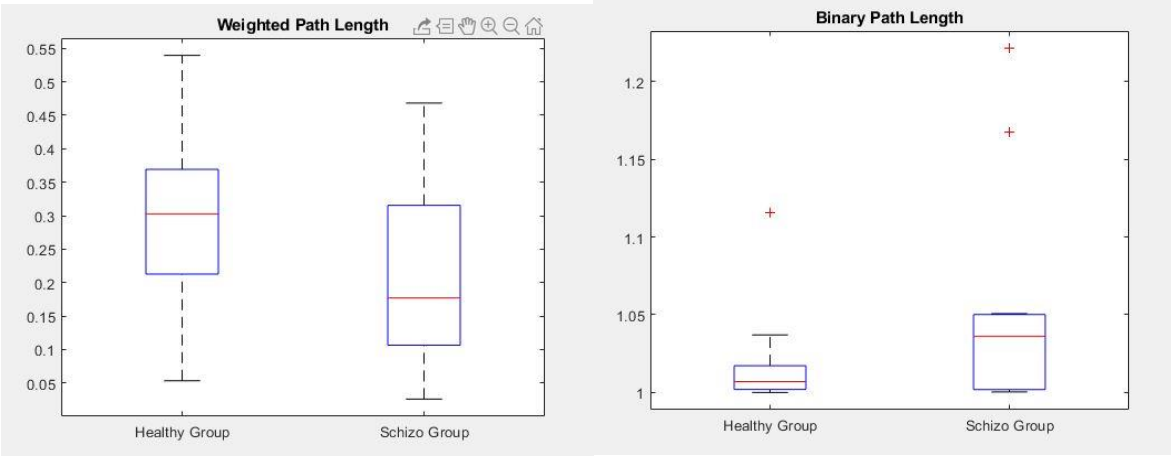


Figure 59 *Distributions of the* ***Weighted*** *(subplot on the left) and* ***Binary*** *(subplot on the right)* ***Path Length*** *of the control (left) and the schizophrenic (right) groups*

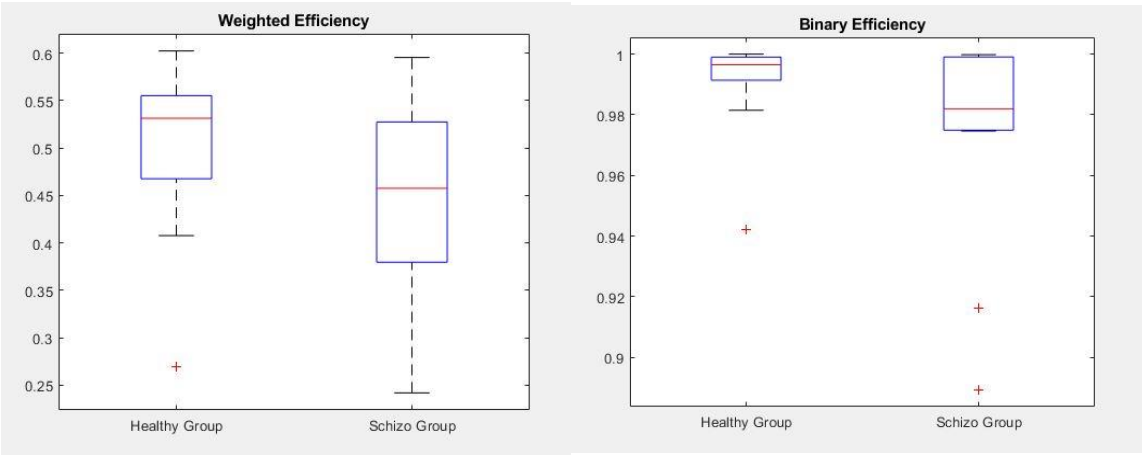


Figure 60 *Boxplots of the distribution of the* ***Weighted*** *(subplot on the left) and* ***Binary*** *(subplot on the right)* ***Efficiency*** *of the control (left) and the schizophrenic (right) groups*

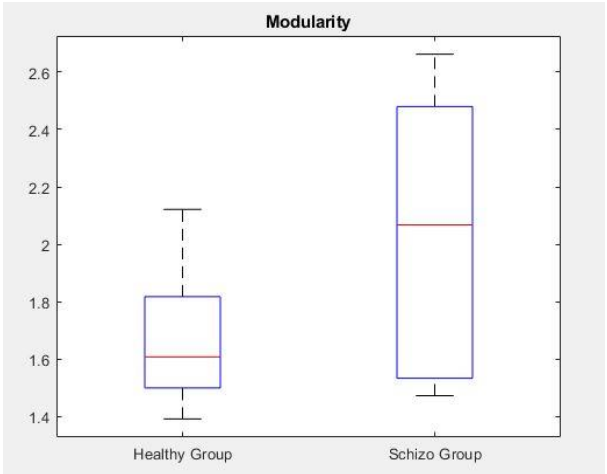


Figure 61 *Distributions of the* ***Modularity*** *(left) and the schizophrenic (right) groups*

The numerical results are summarized in *Table. 10*. In particular, the mean value of the indices is represented, along with the interquantile range (25^{th} and 75^{th} percentile) within square brackets, the Standard Deviation of the index and the p-value. It was obtained by performing a statistical t-test between the values assumed by the metric between control subjects and schizophrenics. The only statistically significant difference was found in the modularity index.

Graph-Based Indexes	Binary			Weighted		
Initial Data	Control	Schizophrenic	p-value	Control	Schizophrenic	p-value
Clustering Coefficient	*Mean*: 0.488 ± 0.096 [0.456; 0.552],	*Mean*: 0.423 ± 0.130 [0.353; 0.518],	0.162	*Mean*: 0.467 ± 0.102 [0.434; 0.527],	*Mean*: 0.397 ± 0.139 [0.321; 0.495],	0.156
Degree/ Strength	*Mean*: 71.750 ± 2.178 [71.74; 72.86]	*Mean*: 69.207 ± 5.149 [69.34; 72.86]	0.132	*Mean:* 35.389 ± 7.165 [33.17; 39.65]	*Mean*: 30.265 ± 9.858 [24.97; 37.40]	0.147
Path Length	*Mean*: 1.017 ± 0.030 [1.002; 1.017]	*Mean*: 1.050 ± 0.071 [1.002; 1.050]	0.132	*Mean*: 0.306 ± 0.141 [0.213; 0.369]	*Mean*: 0.217 ± 0.151 [0.106; 0.316]	0.111
Efficiency	*Mean*: 0.991 ± 0.015 [0.991; 0.999]	*Mean*: 0.974 ± 0.035 [0.975; 0.999]	0.132	*Mean*: 0.502 ± 0.083 [0.468; 0.555]	*Mean*: 0.446 ± 0.109 [0.380; 0.527],	0.158
Modularity	Na	Na	Na	*Mean*: 1.661 ± 0.197 [1.500; 1.812]	*Mean*: 2.016 ± 0.109 [1.534; 2.480]	***< 0.05***

Table 10 *Numerical results of the indices of the two populations (Control and Schizophrenic). In particular, for each index, are shown: Mean Value and Standard Deviation, Interquantile Range (25^{th} and 75^{th} percentiles) within square brackets) and p-value. Na: Index not available for binary/weighted cases.*

These indices were selected considering several findings of functional connectivity disturbances in schizophrenia, reflected in the alteration of the values of the graph metrics defined before. Indeed, for example, in the study conducted by Liu et al., 31 schizophrenic subjects matched by age with the control ones (mean age = 24 years) were acquired with on a 1.5 Tesla GE scanner in the Second Xiangya Hospital with a resting-state fMRI protocol analog to one used to obtain the data

provided by the Zalesky dataset [3]. In this context, the pathological group was characterized by lower Node Degree, Clustering Coefficients and Node Strength, while a higher Path Length was obtained in the schizophrenic group with respect to the control one. Moreover in another study provided by Alexander-Bloch and colleagues in 2010 [121], graph theory was used to investigate the topology of networks derived from resting-state fMRI data on 13 Childhood-Onset-Schizophrenia patients and 19 healthy volunteers, thus, younger subjects (mean age = 19 years) than the ones considered in the dataset under analysis in this work. Bloch and colleagues [121] concluded that the schizophrenic group was characterized by altered Modularity and reduced or unchanged Efficiency with respect to the control group. The same results about the values of the graph connectivity metrics described so far were also confirmed by the analysis of the study conducted by Yu and colleagues in 2012 on 24 schizophrenic and 24 control, age-matched subjects acquired on a 5 minutes resting-state fMRI procedure. Indeed, even in their study the control group had slightly higher connectivity strength, higher clustering coefficient and abnormal modularity [8]. The results of these studies highlight the fact that the schizophrenic networks are characterized by more segregation (lower clustering coefficients, lower degree, lower strength, lower or comparable efficiency, higher path length and altered modularity), reflected in the presence of heavier disconnections in the brain connectivity matrices of the pathological population.

It is possible to state that the same outcomes are obtained with this dataset. Indeed, considering Weighted and Binary Clustering Coefficients, Node Strength, Node Degree and Weighted and Binary Efficiency, it is clear to see the higher segregation of the schizophrenic networks reflected in lower coefficients with respect to the control group, as shown both figuratively in *Fig.57, Fig.58, Fig.60,* respectively, and numerically in *Table.10*. Moreover, bringing back the study conducted by Yu and colleagues [8], in this work, equally, Modularity is abnormal in the pathological group with respect to the control one. In particular, as can be seen by both *Fig.61* and *Table.10*, this index in the schizophrenic group is higher with respect to the control one. Recalling that the modularity measures the number of interconnections in the communities of the network, since there are more disconnections in the schizophrenic group, in keeping with more sparse communities, producing a higher value of the index. Concerning the Path Length, the results of the analysis over this dataset were slightly different with respect to the considered study provided by Liu and colleagues [3] where the considered metric was higher in the schizophrenic group. Indeed, as exhibited in *Fig.59* and *Table.10*, the metric is slightly higher in the pathological case with respect to the control one. In the weighted case the Path Length is greater in the control case

than the schizophrenic one, thus, not resulting in line with the reference study conducted by Liu et al [3].

4.3.3 Probabilistic Normalization Procedure

To further investigate the dataset, the bootstrapping procedure was applied to find the confidence intervals of the proposed graph indices. The creation of the surrogate data was performed as described in the dedicated section (3.3.1) of this work, such that, in the standard case, 1000 surrogate extractions were considered.

First, the density of the dataset obtained from both the populations was analyzed with respect to the resetting of the outlier connections. Analyzing these scatterplots, shown in *Fig.62*, it can be seen that the schizophrenic population is characterized by a higher number of disconnections since the density of the dataset presents an earlier decrease at the increase of the counter used to reset the connections present less times than the value of the counter.

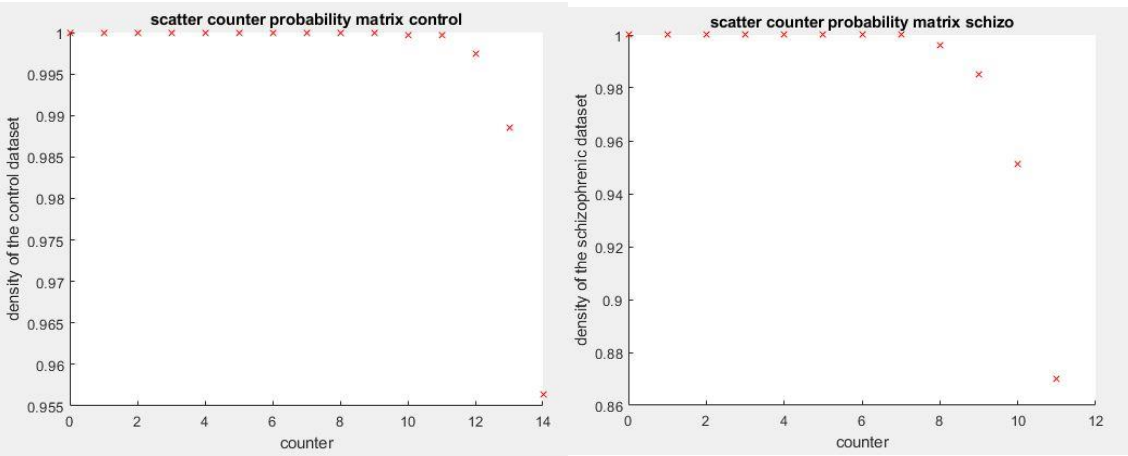


Figure 62 *Scatterplots describing the Density of the dataset for each of the reset counters in the control (left) and schizophrenic (right) datasets.*

Then, the normalization of the initial data was obtained by dividing each of the matrices composing the control and the schizophrenic datasets with respect to the Mean Matrix of each population, without any resetting of connections. Due to the fact that no connection was reset before applying the Normalization, the analysis of the indices is shown only in the Weighted cases (Weighted Clustering Coefficients, Node Strength, Weighted Path Length, Weighted Efficiency and Modularity).

4.3.4 Normalized Graph Metrics Results & Discussion

The distributions of the initial values of the weighted metrics after the normalization of the data are proposed in *Fig.63, Fig.64, Fig.65.*

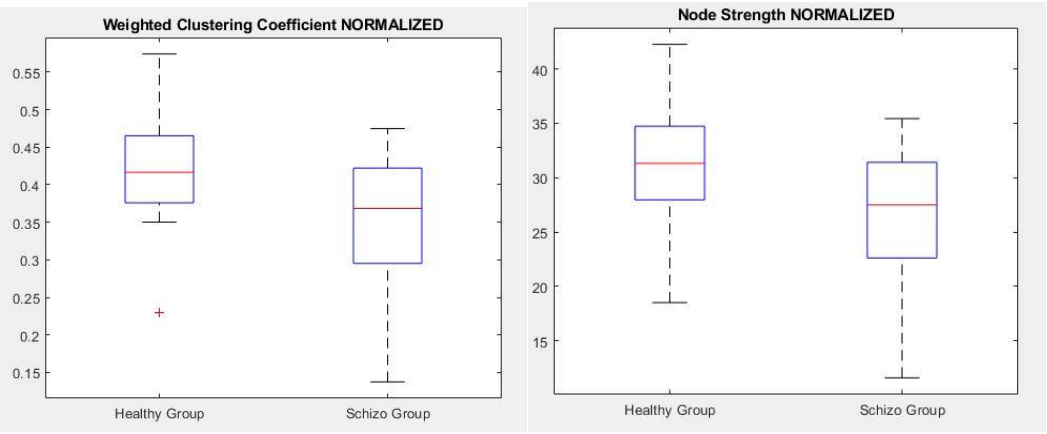


Figure 63 *Boxplots of the distribution of the* ***Weighted Clustering Coefficient (left figure) and Node Strength (right figure)*** *of the control (left boxplot) and the schizophrenic (right boxplot) groups after the normalization of the initial data*

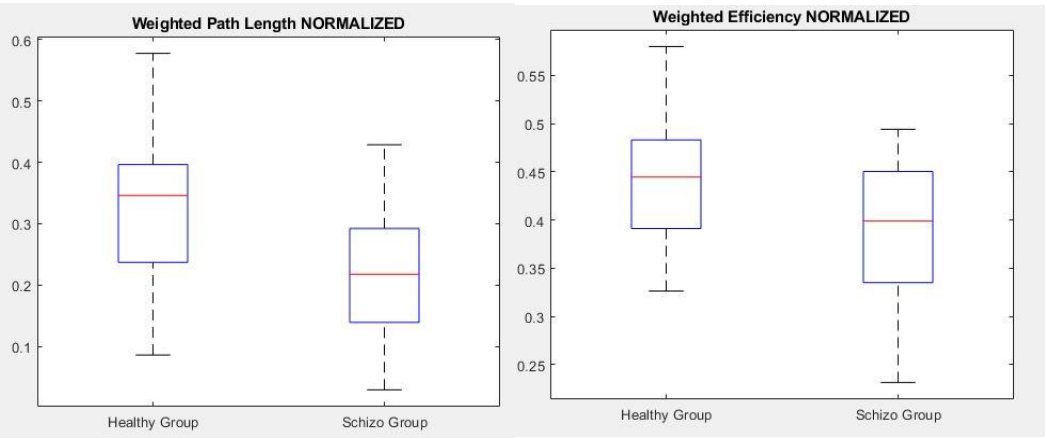


Figure 64 *Distributions of the* ***Weighted Path Length (Left) and Weighted Efficiency (Right)*** *of the control (left boxplot) and the schizophrenic (right boxplot) groups after the normalization of the initial data*

Figure 65 *Distributions of the* ***Modularity*** *of the control (left boxplot) and the schizophrenic (right boxplot) groups after the normalization of the initial data*

Then, all results and significant differences are summarized in *Table 11*.

Graph-Based Indexes	Weighted		
Normalization of data	Control	Schizophrenic	p-value
Clustering Coefficient	*Mean*: 0.419 ± 0.080 [0.376; 0.465]	*Mean*: 0.345 ± 0.103 [0.295; 0.422]	0.054
Node Strength	*Mean*: 31.404 ± 5.612 [27.95; 34.73]	*Mean*: 25.984 ± 7.338 [22.61; 31.42]	***< 0.05***
Path Length	*Mean*: 0.324 ± 0.116 [0.237; 0.396]	*Mean*: 0.220 ± 0.121 [0.139; 0.292]	***< 0.05***
Efficiency	*Mean*: 0.446 ± 0.066 [0.391; 0.483]	*Mean*: 0.385 ± 0.082 [0.335; 0.450]	***< 0.05***
Modularity	*Mean*: 1.496 ± 0.089 [1.449; 1.564],	*Mean*: 1.630 ± 0.251 [1.480; 1.689],	0.103

Table 11 *Numerical results of the indices of the two populations after the normalization of the data (Control and Schizophrenic). In particular, for each index, there are shown: Mean Values and*

Standard Deviation, Interquantile Range (25th and 75th percentiles) within square brackets, and p-values.

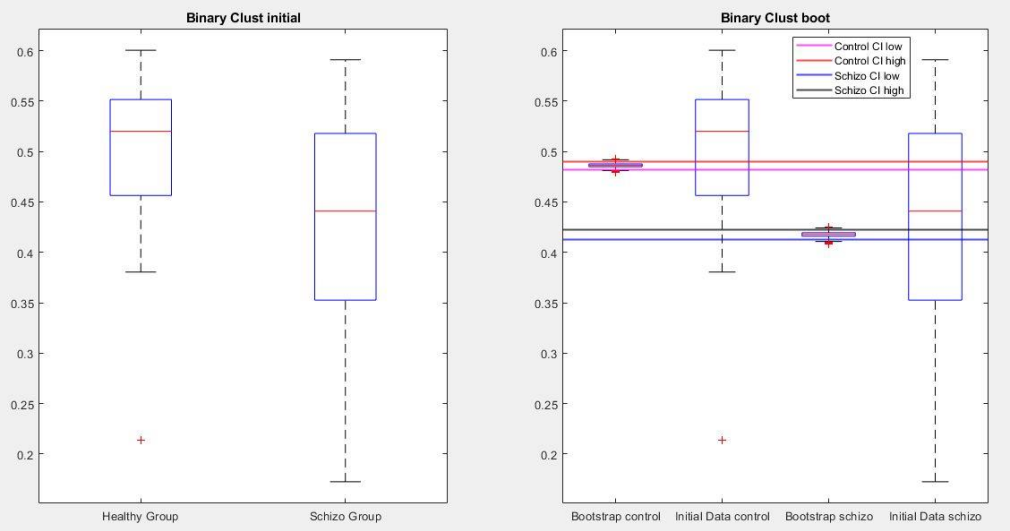
The table confirms the outcome of the previous analysis about a more segregated network in the schizophrenic case as can be seen by the lower Clustering Coefficient, Node Strength, and Efficiency, and higher Modularity, obtained also with the Normalization of the initial data. Again, however, the Path Length, which should have been higher in the schizophrenic population to have a more disconnected network, is slightly higher in the control case with respect to the pathological one. This difference is limited (0.346 in the control group vs. 0.218 in the schizophrenic one) though statistically significant. Seemingly, the higher segregation level confirmed in the schizophrenic group is not related to a loss in the connections weights but in a loss of the overall integration structure.

The normalization of the data also brings advantages in terms of the p-values. Indeed, as shown in *Table.11,* the majority of the p-values are below the threshold at 0.05 (Node Strength, Weighted Path Length and Weighted Efficiency). Without normalization of the data, instead, as shown in *Table.10*, only the Modularity p-value is less than 0.05, while the other weighted indices are characterized by p-value above the threshold, finding new significant differences.

4.3.5 Standard and Probabilistic Normalization Bootstrapping Results

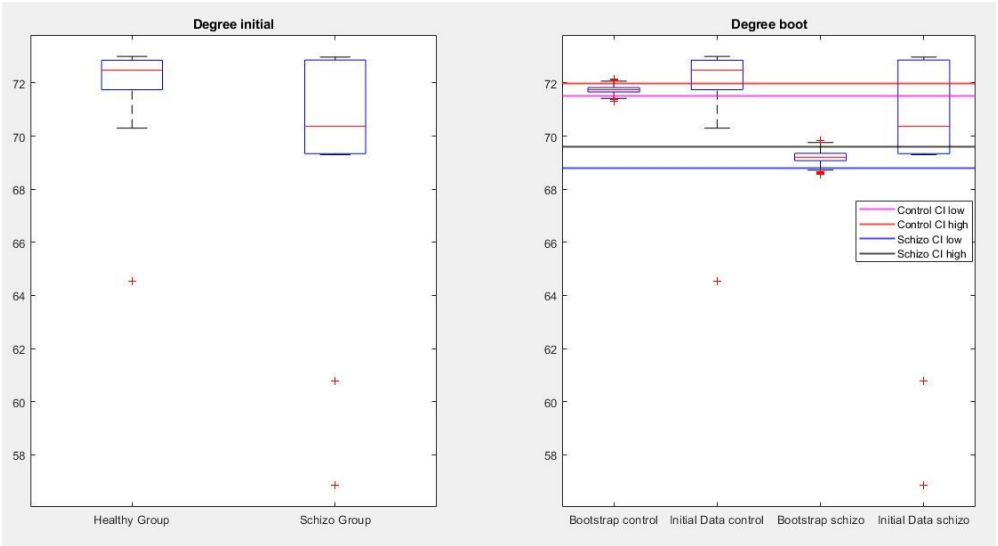
After the application of the bootstrapping procedure, in both cases (standard and normalized), the distributions of the considered indices were plotted. *Figures from 66 to 74 (a, left panels)* recall the previously shown boxplots of the initial distributions in the control group (left) and schizophrenic one (right). In the right panels in the *figures from 66 to 74 (b)*, instead, the bootstrap results are compared with the original distributions, shown again to enhance the bootstrap vs. original distributions both in controls (left) and schizophrenics (right). Confidence Intervals (CI) found by bootstrap are also shown. In particular, the red and the magenta horizontal line represent the upper and lower value (respectively) of the confidence interval (CI) in the control case . The black and the blue horizontal lines, instead, represent the upper and lower values (respectively) of the confidence interval in the schizophrenic group. In the following figures, firstly the binary indices are represented, thus, only the standard bootstrapping case is considered. Following the representation of the boxplots describing the binary indices, the weighted metrics are represented comparing the standard and the normalized bootstrapping procedures.

Binary Indexes (standard Bootstrapping):



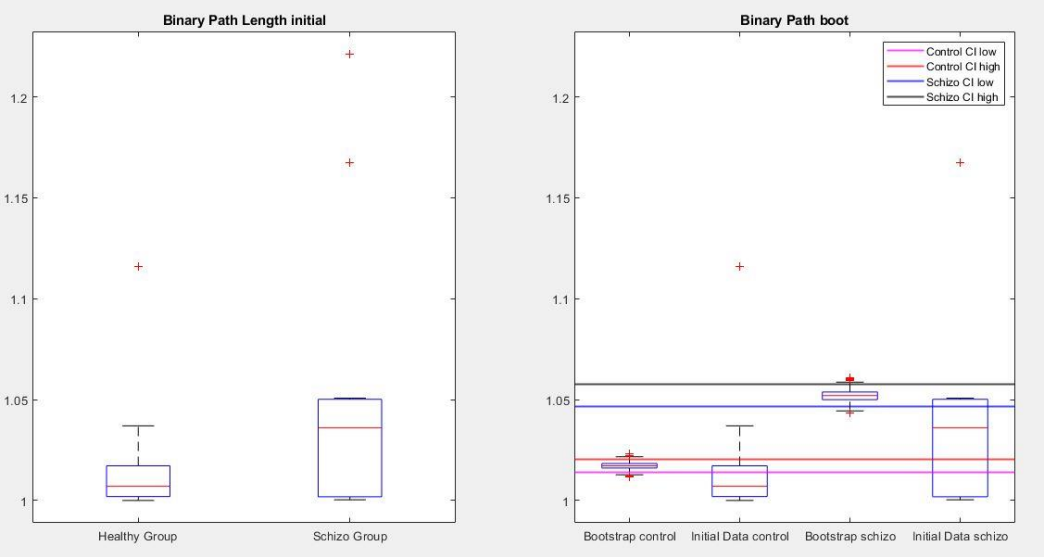
(a) (b)

Figure 66 *Distributions of the* ***Binary Clustering Coefficient*** *for the initial data* ***(a)*** *and after the bootstrapping procedure* ***(b)*** *with the boundaries of the confidence intervals calculated with the percentile method in both populations.*



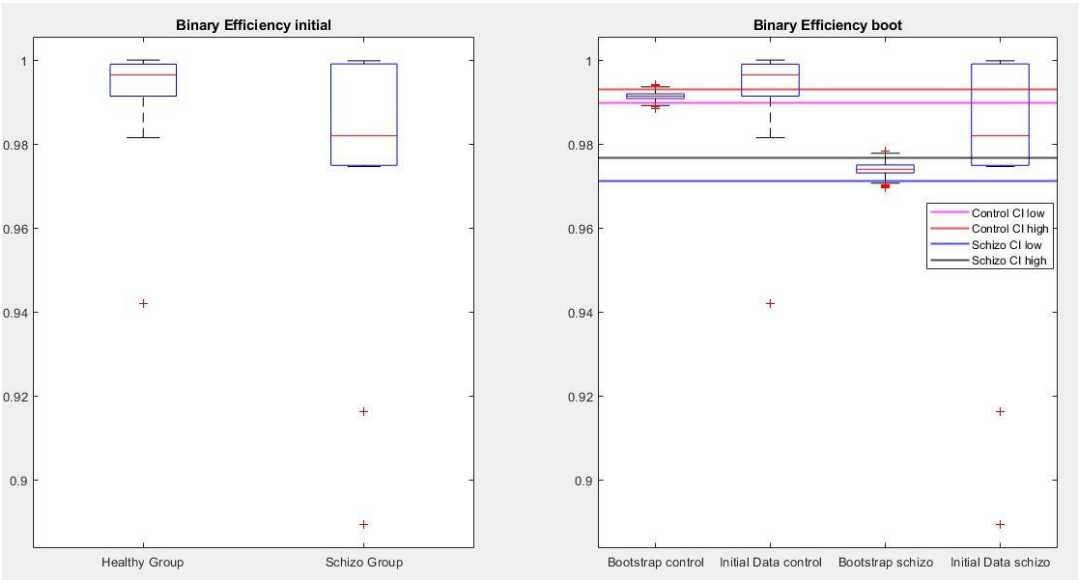
(a) (b)

Figure 67 *Distributions of the* ***Node Degree*** *for the initial data* ***(a)*** *and after the bootstrapping procedure* ***(b)*** *with the boundaries of the confidence intervals calculated with the percentile method in both populations.*



(a) (b)

Figure 68 *Distributions of the **Binary Path Length** for the initial data **(a)** and after the bootstrapping procedure **(b)** with the boundaries of the confidence intervals calculated with the percentile method in both populations.*



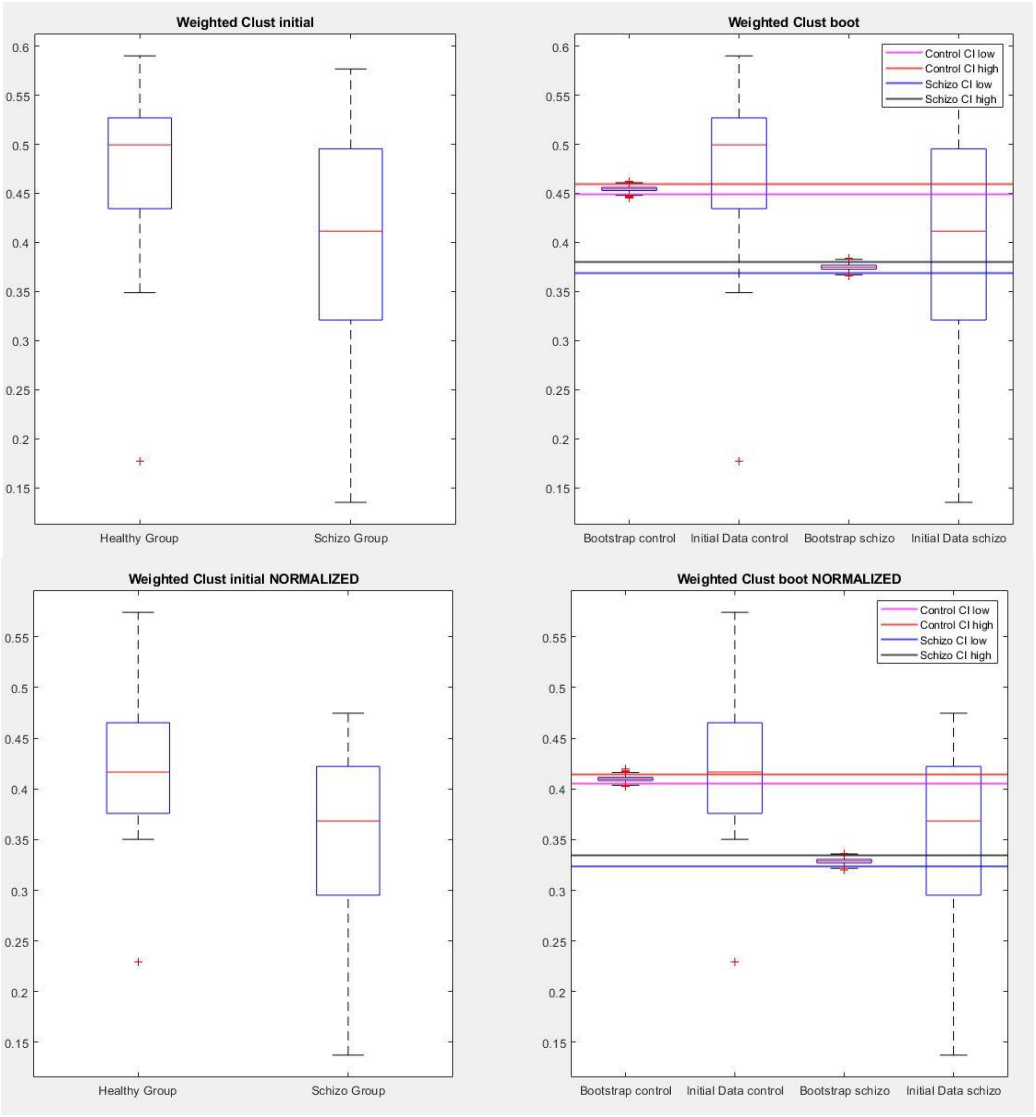
(a) (b)

Figure 69 *Distributions of the **Binary Efficiency** for the initial data **(a)** and after the bootstrapping procedure **(b)** with the boundaries of the confidence intervals calculated with the percentile method in both populations.*

The expected results concerning the boxplots were about a significant shrinking of the index distributions after the application of bootstrapping procedure, which is clearly seen in all the above figures. Remarkably, small biases were present so that the between group contrasts were never changed by the bootstrap. Combining this

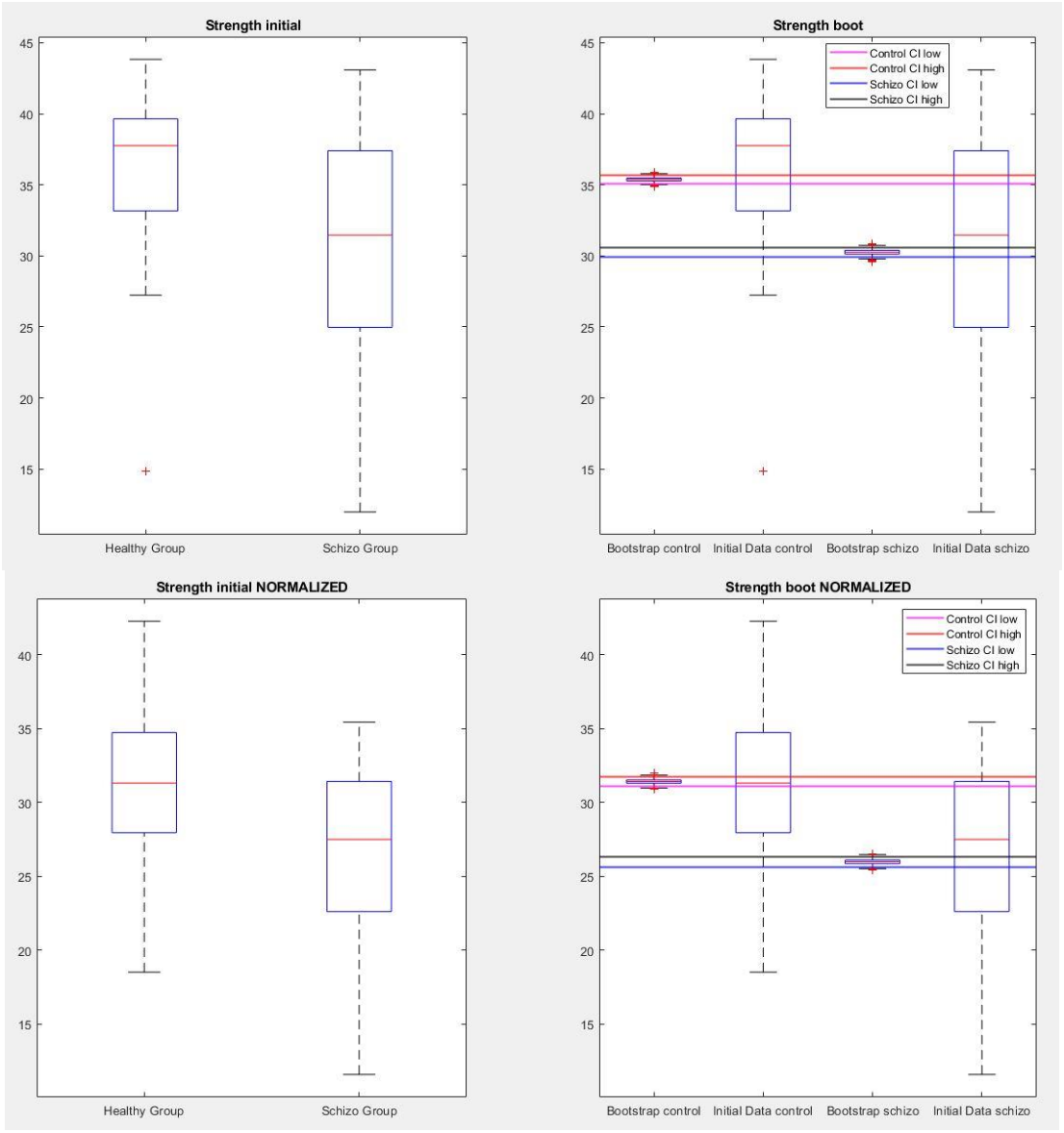
with the shrinking and the narrow CIs it can be concluded that the bootstrap is able to focus graph properties, which survive the random mixing performed by data surrogation.

Weighted Indexes (comparing standard and normalized Bootstraps):



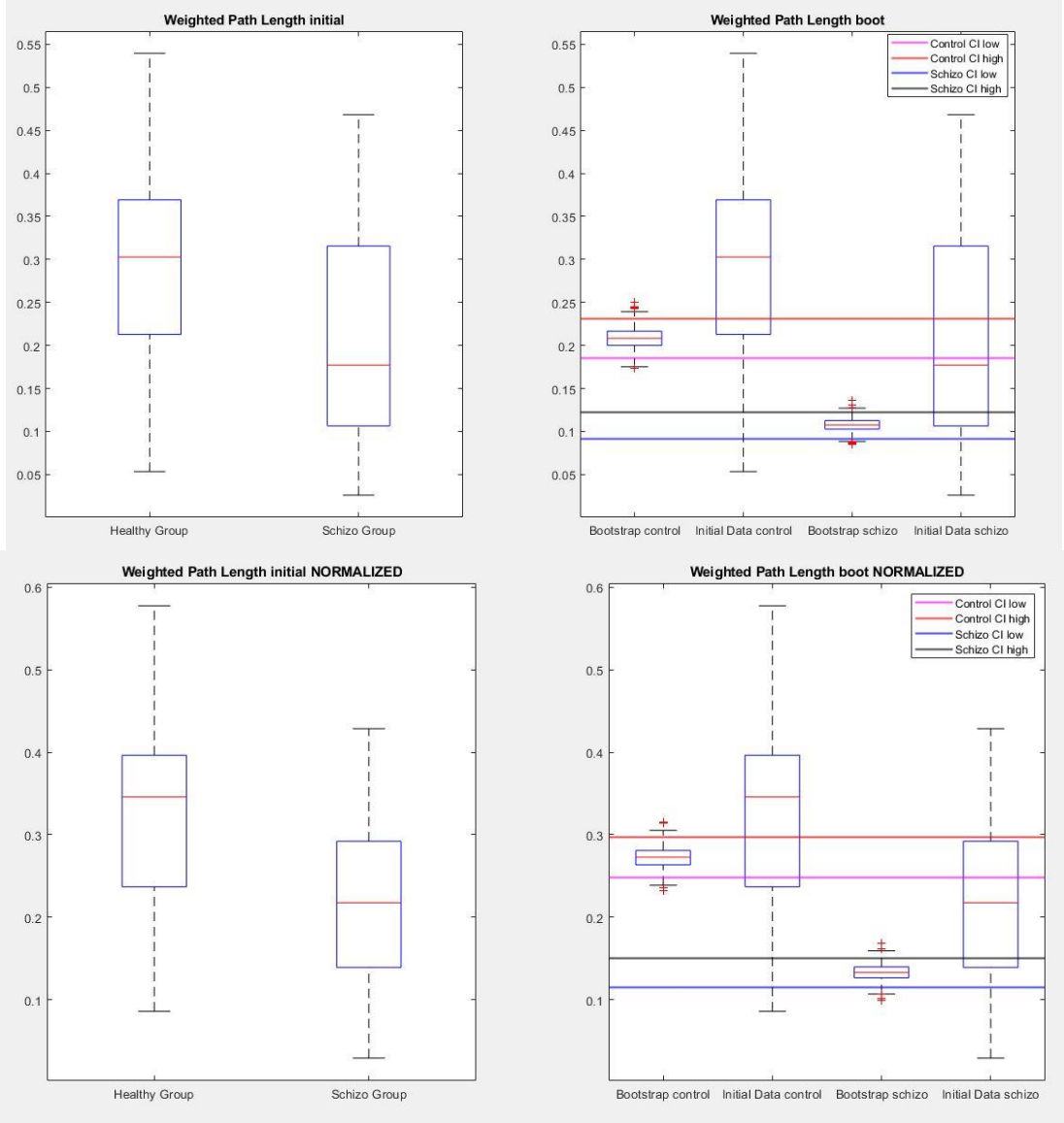
(a) (b)

Figure 70 *Distributions of the* ***Weighted Clustering Coefficient*** *in the standard case (subplots above) and with the probabilistic normalization one (subplots below). Figures* ***(a)*** *represent the distribution of the index evaluated for the initial data. Figures* ***(b)*** *show the boxplots estimated after the bootstrapping procedure with the boundaries of the confidence intervals with different colors calculated with the percentile method in both the control and schizophrenic groups.*



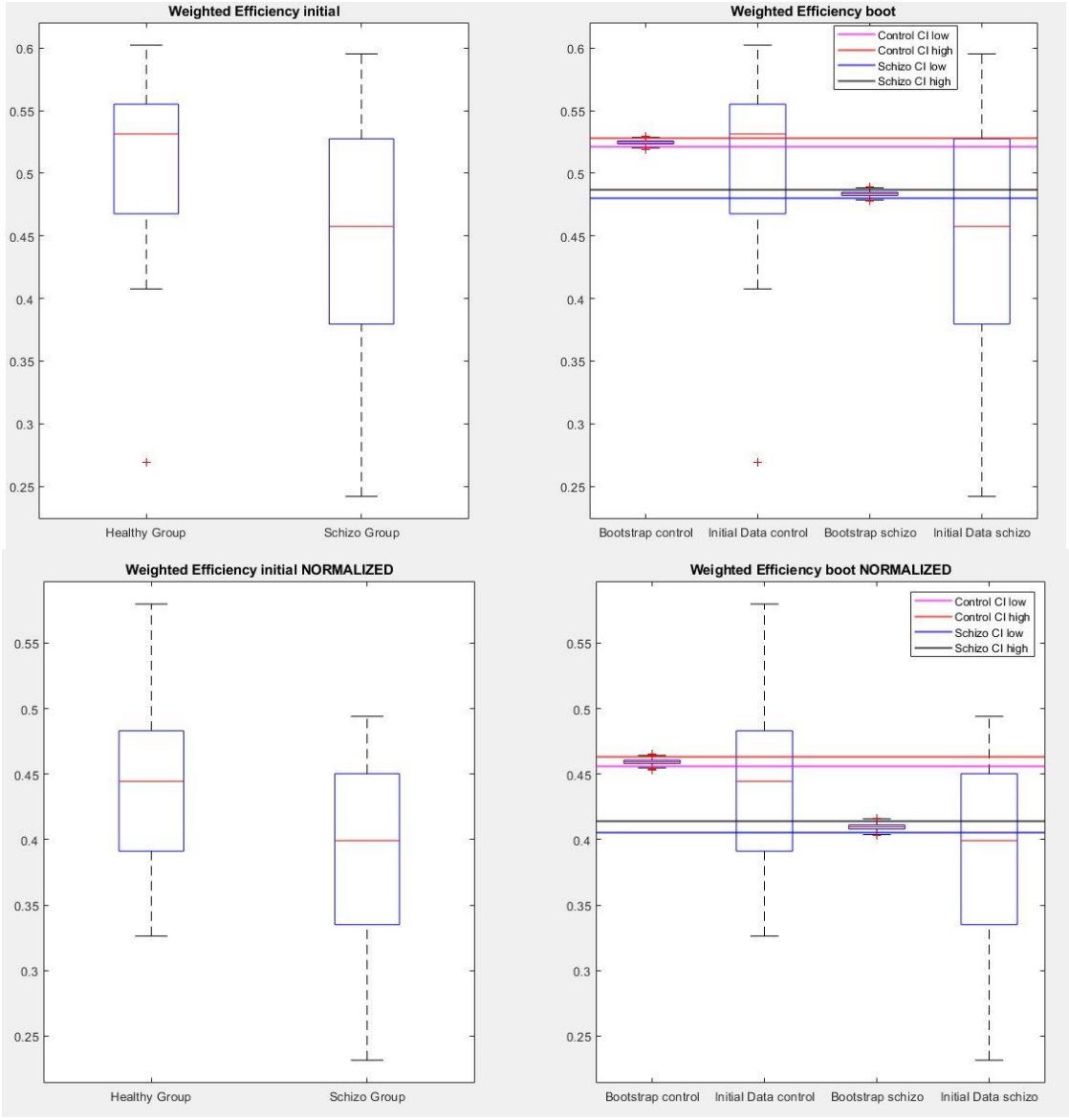
(a) (b)

Figure 71 *Distributions of the* ***Node Strength*** *in the standard case (subplots above) and with the probabilistic normalization one (subplots below). Figures* ***(a)*** *represent the distribution of the index evaluated for the initial data. Figures* ***(b)*** *show the boxplots estimated after the bootstrapping procedure with the boundaries of the confidence intervals with different colors calculated with the percentile method in both the control and schizophrenic groups.*



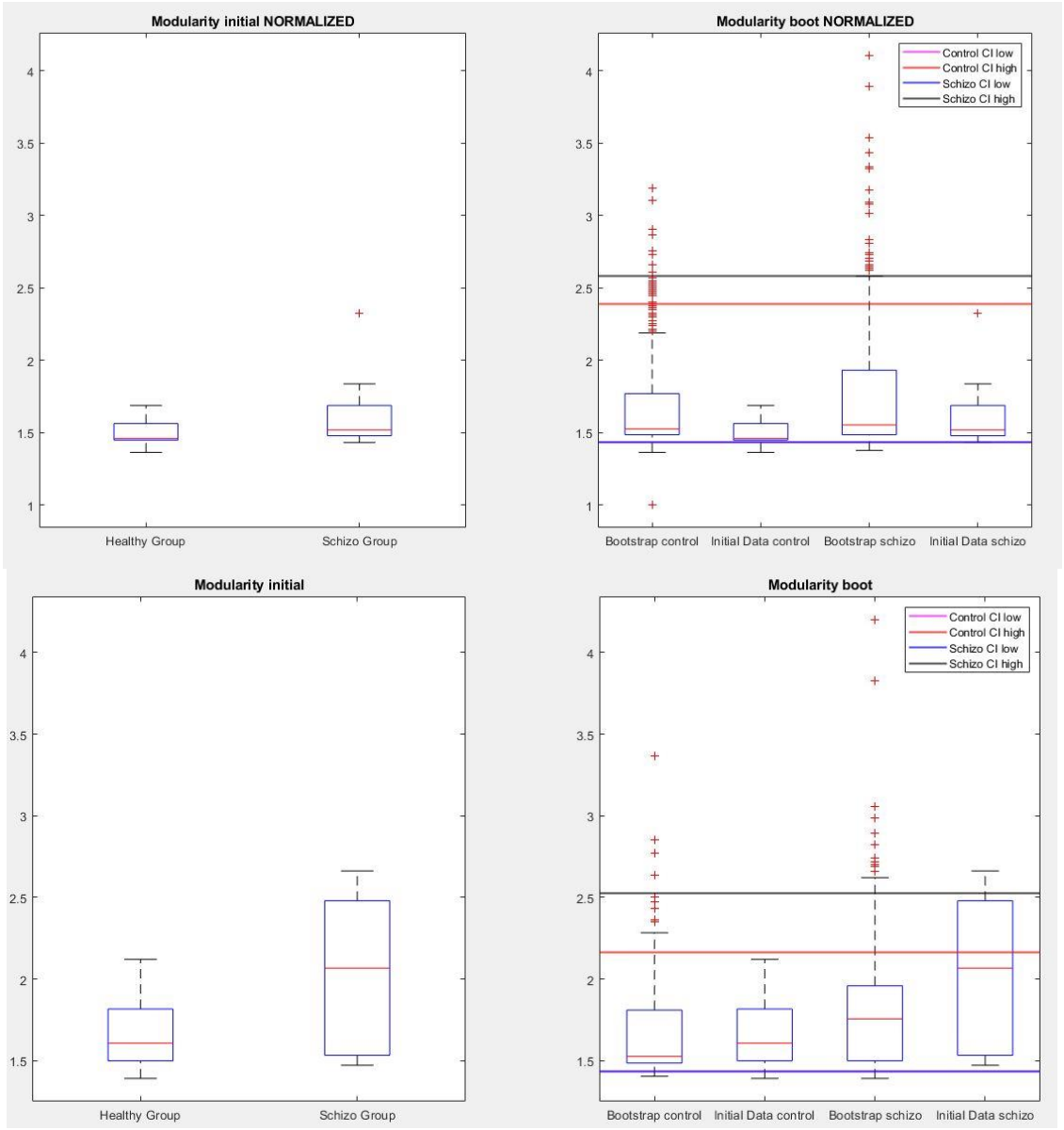
(a) (b)

Figure 72 *Distributions of the* ***Weighted Path Length*** *in the standard case (subplots above) and with the probabilistic normalization one (subplots below). Figures* ***(a)*** *represent the distribution of the index evaluated for the initial data. Figures* ***(b)*** *show the boxplots estimated after the bootstrapping procedure with the boundaries of the confidence intervals with different colors calculated with the percentile method in both the control and schizophrenic groups*



(a) (b)

Figure 73 *Distributions of the* ***Weighted Efficiency*** *in the standard case (subplots above) and with the probabilistic normalization one (subplots below). Figures* ***(a)*** *represent the distribution of the index evaluated for the initial data. Figures* ***(b)*** *show the boxplots estimated after the bootstrapping procedure with the boundaries of the confidence intervals with different colors calculated with the percentile method in both the control and schizophrenic groups*



(a) (b)

Figure 74 *Distributions of the* ***Modularity*** *in the standard case (subplots above) and with the probabilistic normalization one (subplots below). Figures* ***(a)*** *represent the distribution of the index evaluated for the initial data. Figures* ***(b)*** *show the boxplots estimated after the bootstrapping procedure with the boundaries of the confidence intervals with different colors calculated with the percentile method in both the control and schizophrenic groups*

Considering the Weighted Indexes, again, the expected outcomes are in terms of the reduction of the variability of the distributions for all the considered graph metrics. As regards as the differences between the pathological and the control groups, it is possible to notice the disruptions in the schizophrenic networks. Indeed, lower Weighted Clustering Coefficient (*Fig.70*), Strength (*Fig.71*)

Weighted Efficiency (*Fig.73*), Strength and higher Modularity (*Fig.74*). The Path Length does not represent disruptions in the weighted case (*Fig.72*) because, also after the bootstrap, it remains larger in the control case than the schizophrenic one.

As in the previous functional and structural connectivity datasets investigations, the aim was to calculate an appropriate Confidence Interval (CI) for the particular graph metric, which were evaluated as before described in section 3.3.2 of this work, with the standard method. This last one allowed to obtain a restricted interval approaching this kind of analysis to ideally find a normality value for all the indices in both the populations and with both the standard and normalized Bootstrapping procedures.

Graph-Based Indexes	Binary			Weighted		
Standard Bootstrapping	Control	Schizophrenic	p-value	Control	Schizophrenic	p-value
Clustering Coefficient	*Mean*: 0.486 ± 0.002 *Standard CI*: [0.486; 0.487] ***Bias*: 0.432%**	*Mean*: 0.418 ± 0.003 *Standard CI*: [0.417; 0.418] ***Bias*: 1.275%**	***< 0.05***	*Mean*: 0.454 ± 0.003 *Standard CI*: [0.454; 0.455] ***Bias:* 2.738%**	*Mean:* 0.375 ± 0.003 *Standard CI*: [0.374; 0.375] ***Bias:* 5.500%**	***< 0.05***
Degree/Strength	*Mean*: 71.748 ± 0.121 *Standard CI*: [71.740; 71.760] ***Bias*: 0.002%**	*Mean*: 69.213 ± 0.201 *Standard CI*: [69.201; 69.230] ***Bias*: 0.009%**	***< 0.05***	*Mean*: 35.388 ± 0.154 *Standard CI:* [35.382; 35.401] ***Bias*: 0.002%**	*Mean:* 30.261 ± 0.185 *Standard CI:* [30.251; 30.272] ***Bias*: 0.012%**	***< 0.05***
Path Length	*Mean*: 1.017 ± 0.002 *Standard CI:* [1.017; 1.018] ***Bias*: 0.002%**	*Mean:* 1.052 ± 0.003 *Standard CI:* [1.051; 1.052] ***Bias*: 0.008%**	***< 0.05***	*Mean*: 0.208 ± 0.012 *Standard CI*: [0.207; 0.209] ***Bias:* 31.85%**	*Mean:* 0.108 ± 0.008 *Standard CI*: [0.107; 0.108] ***Bias:* 49.03%**	***< 0.05***
Efficiency	Mean: 0.991 ± 8.265e-04 Standard CI: [0.991; 0.992] **Bias: 8.216 * 10^{-4} %**	Mean: 0.974 ± 0.001 Standard CI: [0.974; 0.975] **Bias: 0.004%**	***< 0.05***	Mean: 0.525 ± 0.002 Standard CI: [0.524; 0.525] **Bias: 4.578%**	Mean: 0.484 ± 0.002 Standard CI: [0.483; 0.484] **Bias: 8.497%**	***< 0.05***
Modularity	Na	Na	Na	Mean: 1.647 ± 0.231 Standard CI:	Mean: 1.775 ± 0.312 Standard CI:	***< 0.05***

				[1.633; 1.662] **Bias: 0.844%**	[1.756; 1.794] **Bias: 11.94%**	

Table 12 *Numerical results of the distribution of the indices of the two populations after the bootstrapping procedure applied in the* ***Standard case****. In particular, for each index, there are shown: Mean Value and Standard deviation, Confidence Intervals of the distribution calculated with the Standard Method, p-value and the bias percentage of the distributions of the indices considering before and after the bootstrapping procedure. Na: Index not available for binary/weighted cases.*

Graph-Based Indexes	Weighted		
Probabilistic Normalization Bootstrapping	Control	Schizophrenic	p-value
Clustering Coefficient	*Mean:* 0.410 ± 0.002 *Standard CI:* [0.409; 0.410] ***Bias:* 2.185%**	*Mean:* 0.329 ± 0.003 *Standard CI:* [0.328; 0.330] ***Bias:* 4.697%**	***< 0.05***
Node Strength	*Mean*: 31.41 ± 0.156 *Standard CI:* [31.40; 31.42] ***Bias*: 0.032%**	*Mean*: 25.99 ± 0.177 *Standard CI:* [25.98; 26.00] ***Bias*: 0.023%**	***< 0.05***
Path Length	*Mean*: 0.272 ± 0.013 *Standard CI:* [0.271; 0.273] ***Bias:* 16.06%**	*Mean*: 0.133 ± 0.010 *Standard CI:* [0.133; 0.134] ***Bias*: 39.34%**	***< 0.05***
Efficiency	*Mean:* 0.460 ± 0.002 *Standard CI:* [0.459; 0.460] ***Bias*: 3.064%**	*Mean*: 0.410 ± 0.002 *Standard CI:* [0.409; 0.410] ***Bias*: 6.389%**	***< 0.05***
Modularity	*Mean:* 1.652 ± 0.263 *Standard CI:* [1.635; 1.668] *Bias*: **10.38%**	*Mean:* 1.730 ± 0.352 *Standard CI:* [1.708; 1.752] *Bias*: **6.164%**	***< 0.05***

Table 13 *Numerical results of the distribution of the weighted indices of the two populations after the bootstrapping procedure applied in the* ***Probabilistic Normalization case****. In particular, for each weighted index, there are shown: Mean Value and Standard Deviation, Confidence Intervals of the distribution calculated with the Standard Method, p-value and the bias percentage of the mean of the distributions of the indices considering before and after the bootstrapping procedure.*

As it can be seen by the numerical results shown in the *Table.12* and *Table.13*, the biases are greater in the schizophrenic population considering the majority of the indexes. This highlights the major impact of the more variable connectivity weights for the schizophrenic population, as concluded in the Sensitivity Analysis section 4.3.1.

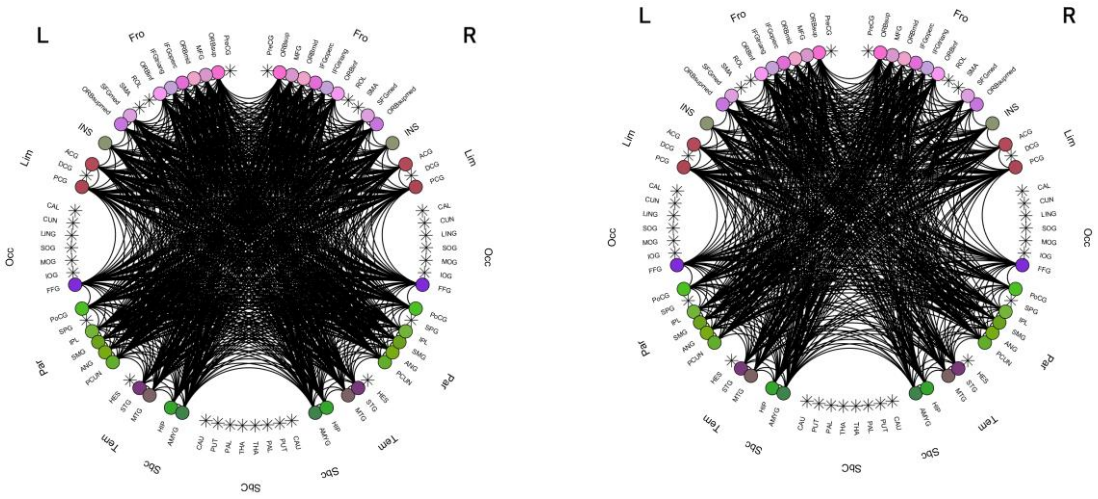
Regarding the differences of the results between the standard and the normalized procedure, the aim was to obtain reduced biases of the mean considering the distribution assumed by the values of the indices before and after the bootstrapping with the normalization of the data than the standard case; since this highlights the fact that normalizing the connectivity weights before applying the bootstrap is beneficial for the numerical results because the impact of those weights which bring a change in the values assumed by the indices is reduced. As can be seen by the numerical results shown in the *Table.12* and *Table.13*, the proposed hypothesis is, again, confirmed since the biases of the weighted graph metrics decrease with the normalization of the data applied before the bootstrap. The Weighted Path Length results are worth to be mentioned; in particular, it is the metric characterized by the greatest bias considering the mean before and after the bootstrapping procedure in both the control and schizophrenic populations (31.85% and 49.03% respectively in the Standard Bootstrapping case, 16.06% and 39.34% in the Probabilistic Normalization Bootstrapping case). The advantages brought by the normalization of the data computed before the bootstrapping procedure in terms of the reduction of the bias are evident, as exhibited by the numerical results described before.

The bootstrap brought advantages considering the p-values. Their numerical values computed on the surrogate data, exhibited in the *Table.12* and *Table.13*, show that the bootstrap is beneficial in the statistical assessment of group differences in the graph metrics. Indeed, after the bootstrapping, the p-values are all below the significance threshold of 0.05. Before the application of the bootstrap, instead, all indexes were not statistically significant except the Modularity. These advantages are related to the possibility of having more robust values for all indexes, with respect to the original population, and to have aligned distribution, with respect to Standard Bootstrap, that can be affected by a bias.

4.3.6 Extraction of the DMN with Spider-Net Results & Discussion

This section highlights the results about the extraction of the connectograms of the Default Mode Network (DMN) with Spider-Net Tool (SNT).

In particular, after the labelling of the appropriate parcels of the AAL atlas belonging to the Default Mode Network, described in Section 3.3.5, the connectograms can be obtained, with an example proposed in *Fig.75.* It shows the DMN subgraph obtained from the 6th subject of the control group and the 6th subject of the pathological group.



.

Figure 75 *Connectograms obtained by extracting the DMN from the sixth subjects in the control (left) and schizophrenic (right) datasets*

Considering *Fig.75*, it can be seen that the disruptions and reduced activity of the network is not straightforwardly identifiable in the schizophrenic subject, although comparing the two figures, it is possible to spot slightly less dense connectivity pathways in the schizophrenic subject's connectogram.

The output matrices given by Spider-Net and describing the DMN subgraph were used to compute the 9 weighted and binary graph metrics, together with the mean density of the network of each population to investigate the possible disruptions in the DMN present in the schizophrenic population. Inspecting the numerical results shown in *Table.14*, the schizophrenic disconnected network in terms of the values of the indices can be, again, confirmed, since the Clustering Coefficients and the Efficiency in both the weighted and binary cases, the Node Strength and the Node Degree are lower in the pathological population, the Modularity is greater and the Path Length slightly greater in the schizophrenic population in the

binary case, while, however, the weighted Path Length remains greater in the control population, as in the previous analyses relevant to the whole graph.

Graph-Based Indexes	Binary		Weighted	
DMN subnetwork	Control	Schizophrenic	Control	Schizophrenic
Clustering Coefficient	*Mean:* 0.489 ± 0.096 [0.450; 0.551]	*Mean:* 0.413 ± 0.121 [0.341; 0.518]	*Mean:* 0.468 ± 0.103 [0.429; 0.531]	*Mean:* 0.381 ± 0.129 [0.306; 0.495]
Degree/Strength	*Mean:* 40.62 ± 1.567 [39.96; 40.86]	*Mean:* 39.24 ± 1.879 [38.24; 40.86]	*Mean:* 20.01 ± 4.082 [18.33; 22.50]	*Mean:* 16.48 ± 5.146 [13.45; 21.05]
Path Length	*Mean:* 1.009 ± 0.038 [1.003; 1.025]	*Mean:* 1.043 ± 0.046 [1.003; 1.067]	*Mean:* 0.336 ± 0.137 [0.233; 0.408]	*Mean:* 0.199 ± 0.162 [0.124; 0.401]
Efficiency	*Mean:* 0.995 ± 0.019 [0.987; 0.998]]	*Mean:* 0.978 ± 0.023 [0.966; 0.998]	*Mean:* 0.504 ± 0.085 [0.461; 0.556]	*Mean:* 0.431 ± 0.107 [0.373; 0.529]
Modularity	Na	Na	*Mean:* 1.524 ± 0.312 [1.452; 1.845]	*Mean:* 1.595 ± 0.602 [1.524; 1.893]
Density	*Mean:* 97.77% ± 3.820%	*Mean:* 95.41% ± 4.580%	Na	Na

Table 14 *Numerical results of the weighted and binary graph metrics of both the control and schizophrenic populations calculated on the DMN subnetwork, showing the mean values and the standard deviations , the Interquantile range (25^{th} and 75^{th} percentiles); together with the mean density and standard deviation of the subnetwork. Na: Index not available for binary/weighted cases.*

The slight difference between densities in the connectograms is also identifiable in *Table.14.* Deeper insight in the visual inspection of the connectograms can be obtained by the density thresholding permitted by the Spider-Net interface, as explained in section 2.2. Besides, by selecting the option "*Show Weights*" it is possible to distinguish the most intense links in terms of connectivity which are colored in red, orange or yellow and designed with different width, according to the strength of connections between two different parcels in the functional connectivity matrices, such that red is for the connectivity weights between 0.66 and 1 representing the strongest functional connectivity, orange between 0.33 and

0.66 representing the middle functional connectivity and yellow between 0 and 0.33 representing the weakest functional connectivity.

From the figures below (*Fig.76* and *Fig.77*) it is possible to identify visually in an undemanding way the weak activation of the Default Mode Network typical of the schizophrenic population, as can be seen by the prevalence of yellow connectivity pathways between parcels in *Fig.77 (left)*, associated with weak connections, compared to the same DMN connectogram represented for the control subject in *Fig.76 (left)*.

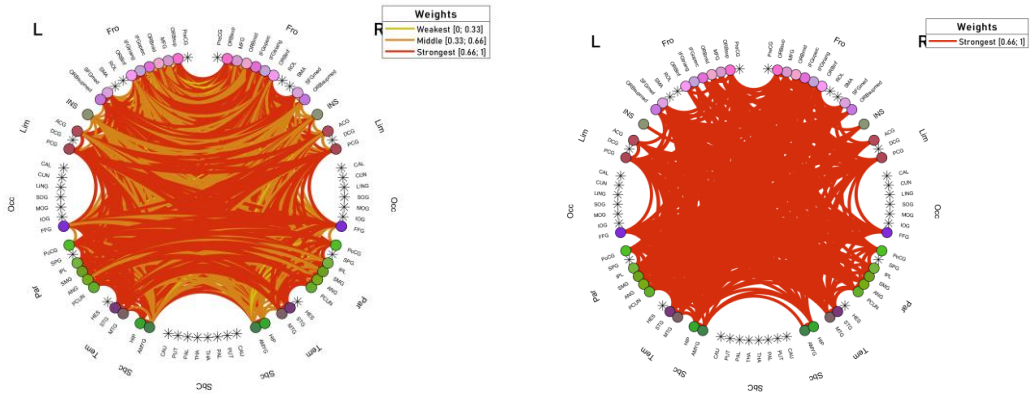


Figure 76 *Connectograms from the sixth subject of the control dataset of the DMN subnetwork extracted from Spider-Net with the option "Show Weights" (left), and the same connectogram with density-thresholding at 40% (right).*

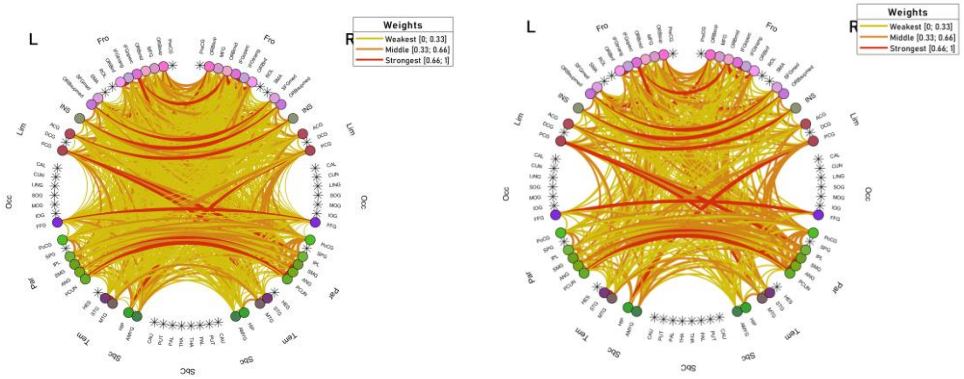


Figure 77 *Connectograms from the tenth subject of the schizophrenic dataset of the DMN subnetwork extracted from Spider-Net with the option "Show Weights" (left), and the same connectogram with density-thresholding at 40% (right).*

This difference in the activation of the DMN subnetwork can be interpreted more easily by density-thresholding the connectivity matrices of the two subjects. Indeed, by considering the *Fig.76 (right)* and *Fig.77 (right)*, the schizophrenic

subject is characterized by the weak connections also after the thresholding, while in the control subject, the DMN is strongly activated as can be seen by the presence of only red connectivity pathways between the parcels in *Fig.77 (right)*. Analyzing more in depth the weaker activation and the disruptions of the DMN of the schizophrenic population, the weakest connections were reset by arbitrarily selecting a threshold at 0.2, which is a weak one, given the range of connection weights from 0 to 1. Once this resetting was done in the whole dataset, the mean density of the DMN subnetwork was computed for each population producing a value of 88.63% for the control group and 79.35% in the pathological one. These results, thus, identify more clearly the disruptions of the DMN in the schizophrenic population. Moreover, plotting the connectograms of the DMN of the reset matrices, it was also possible to visually identify these disruptions, as exhibited in *Fig.78*.

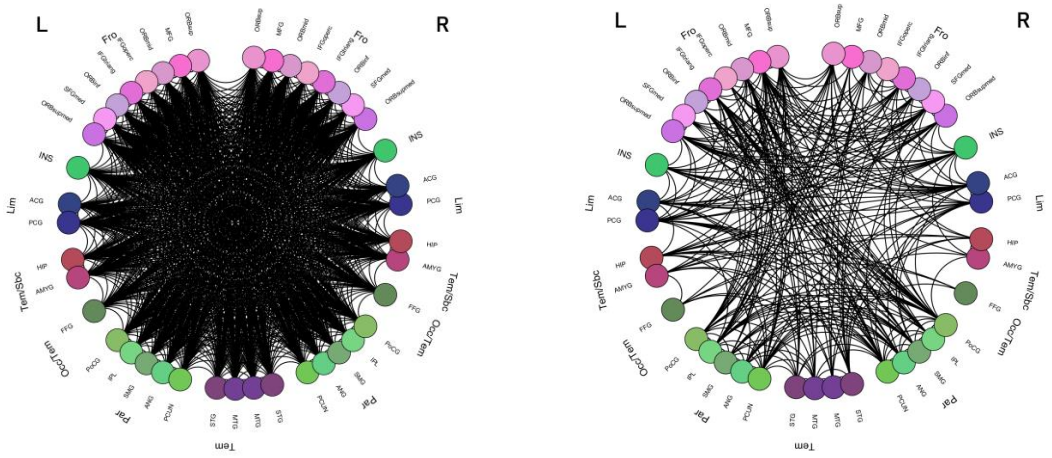


Figure 78 *Connectograms of the Default Mode Network subgraphs extracted from the sixth subject of the control group matrix (left) and the tenth subject of the schizophrenic one (right) whose weakest connections (weights < 0.2) were reset to evidence the disparity in terms of disconnections between the pathological and the control population*

The slight differences in the densities of the DMN between control and schizophrenic groups may be related to other factors. Indeed, as explained in the study provided by Zalesky and colleagues, 4 subjects of the schizophrenic population (12 total) were under psychotropic medication [5]. Although to reduce acute drug effects these patients were not medicated the day of acquisition, still the treatment of the pathology could have influenced the results.

To conclude the analysis of this dataset formed by 27 functional connectivity matrices with 15 control and 12 schizophrenic subjects, it is important to mention

the presence of a probable outlier in the control group, represented by subject number 14. Indeed, considering this last part about the analysis of the Default Mode Network disruptions in the schizophrenic populations, it was seen that by removing this outlier subject from the control group, the disruptions between the healthy and the pathological datasets are more evident. This is due to a higher difference in terms of the mean density of the connectomes from the extraction of the DMN in each subject of the control group. Indeed, the mean density passes from 97.77% to 98.85% on the initial data and from 88.63% to 92.00% after the removal of the weakest connections (weights < 0.2).

Conclusion

The research on the brain networks has been rapidly growing in the last decades. The possibility to represent them according to graph theory and quantify their topological characterization according to graph metrics allow to represent brain networks in an undemanding way. Correspondingly, this increasing interest needs new technical tools for conducting more efficient analysis on connectivity pathways. Moreover, the limitations of the techniques [48], [52], [53], [54], [57], [64] necessitate the introduction of a method to make the connectivity investigation more robust and valid.

In this context, the bootstrap can be applied. In particular, this technique allows to estimate the statistical features of a population characterized by the same condition from a limited number of measurement samples [81]. This procedure is based on a random sampling of the initial dataset to create "surrogate" data which can better explain the statistical properties of the population under analysis. This approach could extract a more stable normality range of values of various graph metrics [82], [83].

The bootstrap was applied in this work with respect to 3 different datasets: a functional connectivity dataset formed by 10 healthy subject, a structural connectivity dataset formed by 17 healthy subjects [4] and a case study involving a functional connectivity dataset formed by 15 healthy and 12 schizophrenic individuals [5].

First, exploration studies and the protocol setting were performed on the structural and functional datasets of healthy control subjects.

Specifically, different graph metrics were calculated and the bootstrap was applied with respect to a different number of iterations to investigate the distributions of the values after the creation of the surrogate data. It was seen that the variability of these values reduces after the bootstrap. Moreover, the advantage of the bootstrapping technique is to provide data-driven confidence intervals, without the need of an a-priori hypotheses. These confidence intervals were computed after the creation of surrogate data in two different methodologies: a) the Percentile method based on the straightforward construction of the 95% confidence interval from the resampled distribution; b) the Standard Method based on well-known formulas to compute the lower and upper boundaries of the confidence intervals. It was seen that both methods allowed to obtain appropriate confidence intervals, with the Standard Method producing narrower boundaries.

The distributions of the bootstrapped graph metrics values highlighted a bias, often very limited, though sensible in some of them. Conversely, from previous studies on bootstrapping, the surrogate distributions were expected to be narrower but unbiased and aligned to the original distributions of values [83], [84]. This points out on the well-known issues related to the brain connectivity analysis. Indeed, one of the main limitations of graph analysis is the lack of standardized procedure regarding network construction and even the most widespread connectivity measures are constantly object of active research [122], [123]. Gold-standard for the definition of matrix weights is still lacking, spurious connections are well-known to be still present and the great variability of the weights can be a confounding factor for the interpretation of the results [114].

For these reasons, before applying the bootstrap, the data were normalized according to a Probabilistic Normalization procedure, based on the group connectivity, to enable a more robust analysis. This method finds the rarest connections in the datasets, resetting them according to a threshold related to the population and normalizing the data by dividing the initial connectivity matrices with the mean matrix of the dataset. The application of this normalization allowed to reduce the outlier weights and, thus, the biases, in particular, for the weighted graph metrics, for all the considered datasets. Even in the case study involving healthy and pathological subjects, where only normalization was applied, the results were promising.

The methods developed and tested on the first two datasets were then validated on a complete second-level (i.e. group) analysis of the third dataset. This involved the comparison of functional connectivity matrixes in healthy controls and schizophrenic subjects. A sensitivity analysis was, first, performed to analyze which were the most variable connectivity weights in each matrix, since these are thought to be the ones causing the biases of the distributions. The sensitivity

analysis was based on graph perturbation [89], [90], [91] to investigate the results of the indexes at the removal of the outlier connections It was seen that the schizophrenic subjects were characterized by an higher number of variable weights, maybe due to the disconnections in brain networks caused by the pathology, resulting in higher values of the biases for all the considered graph metrics.

Moreover, the graph metrics were computed on the initial data to analyze the segregation of the brain networks of the pathological subjects; characterized by disconnections in the connectivity pathways. The results on the initial data highlighted only one statistically significant index ($p < 0.05$), the Modularity. The normalization of the data was applied and it highlighted statistical significances in all the indexes except from Modularity and Weighted Clustering Coefficient, which have anyway p-values close to 0.05, highlighting the advantages of the normalization.

The application of the bootstrapping technique confirmed the sensitivity analysis hypothesis since the pathological population was characterized by higher values of biases in all the considered graph metrics. These biases were reduced by applying the normalization, which allowed to reduce the variability of the connectivity weights. After the application of the bootstrap, all indexes in both standard and normalized cases were statistically significant.

In addition, in the context of providing new technical tools for conducting more efficient analysis on connectivity pathways, a novel software, Spider-Net Tool (SNT) [4], was used for focused analysis on subgraphs of interest.

In particular, SNT was used to investigate the structural connectivity dataset formed by 17 healthy subjects precessed with DTI and HARDI technique, where the latter is thought to bring advantages with respect to DTI. Indeed, in DTI only a single-fiber population is modelled for every voxel which represents a fundamental limitation when the voxels are characterized by multiple crossing fibers [124]. HARDI solves this problem with its acquisition schemes [57]. To analyze the ability of HARDI in detecting connections, the structural connectivity matrices were equalized in terms of density and number of streamlines to investigate which and how many connections were detected by HARDI and not by DTI and viceversa. In particular, the analysis was done at different levels of absolute thresholding [69], [6], and the connectograms were shown with respect to the left and right temporal lobes. The temporal lobe was chosen since DTI often loses several connections within this region due to the crossing fiber issue between the corpus callosum and the cortico-spinal tract [104], [125]. Nevertheless, the results were ambiguous, such that some subjects in the dataset showed more connections detected by HARDI and other more connections detected by DTI.

Moreover, the distributions of some network metrics (Density, Efficiency, Path Length and Node Strength) were examined according to a study developed by Prckosvka and colleagues [115]. However, the structural connectivity dataset under investigation was characterized by the opposite characteristics. There results were unexpected, highlighting the importance of having a coherent preprocessing pipeline in DTI and HARDI [54], [116], [117], [118], [119], [98], [120].

Furthermore, SNT was used to investigate the case study of healthy and schizophrenic subjects to extract the Default Mode Network (DMN) [30]. This subnetwork deeply involved in social behavior, control of the emotional state of the individual, recollection of past memories [30], [106], [108]. Schizophrenia is a mental disorder characterized by altered perception, irregular emotion regulation, hallucinations and weakened working memory; thus justifying studies on the involvement of the DMN [7], [29], [126].

After the extraction of the connectograms composed of all connections related to the DMN subnetwork, the selected graph metrics were computed to, again, show the connectivity disruptions in the schizophrenic subjects, also numerically confirmed by a slight difference in density between the two populations.

To further analyze the differences in the DMN, it was seen that the schizophrenic populations was characterized by the majority of weak connections in the DMN subgraph extracted by SNT. In addition, using Spider-Net with the appropriate threshold settings, it could be easily seen the differences in the strength of activation of the DMN. Moreover, by the resetting of the weakest connections in the dataset, the disruptions in the subnetwork were clearly identifiable in the connectograms, highlighting the advantages brought by this software both in terms of visualization and manipulation of data.

Further developments of this work are related to a deeper validation on different types of data, especially in the structural connectivity case. In particular, to investigate coherent data, it is necessary to maintain the same preprocessing pipeline to reconstruct fibers from the acquisitions. Moreover, considering the investigation of graph indexes, it would be needed to validate bootstrap and sensitivity analysis as a tool to evaluate uncertainty also for other widely used metrics. Concerning the analysis of the biases, of bootstrap vs. original mean values further insight should be gained concerning the effect of the original data dispersion and presence of outliers. Data normalization should be also further studied as alternative or complementary to thresholding in the equalization of data. In the present study, a group-level normalization (dividing the group data by the group Mean Matrix) was successfully applied. However, other aspects could be introduced, such as normalizing by the grand average of groups (thus fully cancelling differences in density) or, conversely, scaling the weight to

compressed scales such as the logarithmic one. The latter, especially in structural connectivity where weights based on the streamline count span over orders of magnitude. Besides, an increased robustness in the analysis of the weights and the indexes could be fundamental for the consideration and investigation of weak connections, whose information is often excluded but that was proved to be potentially powerful in acquiring new knowledge on the connectivity topic [24].

Finally, the investigation of the neuroimaging data was thoroughly treated in this work, analyzing the strength and the weaknesses coming from the information derived from brain networks. This thesis highlighted the fact that it is necessary to implement both graph theory algorithms and visualization techniques to understand the connectivity pathways described by the neuroimaging datasets. These types of evaluations, however, are often hindered by uncertainties and artifacts which need to be sized with appropriate methods and, given the complexity of the brain, require further research in this field.

Bibliography

[1] A. Fornito, A. Zalesky, and E. T. Bullmore, *Fundamentals of Brain Network Analysis*. 2016.

[2] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010, doi: 10.1016/j.neuroimage.2009.10.003.

[3] Y. Liu *et al.*, "Disrupted small-world networks in schizophrenia," *Brain*, vol. 131, no. 4, pp. 945–961, 2008, doi: 10.1093/brain/awn018.

[4] D. Coluzzi *et al.*, "Development and Testing of SPIDER-NET: An Interactive Tool for Brain Connectogram Visualization, Sub-Network Exploration and Graph Metrics Quantification," *Front. Neurosci.*, vol. 16, no. March, 2022, doi: 10.3389/fnins.2022.818385.

[5] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: Identifying differences in brain networks," *Neuroimage*, vol. 53, no. 4, pp. 1197–1207, 2010, doi: 10.1016/j.neuroimage.2010.06.041.

[6] D. S. Bassett, B. G. Nelson, B. A. Mueller, J. Camchong, and K. O. Lim, "Altered resting state complexity in schizophrenia," *Neuroimage*, vol. 59, no.

3, pp. 2196–2207, 2012, doi: 10.1016/j.neuroimage.2011.10.002.

[7] M. P. Van Den Heuvel and A. Fornito, "Brain networks in schizophrenia," *Neuropsychology Review*, vol. 24, no. 1. pp. 32–48, 2014, doi: 10.1007/s11065-014-9248-7.

[8] Q. Yu *et al.*, "Altered topological properties of functional network connectivity in schizophrenia during resting state: A small-world brain Network study," *PLoS One*, vol. 6, no. 9, 2011, doi: 10.1371/journal.pone.0025423.

[9] S. N. Sotiropoulos and A. Zalesky, "Building connectomes using diffusion MRI: why, how and but," *NMR Biomed.*, vol. 32, no. 4, pp. 1–23, 2019, doi: 10.1002/nbm.3752.

[10] L. Euler, "Leonhard Euler and the Koenigsberg Bridges," *Sci. Am.*, vol. 189, no. 1, pp. 66–70, 1953, doi: 10.1038/scientificamerican0753-66.

[11] C. T. Butts, "Revisiting the foundations of network analysis," *Science (80-.).*, vol. 325, no. 5939, pp. 414–416, 2009, doi: 10.1126/science.1171022.

[12] R. C. Craddock *et al.*, "Imaging human connectomes at the macroscale," *Nat. Methods*, vol. 10, no. 6, pp. 524–539, 2013, doi: 10.1038/nmeth.2482.

[13] K. Brodmann, "Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues," vol. 44, no. 0, 1909.

[14] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006, doi: 10.1016/j.neuroimage.2006.01.021.

[15] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *Neuroimage*, vol. 53, no. 1, pp. 1–15, 2010, doi: 10.1016/j.neuroimage.2010.06.010.

[16] J. Zhao, C. Tang, and J. Nie, "Functional Parcellation of Individual Cerebral Cortex Based on Functional MRI," *Neuroinformatics*, vol. 18, no. 2, pp. 295–306, 2020, doi: 10.1007/s12021-019-09445-8.

[17] M. F. Glasser *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016, doi: 10.1038/nature18933.

[18] J. L. Lancaster *et al.*, "Automated labeling of the human brain: A preliminary

report on the development and evaluation of a forward-transform method," *Hum. Brain Mapp.*, vol. 5, no. 4, pp. 238–242, 1997, doi: 10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4.

[19] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002, doi: 10.1006/nimg.2001.0978.

[20] R. Wang, M. Liu, X. Cheng, Y. Wu, A. Hildebrandt, and C. Zhou, "Resting Brain Networks Configure Different Cognitive Abilities," *Pnas*, vol. 36, no. 48, pp. 12083–12094, 2021, doi: 10.1073/pnas.2022288118/-/DCSupplemental.y.

[21] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: emerging methods and principles," *Trends Cogn. Sci.*, vol. 14, no. 6, pp. 277–290, 2010, doi: 10.1016/j.tics.2010.04.004.

[22] E. Bullmore and O. Sporns, "The economy of brain network organization," *Nat. Rev. Neurosci.*, vol. 13, no. 5, pp. 336–349, 2012, doi: 10.1038/nrn3214.

[23] M. Kaiser and C. C. Hilgetag, "Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems," *PLoS Comput. Biol.*, vol. 2, no. 7, pp. 0805–0815, 2006, doi: 10.1371/journal.pcbi.0020095.

[24] D. S. Bassett and E. T. Bullmore, "Small-World Brain Networks Revisited," *Neuroscientist*, vol. 23, no. 5, pp. 499–516, 2017, doi: 10.1177/1073858416667720.

[25] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Phys. Rev. Lett.*, vol. 87, no. 19, pp. 198701-1-198701–4, 2001, doi: 10.1103/PhysRevLett.87.198701.

[26] M.-M. Mesulam, "Large-scale neurocognitive networks and distributed processing for attention, language, and memory," *Ann. Neurol.*, vol. 28, no. 5, pp. 597–613, 1990, doi: https://doi.org/10.1002/ana.410280502.

[27] A. G. Casali *et al.*, "A theoretically based index of consciousness independent of sensory processing and behavior," *Sci. Transl. Med.*, vol. 5, no. 198, 2013, doi: 10.1126/scitranslmed.3006294.

[28] M. Rosanova *et al.*, "Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients," *Brain*, vol. 135, no. 4, pp. 1308–1320, 2012, doi: 10.1093/brain/awr340.

[29] M. L. Hu *et al.*, "A Review of the Functional and Anatomical Default Mode Network in Schizophrenia," *Neurosci. Bull.*, vol. 33, no. 1, pp. 73–84, 2017, doi: 10.1007/s12264-016-0090-1.

[30] M. E. Raichle, "The Brain's Default Mode Network," *Annu. Rev. Neurosci.*, vol. 38, pp. 433–447, 2015, doi: 10.1146/annurev-neuro-071013-014030.

[31] C. M. Lewis, C. A. Bosman, and P. Fries, "Recording of brain activity across spatial scales," *Curr. Opin. Neurobiol.*, vol. 32, pp. 68–77, 2015, doi: 10.1016/j.conb.2014.12.007.

[32] E. M. Purcell, H. C. Torrey, and R. V Pound, "Resonance Absorption by Nuclear Magnetic Moments in a Solid," *Phys. Rev.*, vol. 69, no. 1–2, pp. 37–38, 1946, doi: 10.1103/PhysRev.69.37.

[33] Baselli G., *Slides of the course "Methods for Biomedical Imaging and Computer Aided Surgery"*, "07 – NMR PHYSICS," pp. 1–9.

[34] V. P. B. Grover, J. M. Tognarelli, M. M. E. Crossey, I. J. Cox, S. D. Taylor-Robinson, and M. J. W. McPhail, "Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians," *J. Clin. Exp. Hepatol.*, vol. 5, no. 3, pp. 246–255, 2015, doi: 10.1016/j.jceh.2015.08.001.

[35] Baselli G., *Slides of the course "Methods for Biomedical Imaging and Computer Aided Surgery"*, "08 – BLOCH EQUATIONS OF NMR," pp. 1–6.

[36] A. A. Malayeri *et al.*, "Principles and applications of diffusion-weighted imaging in cancer detection, staging, and treatment follow-up," *Radiographics*, vol. 31, no. 6, pp. 1773–1791, 2011, doi: 10.1148/rg.316115515.

[37] R. Bammer, "Basic principles of diffusion-weighted imaging.," *Eur. J. Radiol.*, vol. 45, no. 3, pp. 169–184, Mar. 2003, doi: 10.1016/s0720-048x(02)00303-0.

[38] A. Goyal, R. Sharma, A. Bhalla, S. Gamanagatti, and A. Seth, "Diffusion-weighted MRI in assessment of renal dysfunction," *Indian J. Radiol. Imaging*, vol. 22, no. 3, pp. 155–159, 2012, doi: 10.4103/0971-3026.107169.

[39] R. Donners, R. S. Z. Yiin, M. Blackledge, and D. M. Koh, "Whole-body diffusion-weighted MRI of normal lymph nodes: prospective apparent diffusion coefficient histogram and nodal distribution analysis in a healthy cohort," *Cancer Imaging*, vol. 21, no. 1, pp. 1–10, 2021, doi: 10.1186/s40644-021-00432-4.

[40] P. B. Kingsley, "Introduction to diffusion tensor imaging mathematics: Part I. Tensors, rotations, and eigenvectors," *Concepts Magn. Reson. Part A Bridg.*

Educ. Res., vol. 28, no. 2, pp. 101–122, 2006, doi: 10.1002/cmr.a.20048.

[41] O. Dietrich, A. Biffar, A. Baur-Melnyk, and M. F. Reiser, "Technical aspects of MR diffusion imaging of the body," *Eur. J. Radiol.*, vol. 76, no. 3, pp. 314–322, 2010, doi: 10.1016/j.ejrad.2010.02.018.

[42] K. R. T. Fink and J. R. Fink, "Principles of Modern Neuroimaging," *Princ. Neurol. Surg.*, pp. 62-86.e2, 2018, doi: 10.1016/B978-0-323-43140-8.00004-4.

[43] M. Descoteaux, C. De Recherche, C. Étienne, L. Bel, C. Poupon, and P. Courier, "Diffusion-weighted MRI," pp. 1–30.

[44] P. Mukherjee, J. I. Berman, S. W. Chung, C. P. Hess, and R. G. Henry, "Diffusion tensor MR imaging and fiber tractography: Theoretic underpinnings," *Am. J. Neuroradiol.*, vol. 29, no. 4, pp. 632–641, 2008, doi: 10.3174/ajnr.A1051.

[45] P. M. Descoteaux, "Diffusion Tractography : Principle & Methods Brain wiring," *OHBM 2016 Educ. Course Present.*, 2016.

[46] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophys. J.*, vol. 66, no. 1, pp. 259–267, 1994, doi: 10.1016/S0006-3495(94)80775-1.

[47] M. Giannelli *et al.*, "Dependence of brain DTI maps of fractional anisotropy and mean diffusivity on the number of diffusion weighting directions," *J. Appl. Clin. Med. Phys.*, vol. 11, no. 1, pp. 176–190, 2010, doi: 10.1120/jacmp.v11i1.2927.

[48] Y. Assaf and O. Pasternak, "Diffusion tensor imaging (DTI)-based white matter mapping in brain research: A review," *J. Mol. Neurosci.*, vol. 34, no. 1, pp. 51–61, 2008, doi: 10.1007/s12031-007-0029-0.

[49] S. Pajevic and C. Pierpaoli, "Erratum: Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: Application to white matter fiber tract mapping in the human brain (Magnetic Resonance in medicine (September 1999) 42:3 (526-540))," *Magn. Reson. Med.*, vol. 43, no. 6, p. 921, 2000, doi: 10.1002/1522-2594(200006)43:6<921::AID-MRM23>3.0.CO;2-I.

[50] K. G. Schilling *et al.*, "Limits to anatomical accuracy of diffusion tractography using modern approaches," *Neuroimage*, vol. 185, no. September 2018, pp. 1–11, 2019, doi: 10.1016/j.neuroimage.2018.10.029.

[51] T. E. J. Behrens, S. N. Sotiropoulos, and S. Jbabdi, *MR Diffusion Tractography*,

Second Edi. Elsevier, 2013.

[52] S. Jbabdi and H. Johansen-Berg, "Tractography: Where Do We Go from Here?," *Brain Connect.*, vol. 1, no. 3, pp. 169–183, 2011, doi: 10.1089/brain.2011.0033.

[53] L. E. Salminen, T. E. Conturo, J. D. Bolzenius, R. P. Cabeen, E. Akbudak, and R. H. Paul, "Reducing CSF Partial Volume Effects to Enhance Diffusion Tensor Imaging Metrics of Brain Microstructure," *Technol. Innov.*, vol. 18, no. 1, pp. 5–20, 2016, doi: 10.21300/18.1.2016.5.

[54] D. K. Jones, T. R. Knösche, and R. Turner, "White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI," *Neuroimage*, vol. 73, pp. 239–254, 2013, doi: 10.1016/j.neuroimage.2012.06.081.

[55] O. Civier, R. E. Smith, C. H. Yeh, A. Connelly, and F. Calamante, "Is removal of weak connections necessary for graph-theoretical analysis of dense weighted structural connectomes from diffusion MRI?," *Neuroimage*, vol. 194, no. March, pp. 68–81, 2019, doi: 10.1016/j.neuroimage.2019.02.039.

[56] T. Wang *et al.*, "Multilevel Deficiency of White Matter Connectivity Networks in Alzheimer's Disease: A Diffusion MRI Study with DTI and HARDI Models," *Neural Plast.*, vol. 2016, 2016, doi: 10.1155/2016/2947136.

[57] M. Descoteaux, "High Angular Resolution Diffusion Imaging (HARDI)," *Wiley Encycl. Electr. Electron. Eng.*, no. October, pp. 1–25, 2015, doi: 10.1002/047134608x.w8258.

[58] E. W. Lang, A. M. Tomé, I. R. Keck, J. M. Górriz-Sáez, and C. G. Puntonet, "Brain connectivity analysis: A short survey," *Comput. Intell. Neurosci.*, vol. 2012, no. iii, 2012, doi: 10.1155/2012/412512.

[59] G. H. Glover, "Overview of functional magnetic resonance imaging," *Neurosurg. Clin. N. Am.*, vol. 22, no. 2, pp. 133–139, 2011, doi: 10.1016/j.nec.2010.11.001.

[60] E. Tagliazucchi and D. R. Chialvo, "The collective brain is critical," no. March, 2011, [Online]. Available: http://arxiv.org/abs/1103.2070.

[61] R. Mohanty, W. A. Sethares, V. A. Nair, and V. Prabhakaran, "Rethinking Measures of Functional Connectivity via Feature Extraction," *Sci. Rep.*, vol. 10, no. 1, pp. 1–17, 2020, doi: 10.1038/s41598-020-57915-w.

[62] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease,"

PLoS Comput. Biol., vol. 4, no. 6, 2008, doi: 10.1371/journal.pcbi.1000100.

[63] C. J. Stam *et al.*, "Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease," *Brain*, vol. 132, no. 1, pp. 213–224, 2009, doi: 10.1093/brain/awn262.

[64] R. M. Hutchison *et al.*, "NeuroImage Dynamic functional connectivity : Promise , issues , and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 2013, doi: 10.1016/j.neuroimage.2013.05.079.

[65] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, no. 4–5, pp. 175–308, 2006, doi: 10.1016/j.physrep.2005.10.009.

[66] L. M. Colon-Perez, M. Couret, W. Triplett, C. C. Price, and T. H. Mareci, "Small worldness in dense and weighted connectomes," *Front. Phys.*, vol. 4, no. MAY, pp. 1–15, 2016, doi: 10.3389/fphy.2016.00014.

[67] A. Serra *et al.*, "Strong-Weak Pruning for Brain Network Identification in Connectome-Wide Neuroimaging: Application to Amyotrophic Lateral Sclerosis Disease Stage Characterization," *Int. J. Neural Syst.*, vol. 29, no. 7, 2019, doi: 10.1142/S0129065719500072.

[68] E. Santarnecchi, G. Galli, N. R. Polizzotto, A. Rossi, and S. Rossi, "Efficiency of weak brain connections support general cognitive functioning," *Hum. Brain Mapp.*, vol. 35, no. 9, pp. 4566–4582, 2014, doi: 10.1002/hbm.22495.

[69] M. P. van den Heuvel, S. C. de Lange, A. Zalesky, C. Seguin, B. T. T. Yeo, and R. Schmidt, "Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations," *Neuroimage*, vol. 152, no. February, pp. 437–449, 2017, doi: 10.1016/j.neuroimage.2017.02.005.

[70] M. Rubinov and O. Sporns, "Weight-conserving characterization of complex functional brain networks," *Neuroimage*, vol. 56, no. 4, pp. 2068–2079, 2011, doi: 10.1016/j.neuroimage.2011.03.069.

[71] J. Daintith and E. Wright, "Dijkstra's algorithm." Oxford University Press, 2008, doi: 10.1093/acref/9780199234004.013.1426.

[72] J. P. Onnela *et al.*, "Analysis of a large-scale weighted network of one-to-one human communication," *New J. Phys.*, vol. 9, 2007, doi: 10.1088/1367-2630/9/6/179.

[73] J. Travers and S. Milgram, "An experimental study of the small world

problem," *Struct. Dyn. Networks*, vol. 9781400841, no. 4, pp. 130–148, 2011, doi: 10.2307/2786545.

[74] D. J. Watts and S. H. Strogatz, "Strogatz - small world network Nature," *Nature*, vol. 393, no. June, pp. 440–442, 1998, [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/9623998.

[75] M. D. Humphries and K. Gurney, "Network ' Small-World-Ness ': A Quantitative Method for Determining Canonical Network Equivalence," vol. 3, no. 4, 2008, doi: 10.1371/journal.pone.0002051.

[76] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 69, no. 2 2, pp. 1–16, 2004, doi: 10.1103/PhysRevE.69.026113.

[77] M. Krzywinski *et al.*, "Circos: An information aesthetic for comparative genomics," *Genome Res.*, vol. 19, no. 9, pp. 1639–1645, 2009, doi: 10.1101/gr.092759.109.

[78] P. Bellec, G. Marrelec, and H. Benali, "A bootstrap test to investigate changes in brain connectivity for functional MRI," *Stat. Sin.*, vol. 18, no. 4, pp. 1253–1268, 2008.

[79] M. Lazar *et al.*, "White matter tractography using diffusion tensor deflection," *Hum. Brain Mapp.*, vol. 18, no. 4, pp. 306–321, 2003, doi: 10.1002/hbm.10102.

[80] G. J. M. Parker and D. C. Alexander, "Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 360, no. 1457, pp. 893–902, 2005, doi: 10.1098/rstb.2005.1639.

[81] M. Lazar and A. L. Alexander, "Bootstrap white matter tractography (BOOT-TRAC)," *Neuroimage*, vol. 24, no. 2, pp. 524–532, 2005, doi: 10.1016/j.neuroimage.2004.08.050.

[82] A. Kulesa, M. Krzywinski, P. Blainey, and N. Altman, "Points of Significance: Sampling distributions and the bootstrap," *Nat. Methods*, vol. 12, no. 6, pp. 477–478, 2015, doi: 10.1038/nmeth.3414.

[83] Y. R. Gel, V. Lyubchich, and L. Leticia Ramirez Ramirez, "Bootstrap quantification of estimation uncertainties in network degree distributions," *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, 2017, doi: 10.1038/s41598-017-05885-x.

[84] A. Green and C. R. Shalizi, "Bootstrapping exchangeable random graphs*,"

Electron. J. Stat., vol. 16, no. 1, pp. 1058–1095, 2022, doi: 10.1214/21-EJS1896.

[85] A. Azzalini and P. Hall, "Reducing Variability Using Bootstrap Methods with Qualitative Constraints," *Biometrika*, vol. 87, no. 4, pp. 895–906, 2000, Accessed: Sep. 14, 2022. [Online]. Available: http://www.jstor.org/stable/2673617.

[86] V. Picheny, N. H. Kim, and R. T. Haftka, "Application of bootstrap method in conservative estimation of reliability with limited samples," *Struct. Multidiscip. Optim.*, vol. 41, no. 2, pp. 205–217, 2010, doi: 10.1007/s00158-009-0419-8.

[87] L. Wei, B. Jing, and H. Li, "Bootstrapping promotes the RSFC-behavior associations : An application of individual cognitive traits prediction," no. October 2019, pp. 2302–2316, 2020, doi: 10.1002/hbm.24947.

[88] C. Pichery, "Sensitivity Analysis," in *Encyclopedia of Toxicology: Third Edition*, 2014, pp. 236–237.

[89] B. Ouyang, Y. Xia, C. Wang, Q. Ye, Z. Yan, and Q. Tang, "Quantifying Importance of Edges in Networks," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 65, no. 9, pp. 1244–1248, 2018, doi: 10.1109/TCSII.2018.2820090.

[90] B. Wang, J. Jia, X. Cao, and N. Z. Gong, "Certified Robustness of Graph Neural Networks against Adversarial Structural Perturbation," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1645–1653, 2021, doi: 10.1145/3447548.3467295.

[91] I. Mishkovski, M. Biey, and L. Kocarev, "Vulnerability of complex networks," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 16, no. 1, pp. 341–349, 2011, doi: 10.1016/j.cnsns.2010.03.018.

[92] "MATLAB version 9.6.0.1335978 (R2019a)." Natick, Massachusetts, 2019.

[93] B. Fischl *et al.*, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002, doi: 10.1016/S0896-6273(02)00569-X.

[94] L. Pelizzari *et al.*, "Combined assessment of diffusion parameters and cerebral blood flow within basal ganglia in early Parkinson's disease," *Front. Aging Neurosci.*, vol. 11, no. JUN, pp. 1–10, 2019, doi: 10.3389/fnagi.2019.00134.

[95] J. L. R. Andersson, S. Skare, and J. Ashburner, "How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor

imaging," *Neuroimage*, vol. 20, no. 2, pp. 870–888, 2003, doi: 10.1016/S1053-8119(03)00336-7.

[96] T. E. J. Behrens *et al.*, "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging," *Nat. Neurosci.*, vol. 6, no. 7, pp. 750–757, 2003, doi: 10.1038/nn1075.

[97] J. L. R. Andersson and S. N. Sotiropoulos, "An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging," *Neuroimage*, vol. 125, pp. 1063–1078, 2016, doi: 10.1016/j.neuroimage.2015.10.019.

[98] F. Dell'Acqua *et al.*, "A modified damped Richardson-Lucy algorithm to reduce isotropic background effects in spherical deconvolution," *Neuroimage*, vol. 49, no. 2, pp. 1446–1458, 2010, doi: 10.1016/j.neuroimage.2009.09.033.

[99] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images," *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002, doi: 10.1006/nimg.2002.1132.

[100] J. Suckling, C. Long, C. Triantafyllou, M. Brammer, and E. Bullmore, "Variable precision registration via wavelets: Optimal spatial scales for inter-subject registration of functional MRI," *Neuroimage*, vol. 31, no. 1, pp. 197–208, 2006, doi: 10.1016/j.neuroimage.2005.11.032.

[101] V. Blasi *et al.*, "Early Life Adversities and Borderline Intellectual Functioning Negatively Impact Limbic System Connectivity in Childhood: A Connectomics-Based Study," *Front. Psychiatry*, vol. 11, no. September, pp. 1–9, 2020, doi: 10.3389/fpsyt.2020.497116.

[102] A. Škoch *et al.*, "Human brain structural connectivity matrices – ready for modelling," pp. 1–9, 2022, doi: 10.1038/s41597-022-01596-9.

[103] S. Tsai, "Reproducibility of structural brain connectivity and network metrics using probabilistic diffusion tractography," no. March, pp. 1–12, 2018, doi: 10.1038/s41598-018-29943-0.

[104] S. De Santis, Y. Assaf, B. Jeurissen, D. K. Jones, and A. Roebroeck, "T1 relaxometry of crossing fibres in the human brain," *Neuroimage*, vol. 141, pp. 133–142, 2016, doi: 10.1016/j.neuroimage.2016.07.037.

[105] Y. Tao *et al.*, "The structural connectivity pattern of the default mode network and its association with memory and anxiety," *Front. Neuroanat.*,

vol. 9, no. November, pp. 1–10, 2015, doi: 10.3389/fnana.2015.00152.

[106] D. Godwin, A. Ji, S. Kandala, and D. Mamah, "Functional Connectivity of Cognitive Brain Networks in Schizophrenia during a Working Memory Task," *Front. Psychiatry*, vol. 8, no. December, 2017, doi: 10.3389/fpsyt.2017.00294.

[107] I. Oliver, J. Hlinka, J. Kopal, and J. Davidsen, "Quantifying the variability in resting-state networks," *Entropy*, vol. 21, no. 9, pp. 1–21, 2019, doi: 10.3390/e21090882.

[108] L. Chen *et al.*, "Topological reorganization of the default mode network in severe male obstructive sleep apnea," *Front. Neurol.*, vol. 9, no. JUN, pp. 1–11, 2018, doi: 10.3389/fneur.2018.00363.

[109] J. R. Andrews-Hanna, J. S. Reidler, J. Sepulcre, R. Poulin, and R. L. Buckner, "Functional-Anatomic Fractionation of the Brain's Default Network," *Neuron*, vol. 65, no. 4, pp. 550–562, 2010, doi: 10.1016/j.neuron.2010.02.005.

[110] Adam Hayes, "T-Test: What It Is With Multiple Formulas and When To Use Them," *investopedia.com*. .

[111] R. Davidson and J. G. MacKinnon, "Bootstrap tests: How many bootstraps?," *Econom. Rev.*, vol. 19, no. 1, pp. 55–68, 2000, doi: 10.1080/07474930008800459.

[112] D. W. K. Andrews and M. Buchinsky, "A Three-Step Method for Choosing the Number of Bootstrap Repetitions Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/2999474 REFERENCES Linked references are available on JSTOR for this article : You may need to log in to J," *Econometrica*, vol. 68, no. 1, pp. 23–51, 2000.

[113] S. Journal, R. Statistical, S. Series, and C. A. Statistics, "Barnard ' s Monte Carlo Tests : How Many Simulations ? Author (s): F . H . C . Marriott Published by : Blackwell Publishing for the Royal Statistical Society Stable URL : http://www.jstor.org/stable/2346816," vol. 28, no. 1, pp. 75–77, 2010.

[114] D. Rangaprakash, G. R. Wu, D. Marinazzo, X. Hu, and G. Deshpande, "Hemodynamic response function (HRF) variability confounds resting-state fMRI functional connectivity," *Magn. Reson. Med.*, vol. 80, no. 4, pp. 1697–1713, 2018, doi: 10.1002/mrm.27146.

[115] V. Prčkovska *et al.*, "Reproducibility of the Structural Connectome Reconstruction across Diffusion Methods," *J. Neuroimaging*, vol. 26, no. 1, pp. 46–57, 2016, doi: 10.1111/jon.12298.

[116] S. Elhabian *et al.*, "Subject-motion correction in HARDI acquisitions: Choices and consequences," *Front. Neurol.*, vol. 5, no. NOV, pp. 1–54, 2014, doi: 10.3389/fneur.2014.00240.

[117] D. Christiaens *et al.*, "Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI," *Neuroimage*, vol. 225, no. September 2020, p. 117437, 2021, doi: 10.1016/j.neuroimage.2020.117437.

[118] B. A. K. Kreilkamp, D. Zacà, N. Papinutto, and J. Jovicich, "Retrospective head motion correction approaches for diffusion tensor imaging: Effects of preprocessing choices on biases and reproducibility of scalar diffusion metrics," *J. Magn. Reson. Imaging*, vol. 43, no. 1, pp. 99–106, 2016, doi: 10.1002/jmri.24965.

[119] X. Zhou *et al.*, "Quantitative quality assurance in a multicenter HARDI clinical trial at 3 T," *Magn. Reson. Imaging*, vol. 35, pp. 81–90, 2017, doi: 10.1016/j.mri.2016.08.022.

[120] S. M. H. Haddad *et al.*, "Comparison of quality control methods for automated diffusion tensor imaging analysis pipelines," *PLoS One*, vol. 14, no. 12, pp. 1–28, 2019, doi: 10.1371/journal.pone.0226715.

[121] A. F. Alexander-Bloch *et al.*, "Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia," *Front. Syst. Neurosci.*, vol. 4, no. October, pp. 1–16, 2010, doi: 10.3389/fnsys.2010.00147.

[122] K. H. Maier-Hein *et al.*, "The challenge of mapping the human connectome based on diffusion tractography," *Nat. Commun.*, vol. 8, no. 1, 2017, doi: 10.1038/s41467-017-01285-x.

[123] J. S. W. Campbell and G. B. Pike, "Potential and limitations of diffusion MRI tractography for the study of language," *Brain Lang.*, vol. 131, pp. 65–73, 2014, doi: 10.1016/j.bandl.2013.06.007.

[124] J. Tournier, "High angular resolution diffusion-weighted imaging (HARDI): beyond the diffusion tensor model The diffusion tensor model and the crossing fibre problem method of choice for assessing white matter ' integrity ' and connectivity . However , there is now a," *Society*, pp. 12–13, 1994.

[125] A. Shah, S. Jhawar, A. Goel, and A. Goel, "Corpus Callosum and Its Connections: A Fiber Dissection Study," *World Neurosurg.*, vol. 151, pp. e1024–e1035, 2021, doi: 10.1016/j.wneu.2021.05.047.

[126] Y. Zhou *et al.*, "Altered resting-state functional connectivity and anatomical

connectivity of hippocampus in schizophrenia," *Schizophr. Res.*, vol. 100, no. 1–3, pp. 120–132, 2008, doi: 10.1016/j.schres.2007.11.039.