Assessment of Brain Connectivity with Deterministic Tractography

Bedömning av hjärnanslutning med deterministisk traktografi

MARIA DEL LLEDÓ NEBOT MARTINEZ

Degree project in medical engineering
First level, 15 credits
Supervisor at KTH: Rodrigo Moreno
Examiner: Mats Nilsson

KTH Royal Institute of Technology
School of Engineering Sciences in Chemistry, Biotechnology and Health
Hälsovägen 11 C
SE-141 57 Huddinge, Sweden
http://www.kth.se/cbh

2023

Abstract

This thesis aim is to assess the performance and effectiveness of different deterministic tractography algorithms in comparison to probabilistic methods in generating tractograms from dMRI data of both control subjects and individuals with mild cognitive impairment (MCI), as well as examine the effects of tractogram filtering through spherical-deconvolution informed filtering of tractograms (SIFT). Subquestenly, connectivity matrices are constructed from the filtered tractograms and evaluated through graph theory measures and other metrics associated with brain connectivity.

Motivated by the limitations observed in current methods for reconstructing brain fibers, which prevent their ability to accurately represent the anatomy of the brain, this research addresses the need for improved methodologies. To overcome this issue, researchers have proposed tractogram filtering algorithms that selectively remove streamlines that do not align well with the actual brain structure, improving the accuracy and reliability of the connectomes.

The experimental results show that the deterministic algorithm chosen for the study yields satisfactory results with a relatively low number of generated streamlines. Nevertheless, the filtering technique significantly affects connectivity measures such as connection density, global efficiency, and clustering coefficients., resulting in values that may fall short when compared to those obtained through probabilistic methods. Furthermore, the filtering of tractograms impacts various graph measures for individual nodes, such as node strength, eigenvector centrality, and betweenness centrality. As a result, new connectomes are formed where specific nodes become more significant. However, it is important to consider the limitations of the study, including a small number of subjects, computation time constraints, and potential errors in image registration and tractography parameterization.

Keywords: deterministic tractography, probabilistic tractography, SIFT, brain connectivity, graph metrics.

Sammanfattning

Syftet med denna avhandling är att bedöma prestanda och effektivitet hos olika deterministiska traktografialgoritmer i jämförelse med probabilistiska metoder för att generera traktogram från dMRI-data från både kontrollpersoner och individer med mild kognitiv funktionsnedsättning (MCI), samt att undersöka effekterna av traktogramfiltrering genom sfärisk dekonvolutionsinformerad filtrering av traktogram (SIFT). Subquestenly, anslutningsmatriser konstrueras från de filtrerade traktogrammen och utvärderas genom grafteoretiska mått och andra mätvärden associerade med hjärnanslutning.

Motiverad av de begränsningar som observeras i nuvarande metoder för att rekonstruera hjärnfibrer, som förhindrar deras förmåga att korrekt representera hjärnans anatomi, tar denna forskning upp behovet av förbättrade metoder. För att övervinna detta problem har forskare föreslagit traktogramfiltreringsalgoritmer som selektivt tar bort strömlinjer som inte passar bra med den faktiska hjärnstrukturen, vilket förbättrar noggrannheten och tillförlitligheten hos kopplingarna.

De experimentella resultaten visar att den deterministiska algoritm som valts för studien ger tillfredsställande resultat med ett relativt lågt antal genererade effektiviseringar. Ändå påverkar filtreringstekniken avsevärt anslutningsmått såsom anslutningstäthet, global effektivitet och klustringskoefficienter, vilket resulterar i värden som kan falla kort jämfört med de som erhålls genom probabilistiska metoder. Dessutom påverkar filtreringen av traktogram olika grafmått för individuella noder, såsom nodstyrka, egenvektorcentralitet och centralitet mellan varandra. Som ett resultat bildas nya kopplingar där specifika noder blir mer betydelsefulla. Det är dock viktigt att överväga studiens begränsningar, inklusive ett litet antal ämnen, beräkningstidsbegränsningar och potentiella fel i bildregistrering och traktografiparameterisering.

Nyckelord: deterministisk traktografi, probabilistisk traktografi, SIFT, hjärnanslutning, grafmetrik.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Rodrigo Moreno, Associate Professor at the Divison of Biomedical Engineering at KTH for his guidance and input during the elaboration of this project. Moreover, I would like to thank the PhD student and at the Division of Biomedical Enginnering at KTH Fabian Sinzinger and the master student Teodor Pstrusiński for offering guidance and valuable insights when I encountered difficulties with particular subjects.

Finally, I am deeply grateful for the unwavering support and belief in my abilities that my close friends and family have consistently provided. Without their encouragement, none of my accomplishments would have been possible.

Contents

Abbreviations

CTL Control Class

MCI Mild Cognitive Impairment

MRI Magnetic Resonance Imaging

dMRI diffusion MRI

WM White Matter

GM Grey Matter

CSF Cerebro-Spinal Fluid

SIFT Spherical-devonvolution Informed Filtering of Tractograms

FA Fractional Anisotropy

MD Mean Diffusivity

RD Radial Diffusivity

ACT Anatomically Constrained Tractography

5TT Five Tissue Type

FACT Fiber Assigned Continuous Tracking

FOD Fiber Orientation Distribution

1 Introduction

Diffusion magnetic resonance imaging (dMRI) is an imaging technique that captures the directional diffusion of water molecules in brain tissues, with a particular focus on the axons composing the white matter (WM). Through this process, dMRI enables the reconstruction of neural fiber pathways within the brain, commonly referred to as tractography, which produces tractograms (see Appendix B.4). There are various factors that can cause inaccuracies such as streamline seeding, selection of fiber orientation and white matter volume. To address these challenges, filtering techniques are applied to the generated tractograms, improving their accuracy (see Appendix B.5). By associating these reconstructed fiber connections with predefined brain regions, a connectome can be computed, representing the brain as a network of interconnected regions (see Appendix B.6).

1.1 Aim

This study seeks to assess the performance and effectiveness of deterministic tractography algorithms in generating tractograms from dMRI data of both Mild Cognitive Impairment (MCI) patients and healthy individuals, improving the specificity of the reconstructed fibers by applying SIFT, a spherical-deconvolution informed filtering technique, to the tractograms obtained.

Moreover, the study strives to compare the outcomes and metrics derived from deterministic tractography with those obtained through probabilistic methods, providing insights into the advantages and limitations of using deterministic approaches. Finally, the study aims to investigate the impact of tractogram filtering on structural connectomes and derived graph metrics.

1.2 Limitation

The study is limited by the relatively small number of subjects available for analysis, which which may make it difficult to apply the findings to a larger group of people. Moreover, the tractography algorithms used in this study may have limitations in accurately representing the underlying anatomical connections in the brain, leading to reduced specificity and interpretability of the resulting connectomes. In regards to the proposed SIFT filtering approach, its performance and reliability may be influenced by various factors such as the quality of the dMRI data and choice of parameters. Finally, other factors that were not directly considered in this study, such as individual anatomical variability or clinical characteristics, could potentially influence the observed connectivity patterns and metrics.

2 Methods

This chapter provides an overview of the methodology and methods applied to achieve the aims of the investigation. First, the dataset of structural and diffusion images including 72 subjects (Section 2.1) is implemented into the pipeline to produce complete brain tractograms (Section 2.2). Subsequently, tractogram filtering is applied (Section 2.2.2). To examine the connectivity between brain regions along the fibers, the FSL Harvard-Oxford parcellation atlas, which is in the MNI152 Standard Space, is utilized. However, it is necessary to align the parcellation atlas with the diffusion space where the tractograms exist. This alignment is achieved through linear and non-linear registrations (Section 2.3). After the registration process, connectivity matrices are generated using various metrics and graph measures (Section 2.4). The schematic diagram illustrating the utilized pipeline can be found in Figure 2.1.

The code for the pipeline applied in this final project is derived from the work carried out by Marvin Köpff in 2020 for his Master's thesis project in the School of Engineering Sciences in Chemistry, Biotechnology and Health of the KTH Royal Institute of Technology (Stockholm, Sweden).

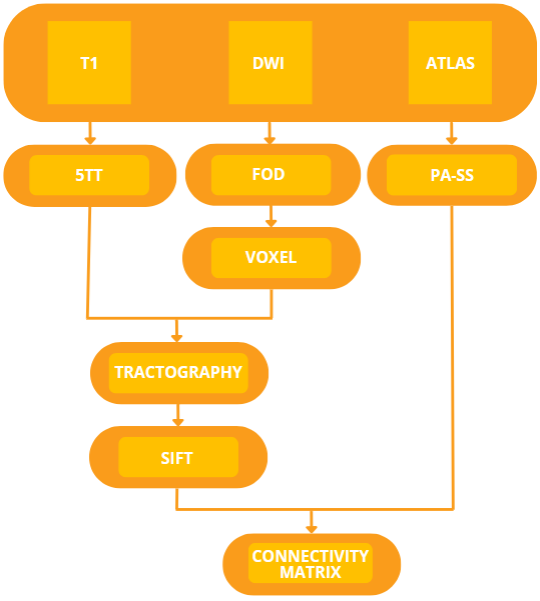

Figure 2.1: Illustrates the processing pipeline employed in this study. Starting from the diffusion images (DWI), the pipeline involves the generation of whole brain tractograms through the Tractography and SIFT steps. The parcellation atlas is then registered into the diffusion space of the subject using PA-SS. Connectivity matrix represents the process of generating connectivity maps based on the filtered tractograms and the parcellation atlas in the diffusion space of the subject.

2.1 Dataset

The dataset used in this study is a subset of structural (T1-weighted) and diffusion images obtained from the Swedish BioFINDER study conducted at Skåne University Hospital and Ängelholm's Hospital (http://biofinder.se/). The dataset includes images from elderly patients with a mean age of 72 years and a standard deviation of 4.35 years. The individuals included in

the dataset are classified as either having Mild Cognitive Impairment (MCI) or normal cognitive abilities (control group, denoted as CTL).

The dataset consists of a total of 73 subjects, with 36 subjects belonging to the CTL class and 37 patients belonging to the MCI class. The T1-weighted images in the dataset contain 176 slices with a voxel size of $0.97 \times 0.97 \times 1.2$ mm³. The diffusion images comprise 65 axial slices with 64 diffusion-weighted directions acquired at a b-value of 2000 s/mm² and an isotropic voxel size of $2 \times 2 \times 2$ mm³.

The diffusion images in the dataset have already undergone preprocessing steps, including eddy-correction and motion distortion correction. Additionally, the dataset includes the parcellation atlas from FSL, which identifies 110 brain regions. This atlas corresponds to the Harvard-Oxford brain atlas (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases).

2.2 Tractography

Tractography involves the reconstruction of the streamlines of axons from previously acquired DWI images and that way recollect data about the connection between brain regions. For the computation of the tractograms, there are mainly two types of algorithms available: deterministic and probabilistic. This study is focused on deterministic algorithms, a streamline-based approach, where a single pathway is traced from a specific seed point. The direction of the pathway is determined by the primary eigenvector at each voxel location, and the tracing process is halted based on predefined criteria, such as the fractional anisotropy (FA) value and angle change.

Additionally, deterministic tractography also allows the use of Anatomically Constrained Tractography (ACT), a technique that utilizes white matter (WM) and gray matter (GM) information to guide the seeding and determination of streamlines. Finally, tractogram reconstruction algorithms are not always able to identify false streamlines (see Appendix B.5), thus, tractogram filtering via the SIFT algorithm is applied to the tractograms computed. The deterministic algorithms and SIFT were applied to the data using the MRTrix3Tissue software package (https://3Tissue.github.io), a fork of MRtrix3.

2.2.1 Tractography algorithm selection

To determine the most suitable deterministic algorithm for the pipeline, a comparative analysis was conducted using the three available deterministic algorithms in the MRTrix3Tissue software package: FACT, SD_STREAM, and Tensor_Det (see Appendix B.4). The evaluation involved examining the remaining streamlines after applying SIFT filtering, as well as the connection density, for different streamline generation values. Based on the results, the FACT algorithm was selected as the preferred choice for generating the whole brain tractograms to be utilized in the subsequent stages of the pipeline.

2.2.2 Tractogram generation

To generate whole brain tractograms using the tckgen script from MRTrix3Tissue, several prerequisites need to be met. These include having five-tissue-type (5TT) images containing cortical grey matter (GM), sub-cortical GM, white matter (WM), cerebro-spinal-fluid (CSF),

and pathological tissue. Additionally, for the FACT algorithm specifically, diffusion images in the form of voxel images, providing spatially resolved information about fixel orientations in the voxel grid, are required.

The 5TT images are generated from the structural image using the 5ttgen tool in combination with the ss3t_csd_beta1 script from MRTrix3Tissue, which estimates the fiber orientation distribution (FOD) images. Unlike the traditional multi-shell multi-tissue (msmt_csd) algorithm provided in the standard MRTrix3 software package, the ss3t_csd_beta1 script is specifically designed for FOD estimation on single-shell images, making it the preferred choice for this study [1]. Even so, to perform FOD estimation, response functions for WM, GM, and CSF need to be extracted for each subject using the dwi2response script and the dhollander algorithm. Furthermore, to meet the input file requirements of the FACT algorithm, additional steps are performed. Specifically, the WM-FOD estimation results are subjected to the fod2fixel and fixel2voxel processes, converting it into voxel images, ensuring compatibility with the FACT algorithm.

Having fulfilled the prerequisites for tckgen, the script requires specific inputs to initiate the process. These inputs include the generated voxel image, the estimated WM-FOD image serving as the seed_dynamic, the 5TT image, and the parameters outlined in table 2.1. The selection of minimum and maximum streamline length was based on several studies performing whole brain tractography. To balance efficiency and accuracy, the number of reconstructed streamlines was set to 1 million, due to the simplicity of the tracking algorithm and the computation time.

Table 2.1: The parameters used for the computation of the whole brain tractograms with tckgen, -select is equal to the number of streamlines which are going to be reconstructed. Additional parameters were *-act* for the use of ACT, *-crop_at_gmwmi*, and *-seed_dynamic*.

tckgen parameters				
-minlength	-maxlength	-step_size	-select	-cutoff
25	350	0.6	1000000	0.03

2.2.2. Tractogram Filtering

In order to filter the generated whole brain tractograms, the SIFT algorithm was implemented, even though SIFT was succeeded by SIFT2. This is due to the fact that SIFT2 assigned lower weights in the central areas of the fiber bundles, where tractography typically generates a greater number of streamlines, leading to the premature elimination of crucial tracts and decreased accuracy (see Appendix B5). The SIFT algorithm was applied using the tcksift script from MRTrix3Tissue, utilizing the WM-FOD image obtained during the preprocessing stage.

2.3 Registration

To produce the connectivity matrices from the generated tractograms, these must be aligned with the parcellation atlas. This process is based on the approach that is described in [2]. To register the individual structural images to the standard space, we first align them to the

ICBM152 structural atlas, which is the standard space for the Harvard-Oxford parcellation atlas. After the affine registration, we account for the non-linear distortions by performing a non-linear registration. By implementing the Fractional Anisotropy (FA) image derived from the raw diffusion image we co-register the diffusion image with the structural image. The previous transformations are performed using an affine transformation. Once the registrations are completed, the parcellation atlas can be transformed to the diffusion space. This is accomplished by applying the inverse transformation obtained from the registrations mentioned. The registrations and transformations are performed using FLIRT and FNIRT, which are tools provided with FSL (FMRIB Software Library) [3, 4, 5, 6].

By following these steps, the parcellation atlas is aligned with the diffusion space, enabling the computation of connectivity matrices. However, the registration process was unsuccessful for subject 19 in the dataset. As a result, the subsequent steps in the pipeline were only performed for the remaining 72 subjects.

2.4 Connectivity Matrix Generation

The *tck2connectome* script is utilized to compute connectivity matrices from both the unfiltered and filtered tractograms, along with the individual parcellation atlases. This connectivity matrix represents the connections between all nodes (brain regions) in the network. The script locates the endpoints of each streamline and attempts to assign them to the nearest available brain regions. By default, this assignment is performed within a search radius of 4 mm, as specified by the *-assignment_radial_search_radius* parameter. The resulting connectivity matrices are represented as symmetric adjacency matrices (110x110), where each element corresponds to an edge connecting brain regions in the respective row and column. Moreover, self-connections of brain regions are excluded from the resulting matrix, to focus on the connections between different nodes and simplify the analysis of network properties.

Furthermore, it is also possible to scale each streamline individually by weights in an external file. To compute the connectivity matrices based on these external weights, the *tcksample* script sampled the image intensities along the streamlines. Images to sample from are derivations of the raw diffusion image, these being fractional anisotropy (FA), mean diffusivity (MD) and radial diffusivity (RD). Since these weighted matrices consisted of values close to zero in the lower diagonal, the 50% quantile of the non-zero value ranges in the connectivity matrix was calculated. This quantile value was then used as a binary threshold, where any value above the threshold was set to 1, while any value below the threshold was set to 0. This binary representation emphasized meaningful connections and reduced noise. However, this technique was not applied to the raw connectomes since their values were already significant.

To compute the connectivity matrices of the filtered and unfiltered tractograms, both raw streamline count, a straightforward count of the number of streamlines that connect two regions, and external weights are implemented. Hence, through the use of tcksample and tck2connectome, a total of eight connectivity matrices per subject were generated, two for the raw streamline count and six for the weighted matrices. These calculations were performed both before and after the implementation of SIFT filtering.

2.4.1 Metrics on the Connectivity Matrices

There are several metrics that can be extracted from the connectivity matrices generated. The initial metric calculated is connection density (Equation 2.1), which is determined by the ratio of nonzero elements in matrix A compared to the total potential connections that could exist in the network [7]. Since the edges in a brain network generated by dMRI data are undirected, the following equation is applied, being M the number of edges and N the number of nodes.

$$d = \frac{2M}{N(N-1)} \tag{2.1}$$

The clustering coefficient measures the likelihood that two nodes, each connected to a third node, will also be directly connected to each other. It quantifies the fraction of a node's neighbors that are connected to each other. The average clustering coefficient of the network provides an indication of the overall presence of clustered connections around individual nodes. In essence, it reveals the extent to which nodes tend to form local clusters of connections within the network [8]. When there are many connections between regions that have similar functions, the clustering coefficient of the graph increases. This means that neighboring nodes within these regions tend to be interconnected [9].

The global efficiency is the average inverse shortest path length [8], focusing on the efficiency of communication between nodes, particularly emphasizing the impact of shorter paths rather than the influence of long paths.

Connectivity matrices can be viewed as graphs, allowing for the assessment of various graph measures. These measures include node strength, degree centrality, eigenvector centrality and betweenness centrality. Node strength, represented as S (Equation 2.2), considers weighted connections and quantifies the sum of a node's connections to other nodes. High node strength indicates a node with a significant number or weight of connections compared to other nodes in the network, highlighting its potential as a network hub [10].

$$s_i = \sum_{j \neq i} w_{ij} \tag{2.2}$$

Degree centrality is the simplest measure of centrality, assuming that nodes with many connections exert more influence over network function and thus have higher topological centrality compared to nodes with fewer connections. To account for its limitations we compute eigenvector centrality, which captures the combined influence and significance of a node's connections, taking into account not only the node's own degree but also the degrees of its neighbors. In other words, it calculates a node's centrality based on the centrality of its neighbors.

The betweenness centrality graph metric involves calculating the shortest paths between all pairs of nodes in the network. Subsequently, each node is assigned a score based on the number of shortest paths that pass through it [11]. Nodes with a greater number of shortest paths passing through them have higher betweenness centrality, indicating their significant role in transmitting information across the network [12]. Graph measures computations were performed using the networkX python library (https://networkx.github.io).

3 Results

3.1 Tractography algorithm selection

The subsequent section presents the results obtained from comparing the three deterministic algorithms offered by MRTrix3: FACT, SD_STREAM, and Tensor_Det. This comparison involved computing the remaining streamlines after applying SIFT, as well as analyzing the connection density and the distribution of streamline lengths before and after filtering.

3.1.1 Streamlines left after SIFT

After performing SIFT on the complete brain tractograms generated using varying numbers of streamlines (100k, 1 million, 3 million, and 5 million), Figure 3.1 depicts the percentage of average streamlines preserved in the entire brain tractograms. Among the different methods employed, the use of FACT with 100k streamlines generated presented the highest retention percentage, whereas Tensor_Det with 5 million streamlines demonstrated the lowest retention percentage, quantifying 31.33% and 1.78%, respectively.

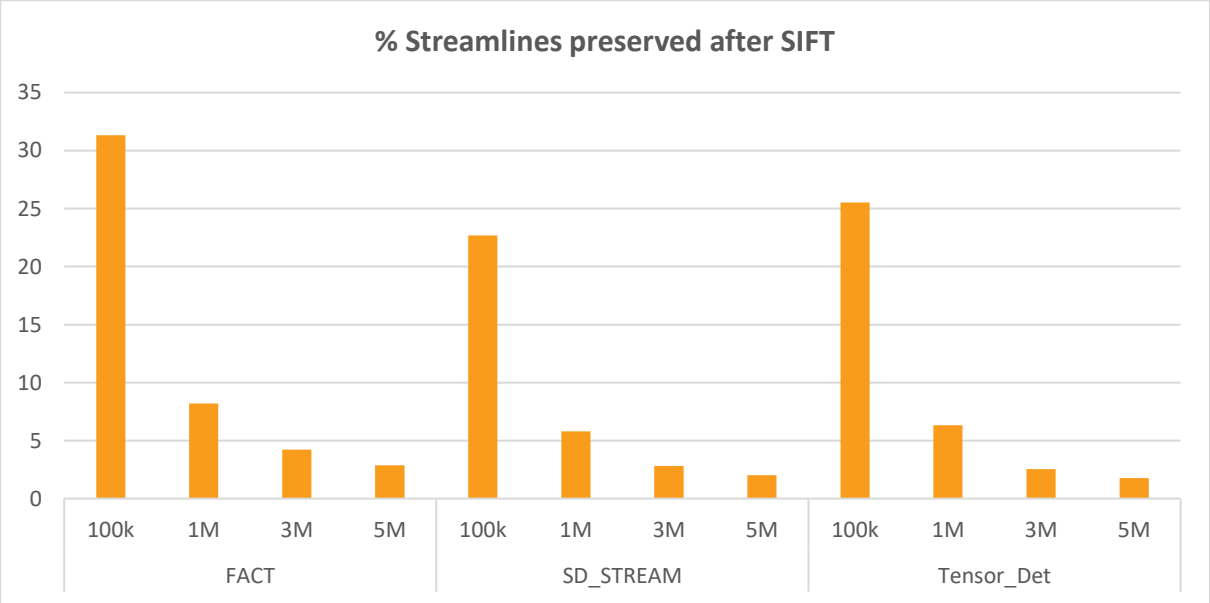

Figure 3.1: Percentage of streamlines preserved after SIFT for the three deterministic algorithms computed.

3.1.2 Connection Density

The connection density, representing the proportion of actual connections (streamlines) existing between pairs of nodes in the network relative to the total number of possible connections, for each algorithm is represented in figure 3.2. In all cases, the connection density is lower after applying SIFT on the tractograms, with its value increasing as more streamlines are generated.

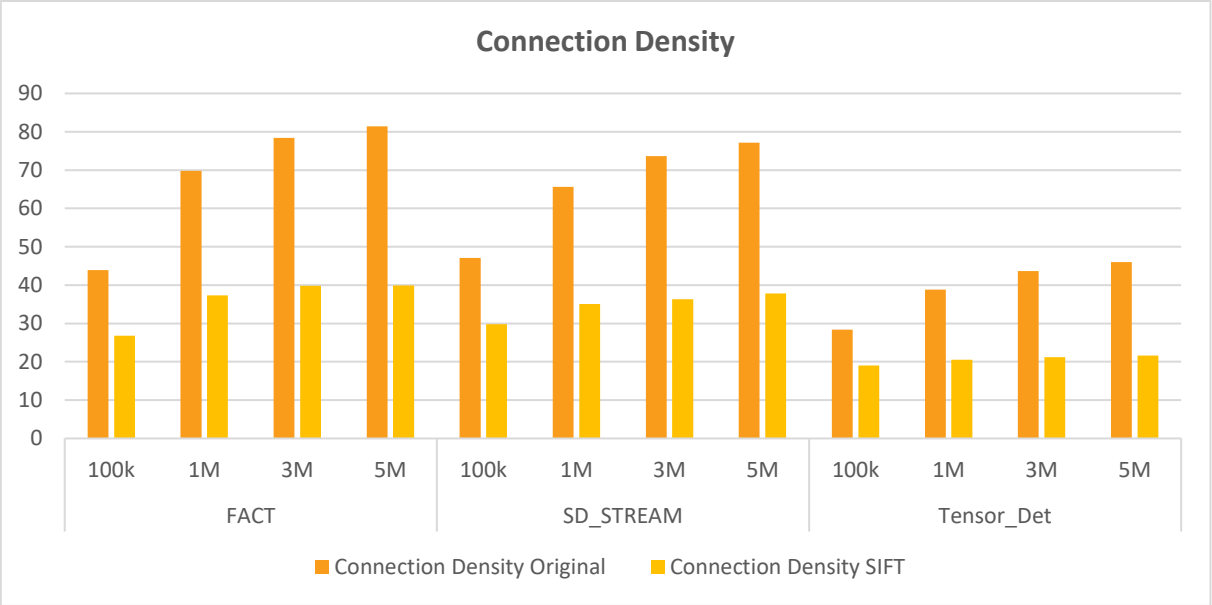

Figure 3.2: Connection density before and after SIFT for the three deterministic algorithms computed.

3.1.3 Streamline lengths

The distribution of streamline lengths before and after applying SIFT on the tractograms generated for 1 million streamlines is represented in figures 3.3, 3.4 and 3.5 in the form of histograms. In all cases the number of longer streamlines is drastically decreased, with FACT having a more uniform distribution before SIFT and having the greatest number of short streamlines after SIFT.

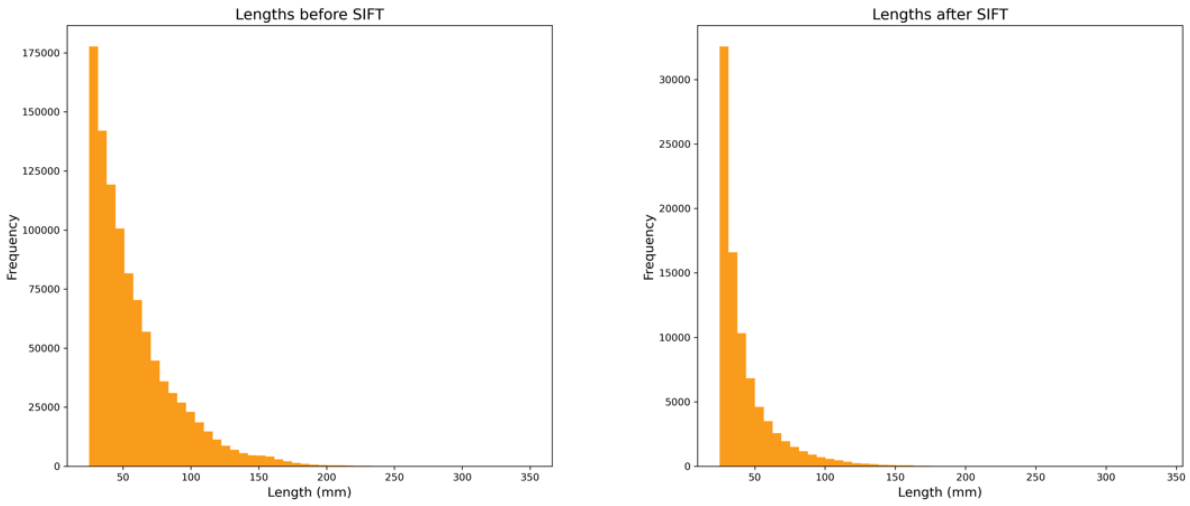

Figure 3.3: Histogram representing the distribution of streamline lengths before and after SIFT for the deterministic algorithm FACT with 1 million streamlines generated.

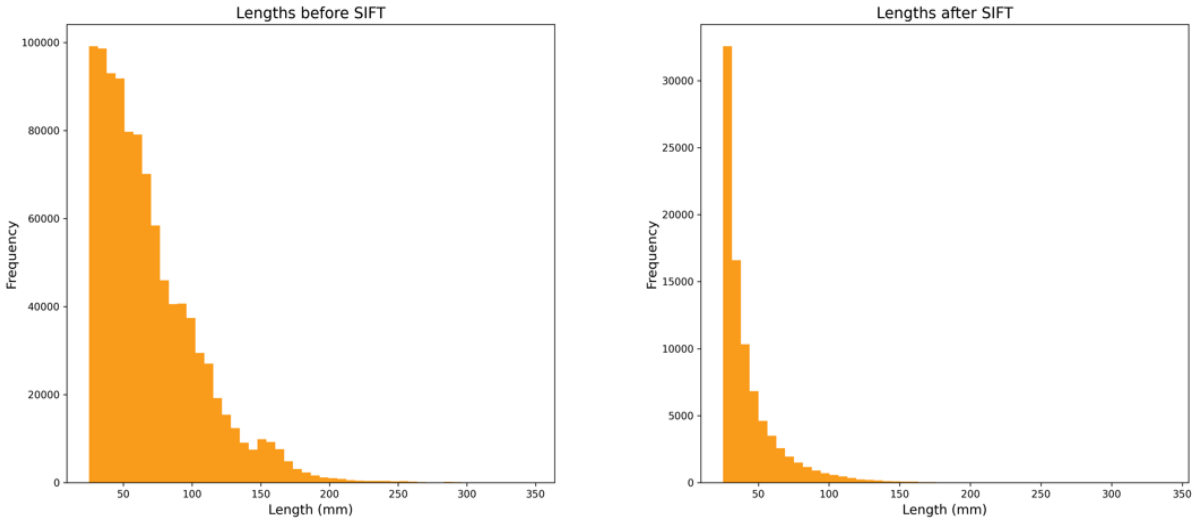

Figure 3.4: Histogram representing the distribution of streamline lengths before and after SIFT for the deterministic algorithm SD_STREAM with 1 million streamlines generated.

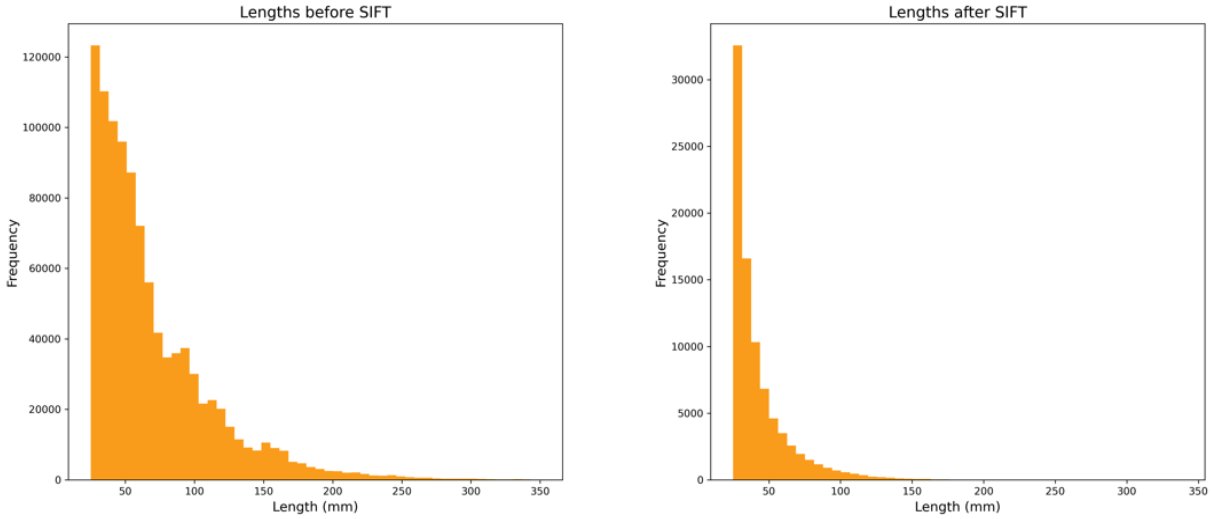

Figure 3.5: Histogram representing the distribution of streamline lengths before and after SIFT for the deterministic algorithm Tensor_Det with 1 million streamlines generated.

3.2 Differences of SIFT on Connectivity Matrices

The following sections describe the differences of connectograms generated without and with applying SIFT filtering to the tractograms.

3.2.1 Connectivity metrics

After applying SIFT to the whole brain tractograms generated with 1 million streamlines, the average streamlines left in the whole brain tractograms is of 92567, resulting in in an average filtering outcome of 9.26%. The connection density, clustering coefficient and global efficiency of the connectivity matrices generated is listed in table 3.1. All values decrease in a drastic manner after SIFT, with connection density having the greatest difference of 0.30.

Additionally, the same connectivity matrices evaluated for the weighted matrices based on FA, MD, and RD images are represented on table 3.1. Compared to the matrices generated with raw streamline count, the same tendency to decrease after applying SIFT can be visualized.

Table 3.1: Difference in connection density, global efficiency and clustering coefficient between original and SIFT filtered connectomes.

	Original				SIFT			
	Raw	FA	MD	RD	Raw	FA	MD	RD
Connection Density	0.68	0.32	0.32	0.32	0.38	0.19	0.17	0.17
Global Efficiency	0.84	0.66	0.65	0.65	0.69	0.57	0.56	0.55
Clustering Coefficient	0.82	0.42	0.49	0.56	0.66	0.33	0.40	0.48

3.2.2 Graph metrics

The importance of the nodes (equivalent to the specific names for the brain regions as seen in Appendix 1) is evaluated using the following graph metrics: node strength, eigenvector centrality and betweenness centrality. The boxplots visualizing the graph metrics in all sections show the mean value of the individual node's graph metric (separating line between boxes), the lower and upper quartile (boxes) as well as the rest of the distribution (whiskers).

Regarding the 10 most important nodes with respect to each metric, applying SIFT changes the order and introduces new more important nodes. For example, in the case of the node strength after SIFT, node 91 replaces node 61 and node 43 is deemed less important.

The figure 3.6 showcases the individual node strength of each node class before and after applying a filtering technique. Observation of the data distribution reveals that prior to the application of SIFT, the distribution appears to be more uniformly spread compared to the distribution after SIFT. This difference is evident in the boxplots, where some exhibit increased length while others become shorter. Furthermore, after applying SIFT, there is a noticeable decrease in the number of nodes with exceptionally high nodal strength. This can be observed from the distribution, which shows a reduction in the occurrence of high node strength values.

Just as with node strength, the centralities have fewer high values after applying SIFT (Figure 3.7). On the other hand, both the eigenvector centralities and the betweenness centralities (Figure 3.8) show a less drastic change in comparison to node strength, resulting in a more uniform distribution after SIFT was applied.

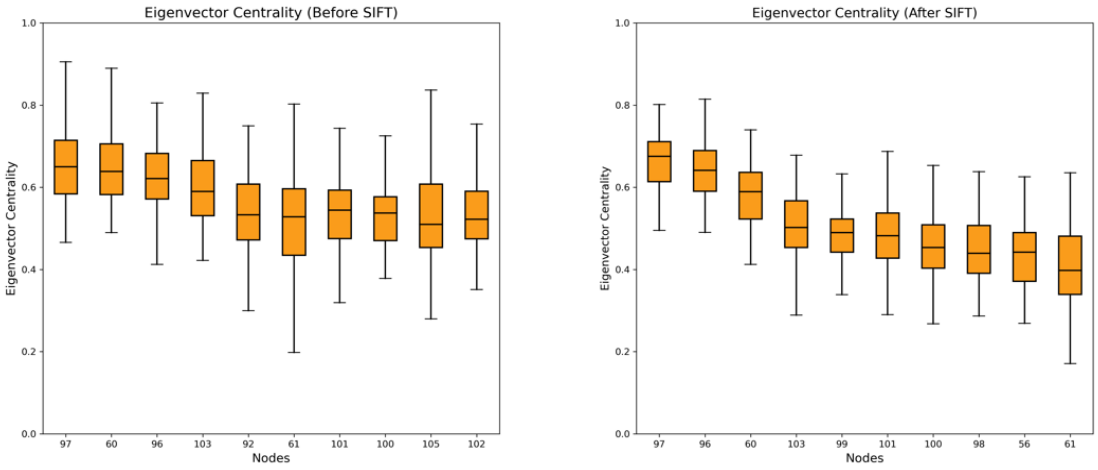

Figure 3.6: Nodestrength of the 10 strongest nodes (normalized) for connectomes based on unfiltered connectograms (left) and filtered connectograms (right).

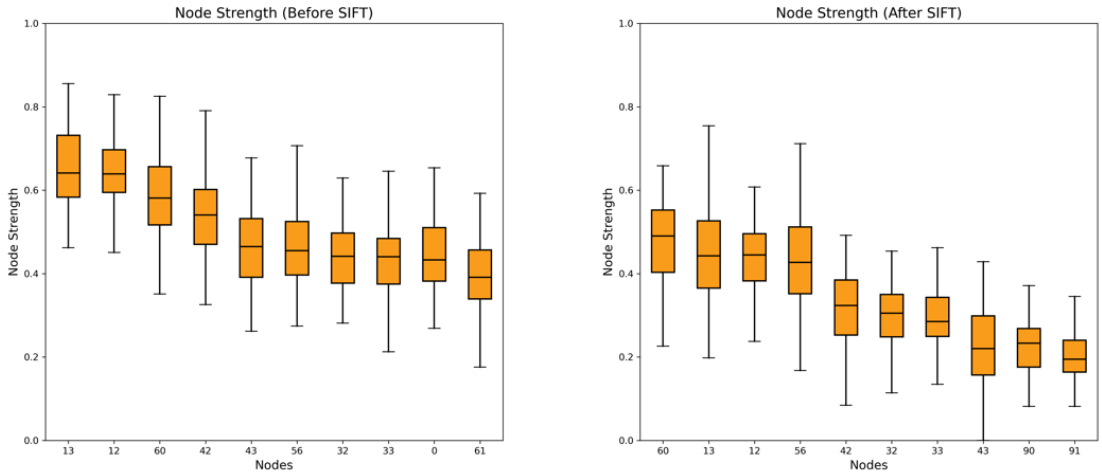

Figure 3.7: Eigenvector centrality of the 10 strongest nodes (normalized) for connectomes based on unfiltered connectograms (left) and filtered connectograms (right).

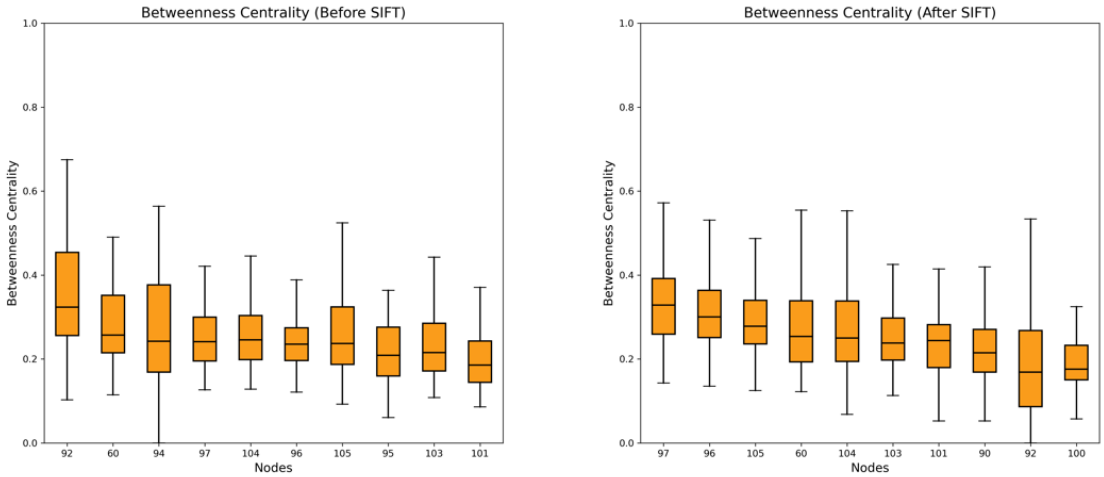

Figure 3.8: Betweenness centrality of the 10 strongest nodes (normalized) for connectomes based on unfiltered connectograms (left) and filtered connectograms (right).

4 Discussion

The following discussion tries to summarize, explain and set conclusions from the results into context. Furthermore, possible sources of error in the pipeline are mentioned.

4.1. Tractography algorithm selection

In terms of the percentage of preserved streamlines after applying SIFT (see Figure 3.1), FACT consistently had the highest percentage among all cases. When considering connection density (see Figure 3.2), both FACT and SD_STREAM exhibited good results. The connection density increased with a higher number of generated streamlines, indicating that more valid streamlines were being produced. However, FACT consistently had a slightly higher connection density across all cases. Moreover, regarding the distribution of streamline lengths before and after SIFT (see Figure 3.3), FACT had a higher number of shorter streamlines after filtering. This could be attributed to FACT's ability to retain a larger number of streamlines.

Overall, all three algorithms showed a satisfactory response to filtering, effectively reducing the number of longer invalid streamlines. However, FACT consistently outperformed the other algorithms in terms of the percentage of preserved streamlines after filtering and the connection density before and after SIFT. Therefore, FACT was selected as the preferred algorithm to proceed with the study.

Additionally, considering the minimal difference in values between a higher number of generated streamlines and 1 million streamlines, it was determined that 1 million streamlines would be sufficient for the scope of this project.

4.2. Connectivity metrics

After applying SIFT filtering to the initial whole brain tractograms, which originally consisted of 1 million streamlines, it was observed that the remaining streamlines accounted for 9.26% of the initial tractograms. This value was found to be significantly higher than the one reported in the previous study conducted [13]. It is worth noting that this difference could be attributed to the generation of a significantly larger number of streamlines in that particular study.

Furthermore, the relatively low percentage of preserved streamlines can be attributed to the nature of the SIFT filtering process. SIFT focuses on optimizing the cost-function and achieving convergence, which takes precedence over retaining a higher percentage of streamlines.

In addition, the application of SIFT also impacts the connection density (see Table 3.1), which is a metric reflecting the proportion of actual connections between pairs of nodes. In particular, there is a notable decrease in the connection density after employing SIFT. This decrease can be attributed to the removal of streamlines that may connect two nodes but fail to meet the filtering constraints, such as high curvature or loops. The decrease in clustering coefficient could suggest that streamlines contributing to local connections within specific brain regions have been affected. Meanwhile, the decrease in global efficiency may be a result of the

removal of streamlines, disrupting the connections between different brain regions. These changes indicate a potential reorganization or disruption of the brain's connectivity patterns due to the SIFT filtering process.

The same decrease in connectivity metrics is seen in the weighted matrices evaluated (see Table 3.2). In all three external weights considered, connection density has similar values, indicating a relatively uniform distribution of connections in the matrices. Moreover, the clustering coefficient metric demonstrates greater variation in each weight considered. This variation can be attributed to its higher sensitivity to local connectivity patterns compared to the other two connectivity metrics analyzed.

4.3. Graph metrics on Connectivity Matrices

The graph metrics of eigenvector (see Figure 3.7) and betweenness centrality (see Figure 3.8) show the distribution of the most important nodes after SIFT filtering being more centralized around their mean. This effect could be due to the more uniform distribution of edge weights (streamline counts) across the subjects after applying SIFT. The graph metric of node strength (see Figure 3.6) for the most important nodes shows the behavior of having a lower mean after SIFT filtering.

Compared to [13], the changes in node strength and betweenness centrality after filtering were less drastic in this analysis. This difference could be attributed to the use of a probabilistic algorithm for tractogram generation, which could be more sensitive to the filtering process. This sensitivity may be due to the inherent nature of probabilistic tractography, which involves generating multiple plausible streamlines between two regions.

The main finding of this study is the significant alteration in the importance of nodes after applying SIFT filtering. In all cases, nodes were replaced and shifted in their importance ranking. This demonstrates the impact of SIFT on the graph properties of the generated networks. A similar result was observed in [13], however, despite the shared dataset, the top 10 important nodes differed between the two algorithms. This discrepancy may be attributed to the choice of streamlines preserved after filtering, which can influence the ranking and identification of important nodes.

The observed differences in the changes before and after filtering for the probabilistic and deterministic methods suggest that these two approaches handle the filtering of streamlines differently. In the case of probabilistic tractography, the algorithm explores a wider range of potential pathways between brain regions, which provides greater flexibility but also introduces more variability in the resulting tractograms. Consequently, when certain streamlines are removed during the filtering process, it can have a more pronounced effect on connectivity measures like node strength and betweenness centrality. This sensitivity to the removal of specific streamlines is a characteristic of the probabilistic method and should be taken into consideration when interpreting the results.

4.5. Sources of error

An area of concern lies in the process of translating the parcellation atlas into individual diffusion spaces, which can potentially introduce errors. The diffusion space possesses a voxel spacing of 2mm, whereas the parcellation atlas in MNI152 space utilizes 1mm spacing. In order to obtain integer labels, nearest-neighbor interpolation becomes necessary. However, this interpolation technique may give rise to interpolation artifacts and lead to the loss of small regions. For instance, region 47, being notably diminutive, tended to be absent following the transformation of the parcellation atlas into the diffusion space.

5 Conclusion

In this work, the effects performance and effectiveness of deterministic tractography algorithms in generating tractograms and connectomes were studied. Additionally, this work investigated whether tractogram filtering improves the accuracy of the tractograms generated. Moreover, the outcomes and metrics on the structural connectomes were studied. This was achieved with a pipeline consisting of libraries and tools such as MRTrix and FSL, where filtering was performed using SIFT.

The difference between the connectivity metrics resulted from both deterministic and probabilistic methods give rise to the conclusion that the deterministic algorithm offers the advantage of simplicity compared to its probabilistic counterpart. It yields satisfactory results even with a relatively low number of generated streamlines, thereby reducing computation time. In order to make a decision about which type of algorithm to use, researchers should take into account the specific objectives of their study, the nature of the data at hand, and the desired balance between flexibility and robustness.

In terms of filtering performance, measured with graph and connectivity metrics, it was concluded that filtering alters the network structure of the tractograms produced, varying the importance of different nodes. This was observed in the three graph metrics of node strength, eigenvector centrality and betweenness centrality since all of them undergo changes in their scores.

Therefore, these findings contribute to our understanding of the strengths and limitations of deterministic tractography algorithms, as well as the effects of filtering on connectivity measures.

6 References

[1] T. Dhollander and A. Connelly, "A novel iterative approach to reap the benefits of multi-tissue csd from just single-shell (+ b= 0) diffusion MRI data," *Proc ISMRM,* vol. 24, p. 3010, 2016.

[2] J. B. Pereira, D. van Westen, E. Stomrud, T. O. Strandberg, G. Volpe, E. Westman and O. Hansson, "Abnormal Structural Brain Connectome in Individuals with Preclinical Alzheimer's Disease," *Cerebral Cortex,* vol. 28, no. 10, pp. 3638-49, 2018.

[3] 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.

[4] J. Andersson, M. Jenkinson and S. Smith, "Non-linear registration aka Spatial normalisation FMRIB Technial Report TRO7JA2," *FMRIB Analysis Group Technical Reports,* 2007.

[5] D. N. Greve and B. Fischl, "Accurate and robust brain image alignment using boundary-based registration," *Neuroimage,* vol. 48, no. 1, pp. 63-72, 2009.

[6] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich and S. M. Smith, "FSL," *Neuroimage,* vol. 62, no. 2, pp. 782-790, 2012.

[7] A. Fornito, A. Zalesky and E. T. Bullmore, "Chapter 3: Connectivity Matrices and Brain Graphs," in *Fundamentals of Brain Network Analysis,* 2016, pp. 98-101.

[8] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage,* vol. 52, pp. 1059-1069, 2010.

[9] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience,* vol. 10, pp. 186-198, 2009.

[10] A. Fornito, A. Zalesky and E. T. Bullmore, "Chapter 4: Node Degree and Strength," in *Fundamentals of Brain Network Analysis,* 2016, pp. 117-121.

[11] U. Brandes, "On variants of shortest-path betweenness centrality," *Elsevier,* vol. 30, no. 2, pp. 136-145, 2008.

[12] A. Fornito, A. Zalesky and E. T. Bullmore, "Chapter 5: Centrality and Hubs," in *Fundamentals of Brain Network Analysis,* 2016, pp. 150-151.

[13] M. Köpff, “Impact of tractogram filtering and graph creation for structural connectomics in subjects with mild cognitive impairment,” 2020.

[14] D. E. Haines and G. A. Mihailoff, “The Cell Biology of Neurons and Glia,” in *Fundamental Neuroscience for Basic and Clinical Applications*, Fifth ed., Elsevier, 2008, pp. 15-23.

[15] B. M. Dale, M. A. Brown and R. C. Semelka, MRI Basic Principles and Applications, Fifth ed., Wiley Blackwell, 2015, pp. 2-27.

[16] J. Pipe, “Pulse Sequences for Diffusion-Weighted MRI,” in *Diffusion MRI*, Elsevier, 2014, pp. 11-13.

[17] P. J. Basser and E. Ö. , “Introduction to Diffusion MR,” in *Diffusion MRI*, Elsevier, 2014, pp. 5-9.

[18] M. Jallais and D. Wassermann, “Single Encoding Diffusion MRI: A Probe to Brain Anisotropy,” in *Anisotropy Across Fields and Scales*, Springer, 2021.

[19] M. C. Garcia Otaduy, “Physics of Diffusion Weighted and Diffusion Tensor Imaging,” in *Diffusion Weighted and Diffusion Tensor Imaging*, Thieme, 2016, pp. 1-7.

[20] D. K. Jones, “Gaussian Modeling of the Diffusion Signal,” in *Diffusion MRI*, Elsevier, 2014, pp. 91-92.

[21] S. Mori and J. Tournier, “Moving Beyond DTI: High Angular Resolution Diffusion Imaging (HARDI),” in *Introduction to Diffusion Tensor Imaging*, Elsevier, 2014, pp. 65-76.

[22] D. J. Tournier, F. Calamante, D. G. Gadian and A. Connelly, “Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution,” *NeuroImage*, vol. 23, no. 3, pp. 1176-1185, 2004.

[23] S. Mori and P. C. van Zijl, “Fiber tracking: principles and strategies-a technical review,” *NMR in Biomedicine*, vol. 15, pp. 468-480, 2002.

[24] F. Siddiqui, T. Höllt and A. Vilanova, “Uncertainty in the DTI Visualization Pipeline,” in *Anisotropy Across Fields and Scales*, Springer, 2021, pp. 125-148.

[25] T. E. Behrens, H. J. Berg, S. Jbabdi, M. F. Rushworth and M. W. Woolrich, “Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?,” *NeuroImage*, vol. 34, pp. 144-155, 2007.

[26] J.-D. Tournier, F. Calamante and A. Connell, “MRtrix: Diffusion Tractography in Crossing Fiber Regions,” *Wiley Periodicals, Inc.,* vol. 22, pp. 53-66, 2012.

[27] M. Susumu, B. J. Crain, V. P. Chacko and P. C. Van Zijl, “Three-Dimensional Tracking of axonal projections in the brain by magnetic resonance imaging,” *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society,* vol. 45, no. 2, pp. 265-269, 1999.

[28] R. E. Smith, J.-D. Tournier, F. Calamante and A. Connelly, “Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information,” *NeuroImage,* vol. 62, no. 3, pp. 1924-1938, 2012.

[29] R. E. Smith, J.-D. Tournier, F. Calamante and A. Connelly, “SIFT: Spherical deconvolution informed filtering of tractograms,” *Neuroimage,* vol. 67, pp. 298-312, 2013.

[30] D. Jörgens, M. Descoteaux and R. Moreno, “Challenges for Tractogram Filtering,” in *Anisotropy Across Fields and Scales,* Springer, 2021, pp. 149-170.

Appendix 1: Parcellation Labels

Node ID	Label
1	Frontal Pole_L
2	Frontal Pole_R
3	Insular Cortex_L
4	Insular Cortex_R
5	Superior Frontal Gyrus_L
6	Superior Frontal Gyrus_R
7	Middle Frontal Gyrus_L
8	Middle Frontal Gyrus_R
9	Inferior Frontal Gyrus, pars triangularis_L
10	Inferior Frontal Gyrus, pars triangularis_R
11	Inferior Frontal Gyrus, pars opercularis_L
12	Inferior Frontal Gyrus, pars opercularis_R
13	Precentral Gyrus_L
14	Precentral Gyrus_R
15	Temporal Pole_L
16	Temporal Pole_R
17	Superior Temporal Gyrus, anterior division_L
18	Superior Temporal Gyrus, anterior division_R
19	Superior Temporal Gyrus, posterior division_L
20	Superior Temporal Gyrus, posterior division_R
21	Middle Temporal Gyrus, anterior division_L
22	Middle Temporal Gyrus, anterior division_R
23	Middle Temporal Gyrus, posterior division_L
24	Middle Temporal Gyrus, posterior division_R
25	Middle Temporal Gyrus, temporooccipital part_L
26	Middle Temporal Gyrus, temporooccipital part_R
27	Inferior Temporal Gyrus, anterior division_L
28	Inferior Temporal Gyrus, anterior division_R
29	Inferior Temporal Gyrus, posterior division_L
30	Inferior Temporal Gyrus, posterior division_R
31	Inferior Temporal Gyrus, temporooccipital part_L
32	Inferior Temporal Gyrus, temporooccipital part_R
33	Postcentral Gyrus_L
34	Postcentral Gyrus_R
35	Superior Parietal Lobule_L
36	Superior Parietal Lobule_R
37	Supramarginal Gyrus, anterior division_L
38	Supramarginal Gyrus, anterior division_R
39	Supramarginal Gyrus, posterior division_L
40	Supramarginal Gyrus, posterior division_R
41	Angular Gyrus_L
42	Angular Gyrus_R
43	Lateral Occipital Cortex, superior division_L

Node ID	Label
44	Lateral Occipital Cortex, superior division_R
45	Lateral Occipital Cortex, inferior division_L
46	Lateral Occipital Cortex, inferior division_R
47	Intracalcarine Cortex_L
48	Intracalcarine Cortex_R
49	Frontal Medial Cortex_L
50	Frontal Medial Cortex_R
51	Juxtapositional Lobule Cortex_L(formerly Supplementary Motor Cortex)
52	Juxtapositional Lobule Cortex_R(formerly Supplementary Motor Cortex)
53	Subcallosal Cortex_L
54	Subcallosal Cortex_R
55	Paracingulate Gyrus_L
56	Paracingulate Gyrus_R
57	Cingulate Gyrus, anterior division_L
58	Cingulate Gyrus, anterior division_R
59	Cingulate Gyrus, posterior division_L
60	Cingulate Gyrus, posterior division_R
61	Precuneous Cortex_L
62	Precuneous Cortex_R
63	Cuneal Cortex_L
64	Cuneal Cortex_R
65	Frontal Orbital Cortex_L
66	Frontal Orbital Cortex_R
67	Parahippocampal Gyrus, anterior division_L
68	Parahippocampal Gyrus, anterior division_R
69	Parahippocampal Gyrus, posterior division_L
70	Parahippocampal Gyrus, posterior division_R
71	Lingual Gyrus_L
72	Lingual Gyrus_R
73	Temporal Fusiform Cortex, anterior division_L
74	Temporal Fusiform Cortex, anterior division_R
75	Temporal Fusiform Cortex, posterior division_L
76	Temporal Fusiform Cortex, posterior division_R
77	Temporal Occipital Fusiform Cortex_L
78	Temporal Occipital Fusiform Cortex_R
79	Occipital Fusiform Gyrus_L
80	Occipital Fusiform Gyrus_R
81	Frontal Operculum Cortex_L
82	Frontal Operculum Cortex_R
83	Central Opercular Cortex_L
84	Central Opercular Cortex_R
85	Parietal Operculum Cortex_L
86	Parietal Operculum Cortex_R
87	Planum Polare_L
88	Planum Polare_R
89	Heschl's Gyrus_L (includes H1 and H2)

Node ID	Label
90	Heschl's Gyrus_R (includes H1 and H2)
91	Planum Temporale_L
92	Planum Temporale_R
93	Supracalcarine Cortex_L
94	Supracalcarine Cortex_R
95	Occipital Pole_L
96	Occipital Pole_R
97	Thalamus_L
98	Thalamus_R
99	Caudate_L
100	Caudate_R
101	Putamen_L
102	Putamen_R
103	Pallidum_L
104	Pallidum_R
105	Hippocampus_L
106	Hippocampus_R
107	Amygdala_L
108	Amygdala_R
109	Accumbens_L
110	Accumbens_R

Appendix 2: State of the Art

B.1 Anatomy

The human brain is the most complex organ of the human body, making up the central nervous system (CNS) together with the spinal cord. Its functions vary from controlling the body to processing the information it continuously receives from every sense organ. The largest part of it is the cerebrum, formed by two cerebral hemispheres, each containing an inner core made of white matter and an outer surface made of grey matter. Grey matter acts in regions involved in muscle control as well as sensory perception and is mostly made of unmyelinated neurons, whereas white matter facilities the transport of information throughout the nervous system and is composed of axons covered in myelin.

Neurones are the primary cells of the brain, they transmit nerve impulses and consist of the cell body (referred to as the soma), the dendrites and one single axon. The soma contains the nucleus and cytoplasm, the dendrites are branches where the neurons receive input signals from other cells, and these allow the electrical signal transmitted by neurons (the action potential) to be conducted down the axon and away from the soma. This axon is a thin structure coated in a myelin sheath, with a thickness proportional to its propagation velocity, that transmits the signal to other neuron's dendrites via synapses. In these synapses, information is transported in the form of chemical substances called neurotransmitters [14].

In essence, the human brain is constituted of around 86 billion neurons interconnected by axons. Moreover, we can study the connectivity pattern of the brain by following the distribution of these neuronal fibers.

B.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is a non-invasive medical imaging technique used to acquire information regarding the function and structure of several sections of the human body. Its fundaments are based on nuclear magnetic resonance (NMR), a physical phenomenon where an applied magnetic field interacts with primarily hydrogen nuclei in a strong magnetic field to generate an electromagnetic signal.

Nuclei present an intrinsic spin angular momentum, always rotating about their own axis at a constant rate, perpendicular to the direction of the rotation. In the case of an absence of a magnetic field, these spin vectors will follow a random distribution, cancelling all magnetic fields and resulting in no net magnetization. Furthermore, when subjecting the atoms to a magnetic field B_0 generated by the MRI scanner, these will start to precess around it, reorienting their axis to its direction. The angular frequency of this precession, commonly referred to as the Larmor frequency (equation B.1), will depend on the gyromagnetic ratio γ, a constant specific to the atom in question.

$$\omega_0 = \gamma B_0 \tag{B.1}$$

As a result of all nuclei being aligned with B_0 there will be a resulting net magnetization M_0, invariant with time in the direction parallel to B_0 and null in its perpendicular direction since all spin vectors will still have a random distribution. This magnetization can be manipulated by exciting the nuclei with a radiofrequency (RF) pulse with the same frequency as the previous Larmor frequency. This will cause the nuclei to absorb energy at this frequency, change their spin direction and conclude in a rotation of the magnetization vector into the perpendicular direction of B_0. This process defines the excitation phase [15].

The moment the pulse is removed, the relaxation phase is set in motion and the nuclei start to return to their original orientation. As this happens, the nuclei will emit energy at ω_0 and M_0 will begin to precess about B_0. Two relaxation times can be considered: T1 and T2. T1 is the longitudinal relaxation time constant that corresponds to the time necessary for the nuclei to go from a high energy to a low energy state. Its value will vary depending on the type of tissue being examined and will depend on the strength of the magnetic field. The magnetization vector will recover 63% of its original value, following an exponential growth process described by T1, where a lower value of T1 will mean a lower strength of the magnetic field and a faster relaxation. On the contrary, T2 is the transverse relaxation time constant that corresponds to the transverse component of the magnetization vector losing 63% of its initial value as a result of the dephasing of the nuclei. Its value will be shorter than T1 and won't vary with the strength of the magnetic field [15].

During the scan, gradients can be applied to make the magnetic field spatially dependent, acting as small perturbations that produce linear variations in any of the three orthogonal directions. They have three main functions when used in MRI: slice selection, image encoding and diffusion weighting [16].

B.3 Diffusion MRI

Diffusion describes the displacement of molecules due to the difference in concentration between two regions and is explained by Fick's first law, following the molecules from a high to a low concentration area.

Water diffusion can be either hindered or restricted, depending on its position, outside or inside the axon respectively. If the water is trapped inside myelinated axons, as it's the case in most of the white matter, its displacement will be confined by the diameter of the axon and the water molecules will be restricted, resulting in anisotropic diffusion. Contrary to grey matter, where water is free to move in any direction in somas, resulting in isotropic diffusion. Anisotropic water diffusion can infer a substantial amount of information when using the image acquisition technique of diffusion weighted imaging, a form of MRI.

Diffusion MRI measures the dispersion of water molecules within tissue over a time interval following the pulsed gradient spin echo (PGSE) sequence. This sequence is formed by a first 90º RF pulse, causing a dephasing of the spins, followed by a 180º RF pulse that inverts those spins and causes them to begin rephasing. Additionally, a pair of pulsed gradients are applied before and after the second pulse to make the spins completely rephase. If the spins do not change their position during the time interval between the pulsed gradients, they all produce

the echo signal, however, water molecules are in constant motion and the phase they gain during the first pulse won't completely cancel the phase they lose in the second pulse, resulting in phase dispersion and an attenuation of the signal [17].

To quantify this attenuation in white matter, diffusion tensor imaging (DTI) is used, a special kind of DWI that describes the fiber orientations using a Gaussian distribution. This way we use the signal intensity of the sequence with (S) and without (S_0) the application of the gradients, being D the diffusion coefficient [18].

$$\frac{S}{S_0} = e^{-bD} \tag{B.2}$$

Diffusion weight can be increased in images by increasing the following b factor,

$$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) \tag{B.3}$$

with G being the gradient amplitude, δ the duration of the gradient and Δ the interval between the pulse gradients. Therefore, larger b values can be acquired by increasing the values of the parameters listed.

Apparent diffusion coefficient (ADC) is the parameter used to quantify diffusion. In grey matter we can describe the diffusion characteristics with a singular ADC. However, in white matter, as a result of anisotropic diffusion, ADC will vary depending on its direction, making a single value not enough to describe water diffusion. Therefore, its value is replaced by the diffusion tensor D. [19]

$$ADC = -\frac{\ln\left(\frac{S}{S_0}\right)}{b} \tag{B.4}$$

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \tag{B.5}$$

When the off-diagonal elements are zero, there will be no correlation between the displacements in orthogonal directions and the tensor will be aligned with the principal axes. In this case, the diagonal elements will be referred to as eigenvalues (λ_1, λ_2 and λ_3) and the orientation of the principal axes will be given by three eigenvectors ($\varepsilon_1, \varepsilon_2$ and ε_3). Furthermore, the principal eigenvector will be determined by the largest eigenvalue and will be parallel to the orientation of the tensor. This vector will represent the direction of the diffusion of the water molecules [20].

Implementing this diffusion tensor we can determine various parameters, one of these being fractional anisotropy (FA), which will quantify the fraction of the tensor attributed to anisotropic diffusion**,** with a high value (FA = 1) representing total anisotropic diffusion and a lower value (FA = 0) total isotropic diffusion [18].

$$FA = \sqrt{\frac{1}{2}}\sqrt{\frac{(\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_3-\lambda_1)^2}{{\lambda_1}^2+{\lambda_2}^2+{\lambda_3}^2}} \quad \text{(B.6)}$$

As a result of FA, DTI measures diffusion anisotropy and determines an estimate of the principal direction of axon fibers, which will later enable tractography. However, DTI is limited due to its inability to recover more than one fiber orientation per voxel, producing wrong estimations in zones with crossing fibers and reducing its accuracy in its estimation of fiber pathways. The imperfections regarding the tensor model are considered and rectified in the model described by [21], high angular resolution diffusion imaging (HARDI), since it considers various fiber orientations in a single voxel.

We can also estimate the fiber distributions in a voxel from HARDI data with the fiber orientation density (FOD) model proposed by [22]. This model assumes that the diffusion profiles of two fiber populations are identical and can be distinguished by representing their DW signal attenuation with an axially symmetric response function R(θ), being θ the elevation angle in spherical coordinates. Therefore, every fiber group in a singular voxel of white matter causes an identical but rotated response function in the direction (θ, ϕ). Hence the signal from a sample containing several distinct fiber populations can be measured by the sum of the signals multiplied by their individual volume fractions, being this equivalent to the spherical convolution of the FOD F(θ, ϕ) with the response function R(θ) [22].

$$S(\theta, \phi) = \mathrm{F}(\theta, \phi) \otimes \mathrm{R}(\theta) \quad \text{(B.7)}$$

Thus, we can directly estimate the FOD from the given signal by implementing the spherical deconvolution of the acquired signal with the response function. This can be represented by simple matrix multiplications of the spherical harmonics of the FOD and the response function [22].

$$S^n = R^n F^n \quad \text{(B.8)}$$

Being the parameters of Equation B.8 the nth order spherical harmonics of the signal, the response function and the FOD, accordingly. The biggest advantage of this model is that it does not rely on any model of diffusion and the streamlines from the FODs can be obtained with both deterministic and probabilistic algorithms.

B.4 Tractography

The main objective of tractography is to reconstruct the streamlines of the axons from previously acquired DWI images and that way recollect data about the connexion between brain regions. As explained previously, due to the limitations surrounding DTI, more complex tractography models and algorithms like HARDI and FOD have been developed.

To produce the streamlines of a tractogram we can use two different algorithmic approaches: deterministic and probabilistic tractography. Deterministic tractography follows a streamline approach, following a singular pathway starting from a determined seed point. The

direction is determined by the main eigenvector at each voxel position and the tracing is stopped following a certain set of criteria, such as FA value and angle change. Typically, the fractional anisotropy in gray matter is in the range of 0.1-0.2. To address these issues, a straightforward approach is to set the tracking termination threshold at 0.2, allowing for a more reliable termination point and reducing the impact of noise on the tracking process [23]. The angle change between pixels is another critical factor in tractography accuracy, due to large errors occurring if the angle transition is significant. Hence, it is advisable to impose a threshold that prevents a sharp turn during line propagation. The impact of this angle-transition threshold depends on the tracts of interest and the image resolution [23].

Furthermore, another source of uncertainty in this algorithm is caused by region definition since defining regions in this process can introduce variability in the outcome. The definition is often done manually, which can introduce an implicit user bias. Even a minor variation in the definition can lead to substantially different pathways. This leads to the algorithm being sensitive to what initial seed point is chosen and what estimated principal directions are produced [24].

Alternatively, probabilistic tractography is a variant of streamline tractography, with the key difference being that at each step, instead of following the most probable principal diffusion direction, a direction is sampled from the posterior distribution on principal diffusion directions and followed instead. This sampling process generates a set of streamlines that collectively represent the distribution of possible connectivity pathways. The multiple samples generated by this method account for the uncertainty in local fiber orientation, as two streamlines arriving at the same voxel will follow different directions based on their individual samples from the posterior distribution on local orientation [25].

The MRtrix3 software provides us with three deterministic algorithms. First, the SD_STREAM algorithm is based on Spherical Deconvolution (SD) and uses the FOD, and a peak-finding procedure based on the Newton-Raphson gradient ascent algorithm to estimate the local fiber orientation, and then tracks the fiber pathways by stepping along the estimated orientations [26]. Secondly, the Tensor_Det algorithm takes local diffusion data, fitting a diffusion tensor, and determining the principal eigenvector. By following the principal eigenvector direction, it enables the estimation of white matter fiber pathways within the brain based on the diffusion properties of water molecules.

Finally, the deterministic algorithm Fiber Assigned by Continuous Tracking (FACT) proposed by [27], tracks a continuous vector field, rather than a discrete one, removing the issue of limited angle ranges. This algorithm starts from the centre of the voxel and continues depending on the vector direction, determining the endpoint of the fiber projection as sudden changes in fiber orientation are used as a criterion. These abrupt changes are quantified using a parameter, denoted as R, which is calculated as the sum of inner products of neighbouring data points, being large when there is a strong alignment between adjacent fibers and small when there is no continuity in fiber direction.

$$R = \sum_i^s \sum_j^s \frac{abs(v_{\lambda_{1i}} * v_{\lambda_{1j}})}{s(s-1)} \quad \text{(B.9)}$$

Being v_{λ_1} the unit vector indicating the longest principal diffusion axis (λ_1) and s the number of data points applied. Thus, the neuronal connections are mapped by inputting an arbitrary point in 3D space. The algorithm then traces the extent of axonal projections in both the forward (orthograde) and backward (retrograde) directions to determine their connections with functional regions [27].

However, MRI provides information on the average axonal orientation within a voxel, which is resolution-dependent and cannot differentiate between very small projections that are near each other. Furthermore, it is also unable to discern afferent from efferent fibers and issues may arise when two pathways are in close proximity to each other, as the program may mistakenly switch pathways [27].

In situations where there are complex fiber structures such as crossing fibers, or partial volume contamination between adjacent fiber populations, the diffusion tensor, which can only represent a single fiber orientation per voxel, is insufficient to accurately represent the underlying structure. This limitation can result in errors in the estimation of fiber orientations, ultimately leading to errors in the reconstruction of the underlying fiber pathways. To improve the results of the reconstruction and regardless of the algorithm used to produce the streamlines of the tractograms, [28] proposed the application of anatomically constrained tractography (ACT). This addition to streamlines tractography incorporates biologically realistic priors obtained from the segmentation of an anatomical contrast image. Rather than relying solely on a binary mask to restrict streamline propagation, we consider the properties of the brain tissue and fluid as well as the axons of the white matter being reconstructed to tailor the streamline criteria. Through the use of random streamline seeding, this algorithm was able to generate 5 million streamlines that satisfied all relevant streamline acceptance criteria until the target number of streamlines was achieved [28].

B.5 SIFT

The tractogram generated by both a probabilistic or deterministic approach may have attributes that are caused by the reconstruction approach and are not accurate to the actual biology of the brain. There are various factors that can cause these inaccuracies such as streamline seeding, selection of fibre orientation and white matter volume. Various methods have been developed to address these reconstruction issues, one of them being SIFT, an acronym for spherical-deconvolution informed filtering of tractograms, a novel algorithm submitted by [29] that reconstructs the fiber orientation distribution function (fODF) in each voxel without the need to utilize the original dMRI image, showing a reduction in known reconstruction biases, as well as an improvement in plausibility in terms of reconstructing the underlying biology.

To begin with, it obtains the fODF through constrained spherical deconvolution (CSD), continuing by assessing the contribution of every streamline to the fODF and using them to decide whether a streamline is dismissed. The contributions are sorted to dismiss the less relevant streamlines and the last two steps are iterated until a specific number of streamlines is achieved or a certain residual level is reached. The accuracy of the reconstructions is evaluated with the coefficient μ (Equation B.10), which correlates the streamline density assigned to each FOD lobe by attributing a value with the integral of that lobe [29].

$$\mu = \frac{\sum_V (PM_V \sum_{l=1}^{L_V} FOD_{V,l})}{\sum_V (PM_V \sum_{l=1}^{L_V} TD_{V,l})} \quad \text{(B.10)}$$

Being $FOD_{V,l}$ the FOD integral of lobe number l in voxel V, $TD_{V,l}$ the track density associated to that specific lobe, L_V the entire number of FOD lobes in voxel V, and PM_V the value of the processing mask in voxel V. To consider any possible biases during the process of reconstruction, the cost function f (Equation B.11) is defined, measuring the accuracy of the reconstruction containing the underlying diffusion data.

$$f = \sum_V (PM_V \sum_{l=1}^{L_V} (\mu * TD_{V,l} - FOD_{V,l})^2) \quad \text{(B.11)}$$

Thus, removing inefficient streamlines improves the model by reducing the cost function but it also affects it in two ways, reducing $TD_{V,l}$ and increasing μ. Consequently, this will alter the way all the FOD lobes within the processing mask contribute to the cost function. This is one the reasons why the SIFT2 method was developed, assigning weights to streamlines with a high contribution to the reconstructed FODs. Regardless, it has been proven [30] that although SIFT2 can provide valuable information on how streamlines contribute to acquired data, its direct application for filtering tractograms is not as straightforward as SIFT. It was shown SIFT2 assigned lower weights in the central areas of the fiber bundles, where tractography typically generates a greater number of streamlines, leading to the premature elimination of crucial tracts. This caused some noisy streamlines to be classified as acceptable simply because they reached distant regions. This problem arises because the weights in SIFT2 are intended for fitting the acquired data and not for filtering.

In general, SIFT is used because streamline tractography often produces wrong pathways that do not represent actual white matter fibers. These incorrect pathways can arise due to several factors, such as noise in the data, partial volume effects, and errors in the tracking algorithm. They produce a negative impact on later analysis, such as connectivity-based analyses, as they can introduce false connections or distort the estimated properties of the white matter pathways. Therefore, SIFT is used to remove them by filtering out streamlines that do not have sufficient evidence in the data. This helps to improve their accuracy and specificity, and thus leads to more reliable and interpretable results. Overall, the statistical framework used in SIFT allows for a more robust and accurate reconstruction of the underlying fiber pathways in the brain [29].

B.6 Connectivity matrices

Brain networks can be described as complex systems composed of nodes and edges, where the nodes represent regions of the brain and the edges represent the structural or functional connections between them. The edges may be weighted to represent the strength of the connections, and the networks can be represented as graphs or matrices. The networks can be analyzed using graph theory measures to gain insights into their organization and function, and to compare differences between healthy and diseased states [8].

In general, edge weights in brain networks can be organized in an adjacency matrix, which is a square matrix that represents the connections between all nodes (brain regions) in

the network. The diagonal of the matrix represents the self-connections of each node, while the off-diagonal elements represent the connections between nodes. The value of each element in the matrix corresponds to the strength of the connection between the two nodes it connects, according to a chosen connectivity measure, such as the number of streamlines connecting two regions in a tractography-based network. In some cases, the adjacency matrix may be symmetrical, indicating that the connections between two nodes are undirected, while in other cases, it may be asymmetrical, indicating that the connections are directed, and therefore, have a specific source and target region [9].

Various network measures can be computed using the adjacency matrix, such as degree centrality, clustering coefficient, and betweenness centrality. These measures can characterize different aspects of brain connectivity and identify changes in connectivity patterns associated with various neurological and psychiatric disorders. Specifically, betweenness centrality can be used to identify hubs or highly connected regions in the brain, while measures such as clustering coefficient and modularity can be used to identify communities or groups of regions that are densely connected to each other. Other measures such as efficiency and small-worldness can be used to characterize the overall efficiency and organization of the brain network [8]. Moreover, network measures can be used to identify alterations in brain connectivity in disorders such as Alzheimer's disease, schizophrenia, and multiple sclerosis, arguing that graph theoretical analysis can provide valuable insights into the underlying pathophysiology of these disorders and can be used to identify biomarkers for early diagnosis and prognosis [9].

Overall, structural and functional networks are typically obtained from histological techniques like tract tracing or neuroimaging techniques like diffusion MRI and fMRI, respectively. In order to facilitate computational analysis, these networks are often represented as connectivity matrices with nodes represented by rows and columns, and links represented by matrix entries, with networks being transformed into a sparse binary undirected form through thresholding, binarizing, and symmetrizing [8].