Universidad de Concepción
Dirección de Postgrado
Facultad de Ingeniería - Programa de Doctorado en Ciencias de la Computación

ALGORITHMS AND METHODS FOR THE STUDY OF BRAIN CONNECTIVITY BASED ON DMRI

Tesis para optar al grado de
DOCTOR EN CIENCIAS DE LA COMPUTACIÓN

POR
Narciso López López
CONCEPCIÓN, CHILE
Agosto, 2020

Directoras: Pamela Guevara Alvez
Susana Ladra González
Departamento de Ingeniería Informática y Ciencias de la Computación
Facultad de Ingeniería
Universidad de Concepción

A Andrea

ACKNOWLEDGEMENTS

I would like to thank Pamela for letting me participate in the area of computational neuroscience and for being there when needed. Thank you for making this possible. Thanks to Susana for being available from afar and for achieving the international agreement. Thanks to Nieves for giving me the opportunity to start my Ph.D.

Thanks to Diego for being such a good teacher, to Roberto for his wise advice, to Gonzalo, Lilian, Andrea, Guillermo, Cecilia and the rest of the DIICC for the moments shared and lessons learned. Also, thank the thesis reviewers Julio and Mauricio for the corrections. To my colleagues, Mabel, Luis, José, Isyed, Pedro, Pepe, and Erick for all the moments lived during these years. Also to my colleagues from the LBD and the Medical Image Analysis Group.

To my friends Ramón and Santi for making my days more bearable and making me laugh. Koff! We did it! I do not forget you. Thanks to little Champa for being with me every day. To my family and Andrea's family for supporting me in the distance, especially my mother who supports me daily and has always had faith in me. Dad, I dedicate it to you.

Especially thanks to Andrea, who crossed half the world and was by my side in all the moments that I have been bad in the Ph.D.[1] Besides taking care of me and always supporting me, I love you very much.

Thank you all.

[1]This work received funding from the ANID PFCHA/DOCTORADO NACIONAL/2016-21160342, ANID FONDECYT 1190701, ANID PIA/ Anillo de Investigación en Ciencia y Tecnología ACT172121, ANID-Basal Project FB0008 (AC3E), ANID-Basal Project FB0001 (CeBiB), and from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 690941. This work was also partially funded by the Human Brain Project, funded from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 945539 (HBP SGA3), No. 785907 (HBP SGA2) and No. 604102 (HBP SGA1). It was also partially funded by Ministry of Economy and Competitiveness (MINECO, AEI co-founded with ERDF) [grant number TIN2016-78011-C4-1-R] and the Consellería de Educación, Universidade e Formación Profesional from Xunta de Galicia and European Union (European Regional Development Fund - FEDER Galicia 2014-2020 Program) [grant numbers ED431C 2017/58; ED431G 2019/01].

Abstract

A main goal for understanding how the brain works is the description of the network of brain connections or the human brain connectome. The basic elements of the connectome are regions of gray matter (GM) formed by neuronal brain bodies (somas) and the connections (axons) between them formed by the white matter (WM) neuronal fibers. Currently, diffusion-weighted Magnetic Resonance Imaging (dMRI) techniques with High Angular Resolution Diffusion Imaging (HARDI) have improved the quality of tractography datasets concerning Diffusion Tensor Imaging (DTI). Tractography data are complex and contain noise and artifacts, so they require computational methods capable of processing them efficiently and extract useful information.

Therefore, this thesis proposes the development and implementation of several methods to process the brain tractography data and through them provide analysis tools for study the brain structural connectivity. Among them, we contributed to the improvement of a clustering of white matter fibers, to the labeling of the superficial white matter bundles, to the development of a method for the parcellation of the cortical surface from short and long segmented bundles for a group of subjects, and to the development of complementary methods to carry out the individual parcellation.

We collaborated in the development of an efficient clustering of white matter fibers that was evaluated in terms of quality and execution time against other state-of-the-art methods, giving as a result 8.6 times more speed than the most efficient method. Moreover, we created a method which performs intra-subject labeling of superficial white matter fibers in 3.6 min, and two inter-subject labeling methods. One is based on matching and obtains good correspondence but little reproducibility with an execution time of 96 s. The other one focused on clustering that obtains good correspondence and reproducibility among subjects achieving a short execution time of 9 s for a subject. On the other hand, a method for cortical surface parcellation which creates parcel atlases was developed. Then, two generated parcellations were compared with state-of-the-art methods, finding a degree of similarity with dMRI, functional, anatomical, and

multi-modal atlases. The best comparison was between our parcellation composed of 185 sub-parcels and another dMRI-based parcellation, obtaining 130 parcels in common for a Dice coefficient $\geq$0.5. The parcellation composed of 160 parcels achieves a reproducibility across subjects of $\approx$0.74, based on the average Dice's coefficient between subject's connectivity matrices, rather than $\approx$0.73 obtained for a macroanatomical parcellation of 150 parcels. In addition, two complementary methods were developed to perform individual cortical parcellations, one based on clustering and the other based on the geodesic distance, the latter obtains a parcellation of 350 parcels in 18 s for the atlas-based mode and 82 s for the whole cortex mode. Moreover, it obtains better reproducibility against two state-of-the-art methods with a difference of 0.024 and 0.043 according to the Dice coefficient.

This thesis contributes to the development of efficient computational methods for the study of brain connectivity that can be applied to high-quality and large databases, capable of dealing with the noise present in the tractographies. In addition, thanks to this research, these methods can be used by neuroscientists, neuroanatomists, and neurologists to study and develop new studies of brain connectivity and to obtain more answers about the structure of the brain and its connectivity.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Motivation

The human brain is highly complex and its complete functioning has not been deciphered at present. Its study dates back to ancient Egypt, and although there has always been a desire to understand how it works and its anatomical structure, its anatomy has not been known until recently, especially concerning the connectivity of white matter (WM) and its structure [103]. Thanks to the study of the different brain areas based on different modalities it is possible to better understand how it works, as well as the multiple associated pathologies. Brain structure and function are strongly linked.

To understand how the brain works, it is necessary to know the network of brain connections or the human brain connectome. The basic elements of the connectome are a composition of bodies of brain neurons (nodes) and the connections (edges) that exist between them, formed by the neuronal fibers of white matter (WM) [118, 59, 13].

Currently, different techniques are known to be able to non-invasivelly observe the brain. One of them is Magnetic Resonance Imaging (MRI), through which it has been possible to obtain high-resolution images to study different areas of the brain. This technique has different modalities, one of which is known as diffusion Magnetic Resonance Imaging (dMRI) [90]. This allows us to characterize the diffusion process of water molecules [6, 75]. Thanks to this, the trajectories of the fibers that form the main bundles connecting the different areas of the brain can be obtained [101] and, by means of tractography algorithms, they can be reconstructed in 3D [92]. Thus, the structure of the brain can be studied in vivo and in a non-invasive way [91, 129, 53].

However, results based on dMRI tractography have to be analyzed carefully because of the limitations of this modality. Analysis of the results of tractography algorithms contain noise due to the artifacts in the tractography which may lead to differences in connectivity profiles in studies with populations of subjects [15]. Despite the fact that tractography algorithms generate valid bundles for the study of brain connectivity, the large number of streamlines or fibers that pass through the voxels have the spatial limitation of the size of the voxels, thus generating a number of false positives and negatives. One of the challenges to be achieved in the coming years will be to reduce these false positives and improve the resolution of the white matter tracts [82]. Moreover, the high variability of white matter fibers between subjects highly difficults the study of the brain connectivity as tractography results generate complex datasets of at least 1 million fibers, especially when superficial white matter fibers are taken into account.

The study of the human connectome is a key and growing area of research [121]. It is also the intersection of the following fields: Biology, Electronic Engineering, Computer Science, Physics, and Neuroscience. In particular, Computer Science is a very important discipline because it provides efficient and robust methods that can work with large datasets in a reasonable time. In addition, these methods are able to deal with the noise produced in the tractographies, as well as handle the high variability that exists among subjects. Handling tractography data is not a trivial task since it is more complex than it seems beforehand because it requires the manipulation of many formats, the data contains noise, and it is necessary to apply transformations on the data. From that point of view, further validation of the results through diffusion tractography is needed and should be contrasted with anatomy, functional Magnetic Resonance Imaging, and post-mortem studies. This work focuses on methods to analyze tractography data in order to obtain the best information about connections and to contribute to the understanding of the human connectome.

In this thesis, we propose efficient algorithms and methods for the study of brain connectivity capable of working with large sets of fibers generated by tractography

algorithms. Therefore, we contributed to the development of a white matter fiber clustering method to deal with tractography data. Currently, there are many fiber clustering methods [42, 43, 55, 106]. However, not all of them perform such a thorough validation as the one we propose in this work. Fiber clustering does not validate whether the fibers are true or not since it seeks to represent the white matter bundles that are present in the subjects in order to later perform brain connectivity analyses and studies. Moreover, we developed a method that performs the labeling of superficial white matter bundles. There exists extensive knowledge about long association and projection fibers, used to identify functional regions through their connectivity, but superficial white matter bundles are less studied in the state of the art, due to the fact that fiber bundles have a great variability among subjects [49, 106] and have a small size, which makes labeling and identifying them a complex tasks. Therefore, by means of the intra- and inter-subject methods developed, we have been able to identify the inter-subject bundles with a high degree of reproducibility and give them an automatic name since they are not described in detail in the anatomy. Finally, in this thesis, three methods of parcellation are proposed, the main one being based on a fiber-bundle atlas of white matter, and two secondary methods based on clustering and brain topology. A cortical parcellation is a method to subdivide the cerebral cortex into parcels [141]. There exist parcellations based on different modalities, such as anatomical, diffusion-based information, functional magnetic resonance imaging, and multimodal. As there are currently many different types of parcellations, there is no consensus on which parcellation is better than another, or which is the best method to carry out a parcellation. For this reason, we have developed a method of parcellation based on an atlas of white matter fibers which can help in the study and deciphering of the human connectome. In addition, as a proof of concept, we have carried out an individual parcellation based on fiber clustering to deal with the limitation found when using a fiber bundle atlas. By using the whole tractography, the number of parcels obtained is larger since all the brain fibers are used and not just a set of fibers that form the atlas. Finally, we have developed a method of individual parcellation based on the topology of the mesh. This method is able to carry out the cortical parcellation in an efficient time and with a good homogeneity in the obtained parcels, being highly competitive with the methods of the

state of the art based on macroanatomy.

To conclude, this research is likely to be of high interest to neuroscientists, neuroanatomists, and neurologists, as high-quality pre-processing methods and software can be used to study brain connectivity. Thanks to this work it will be possible to develop modern clinical studies by analyzing new databases and thus be able to obtain answers regarding brain functions and structure.

1.2. Contributions

This section presents the main contributions of the thesis. They all deal with the developments of methods for the study of brain connectivity. These contributions are listed below:

- Fiber clustering algorithms in neuroanatomy research generate data that allow the study of the structure of white matter. The first contribution of this thesis is the collaboration to the development of an automated fiber clustering method called *FFClust*, which identifies white matter bundles from large tractography datasets. The first version of this work was developed in another thesis [125]. The main goal is to develop an efficient clustering to group fibers into compact and regular clusters, representing the whole-brain WM structure. Resulting clusters describe the whole set of main white matter fascicles present on an individual brain. In individuals, the clusters can be used to study the local connectivity in pathological brains, while at population level, the processing and analysis of reproducible bundles, and other post-processing algorithms can be carried out to study the brain connectivity and create new white matter bundle atlases. A special interest is its use for the study of short association bundles and their segmentation, as well as of subdivisions of long anatomical bundles. The proposed method is about 8.6 times faster than the state-of-the-art method, which enables a fast processing and visualization of main white matter fiber clusters.

  My contributions to this work are the collaboration in the following tasks: writing the first draft, reviewing all versions and subsequent drafts, study and review of the literature in fiber clustering, creating schemes and most images of the paper, changing concurrency by parallelism in steps 1 and 4, parallelizing step 2 mapping which had no parallelism, implementation of the entire code, giving an improvement of at least 2.5 times faster than its previous version, input/output optimization and format change, testing and error search and implementation of the two codes of the centroid calculation, re-execution of tests with 50 subjects

(before they were made with only one subject), use of the mean and the standard deviation in all the graphics, addition to the comparison with the state-of-the-art methods of *QuickBundlesX* (*QBX*) [43] and *Guevara* [55], implementation of maximum distance in *QuickBundles* (*QB*) [42] and *QBX* (which requires execution times of more than one day) for distance tests with all methods, runtime tests of *QB* and *QBX* with maximum distance (hours of tests), implementation of inter-cluster distances (not existing), addition of the Davies-Bouldin (DB) index [9] and comparison with all methods, qualitative analysis using segmentation to demonstrate that the clusters obtained have biological significance, addition of *Guevara*'s method to all qualitative analysis, the comparison of the most similar clusters and the execution time between the methods.This paper was accepted in the journal NeuroImage [127].

The following contributions are a direct result of this thesis work.

- Fiber clustering methods have been used to automatically group similar fibers into clusters. However, due to inter-subject variability and artifacts, the resulting clusters are difficult to process for finding common connections across subjects, especially for superficial white matter. The second contribution of this thesis is an automatic labeling method of short association bundles on a group of subjects. The method is based on an intra-subject fiber clustering that generates compact fiber clusters. After that, taking the *Desikan-Killiany* atlas [31] as a reference, the clusters are labeled based on the cortical connectivity of the fibers, and named according to their relative position along one axis. The labels provide useful information for the visualization and analysis of individual connections, which is very difficult without any additional information. Finally, we compared and applied two different strategies for the labeling of inter-subject bundles: one of them based on a famous state-of-the-art clustering algorithm, and the other based on a matching algorithm. The performance of both implementations is compared in terms of reproducibility and inter-subject bundle distance. The obtained clusters could be used for performing manual or automatic connectivity analysis in individuals or across subjects. The resulting work is published in the journal Biomedical

Engineering Online [126].

- The third contribution of this thesis, and also the most important one, is a method to create fine-grained parcellations of the cortical surface from an anatomical atlas. Connectivity information is obtained from an atlas of white matter bundles, instead of the whole tractography. This atlas is composed of a deep white matter fiber atlas and two superficial white matter atlases. Thus, a direct correspondence is obtained between the bundles and the cortical regions among subjects, being able to get a good representation of the human brain connectome, since the bundles obtained are based on bundle atlases that contain the most reproducible short and long connections found on a population of subjects. Furthermore, the method produces another output containing the probabilistic representation of the preliminary sub-parcels. In this way, the information could be used with segmented bundles of each individual and thus create individual parcellations adapted to each subject that should lead to increased consistency in structural connectome across subjects. This work has been published in the journal Frontiers in Neuroinformatics [79].

- The fourth contribution is a complementary method to the parcellation method and performs the complete cortical parcellation of an individual taking into account the connectivity information of the white matter fibers. Our goal is to perform a good quality individual cortical parcellation to be used for a group-wise parcellation in the future. The output is the complete labeling of cortical mesh vertices, representing the different cortex sub-parcels, with strong connections to other sub-parcels. We used brain network metrics to evaluate the method on a set of subjects. These metrics comply with segregation and functional integration, as well as with the definition of small-world. The result of this work was published at the IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) [80].

- Finally, the last contribution of this thesis is a parallel method for the complete parcellation of the cortical surface, based on the geodesic distance. The method has two modes of use, the first subdivides each anatomical parcel given by the

Desikan-Killiany atlas. The second mode of use performs the complete cortical division of a subject. The proposed method will be available to the community to perform the evaluation of data-driven cortical parcellations. As an example, we compared *GeoSP* parcellation with *Desikan-Killiany* and *Destrieux* atlases in 50 subjects, obtaining more homogeneous parcels for *GeoSP* and minor differences in structural connectivity reproducibility across subjects. The resulting paper was published at the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [81].

1.3. Structure of the thesis

The thesis is organized as follows:

- In Chapter 2, we expose the research carried out in the thesis, that is, the hypotheses, general and specific goals, as well as the methodology, available resources, and database used in the experiments conducted.

- In Chapter 3, we introduce the method of clustering of white matter fibers. Moreover, we show the results of the comparison with other methods of the state of the art, as well as a qualitative analysis and the execution times of the steps of the method.

  On the other hand, we present the automatic labeling of superficial white matter bundles. In addition, we use two methods to perform inter-subject clustering and compare the results obtained by both methods.

- Chapter 4 contains methods for the parcellation of the cerebral cortex. The first and main method uses an atlas of white matter bundles capable of generating atlases of parcels with different granularities based on the *Desikan-Killiany* atlas. Afterward, we carried out a reproducibility analysis among the atlases generated by our method and a comparison with the most relevant parcellation atlases of the state of theart based on macroanatomy, multimodal, diffusion, and functional Magnetic Resonance Imaging.

  Moreover, in this chapter, we present two complementary methods for cortical surface parcellation, these methods are:

  1. An individual cortical surface parcellation made with fiber clustering, which was evaluated with different network metrics and a qualitative analysis for each subject.
  2. A cortical surface parcellation for an individual based on the geodesic distance. The results show more reproducibility than two atlases based on the macroanatomy of the state of the art.

- The final conclusions and future work are presented in Chapter 5.
- Appendix A lists the publications obtained in the thesis and other research results.
- Appendix B contains the supplementary material related to the method for the parcellation of the cerebral cortex based on the atlas of white matter fibers.
- Finally, Appendix C contains the summary of the thesis in Spanish, specifically, the introduction, the methodology, the most relevant results, and conclusions.

Chapter 2

Research conducted

In this chapter, we present the hypotheses formulated for the development of the different algorithms and methods covered by the thesis. Next, we will focus on the main and specific goals, as on the methodology. Finally, the characteristics of the database used for all the experiments carried out in the thesis are described.

2.1. Research hypotheses

To carry out the thesis, research hypotheses were formulated for the development of methods for the study of brain connectivity. These are:

1. It is possible to improve the computation time in the task of clustering brain white matter fibers with respect to other existing algorithms by designing and implementing parallel clustering algorithms while maintaining the quality in the clusters.

2. It is possible to produce an automatic method for the labeling of superficial white matter fibers thanks to the use of intra-subject clustering and the connectivity of the clusters with the cerebral cortex, providing information for the visualization and analysis of the individual connections.

3. It is possible to develop a hybrid cortical parcellation method, from segmented bundles using two SWM fiber bundle atlases and a DWM bundle atlas, with a good correspondence between subjects.

4. It is possible to develop other individual parcellations based on clustering or on the topology of the brain.

2.2. Main goal

This thesis focuses on the development of algorithms and methods to analyze and study brain anatomical connectivity using dMRI data, and thus cooperate to the decoding of the human brain connectome.

2.3. Specific goals

The following specific objectives are proposed to address the thesis work:

- **G1**: To improve an efficient method for the clustering of fibers.
- **G2**: To evaluate the fiber clustering by performing qualitative and quantitative analysis and comparisons with other existing algorithms.
- **G3**: To design and implement a method for automatic labeling of superficial white matter bundles.
- **G4**: To perform the evaluation of the labeling by intra- and inter-subject clustering.
- **G5**: To create a method for the generation of atlases of parcels for the population of subjects studied.
- **G6**: To evaluate the parcellation method with reproducibility metrics and to make the comparison with other atlases of parcels of the state of the art.
- **G7**: To generate a method of individual cortical parcellation based on white matter connectivity information from a tractography dataset and to perform the evaluation of the individual parcellation method with brain network connectivity metrics.
- **G8**: To create a method to perform individual cortex parcellation taking into account the topology and to evaluate the reproducibility of the generated method with other atlases based on macroanatomy.

2.4. Methodology

To achieve the specific goals of this thesis, the following tasks must be completed:

- **T1**: Review and study of the state of the art of white matter fiber clustering methods, automatic labeling of superficial white matter fibers, and cortical surface parcellation based on different modalities (**G1, G3, G5, G7, G8**).

- **T2**: Collaboration in the implementation of a fast and efficient fiber clustering method. To achieve this, a clustering algorithm will be used to reduce the number of input elements from the white matter fibers. On the other hand, a fiber segmentation algorithm will be adapted for reassigning small clusters to larger clusters. Finally, the clusters will be mixed, thus reducing the number of clusters generated (**G1**).

- **T3**: Qualitative evaluation of the fiber clustering method with state-of-the-art algorithms. We will analyze the clusters with different configurations. In addition, images of anatomically significant clusters will be extracted, checking that the method presents better quality and cluster delimitation (**G2**).

- **T4**: Quantitative evaluation to measure the execution time of the fiber clustering method, as well as the intra- and inter-cluster distances (**G2**).

- **T5**: Design and implementation of the intra-subject superficial white matter bundle labeling method. First, clustering will be applied to the subject's tractography, and then will be filtered for short bundles. The fibers will be intersected with the mesh and, finally, the clusters will be labeled, based on an atlas (**G3**).

- **T6**: Design and implementation of the inter-subject superficial white matter labeling method. More specifically, a group of subjects will be labeled, maintaining correspondence among them. To achieve this, a matching algorithm and a clustering algorithm will be used to search for the correspondence of clusters of different subjects (**G3**).

- **T7**: Evaluation of the method of intra-subject labeling. First, the correspondence of bundles will be found across subjects. Moreover, the distances of the bundle centroids between pairs of subjects will be measured (**G4**).
- **T8**: Evaluation of the inter-subject labeling methods to measure the reproducibility of the clusters with different thresholds among the subjects (**G4**).
- **T9**: Fusion of a white matter atlas by means of two short fiber atlases and one long fiber atlas. Moreover, the anatomical information will be generated from the cortical segmentation to rule out badly segmented fibers for each of the subjects in the ARCHI database. Then, the BrainVisa software pipeline will be used to perform the conversion between the obtained mesh spaces (**G5**).
- **T10**: Design, implementation, and evaluation of the white matter fiber segmentation algorithm optimized to obtain stable fiber bundles present in most subjects (**G5**).
- **T11**: Design and implementation of a method to generate atlases of parcels in different granularities and thus obtain a representative parcellation for a group of subjects. This method will take into account the overlaps that occur among the different sub-parcels of the different subjects. In addition, it will obtain the probabilities of the underlying connections to generate the sub-parcels. Finally, a post-processing of the sub-parcels will be applied in order to obtain a better definition of them (**G5**).
- **T12**: Evaluation of the cortical parcellation method through analysis of brain connectivity and reproducibility of the different generated atlases. Then, two atlases of the method will also be compared with other state-of-the-art atlases based on MRI modalities, looking for common sub-parcels and functionalities (whenever possible). On the other hand, an analysis of cross-validation will be performed and network compute metrics will be used for the study of brain connectivity. (**G6**).
- **T13**: Design and implementation of a method to carry out the individual parcellation of the cerebral cortex. Then, fiber clustering will be applied to the whole tractography. Finally, the algorithm of individual parcellation will be used and a

post-processing will be performed to the sub-parcels. Moreover, brain connectivity metrics will be used for the evaluation. Finally, a qualitative analysis will be performed with the result of the parcellation for the subjects (**G7**).

- **T14**: Design and implementation of a parallel cortical parcellation method based on the brain topology which will take into account the gyri and sulci. Furthermore, a qualitative analysis will be carried out and the execution time of the cortical parcellation method will be measured. Finally, the reproducibility with other state-of-the-art methods will be evaluated (**G8**).

2.5. Available resources and database

The Medical Image Analysis group of the Electrical Engineering Department of the Universidad de Concepción has the adequate computing resources required by the developments and subscriptions to the main journals. For the development of this thesis, the Linux platform was used, specifically Ubuntu 18.04 LTS, mainly using the following programming languages: Python, C, and C++.

Thanks to Neurospin (CEA-Saclay, France), the center for the study of the brain, it is possible to work on this project with a high-quality MRI database called ARCHI, based on a HARDI model.

The ARCHI [113] database contains anatomical MRI, HARDI and fMRI data acquired with special acquisition sequences from a 3T MRI scanner (with a total of 12 hours of acquisition per subject) with a 12 channel antenna (Siemens, Erlangen). The MRI protocol included the acquisition of T1 image datasets using a MPRAGE sequence [10] (160 slices; TH = 1.10 mm; TE/TR = 2.98/2300 ms; TI = 900 ms; FA = 9 angle of deviation; matrix = 256$\times$240; RBW = 240 Hz/pixel; voxel size = 1$\times$1$\times$1 mm), a B0 field map, and a HARDI SS-EPI sequence [84] single-shell dataset along 60 optimized broadcast directions (DW), b = 1500 s/mm^2 , (70 cuts; TH = 1.7 mm, TE = 93 ms; TR = 14,000 ms; FA = 90; matrix = 128$\times$128; RBW = 1502 Hz/pixel; echospacing ES = 0.75 ms; partial Fourier factor PF = 6/8; GRAPPA = 2 [48]; total scan time = 16 min and 46 s). The database has the transformation matrices to convert the data between T1,

T2 (diffusion space) and Talairach spaces. The meshes of each hemisphere can also be found.

The main artifacts such as noise, susceptibility effects, geometric distortions, and eddy currents have been corrected. Further, defective slices were removed. The analytic Q-ball model was also computed [30]. Whole-brain streamline deterministic tractography was calculated [101], using a T1-based brain propagation mask [52], with one seed per voxel at T1 resolution, a maximum curvature angle of 30° and tracking step of 0.2 mm.

Special attention must be given to the propagation mask, by the use of a low FA threshold and a visual inspection of the mask, to prevent the remotion of superficial white matter voxels. More details about tractography parameters for the reconstruction of superficial white matter bundles and their reproducibility can be found in [49].

Resulting tractography datasets contain about one million fibers per subject. These calculations were performed using Brainvisa/Connectomist software[1] [33]. In addition, cortical surfaces were calculated with FreeSurfer[2]. As a post-processing step, all the fibers were resampled using 21 equidistant points, as showed in other works [55, 53].

[1] http://brainvisa.info/web/index.html
[2] https://surfer.nmr.mgh.harvard.edu/

Chapter 3

Clustering and labeling of white matter fibers

This chapter presents in Section 3.1 an efficient clustering method to deal with the white matter fiber bundles from large tractography datasets and represent the main structure of white matter fibers. The other method presented in this chapter, specifically in Section 3.2, is an automatic method for labeling superficial white matter bundles. Due to the high variability in the fibers of the superficial white matter among subjects, it is a complicated task to carry out the labeling of the white matter bundles. To achieve this goal we have applied two strategies, one focused on a matching algorithm and the other based on a popular state-of-the-art clustering method. These methods are presented in Section 3.2.3.

3.1. White matter fiber clustering

In this section, we present a new clustering of white matter fibers. This clustering, *FFClust*, efficiently groups large datasets from tractography. The resulting clusters describe the white matter bundles of an individual and thus study the structural connectivity in brains with pathologies. First, we expose the related work of white matter fiber clustering. Then, we describe the method, which is composed of four steps. The results are shown in the next section, performing a quantitative and qualitative analysis in comparison with other state-of-the-art methods. Finally, a discussion of the proposed work is made.

3.1.1. Related work

Diffusion Magnetic Resonance Imaging (dMRI) is an in-vivo and non-invasive technique that estimates the structure of white matter (WM) through the measurement of water molecules diffusion [6, 74]. The main trajectories of WM can be reconstructed

using tractography algorithms based on local orientation fields estimated from dMRI. The generated datasets consist of a 3D representation of the main WM fiber tracts [7]. Streamline deterministic tractography follows the preferred direction of water diffusion in each voxel to reconstruct trajectories or lines represented by a sequence of point coordinates in 3D space. These lines are called “streamlines” or simply “fibers”, though they do not represent real neural fibers but an estimation of the main trajectory of WM fascicles.

Tractography datasets contain fibers belonging to well known anatomical bundles, and also a set of bundles, mostly short association bundles, which have not yet completely described. They also have noisy fibers and artifacts, coming from dMRI intrinsic limitations and uncertainty, producing an incomplete reconstruction of the fibers [82]. The application of clustering methods has helped to develop methods for the study of deep white matter bundles (DWM), in particular, the construction of deep white matter bundle atlases [95, 53]. More recently, the study of superficial white matter (SWM) or short association fiber has been carried out, with the study of reproducibility of these bundles and the construction of two SWM bundle atlases [50, 106].

In general, there are two main strategies to study the brain connections given by tractograms. One is the segmentation based on anatomical Regions Of Interest (ROI) of the brain, which takes into account information on the morphology of the folding patterns of the cerebral cortex or other grey matter structures [19, 18], to extract fibers connecting two regions. The second strategy is the clustering of fibers used to obtain bundles of similar fibers, considering their shape and position [97, 55]. Both strategies can be combined resulting in a hybrid approach that can improve the definition of anatomical bundles, since more information is included in the analysis [93].

Typically, exploratory clustering methods find a great amount of bundles that characterize the structure of the white matter in its totality by using representatives such as clusters and centroids [42, 43, 55].

The method proposed by *Guevara* et al. [55] consists of several processing steps to subsequently subdivide the fibers into groups based on different criteria, like brain

masks, voxel connectivity, fiber length and point-wise fiber distance. This method obtains compact and thin clusters that can be represented by a centroid. Another important work that performs clustering with large tractographies is *QuickBundles* (*QB*) [42]. This is an unsupervised clustering algorithm that groups the fibers into clusters, without recalculating the clusters, like classical methods such as K-means. The algorithm uses a distance threshold to define whether a new fiber will be assigned to the closest cluster or will start a new cluster. It is based on the Minimum average Direct-Flip (MDF) fiber distance, although other fiber distance measures can be used. The clustering results of this method depend on the initial permutation of input fibers. This algorithm is very fast, taking about 30 minutes for a set of one million fibers.

Using a clustering method aids to process the tractography data, to subsequently apply other analyses on the resulting clusters. Example of analyses are the construction of WM bundle atlases [95, 53, 50, 106, 142]. Another application is the segmentation of bundles, also called virtual dissection, that seeks to label the anatomical bundles, already described by anatomists. Clustering-based segmentation methods use a fiber similarity or distance measures to group similar fibers along with anatomical information to identify known bundles. The algorithms embed anatomical knowledge commonly in the form of a bundle atlas or model [95, 53, 107, 65, 140, 44], or use a ROI atlas to guide the identification of anatomical bundles [130, 78, 21]. Recently, methods using Deep Learning have been proposed for the segmentation of anatomical fascicles with promising results [57, 56, 132]. Other applications are the study of reproducibility of white matter bundles [50, 49], the creation of diffusion-based cortex parcellations [89, 80] and the study of the human brain connectome [145]. Furthermore, the segmentation of anatomical bundles have been extensively applied to perform clinical studies. These methods in general compare features extracted from the bundles, such as mean FA (Fractional Anisotropy) or other diffusion-based indices, bundle volume or bundle shape descriptors. For example, studies have been carried out to study bipolar disorder [110], schizophrenia and autism spectrum disorder [68], parkinson [25] and major depressive disorder [136]. Also, white matter fiber tracts can be identified in patients with brain tumors for neurosurgical planning [94, 44].

The size increment of tractography datasets from new high quality MRI databases, and various analysis that can take advantage of clustering results, has imposed the challenge to develop high quality and optimized fiber clustering methods.

This work proposes a new method for the clustering of large tractography datasets. The main goal is to develop an efficient clustering to group fibers into compact and regular clusters, representing the whole brain WM structure. The representation must be of good quality, i.e., the clusters must be compact along all the fibers, so that several analyses can be performed, using as input the resulting clusters or cluster centroids. A special interest is its use for the study of short association bundles and their segmentation, as well as of subdivisions of long anatomical bundles.

3.1.2. Our approach

The proposed approach consists of four steps. The first step reduces data dimensionality by applying a partitioning clustering algorithm on fiber points instead of whole fibers. The second step groups fibers sharing the same cluster points into preliminary streamline clusters. Next, small preliminary streamline clusters are reassigned to larger clusters based on their direct or flipped distance. The last step builds compact clusters by merging candidate clusters based on a maximum Euclidean distance threshold using a graph representation of candidate cluster centroids. All the steps of the algorithm are executed in parallel. The experimental evaluation and comparison with the state of the art shows that the proposed method is effective in the creation of compact clusters, with a low intra-cluster distance, while keeping an inter-cluster distance not excessively large, and it provides high performance. The proposed method is about 8.6 times faster than the state-of-the-art method, which enables a fast processing and visualization of main white matter fiber clusters.

Let T_s be a tractography dataset of an individual subject consisting of a collection of fibers or streamlines, where each fiber is formed by 21 points in $\mathbb{R}^3$. The streamlines on a dataset are loaded into the main memory following the order of points calculated during the tracking. Hence, two orientations are possible: direct, or reverse (flipped),

which must be considered in the streamline analysis. In this work, the similarity between fibers is defined by the maximum Euclidean distance between corresponding points (d_{ME}) [55, 53, 50, 106]. It is a restrictive distance, since any relevant local difference between the fibers, on the fiber extremities and along all their shape, will be successfully captured.

We denote a fiber in direct order as a with 21 points in $\mathbb{R}^3$, that is each $a_i \in a$ is a 3D point, with coordinates x, y and z. In addition, we denote the fiber flipped representation as a^F, which contains 21 points in $\mathbb{R}^3$ in reverse order, that is, the first 3D point in a^F is the last in a, and so on. We denote $d_E(a_i, b_i)$ as the Euclidean distance between corresponding points a_i and b_i of fibers a and b. We assume a direct (d_E), as the maximum distance between any of the 21 points in fibers a and b, in direct order, and a flipped Euclidean distance (d_{EF}) with one fiber in inverse order. We consider the minimum of distances d_E and d_{EF}, denoted as d_{ME}, to measure the distance between streamlines a and b, as defined in Eq.3.1.

$$
\begin{aligned}
d_E(a_i, b_i) &= ||a_i - b_i|| = \sqrt{(a_{i_x} - b_{i_x})^2 + (a_{i_y} - b_{i_y})^2 + (a_{i_z} - b_{i_z})^2} \\
d_E(a, b) &= max_{i \in 21}(d_E(a_i, b_i)) \\
d_{EF} &= d_E(a, b^F) = d_E(a^F, b) \\
d_{ME} &= min(d_E(a, b), d_{EF}(a, b))
\end{aligned} \tag{3.1}
$$

Hence, distance d_{ME} calculates the maximum Euclidean distance between corresponding points, taking into account the two possible fiber orientations.

Our approach aims at improving the final clusters quality and the algorithm time complexity. A special interest is to keep a good similarity between fibers along all the fiber shape, in particular, on the extremities of the fibers. This feature is crucial for the study of superficial white matter, where short association fibers connect small gray matter regions, and a difference in the fiber end points for a group of fibers must lead to different clusters. To achieve this goal, the algorithm is based on a partition clustering applied separately to a subset of points along the fibers in parallel. Then, fibers with

points sharing the same cluster points are merged. This strategy is more efficient than using the complete fiber data representation which requires more expensive fiber distance computation.

The algorithm proceeds in the following four steps: *(1)* building point clusters, *(2)* generating preliminary streamline clusters, *(3)* reassigning small preliminary streamline clusters and *(4)* merging candidate streamline clusters. Figure 3.1 displays the complete workflow.

STEP 1: Building point clusters

This step aims at reducing the dimensionality of the input data by applying a partition clustering on a subset of streamline points. Applying the clustering locally on a subset of fiber points reduces the number of dimensions of the input elements and then the number of pairwise distance computations needed to form clusters. Distance computations are performed on three dimensions fiber points instead of fibers formed by 21 points, where each point has three dimensions. The method uses a subset of five points, including the two ending points (1, 21), the central point (11) and two intermediate points (4, 18). The decision of using these points is given because it has been shown that this sampling strategy is efficient and significant to estimate the maximum distance between fibers [72] and hence, to discriminate fiber differences.

The Minibatch K-means [114] (MK) was chosen as a partition algorithm because it is known to provide good quality, and low time and space complexity. Moreover, given that the clustering algorithm is applied on each streamline point independently, the number of clusters do not need to be the same in all streamline points. In fact, the proposed algorithm uses different numbers of clusters for streamline ending points and central points. We denote the number of clusters for ending points as Kp_o, and the number of clusters for intermediate and central points as Kp_c. We apply the elbow method to find out the best number of clusters on each point [70].

Figure 3.1-(a) illustrates with an example this STEP, using the five streamline points 1, 4, 11, 18, and 21. The intermediate and central points of the clusters are shown as p_c and the ending points as p_o. The central and intermediate point clusters are

identified by the membership labels C, B, G, D, and H; and the ending points clusters by the labels A, F, E and I.

STEP 2: Generating preliminary streamline clusters

This step builds preliminary streamline clusters by grouping streamlines based on the membership labels of point clusters obtained in the previous step. A preliminary streamline cluster will contain all the fibers which points share the same point cluster labels. The method uses a dictionary data structure in which the key is the set of membership labels where each streamline point belongs to, and the value contains all fiber IDs that share the same set of point cluster membership labels. A preliminary streamline cluster contains all fiber IDs stored in the value associated to a key in the dictionary.

Figure 3.1-(b) shows two streamline preliminary clusters, one is formed by streamlines (p, q) and the other by streamlines (r, s). As seen in Figure 3.1-(a), points 1, 4, 11, 18 and 21 of fibers p and q belong to the same point clusters, then such streamline cluster is defined by the corresponding point cluster labels (A, B, C, D, E). In the same way the streamlines (r, s) formed a second preliminary streamline cluster identified by the point cluster labels (F, G, C, H, I).

STEP 3: Reassigning small preliminary streamline clusters

This step reassigns small preliminary streamline clusters that could be separated from large clusters in the previous steps. A small cluster is reassigned to the nearest large cluster, given a maximum distance threshold. In addition, with this processing, small noisy clusters are identified and discarded.

In order to do this, we first divide the preliminary clusters in two sets based on their number of streamlines. A set S_L contains all preliminary clusters with number of streamlines equal or greater than 6, and set S_S contains all preliminary clusters with 5 or fewer streamlines. Centroids for each preliminary cluster in both sets are computed as the arithmetic mean of each streamline point. Then, a preliminary cluster in set S_S is reassigned to the closest preliminary cluster in set S_L, only if the distance between

their centroids is below the threshold d_{Rmax}, that is, if $d_{ME}(a, b) < d_{Rmax}$ (see Eq. 3.1). Otherwise, corresponding clusters are not reassigned.

At the end of this step, if there still are preliminary clusters in set S_S containing one or two streamlines, these are considered noise and eliminated by default. Figure 3.1-(c) shows preliminary streamline cluster set separation (Figure 3.1-(c).3.1.), centroid computation (Figure 3.1-(c).3.2.), reassignment, not reassignment and elimination cases (Figure 3.1-(c).3.3). At the end of this step we obtain *candidate clusters*.

STEP 4: Merging candidate streamline clusters

This is a global refining step that aims to merge *candidate clusters* which still might be close based on a maximum distance parameter d_{Mmax}, in particular, clusters with flipped streamlines. This step first makes candidate cluster groups, where each group consists of clusters that share the same central point membership label obtained during the *STEP 1*. This makes groups that are close only by the central point. This processing is done only to avoid the pairwise comparison among all cluster centroids, however, this processing adds no error to the merging computation.

Then candidate cluster centroids in each group are merged based on the maximum distance parameter d_{Mmax}. If candidate cluster centroids are below d_{Mmax} for a maximum Euclidean direct or flipped distance, then such clusters are merged. To avoid the computation of all possible configurations of multiple candidate clusters that can be merged we formulate the problem using a graph representation and approximate the solution using a graph algorithm. The graph representation considers that each candidate cluster centroid is a vertex, u, in an undirected graph, $G(u, v)$, and there is an edge, e, between two vertices, u and v, only if the cluster centroids they represent are below a maximum Euclidean distance threshold d_{Mmax}, that is, only if $d_{ME}(u, v) < d_{Mmax}$. Figure 3.1-(d) shows an example where there are six candidate cluster centroids (c_1, c_2, c_3, c_4, c_5, c_6) represented with corresponding vertices (v_1, v_2, v_3, v_4, v_5, v_6). In this case, there are four edges (e_1, e_2, e_3, e_4), which exist only because the distance threshold is satisfied. Then, a graph algorithm is applied to find groups of centroids, where each group contains all of its centroids close to each other. In a graph

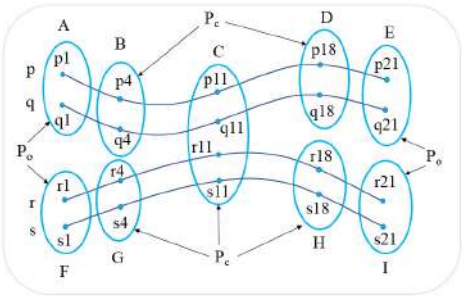
representation this is called a *clique*, where each group consists of vertices where all pair of vertices are connected by an edge. Specifically, the proposed method aims at finding *maximal cliques*, which are cliques that cannot grow by adding another vertex. Figure 3.1-(d) shows three maximal cliques, one is formed by vertices (v_1, v_4, v_6), another is formed by vertices (v_2, v_3) and the other has vertex (v_5). Next, the method sorts all maximal cliques by decreasing size and merges all candidate clusters represented in cliques having at least two vertices, which clusters have not been previously merged. Note that a clique of size one means that the centroid is not close to any other centroid and then no merge should be performed. Given that a vertex in a clique represents a cluster centroid, and that the graph representation does not include the actual distance values among centroids, merging largest cliques first aims at joining more clusters that are close to each other based on the given threshold d_{Rmax}.

Figure 3.1-(d) shows that candidate clusters represented by the centroids c_1, c_4 and c_6 are merged into the final *green cluster*, candidate clusters represented by the centroids c_2 and c_3 are merged into the final *red cluster* and candidate cluster represented by the centroid c_5 becomes the final *orange cluster*.

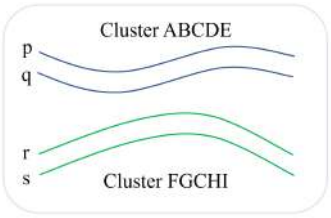
3.1.3. Experimental results

This section describes the experimental evaluation and the results obtained in terms of clustering quality and performance of the method. It also performs a comparison analysis with the state-of-the-art techniques including the methods: *Guevara* [55], *QuickBundles* [42], and *QuickBundlesX* [43]. In addition, it presents the results obtained for the segmentation of bundles based on a recent SWM atlas [50]. All results are obtained by using 50 subjects.

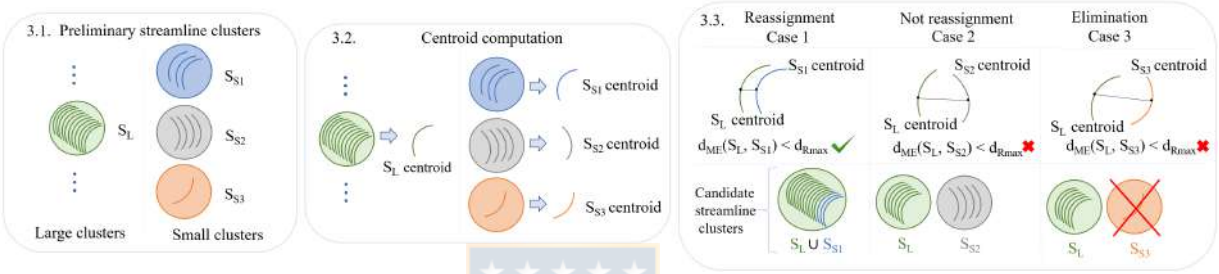
The method proposed by *Guevara* et al. [55] consists of several processing steps to subsequently subdivide the fibers into groups based on different criteria, like brain masks, fiber length, voxel-based connectivity, and point-wise fiber distance. The method provides high-quality clusters, but it has about 13 configuration parameters and it is time consuming. *QuickBundles* [42] (*QB*) is a clustering method specialized in grouping white matter fibers from tractography datasets quickly and with good quality. This



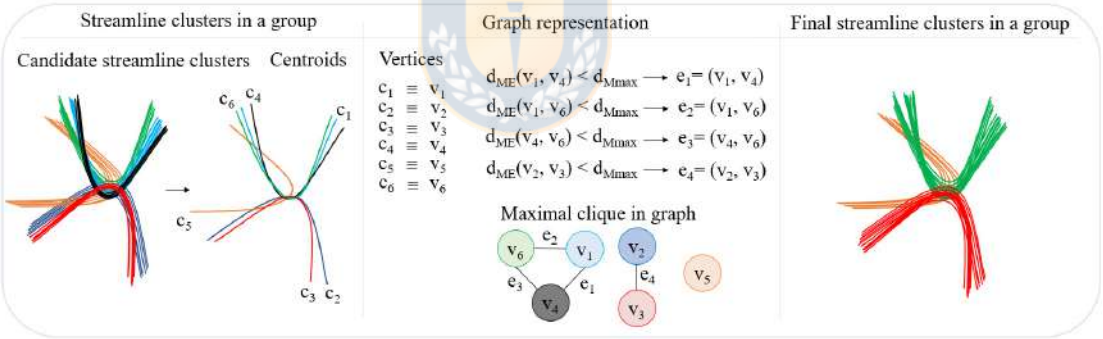
(a) STEP 1: Building point clusters.



(b) STEP 2: Generating preliminary streamline clusters.



(c) STEP 3: Reassigning small preliminary streamline clusters.



(d) STEP 4: Merging candidate streamline clusters.

Figure 3.1: *FFClust*. (a) STEP 1: Building point clusters. MK is applied on the marked points. (b) STEP 2: Generating preliminary streamline clusters. Fibers sharing the same label form a cluster. (c) STEP 3: Reassigning small preliminary streamline clusters. Clusters are separated into large and small (3.1.). Centroids are calculated (3.2.). Reassignment (3.3.). Case 1: $d_{ME} < d_{Rmax}$, clusters are joined in a candidate. Case 2: $d_{ME} \geq d_{Rmax}$ and clusters with sizes greater than one fiber. Case 3: $d_{ME} \geq d_{Rmax}$ and one cluster with a fiber. (d) STEP 4: Merging candidate streamline clusters. Centroids are vertices, close vertices are edges based on d_{Mmax}. Maximal cliques are found to merge candidate clusters.

unsupervised clustering algorithm groups the fibers into clusters, without recalculating the clusters, like classical methods such as K-means. The algorithm uses a distance threshold to define whether a new fiber will be assigned to the closest cluster or will start a new cluster. The algorithm has a single parameter, which is the Minimum average Direct-Flip distance (MDF) between two pairs of fibers [42], although other fiber distance measures can be used. It is one of the fastest methods that exist today, with runtime $\mathcal{O}(N^2)$, being N the size of the dataset. *QuickBundlesX* [43] optimizes the *QuickBundles* algorithm using a tree data structure.

FFClust method was implemented in Python version 3.6 and in C language using compiler g++ version 7.4.0. The method supports sequential and parallel execution using OpenMP. All experiments were executed in a machine consisting of a 12-core Intel Core i7-8700K CPU with 3.70GHz,680, and 32GB of RAM, using Ubuntu 18.04.2 LTS with kernel 4.15.0-51 (64 bits).

The experiments were performed using deterministic tractography datasets from the ARCHI database [113], with one million of streamlines. To compare execution times, we use tractography datasets with resampled subjects from 330,000 to 2,729,000 streamlines.

Parameter configuration for quantitative analysis

The method has three configurable parameters. First, the number of clusters (Kp_c and Kp_o) for each of the five streamline points on which the MK algorithm is applied (*STEP 1*). Second, the maximum Euclidean distance threshold (d_{Rmax}) for the reassignment of small to large preliminary clusters (*STEP 3*). Third, the maximum Euclidean distance threshold (d_{Mmax}) for merging candidate clusters into final clusters in the last step (*STEP 4*).

- **Finding the number of fiber point clusters**

  First, the number of clusters for applying MK in *STEP 1* is determined using the Elbow method [70]. To do this, the MK algorithm is executed for each of the five streamline points of a subject of one million fibers of a subject of the ARCHI

database with 50, 150, 200, 250, 300, 350, 400 and 450 clusters. Figure 3.2 displays the total intra-cluster variation or total within-cluster sum of squares (WCSS) for different number of clusters for the five streamline points. The elbow method shows the total WCSS as a function of the number of clusters. As the number of clusters increases the WCSS decreases, which indicates that clusters get more compact. The idea of the elbow method is to choose a number of clusters where the WCSS does not decrease much when using more clusters.

Figure 3.2 shows that a good number of clusters for the intermediate and central points (Kp_c) is between 150 and 200, and for the ending points (Kp_o) is between 150 and 300. Finding the best values for the parameters d_{Rmax} and d_{Mmax} were considered using intra-cluster maximum distance.

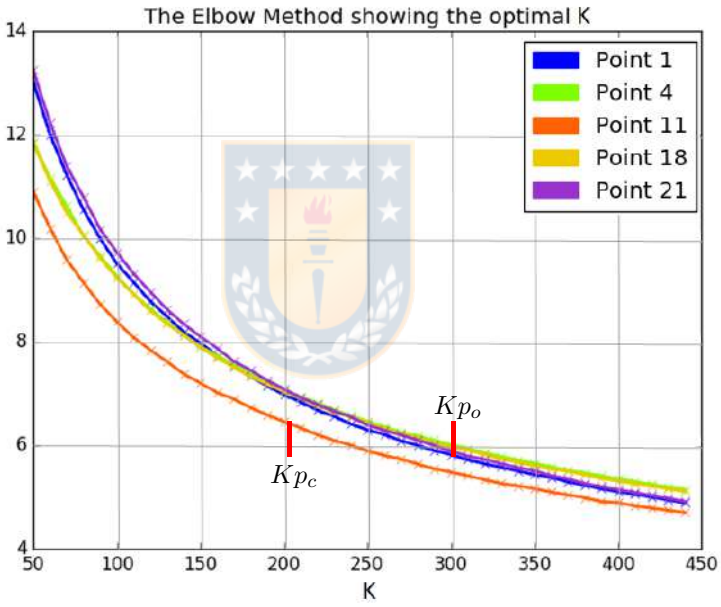


Figure 3.2: Elbow method showing the optimal number of clusters K. The x-axis shows the number of clusters, y-axis shows the inertia. K's optimal values are located at the elbow of the line.

- **Best configuration**

  The best parameter configuration for the proposed method consists of the number of clusters for the intermediate and central points $Kp_c = 200$ and the ending points

$Kp_o = 300$ (*STEP 1*), and the value of 6 mm for the threshold distances d_{Rmax} in *STEP 3*, and d_{Mmax} in *STEP 4*.

Comparison with the state-of-the-art methods

This section provides a quality comparison with the state-of-the-art methods including *Guevara* [55], *QuickBundles* [42], and *QuickBundlesX* [43]. *QuickBundlesX* is the successor of *QuickBundles*, and it performs clustering on streamlines by building a tree at different levels, with different distance thresholds. The experimental evaluation considers both the default and best parameter configurations for all methods. Specifically, *QBmdf6* refers to using *QuickBundles* using MDF distance of 6 mm; *QBmdf10* using MDF distance with 10 mm; *QBXmdf6* using *QuickBundlesX* with distance 6 mm; and *QBXmdf10*, using *QuickBundlesX* with distance 10 mm. Finally, *Guevara* refers to the method proposed in [55], which uses a maximum Euclidean distance of 10 mm.

The first evaluation, showed in Figure 3.3, considers the quality of the clustering approaches, by computing the intra-cluster, inter-cluster maximum Euclidean distances and the cluster sizes obtained by all the alternatives. The comparison includes the error bars using 50 subjects, with tractography datasets of 1,045,676 fibers in average.

Figure 3.3-(top) shows the number of clusters with corresponding intra-cluster distance, Figure 3.3-(middle) shows the number of clusters with inter-cluster distance, and Figure 3.3-(bottom) shows the number of clusters with cluster sizes. As observed, *FFClust* provides clusters with small intra-cluster distance, where all clusters have distances below 60 mm. *QuickBundles* with MDF of 6 mm also provides clusters with small intra-cluster distance. All other methods produce considerable more clusters with intra-cluster distance greater than 45 mm. Note that the intra-cluster distance is measured with the maximum Euclidean distance between the corresponding points. This distance is more restrictive that MDF distance, based on the mean Euclidean distance, and employed by *QB*. Furthermore, the distance threshold used to compare and fuse the clusters in all the analyzed methods (*QB*, *Guevara* and *FFClust*) is applied to the cluster centroids. Hence, the distance between the fibers of the clusters can be higher than the threshold.

On the other hand, as seen in Figure 3.3-(middle), the inter-cluster distance of *FFClust* is similar to the inter-cluster distance of *QBmdf6* and *QBXmdf6*. The other methods have more clusters with a greater inter-cluster distance than *FFClust*, but as mentioned, they also have greater intra-cluster distances. Also, Figure 3.3-(bottom) shows that *FFClust* and *Guevara* generates smaller clusters than the other methods, however, they are also able to find large clusters. The maximum cluster size is about $8,000$ for *FFClust* and $9,000$ for *Guevara*. In contrast *QBmdf10*, *QBXmd6*, and *QBXmdf10* have clusters of sizes over $15,000$. Note that Figure 3.3 shows the results in logarithmic scale.

Both *FFClust* and *Guevara* methods eliminate fibers, but *QB* and *QBX* do not. The *Guevara* method eliminates 37%, and *FFClust* eliminates 13% of the total number of fibers. Note that *FFClust* eliminates only small clusters containing 1 or 2 fibers.

Another experimental evaluation was performed to consider *QuickBundles* and *QuickBundlesX* using maximum Euclidean distance instead of the MDF. However, both methods become time consuming, taking days to complete on a dataset of 100,000 fibers. Experiments showed the quality of clusters, considering the same measures, i.e., intra-cluster and inter-cluster distances, and cluster sizes, are similar to *FFClust*'s.

A third quality clustering evaluation uses the Davies–Bouldin (DB) index [137]. The DB index is defined as the average similarity between each cluster with its most similar cluster, where similarity refers to the ratio of intra-cluster and inter-cluster distances. The DB index is computed by Equation 3.2, where n is the number of clusters; α_i and α_j are the average distances between all elements of cluster i and j respectively; c_i and c_j are the centroids of cluster i and j; and $d(c_i, c_j)$ is the average distance between both centroids. A DB index has normalized values between 0 and 1, where a value closer to zero indicates a better separation between the clusters.

$$DB = \frac{1}{n}\sum_{i=1}^{n} \max_{i \neq j} \left(\frac{\alpha_i + \alpha_j}{d(c_i, c_j)} \right) \tag{3.2}$$

Figure 3.4 displays the DB index for all methods, which shows that *QBmdf6* provides the best score, and *FFClust* is the second best method. Then, closely follows

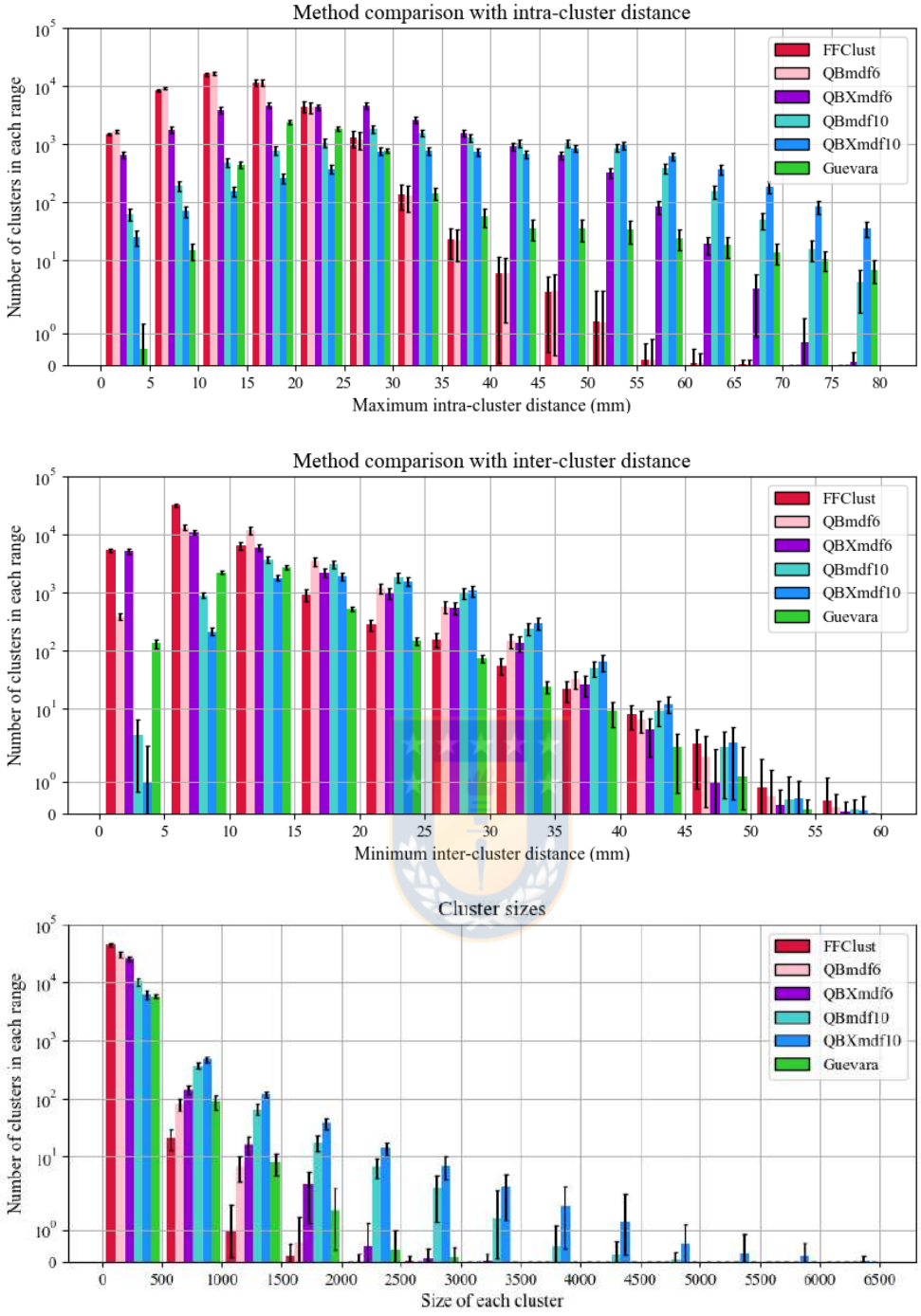


Figure 3.3: Method comparison with intra-cluster, inter-cluster maximum distances, and cluster sizes. Error bars are computed by using 50 subjects with approximately one million fibers. Top figure shows intra-cluster maximum distance, middle figure shows inter-cluster maximum distance, and bottom figure shows cluster sizes.

the *Guevara* method. The figure also shows that *QB* provides better DB score than *QBX*.

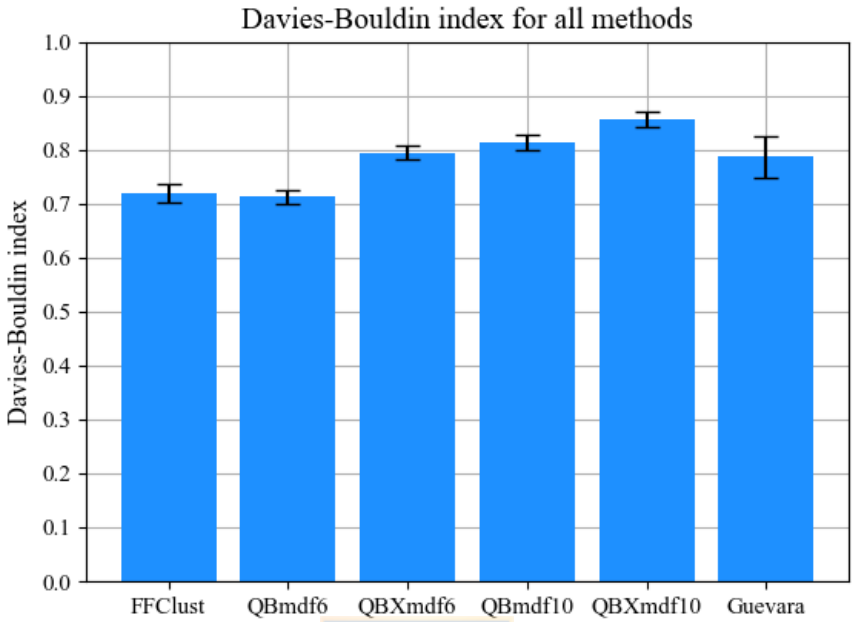


Figure 3.4: Davies-Bouldin index for each method. X-axis contains each method. *FF-Clust* ($d_{Rmax} = 6mm$ and $d_{Mmax} = 6mm$). Y-axis shows the DB index, the closer to zero the better value.

Qualitative analysis using segmentation

This section describes a qualitative analysis based on the segmentation of bundles using an atlas of superficial white matter bundles [50]. This analysis compares the segmentation results obtained by the state-of-the-art methods. The segmentation method selects and labels the closer cluster to each atlas bundle, based on a maximum Euclidean distance threshold [72, 128].

Figure 3.5 shows the segmentation comparison with the SWM atlas [50], obtaining the closest clusters for each method using the four atlas bundles connecting the postcentral (PoC) and precentral (PrC) gyri. To obtain the closest clusters, the strategy uses a segmentation threshold distance of 6 mm, and then if a method does not find all

bundles, it is increased to 8 mm. The *Guevara* and *FFClust* methods are able to identify the four bundles with a threshold distance of 6 mm, and *QB* alternatives are able to identify the four bundles using the segmentation threshold of 8 mm. Figure 3.5 also shows an error percentage on the bottom right of each image. This error measures the percentage of fibers that are out of the correct regions. Since there is no ground truth, this is not a real measure of quality but provides an insight into the fibers that are included in the clusters but differ from the main fiber shape, with fibers that connect surrounding cortical regions. The *Guevara* method achieves the best streamline endpoint error (5.4%) and the second best is achieved by *FFClust* (6.2%), and QBmdf6 follows with 6.5%.

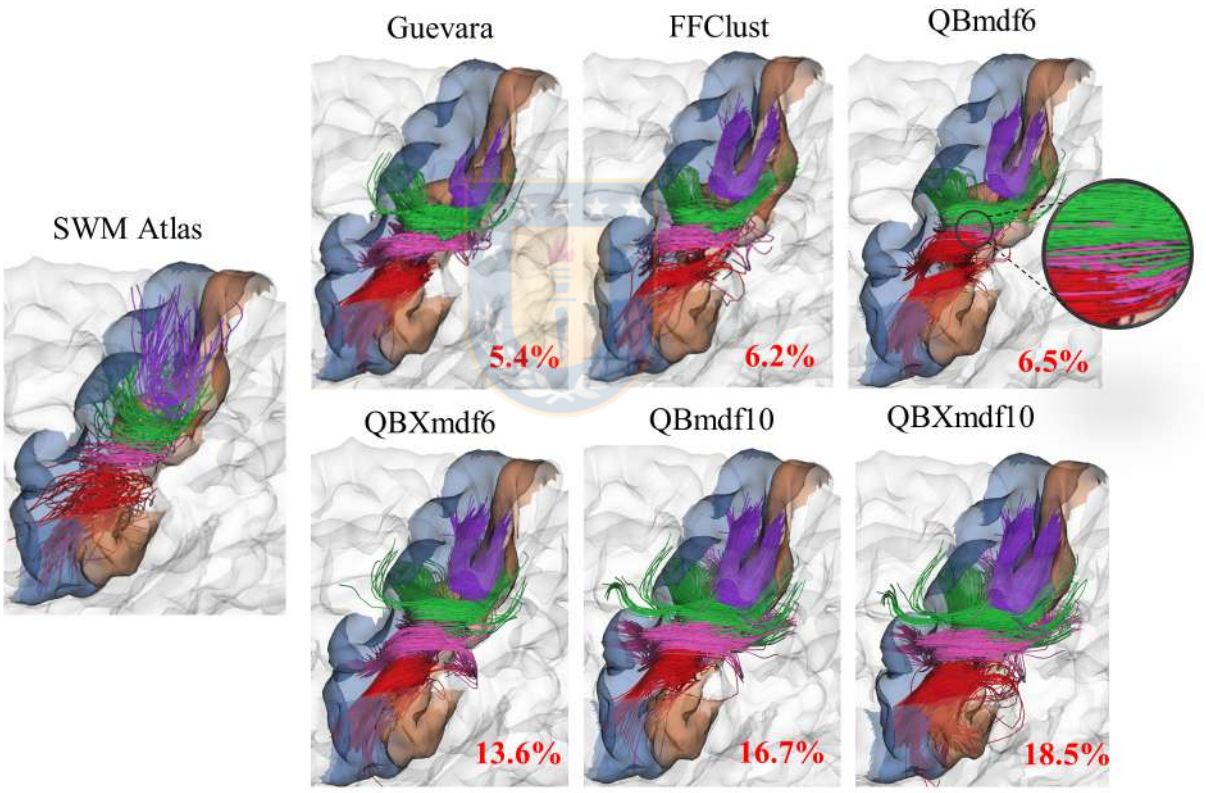


Figure 3.5: Segmentation comparison with bundles of a SWM atlas [50] connecting the postcentral (PoC) and precentral (PrC) gyri. Top image show the *Guevara*, *FFClust* ($d_{Rmax} = 6mm$ and $d_{Mmax} = 6mm$), and *QBmdf6* methods. The circular image displays a zoom for *QBmdf6*, to better show the pink cluster, that is under the red one. Bottom image show *QBXmdf6*, *QBmdf10*, and *QBXmdf10* methods. The criterion for segmentation was a greedy approach with heuristics to the nearest bundle.

Figure 3.6 shows the segmentation comparison with the SWM bundle atlas [53] for the corticospinal tract (CST), the inferior fronto-occipital fasciculus (IFO), and the inferior longitudal fasciculus (IL). We can observe that all the methods are able to re-group the fibers belonging to these long bundles, with a variable number of clusters, depending on the threshold and the method itself.

Comparison with QuickBundles by qualitative analysis

This section evaluates the final clusters obtained in *FFClust*, *QBmdf6*, *QBmdf10*, and *Guevara* using a visual inspection to observe the quality of the final clusters. *QBX* is not considered because both the DB index and the segmentation evaluation show that *QB* achieves better quality than *QBX*.

The evaluation considered five cases: thinner clusters (50 thinnest clusters that present between 2 and 5 fibers), thicker clusters (50 clusters with the most fibers), short fiber clusters (200 short fiber clusters, between 30 and 60 mm), long fiber clusters (50 clusters with the longest fibers, starting at 80 mm), the least homogeneous clusters, i.e., clusters with the largest maximum intra-cluster distance, and the most similar clusters (94 clusters with the most similar ones among them). The following is a detailed description of each of the cases:

1. **Thinner clusters**. The thinnest clusters are those with the least number of streamlines. We considered such clusters are those having between 2 and 5 streamlines. Figure 3.7 shows the results of *FFClust*, *QBmdf6*, *QBmdf10*, and *Guevara* with the 50 thinner clusters. Visual inspection shows that the methods provide similar clusters, however *QB* and *Guevara* clusters seem to be more scattered than clusters obtained by *FFClust*.

2. **Thicker clusters**. The thickest clusters are those having the largest sizes, that is, clusters with the largest number of streamlines. We consider the 50 thickest clusters for visual inspection. Figure 3.8 shows the results for both algorithms. We provide each of the three views of the brain (coronal, axial and sagittal). We note that *FFClust*, *QBmdf6* and *Guevara* obtain clusters that look uniform. As for

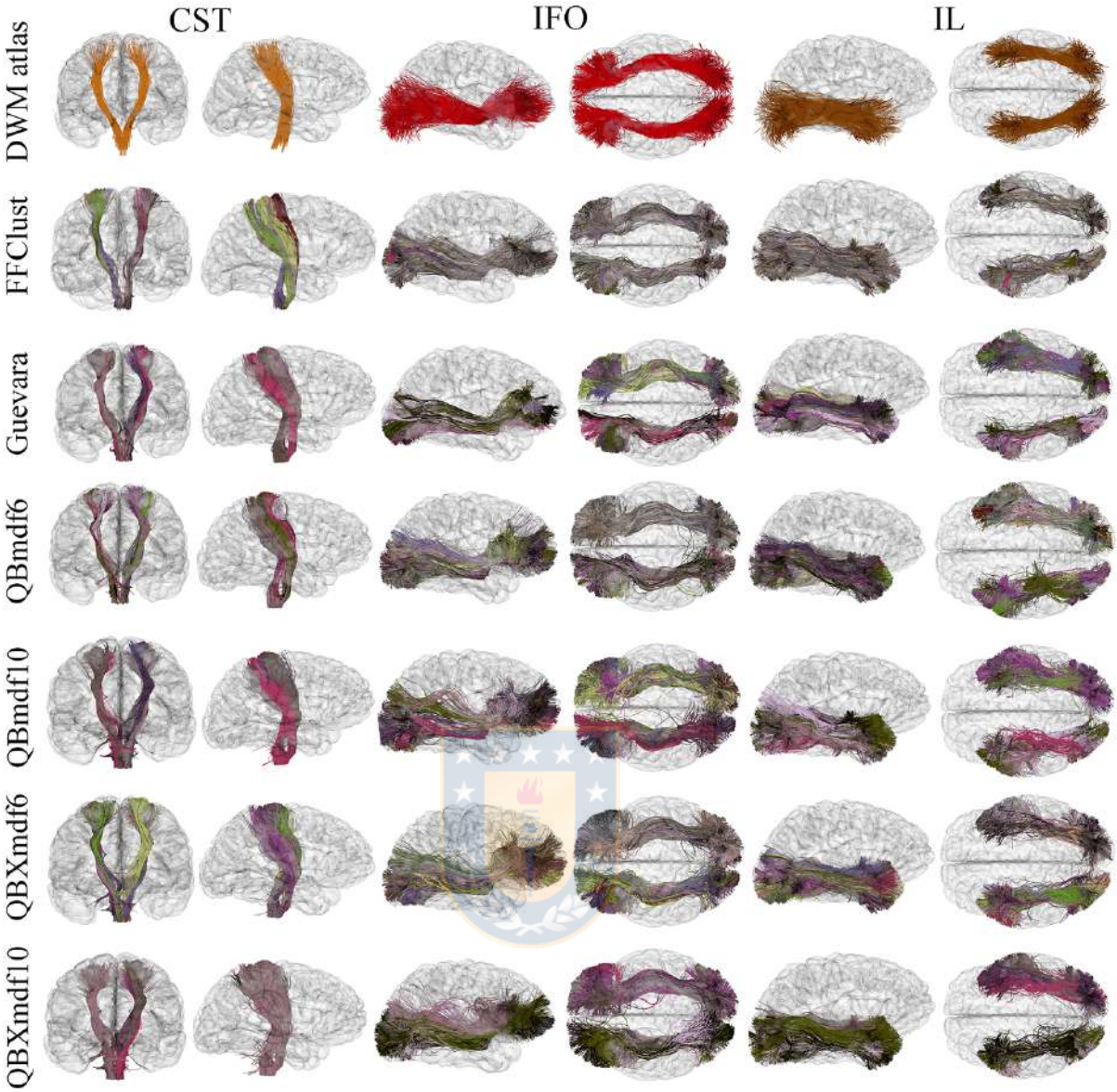


Figure 3.6: Segmentation comparison based on a DWM bundle atlas [53] for the corticospinal tract (CST), the inferior fronto-occipital fasciculus (IFO), and the inferior longitudal fasciculus (IL). The first two columns show the CST. The third and fourth columns show the IFO, and the last two columns show the IL. The first row shows the DWM atlas bundles. The following rows show the bundles obtained by the different methods used in the comparison of long fiber segmentation for the mentioned bundles.

QBmdf10, we see that the clusters have less homogeneous and scattered ending points.

3. **Short fiber clusters**. We visualize short fiber clusters having fibers of length up

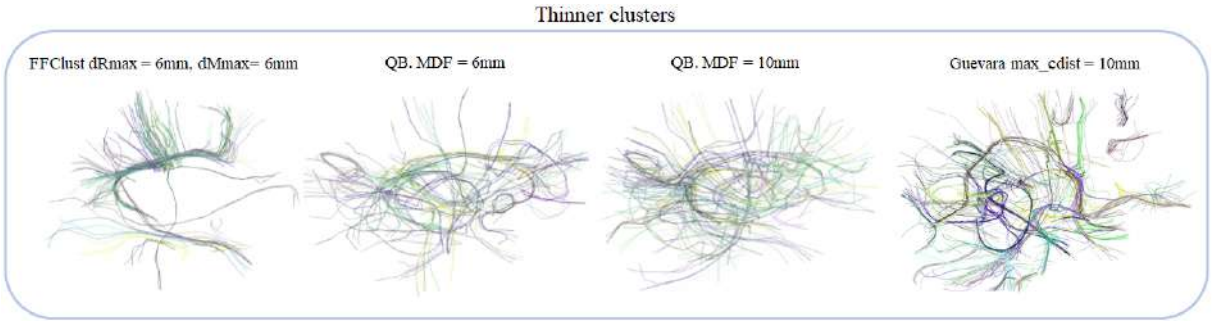


Figure 3.7: Thinner cluster comparison. In the comparison we show *FFClust* with $Kp_c = 200$, $Kp_o = 300$, d_{Rmax} = 6 mm, d_{Mmax} = 6 mm, together with *QB* with MDF = 6 mm, MDF = 10 mm, and *Guevara* max_cdist = 10 mm.

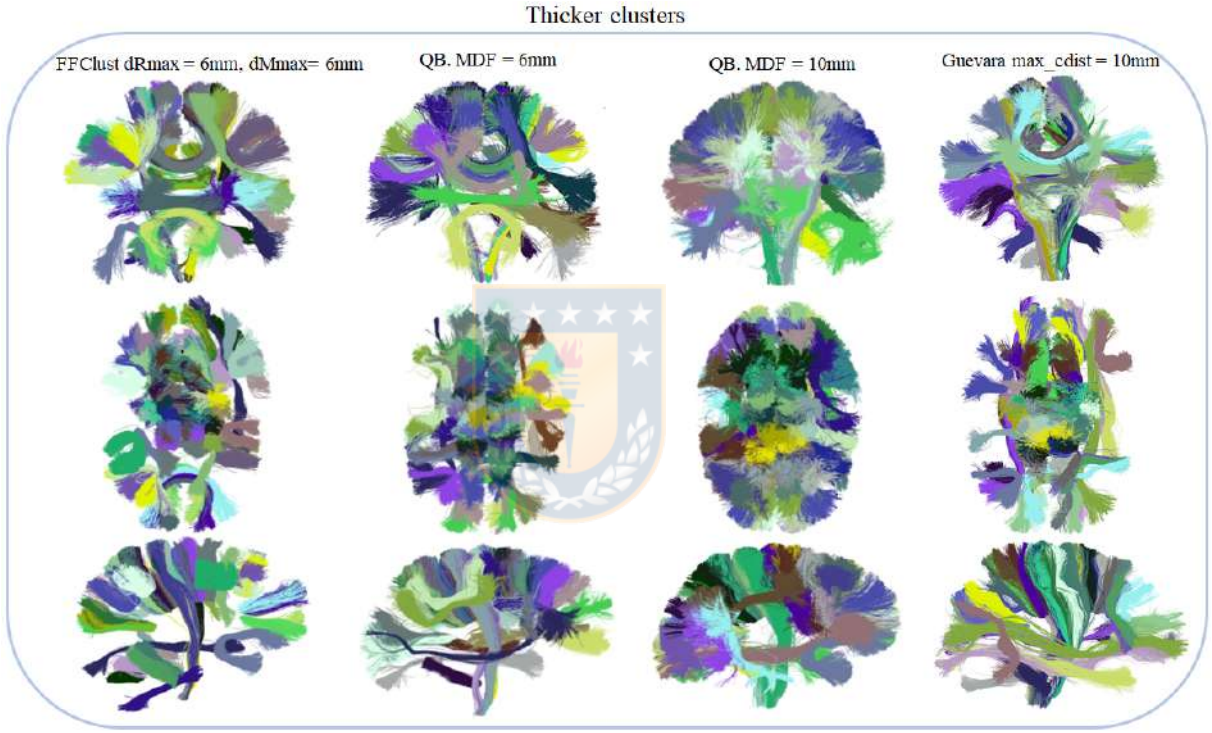


Figure 3.8: Thicker cluster comparison. The comparison shows *FFClust* with Kp_c = 200, Kp_o = 300, d_{Rmax} = 6 mm, d_{Mmax} = 6 mm, together with *QB* with MDF = 6 mm, MDF = 10 mm, and *Guevara* with max_cdist = 10 mm. Display of coronal, axial and sagittal views.

to 60 mm. Figure 3.9 shows the comparison of short fibers for the three algorithms. We observe that the quality of *QBmdf6*, *FFClust* and *Guevara* are very similar, but we see some clusters with more scattered ending points for *QB*. Again *QBmdf10* presents clusters too wide with frizzy ending points. Short fibers are

usually analyzed based on the regions they connect [49]. Hence, having endpoints from different streamlines close to each other in a cluster will help to minimize the number of streamlines connecting neighboring regions. However, this does not ensure that all the cluster streamlines will land on the same anatomical structure.

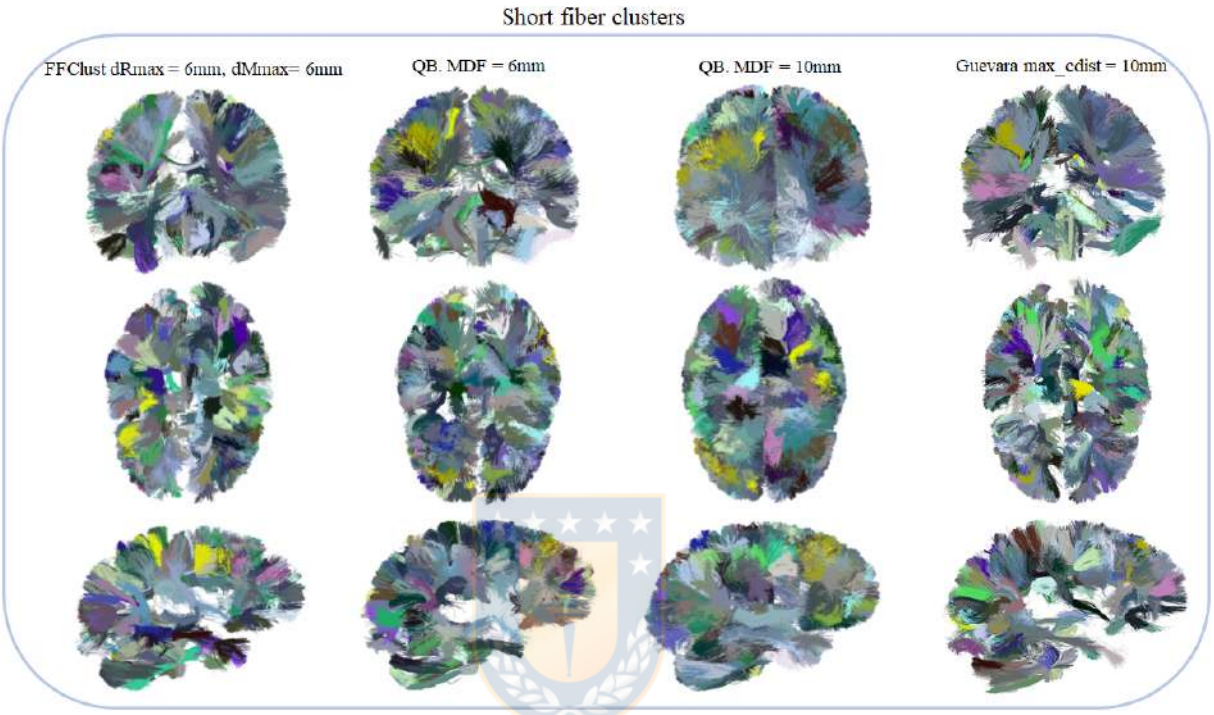


Figure 3.9: Comparison of short fiber clusters. The comparison contains the 200 thickest clusters with fiber lengths between 30mm and 60mm. It shows *FFClust* with Kp_c = 200, Kp_o = 300, d_{Rmax} = 6 mm, d_{Mmax} = 6 mm, together with *QB* with MDF = 6 mm, MDF = 10 mm, and *Guevara* max_cdist = 10 mm. Display of coronal, axial and sagittal views.

4. **Long fiber clusters**. We denote long fiber clusters those clusters with fibers of length greater than 80 mm. To facilitate visual inspection we present the 50 longest clusters. Figure 3.10 shows the comparison between *FFClust*, *QBmdf6*, *QBmdf10*, and *Guevara* for coronal, axial and sagittal views. As in the previous experiments, we observe that the clusters generated by *FFClust*, *QBmdf6* and *Guevara* are very similar and compact, whereas clusters for *QBmdf10* have frizzy ending point.

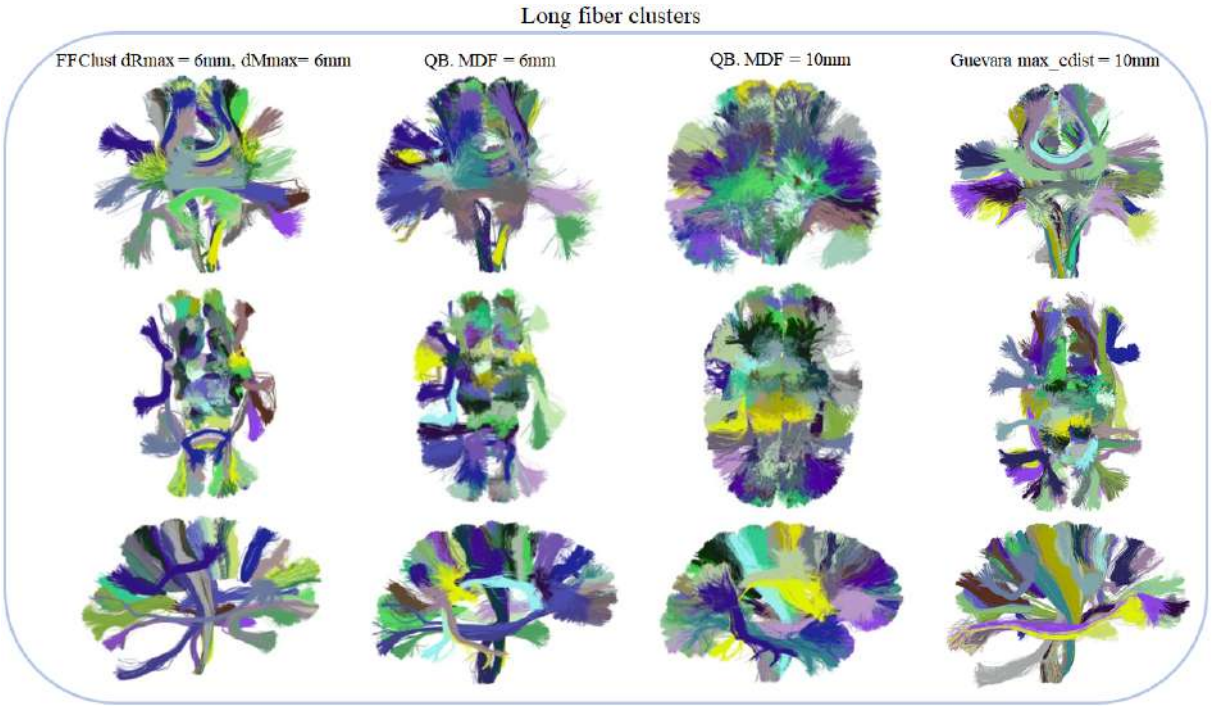


Figure 3.10: Comparison of long fiber clusters. In the comparison appears the 50 thickest clusters with fibers longer than 60 mm long. We show *FFClust* with Kp_c = 200, Kp_o = 300, d_{Rmax} = 6 mm, d_{Mmax} = 6 mm, together with *QB* with MDF = 6 mm, and MDF = 10 mm, and *Guevara* max_cdist = 10 mm. Display of coronal, axial and sagittal views.

5. **Clusters with largest maximum intra-cluster distance**. This experiment studies the final clusters that have a maximum intra-cluster distance over 40 mm. *FFClust* obtains three clusters over this maximum distance, shown in Figure 3.11-(left). *QBmdf6* obtains about 15 clusters over 50 mm, shown in Figure 3.11-(middle left).

   QBmdf10 and *Guevara* obtain about 140 over 50 mm, hence we display only the clusters with maximum intra-cluster distance over 70 mm (see Figures 3.11-(middle right) and Figure 3.11-(right)).

   Figure 3.11 shows that *FFClust* has only a few clusters with small number of streamlines and atypical forms. We suggest that they are noise in the tractography. Some clusters found in *QB* with intra-cluster distance were probably divided into several small clusters by *FFClust*. *FFClust* only discards small and isolated

clusters, that are dissimilar to all the other large clusters. This is performed in its third step, which tries to reassign the small clusters (with five or fewer streamlines) to the largest clusters. All small clusters, with one or two streamlines that are not reassigned to a large cluster are eliminated. It should be noted that none of the compared methods were designed to eliminate erroneous clusters of medium or large size. Further analyzes with additional information are required to perform this kind of filtering.

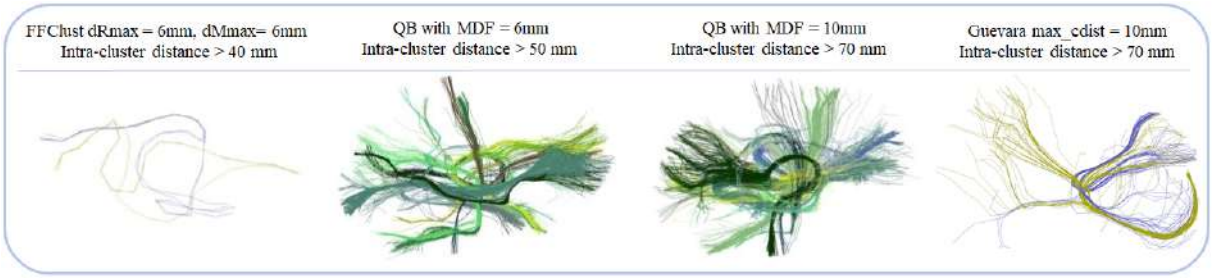


Figure 3.11: Images of clusters with greater intra-cluster distance. Clusters with large intra-cluster distance, *FFClust* with d_{Rmax} = 6 mm and d_{Mmax} = 6 mm are shown with distance $>$ 40 mm, *QB* with MDF = 6 mm with distance $>$ 50 mm, *QB* with MDF = 10 mm with distance $>$ 70 mm, and in *Guevara* with max_cdist = 10 mm.

6. **The most similar clusters**. We also analyze the most similar clusters, i.e., those that most resemble each other for *FFClust*, *QBmdf6*, *QBmdf10*, and *Guevara*. To identify similar clusters, we used the 100 bundles of a SWM atlas [50] as a reference. Those bundles where found to be the short association bundles most stable across subjects [50]. We identified the clusters that most resemble them, using a maximum threshold of 6 mm, obtaining 94 bundles for the four configurations. Figure 3.12 shows the results of *FFClust*, *QBmdf6*, *QBmdf10*, and *Guevara* with the 94 most similar clusters. Those clusters appear very similar for all the methods.

Runtime comparison

This section evaluates the execution time of *FFClust* with state-of-the-art methods. *FFClust*-seq denotes the sequential version of *FFClust*, where the complete algorithm is executed using only one thread. *FFClust*-par denotes the parallel implementation

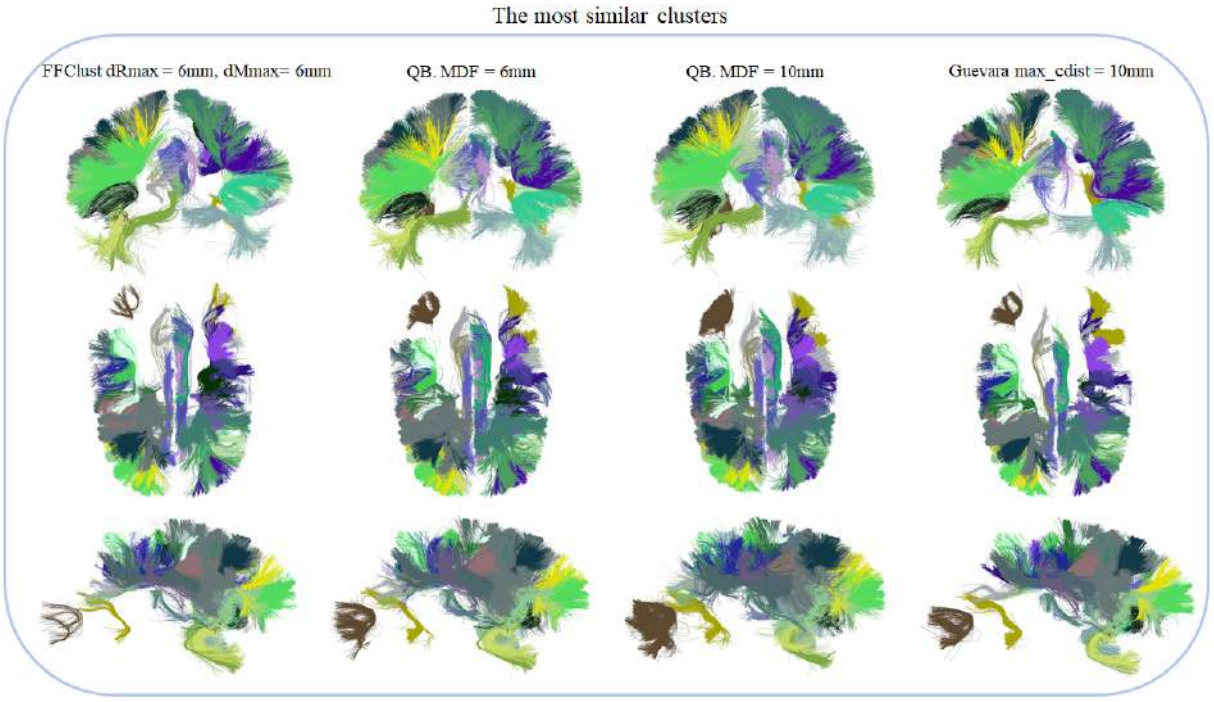


Figure 3.12: Comparison of most similar clusters. In the comparison appears the 94 most similar clusters taking as a reference a SWM atlas [50]. We apply *FFClust* with d_{Rmax} = 6 mm and d_{Mmax} = 6 mm, together with *QB* with MDF = 6 mm, and MDF = 10 mm, and *Guevara* max_cdist = 10 mm. Display of coronal, axial and sagittal views.

of *FFClust*, where each of the steps of the algorithm is executed using five threads in the first step and 12 in the other steps. The methods used for comparison are: *QuickBundles* with MDF of 6 mm (*QBmdf6*), with MDF of 10 mm (*QBmdf10*), *QuickBundlesX* with MDF of 6 mm and 10 mm (*QBX*). However, we did not include the *Guevara* [55] method for this experiment because its execution times are longer than two hours for any dataset, far exceeding the execution times of algorithms such as *QB* and *FFClust*, which were designed to be efficient. The evaluation was performed using subjects from the ARCHI database with a number of streamlines varying from 330,000 to 2,729,000.

Table 3.1 shows the execution times in seconds for all considered methods. Figure 3.13 shows the execution times in logarithmic scale of the methods. *FFClust*-seq and *FFClust*-par provide the best execution times. It is at least an order of magnitude faster than *QuickBundles*. Also, we can see the trend of the *QB* algorithm, where the execution time increases when MDF is set to 6 mm, instead of using the *QB* default

value of 10 mm. In addition, the speed up of *FFClust-par* has been calculated according to the number of fibers used. The speed up equation is given by $S_N = T_1/T_N$, where T_N is the parallel time, T_1 is the time in a processor and S_N is the speed up algorithmic. Table 3.2 shows the speed up for each number of fibers.

Total fibers	FFClust-par time (s)	FFClust-seq time (s)	QBmdf6 time (s)	QBmdf10 time (s)	QBX time (s)
330K	9.92	28.13	334.03	108.77	32.24
659K	24.61	70.09	3,594.02	1,031.37	217.41
955K	45.96	119.84	8,032.76	2,318.02	399.61
1296K	76.82	181.08	13,404.17	4,186.45	604.14
1634K	94.75	250.13	21,813.78	7,253.64	850.67
1945K	125.91	317.59	30,394.23	10,002.94	1,125.90
2338K	193.58	460.76	50,946.22	14,312.24	1,540.61
2729K	264.30	623.13	71,243.69	19,861.15	2,194.00

Table 3.1: Execution times in seconds for *FFClust*-par, *FFClust*-seq, QBmdf6, QBmdf10 and *QBX*, varying the number of streamlines in the range of 330,000 and 2,729,000.

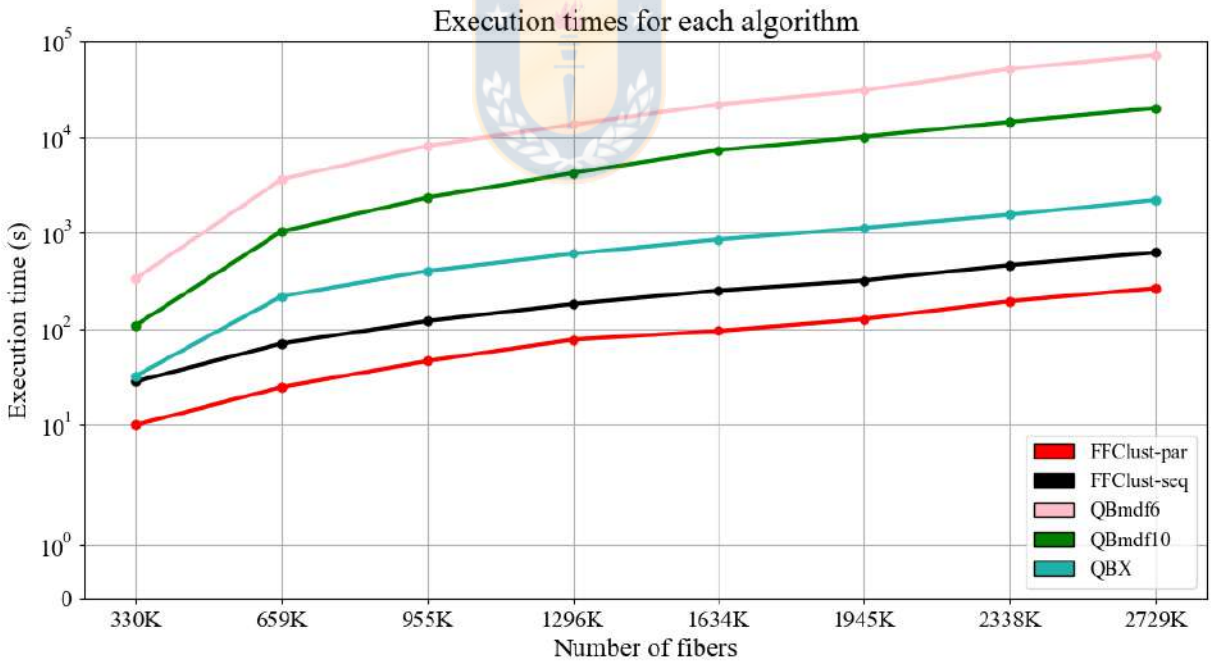


Figure 3.13: Execution times for *FFClust*-par, *FFClust*-seq, QBmdf6, QBmdf10, and *QBX*. The streamlines range is from 330,000 to 2,729,000.

Speed up	2.84	2.85	2.61	2.36	2.64	2.52	2.38	2.36
Total fibers	330K	659K	955K	1296K	1634K	1945K	2338K	2729K

Table 3.2: Speed up for *FFClust*-par, varying the number of streamlines in the range of 330K and 2,729K.

Execution time complexity

FFClust's time complexity is made up of four stages, these are:

1. The time complexity of the *STEP 1: Building point clusters* is based on MK. Its upper bound is $\mathcal{O}(tKDN)$, where N is the number of elements, D is the dimensionality of the elements, K is the number of clusters and t is the number of iterations or until convergence. *FFClust* sequential implementation executes the MK algorithm on five streamline points, as *FFClust* uses five points. *FFClust* parallel implementation executes MK on each streamline point in parallel.

2. The *STEP 2: Generating preliminary streamline clusters* of the algorithm builds a dictionary data structure to map the clusters of points to streamline clusters. It is the fastest step of all and its complexity is $\mathcal{O}(N)$. This step is also parallelized in the *FFClust* parallel implementation.

3. The time complexity of the *STEP 3: Reassigning small preliminary streamline clusters* is $\mathcal{O}(|S_s| \times |S_l|)$, where both $|S_s| << N$ and $|S_l| << N$ because they are centroids not fibers. The parallel version of this step is described by Vazquez et. al [128] and it is included in *FFClust* parallel implementation.

4. Finally, last *STEP 4: Merging candidate streamline clusters* time complexity is determined by the maximal clique algorithm. Although the problem is NP-hard, when using parameterized complexity in sparse graphs, algorithms can be near linear [37]. This step is also parallelized in the *FFClust* parallel implementation.

3.1.4. Discussion

We propose *FFClust*, a new fast clustering algorithm for large whole-brain tractography datasets of the brain's white matter. We compare our clustering results with the

state-of-the-art clustering using the *QuickBundlesX* (*QBX*) [43], *QuickBundles* (*QB*) [42] and *Guevara* [55] methods.

After tuning the parameters of all methods, our experimental evaluation shows that *FFClust* identifies homogeneous clusters with a moderate maximum intra-cluster Euclidean distance and still it is able to find large clusters. Using the DB index as a metric of clustering quality, we found that *QB*, using MDF of 6 mm, is the best and *FFClust* is the second best, whereas using *QB* with the default distance MDF of 10 mm its DB index is less competitive. Based on the DB index, *QBX* does not improve the quality of *QB*.

We also compare the resulting clusters using as reference bundles connecting the postcentral (PoC) and precentral (PrC) gyri of a superficial WM bundle atlas. The results show that only the *Guevara* method and *FFClust* are able to find all bundles with a small error, i.e., with fewer fibers connecting surrounding regions. On the other hand, *QB* and *QBX* are able to identify the bundles, but with higher error. This analysis was performed to evaluate the potential applications of *FFClust*. It was designed to create compact clusters, with the purpose to be used in applications like bundle segmentation [53, 72, 128] and inter-subject analyses for the creation of WM bundle atlases [53, 106, 50, 142] and connectivity-based parcellations [89, 80]. However, for some applications, bigger clusters would be more suitable. For example, if the main goal is the segmentation of large anatomical bundles, large clusters would be more useful, or easier to handle. On the other hand, if the clusters will be used for the study of short association bundles, small clusters are more suitable, since large clusters could connect neighboring anatomical regions. Another application is the diffusion-based parcellation, where, the size of the clusters depends on the size desired for the final parcels (or the number of parcels). Hence, the utility of each method must be evaluated by the user, in function of the particular requirements of the analysis to be performed.

Another advantage of *FFClust*, in comparison with the state-of-the-art methods, is the improvement in execution time. *FFClust* is at least an order of magnitude faster than *QB*. For instance, with a subject of 1 million of fibers, the sequential version of *FFClust* takes 1.99 minutes and its parallel implementation takes 45 seconds. *QB*, on

the other hand, takes 2.2 hours using its best quality configuration (MDF of 6 mm). Also, at the expense of decreasing quality for some applications, using *QB* with its default value for MDF of 10 mm still takes 38 minutes, and its optimized version, *QBX* takes 6.6 minutes, which makes *FFClust* parallel implementation at least 50.4 times faster than *QB* and 8.6 times faster than *QBX*.

In summary, in addition to its reduced computation time, *FFClust* presents the advantage of producing good quality clusters, with a compact shape and without frizzy ending points. This feature will enable a more detailed study of brain connectivity, in particular, short association fibers, and could enable the development of diffusion-weighted parcellations. We notice that *FFClust* provides similar results to the *Guevara* method. In particular, they achieve a similar DB index score, and both are able to identify all bundles in the segmentation application. Moreover, both provide the lowest error percentages in the quality of such identified bundles. However, *FFClust* is more simple than the *Guevara* method, requires fewer parameters and it is faster. Then, we suggest that *FFClust* can be used in similar applications where the *Guevara* has been successfully used [53, 50].

3.2. Labeling of superficial white matter bundles

This section describes a method of labeling of superficial white matter bundles on a group of subjects. First of all, we discuss the related work on fiber labeling. Then, we explain the method based on intra-subject clustering. In addition, we present two algorithms to perform the inter-subject labeling. One is based on clustering and the other is based on matching. Then, we present the experimental results for the intra- and inter-subject methods. Finally, we discuss the fiber labeling, the methods developed, and the results.

3.2.1. Related work

The preferred technique to non-invasively study structural brain connections is diffusion weighted Magnetic Resonance Imaging (dMRI), based on the measurement of water molecules movement [74, 6]. Diffusion tractography estimates the main white matter (WM) tracts, obtaining a set of 3D paths, called streamlines or fibers [7]. Tractography datasets contain a large number of streamlines, some of which represent the trajectory of known WM bundles, with anatomical meaning. Such bundles have been described in the literature by neuroanatomists [17], and have been validated with other techniques like post-mortem dissections [85]. However, these datasets also include artifacts or false positives, some of which can occur systematically across subjects [82]. Hence, tractography datasets can be analyzed to extract or segment known WM bundles, which requires the inclusion of anatomical information in the processing. One strategy can be the manual delineation of regions of interest (ROIs) in the cortex, and the extraction of fibers connecting a pair of cortical regions for a specific bundle. This analysis has been recently used to study short association bundles [18]. The bundle segmentation can be performed automatically by applying an atlas of gray matter and WM ROIs, and then using anatomical descriptions of the bundles to segment fibers connecting or passing through specific ROIs [131].

Automatic methods based on ROIs allow an easy modification or addition of bundle extraction rules, but do not include an analysis based on the trajectories of the fibers as a whole. Another strategy is based on clustering to group fibers with similar shape

and position, commonly based on a fiber pairwise distance measure that considers the Euclidean distances between the corresponding points (or closest points) of the two fibers. To extract anatomical bundles, some methods use a clustering algorithm and an atlas embedding anatomical bundle information [95, 94, 44]. Also, other simpler algorithms have been implemented to extract bundles based on a multi-subject bundle atlas [53, 72, 128]. Several atlases have been created to represent main deep white matter (DWM) bundles, which have been well described by anatomists and are very stable across subjects [95, 53, 107], i.e., present high similarity and can be found in all the subjects on medium- to high-quality databases. However, there exist several WM fiber bundles still unknown or not sufficiently described, because of their higher inter-subject variability and fewer reproducibility [49]. This is the case of short association bundles, where only a few works have been focused on their description for the whole-brain [50, 106]. Short association fibers are placed immediately underneath the gray matter of the cortex and connect adjacent or close gyri. They can present different sizes, where the shortest ones are the nearest to the cortex and present the typical U-shape, due to their closeness to the walls of the convolution depression [86]. Their description is still incomplete [49], however, post-mortem dissections have been used to validate the largest and reproducible bundles [18, 66]. Superficial white matter (SWM) fibers can be studied using exploratory fiber clustering methods that aim to detect fiber tracts without having any reference to the start or end of WM fibers [93]. This type of algorithm, applied to a whole-brain tractography dataset, generates a set of fiber clusters representing the main WM connections in the analyzed brain. In the case of the works in [50, 106], superficial white matter bundle atlases were obtained using different methods based on fiber clustering and the addition of anatomical information. Also, a recent work found a great amount of SWM bundles [142], but those were not labeled, requiring a posterior analysis for their study. Hence, existing methods have been focused on finding reproducible bundles across subjects, but not on the development of an automatic labeling of individual or inter-subject SWM clusters. Whole-brain fiber clustering methods, applied to individuals or to a population of subjects, do not return directly the identification of the obtained clusters, e.g., information about the anatomical areas connected by the fibers and their relative position in the

cortex. Such identification or labeling could be very useful for the study of the human brain connectome in individuals and different populations. The labeled clusters could then be used to perform detailed analyses of known bundles, i.e., subdivisions of the main bundles, and also of unknown fascicles, such as short association bundles. Furthermore, multi-subject analyses could be applied to create new bundle atlases.

Superficial white matter bundles are more variable across subjects and more susceptible to noise than deep white matter bundles, due to their smaller size and location in the brain, which presents partial volume effect. Hence special attention must be given to the diffusion local model and tractography methods. Due to the improvement of imaging quality, i.e., more signal-to-noise ratio, higher resolution, better distortion correction methods, between others, have allowed a better reconstruction of short association bundles. The most stable bundles can be reconstructed using deterministic tractography, with adapted parameters, in particular, a larger number of streamlines and a low fractional anisotropy (FA) threshold or an adapted propagation mask, to prevent the removal of voxels in the superficial white matter [49].

We propose a method that automatically labels the clusters of a subject obtained from an intra-subject clustering, based on the regions connected by the clusters. This information is based on a cortical surface mesh, labeled with the *Desikan-Killiany* atlas (35 gyri per hemisphere). Direct correspondence between subjects is obtained for the connected anatomical regions. Furthermore, within each region, the clusters on individuals are labeled following and ordering criterion. Moreover, we apply two strategies for inter-subject cluster labeling. First, a matching method is implemented based on the *Hungarian* algorithm to find correspondence between bundles across subjects and subsequently apply a labeling that gives the same names to bundles identified in several subjects. Also, a clustering algorithm is applied to group similar bundles on a set of subjects and perform the labeling across them. While the matching algorithm finds the best matching for single bundles, the clustering may group similar bundles on some subjects and identify similar bundles (or groups of bundles) across subjects. Both inter-subject implementations are fast, taking, respectively, about 96 and 9 s, over

20 subjects. The performance of both implementations was compared in terms of reproducibility and inter-subject bundle distance.

3.2.2. Methods for intra- and inter-subject labeling

Automatic labeling of SWM bundles

To perform the automatic labeling of bundles of superficial white matter, a method consisting of four stages (see Figure 3.14) was developed, these are: *(1)* fiber clustering, *(2)* cluster filtering, *(3)* fiber intersection and *(4)* cluster labeling.

Stage 1: Fiber clustering

This first stage performs an automatic clustering of a whole-brain tractography dataset, which returns a set of clusters of similar fibers (see Figure 3.14-(1)). The clustering method [125] is an improved version of an algorithm proposed in [109]. To apply the clustering, fibers must be first resampled with 21 equidistant points, as in [55, 53]. The method consists of 4 steps: *(1) Building clusters on a subset of fiber points*, where mini batch K-means is applied in parallel on a subset of fiber points, obtaining groups of points; *(2) Generating preliminary clusters*, which groups fibers sharing the point cluster labels from the previous step; *(3) Defining candidate clusters by reassigning small preliminary clusters*: reassigns small clusters to larger clusters based on a maximum distance threshold between clusters; *(4) Computing final clusters by merging close candidate clusters*: merges close clusters that share the central label obtained from step 1, according to a criterion of maximum Euclidean distance between clusters. Finally, a representative fiber of each cluster is selected, as its centroid, and resampled with 21 equidistant points.

Stage 2: Cluster filtering

The second stage automatically filters out the small and long fiber clusters (see Fig. 3.14-(2)). Clusters are denoted as C_i, with $i = 1, ..., n$ the index of the cluster. The filter receives a minimum size of the cluster $min_{nf}(C_i)$ (number of fibers), to remove small fibers, and a minimum $min_{len}(C_i)$ and maximum cluster length $max_{len}(C_i)$ to only keep

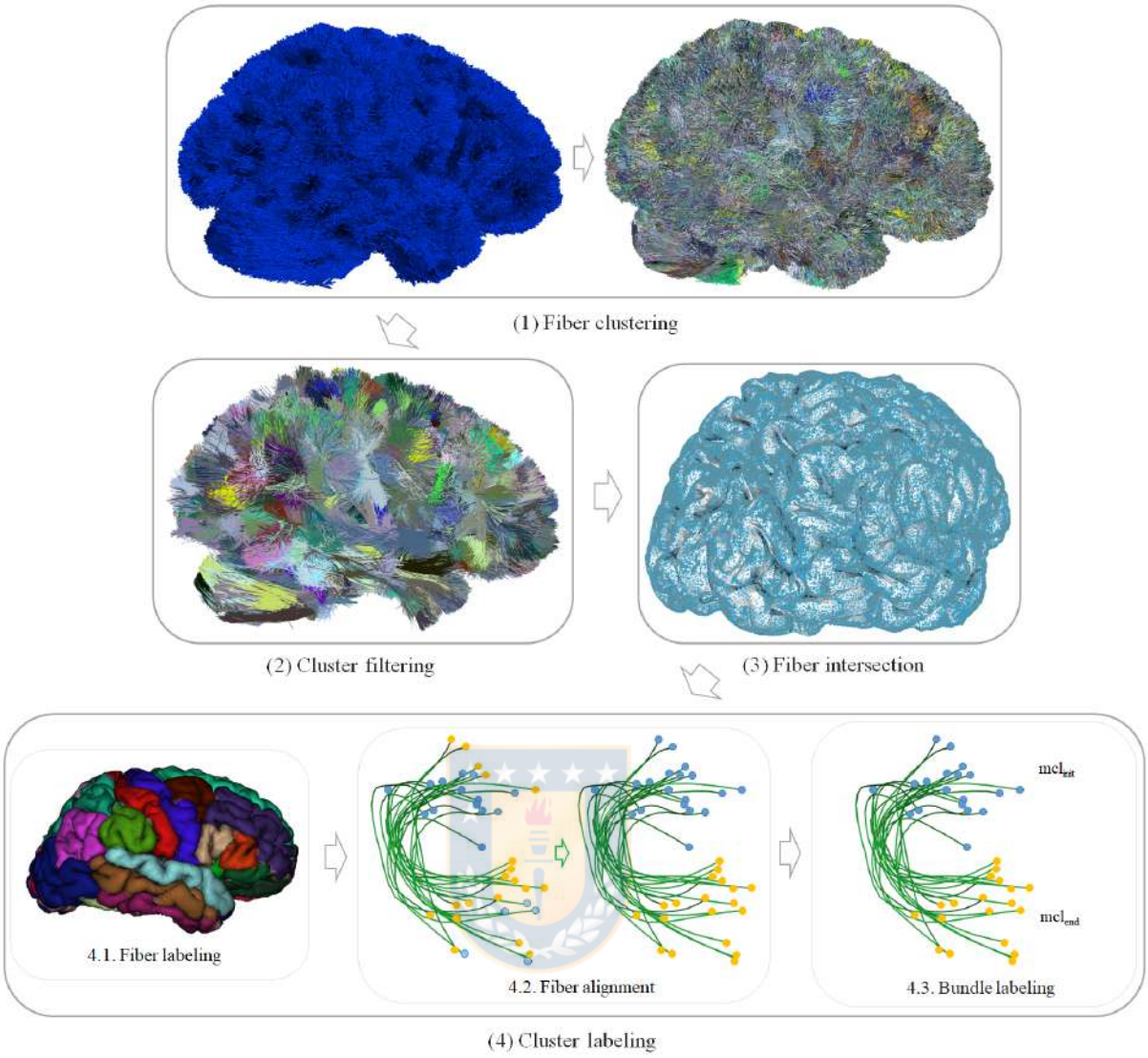


Figure 3.14: Schematics of the labeling method. Stage 1: Fiber clustering. Performs the clustering of the entire tractography. Stage 2: Cluster filtering. Filters out the small clusters and only keep the short bundles, obtained in the previous stage. Stage 3: Fiber intersection. Calculates the fiber bundle intersection with the cortical mesh. Stage 4: Cluster labeling. Renames the clusters based on the two connected regions of the cortex and their position.

short fibers within a reasonable range. The length of each cluster is measured using the Euclidean distance between two adjacent points of the cluster's centroid.

Stage 3: Fiber intersection

This step automatically calculates the intersection of the fibers with the cortical mesh, based on the algorithm proposed in [115] (see Fig. 3.14-(3)). The method first performs a subdivision of $3D$ space into 1.5 mm size cells, which speeds up searches in the mesh. Next, each fiber endpoint is projected one point backward and two points forward to extend the search along the fiber trajectory on ending points and avoid missing intersections. All cells that include these points and their neighboring cells are selected. Finally, the intersection point of each fiber extremity with the cortical mesh triangles is determined using Möller-Trumbore equation [87], based on the analysis of the triangles contained in the selected cells. The intersection algorithm is given by Equation 3.3:

$$O + tD = (1 - u - v)V_0 + uV_1 + vV_2 \tag{3.3}$$

where (u, v) are the exact coordinates of the intersection with the mesh triangle, V_0, V_1 and V_2 are the vertices of a triangle, t is the direction, D is the normalized ray trajectory, and O is the ray from the point of origin.

Finally, for each hemisphere, the indexes of the start (Tri_{init}) and end triangles (Tri_{end}) where the fiber intersects the mesh are obtained, as well as the coordinates of the two exact points of the initial ($Point_{init}$) and final ($Point_{end}$) intersection.

Stage 4: Cluster labeling

This stage performs an automatic labeling of all the clusters based on the cortical regions they connect, by using a cortical ROI atlas. For testing, we use the *Desikan-Killiany* atlas [31], consisting of 35 regions (gyri) per hemisphere. We use the cortical meshes, containing a list of vertices and triangles, and a labeling file, containing the cortical region label of each mesh vertex.

- *4.1. Fiber labeling*. First, for each cluster, each fiber is labeled according to the triangle of the mesh that the fiber intersects, based on the region labels of the triangle vertices. The labeling of each triangle is defined as the most repeated label between its three vertices (see Fig. 3.14-(4.1.)).

- *4.2. Fiber alignment*. Next, the fibers require to be aligned, since, in a tractography dataset, there is no unique direction and fibers can be stored in direct or inverse direction. Since after the clustering the fibers are grouped on compact clusters, these can be aligned so that the starting and ending points have the same orientation in a cluster (see Figure 3.14-(4.2.)). Hence, the fibers of a cluster are oriented based on the cluster centroid. To perform the alignment, we verify if the fibers are inverted with respect to the centroid. We denote f_i as the fiber i of the bundle, with $i = 1, ..., n$, and the centroid of the bundle as c_j, with $j = 1, ..., m$. Then, the Euclidean distance (d_E) is calculated between the first point of the fiber (f_{i1}) and both endpoints of the centroid (c_{i1} to c_{j21}). If $d_E(f_{i1}, c_{j1}) > d_E(f_{i1}, c_{j21})$, the fiber is inverted by flipping its fiber points.

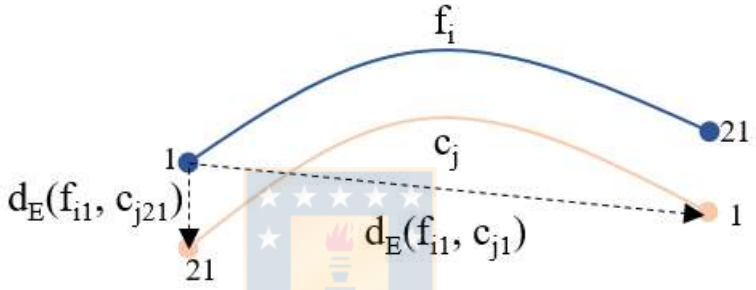


Figure 3.15: Fiber bundle alignment with respect to its corresponding bundle centroid. The Euclidean distance (d_E) is calculated between the first point of the fiber (f_{i1}) and both end points of the centroid (c_{i1} to c_{j21}). If $d_E(f_{i1}, c_{j1}) > d_E(f_{i1}, c_{j21})$, the fiber is reversed.

- *4.3. Bundle labeling*. Next, each cluster (or bundle) is labeled according to the most connected regions. For each bundle, the labels of both bundle extremities, i.e., the beginning and end of each bundle, are processed separately. The most common label (*mcl*) for bundle start (mcl_{init}) and end (mcl_{end}) is determined and used to name each bundle, with format mcl_{init}-mcl_{end} (see Figure 3.14-(4.3.)). For instance, a bundle connecting the postcentral and precentral anatomical regions will have the label *PoC-PrC*. Note that several bundles may connect the same pair of anatomical regions (gyri), as each cluster extremity only intersects a portion of a gyrus. Then, an order is assigned to each pair of bundles defined by the index of the regions in the cortical region label file. For example, PrC has index 24 and

PoC has index 22, then, the bundle is named as *PoC-PrC*. Subsequently, bundles with inverted names are flipped. For example, all the bundles labeled with *PrC-PoC* are inverted and named as *PoC-PrC*. Finally, as several bundles may have the same name, but connecting different specific sub-regions of the gyri. These are labeled with an extra index, indicating the relative position according to an axis in the brain in MNI space. The intersection points of all the bundle centroids in a gyrus are sorted based on a spatial coordinate (x, y, or z), in ascending or descending order. By default the order is ascending according to y axis, i.e., from the bottom-up.

Inter-subject labeling

In this section, two methods are presented to obtain automatic group-wise bundle labels of superficial white matter bundles, among the subjects of a population. Intra-subject labeling, presented in the previous section, labeled the bundles of a subject based on the connected brain regions individually, and an order based on the coordinates, producing a certain similarity between the subjects' bundles. However, this was not the main objective of the intra-subject labeling method and the correspondence between subjects can be improved by applying inter-subject methods. The methods used to perform this processing are a matching algorithm and a clustering algorithm. To apply these methods, the tractography datasets need to be in a common space. All subjects were aligned to Talairach space using the affine transformation of the database, and then a rigid transformation to MNI space. Both methods use a maximum distance threshold to define the similarity of bundles across subjects. The inter-subject labeling renames all the bundles according to the correspondence found in the analyzed group of subjects. The part of the name related to the connected cortical regions is kept, but the index is assigned again to all the bundles. Bundles found similar in several subjects will have the same label.

Matching algorithm for inter-subject labeling

The aim of this step is to apply a matching [88] for finding a correspondence between similar bundles in the different subjects. Bipartite matching algorithms find correspondence between pairs of elements from two distinct sets. These algorithms are based on graph theory to find connections in two sets of vertices, where vertices in one set must match with vertices in the other set [134].

A well-known algorithm for a bipartite matching problem is the *Hungarian* algorithm, that solves the minimum weight matching, i.e., the minimum distance between vertices from the two sets, A and B [40]. Being *V* the number of vertices from the two sets, the algorithm receives a matrix M, containing the distances between the vertices from the two sets. In our application, *V* is the total number of bundles from a pair of subjects and matrix M contains the distances between the bundle centroids from the two subjects, being one set represented in the rows, and the other in the columns. The original algorithm performs a perfect matching, i.e., each vertex (or bundle) in set A is matched with a vertex in set B, which requires an equal number of vertices in both sets and produces a square matrix M. Our problem presents a different number of vertices in each set, as different number of bundles are found in each subject. Hence, we used an adapted algorithm that performs the analysis over non-squared matrices and leaves unmatched the most dissimilar elements.

More formally, each element $M[i, j]$ in matrix M represents the distance between bundle i of set A (subject A) and bundle j of set B (subject B), being the cost of matching between the two vertices. The result is an assignment of the elements of set A with set B by using the minimum assignment cost. The distance used is the minimum average direct-flip distance (MDF) between two pairs of fibers [42] (Equation 3.4), a distance commonly used for tractography fiber comparison. This distance calculates the mean Euclidean distance between corresponding points of a pair of centroids or fibers. To be used, fibers are resampled with a defined number of *K* equidistant points (21 in our case). Since fibers can be ordered in memory in opposite directions, the distance must be calculated with fibers in both directions (direct and flipped order), and then the minimum value must be selected, which will correspond to the correct order.

$$d_{direct}(a,b) = d(a,b) = \frac{1}{K}\sum_{i=1}^{K}|a_i - b_i|$$
$$d_{flipped}(a,b) = d(a,b^F) = d(a^F,b)$$
$$MDF(a,b) = min(d_{direct}(a,b), d_{flipped}(a,b)) \quad (3.4)$$

The *Hungarian* algorithm has a complexity of $\mathcal{O}(V^3)$, however, as we perform the analysis separately for each pair of anatomical regions, the analyzed datasets are small with low execution time.

The matching algorithm applied to inter-subject bundle labeling first performs a bundle pre-processing. For each subject, previously labeled bundles with the intra-subject labeling, are separated into different groups depending on the pair of anatomical regions they connect. Then, for each region a map is created, whose key is the subject and the value is a list of the bundles that belong to the subject and region. For instance, for region *PoC-PrC* the bundles for *Subject001*, will be stored in the key-value pair: *Subject001: [bundle0, bundle1, ..., bundleN]*. Next, the algorithm consists of four steps:

- *Step 1*. Once the maps of all the regions are obtained, the bundles of each region are processed sequentially. First, the subjects are ordered from highest to lowest, based on the number of bundles they contain. For each bundle, its centroid is calculated using the mean of the streamline point coordinates.

- *Step 2*. The analysis begins with the first subject on the list as a reference subject. This subject is compared with each of the following subjects using the *Hungarian* algorithm, receiving as input the distance matrix. This returns a matching based on the distance of one bundle centroid with another. The *Hungarian* algorithm receives as input the matrix of distances, which are calculated using the MDF distance (Equation 3.4) between all the bundle centroid pairs of all subjects. For each bundle, the algorithm returns the bundle that best matches it, according to the solution of the minimization problem. However, the distance between a

pair of bundles could be higher, hence, the method evaluates all the distances between the matched bundle centroids and only keeps the pairs of bundles in which distances do not exceed the established maximum distance threshold. This avoids the assignment of distant bundles, leaving them unassigned. Bundles that match each other are labeled with the same indexes, based on the label of the reference subject. For example, for two corresponding bundles, they would be called *PoC-PrC_0* even if they are from different subjects.

- *Step 3*. Two cases can happen with unassigned bundles: *(i) Bundles of the reference subject*. They are not similar to any other bundle in the dataset and they are labeled with a new index. *(ii) Bundles of the remaining subjects*. The bundles are stored. In the iteration in which the subject is taken as a reference, comparisons are made again with the rest of the subjects.

- *Step 4*. Repeat *Step 2* with the unassigned bundles of the following subjects, taking as reference the next subject in the list with unassigned bundles.

Figure 3.16 shows an example scheme of the algorithm for three subjects and the bundles connecting *PoC-PrC* regions. Each circle corresponds to a bundle. First, the subjects are ordered from highest to lowest number of bundles (see Figure 3.16-(1.)). Second, *Subject001* that is being compared with the rest is the reference. This step receives a distance matrix between two subjects' bundles (see Table 3.3), and then, the bundles of *Subject001* and *Subject002* are matched using the *Hungarian* algorithm. In the example, it matches only the first two bundles, leaving an unassigned bundle in *Subject002*, which will be saved for later comparison (see Figure 3.16-(2.)). Third, *Subject001* continues to be compared with the remaining subjects, in this case, with *Subject003*, which leads to the matching of bundles 1 and 2. In *Subject003* there remains an unassigned bundle (see Figure 3.16-(3.)). Finally, once *Subject001* is compared with all subjects, the reference subject becomes the next one, in this case, *Subject002*. Then, unassigned bundles are compared, for example, *Subject002* is compared with *Subject003* and the two unassigned bundles are matched (see Figure 3.16-(4.)). The bundle with the highest reproducibility in the example is the 1, since it is present in all

subjects, and would be named as *PoC-PrC_1*, according to the label of the first reference subject of the bundle. We used the implementation of the *Hungarian* algorithm available at scipy library[1].

	Subject002_b0	Subject002_b1	Subject002_b2
Subject001_b0	8 mm	5 mm	7 mm
Subject001_b1	5 mm	9 mm	10 mm
Subject001_b2	12 mm	13 mm	14 mm
Subject001_b3	18 mm	19 mm	17 mm

Table 3.3: Example of distance matrix between subjects' bundles. Step 2 uses the Hungarian algorithm which receives as input the distance matrix (in mm) between the PoC-PrC bundles for two subjects. This returns a matching based on the distance of one bundle centroid with another.

Clustering algorithm for inter-subject labeling

Clustering is an unsupervised classification method, which groups similar elements into subsets called clusters. Each cluster is made up of elements that have similar characteristics, however, the elements of each cluster are different from each other [138].

The clustering method used to group the clusters is a well-known fiber clustering algorithm called *QuickBundles (QB)*[2] [42]. Before applying *QB*, we apply the same bundle pre-processing as for the matching, to create a map for each pair of regions, with the bundles of each subject. Next, the *QB* algorithm is performed sequentially to each pair of regions. For each pair of regions and all the subjects, the centroids of all clusters are calculated. The algorithm is applied to the complete set of clusters, i.e., from all subjects for the pair of regions. Once the inter-subject clusters are obtained, all intra-subject clusters belonging to the same inter-subject cluster are labeled with the same label. If several clusters of the same subject belong to the same inter-subject cluster, they are merged.

Figure 3.17 shows an example scheme for *QB* application to three subjects on

[1]https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.optimize.linear_sum_assignment.html

[2]https://dipy.org/documentation/1.0.0./examples_built/segment_quickbundles/

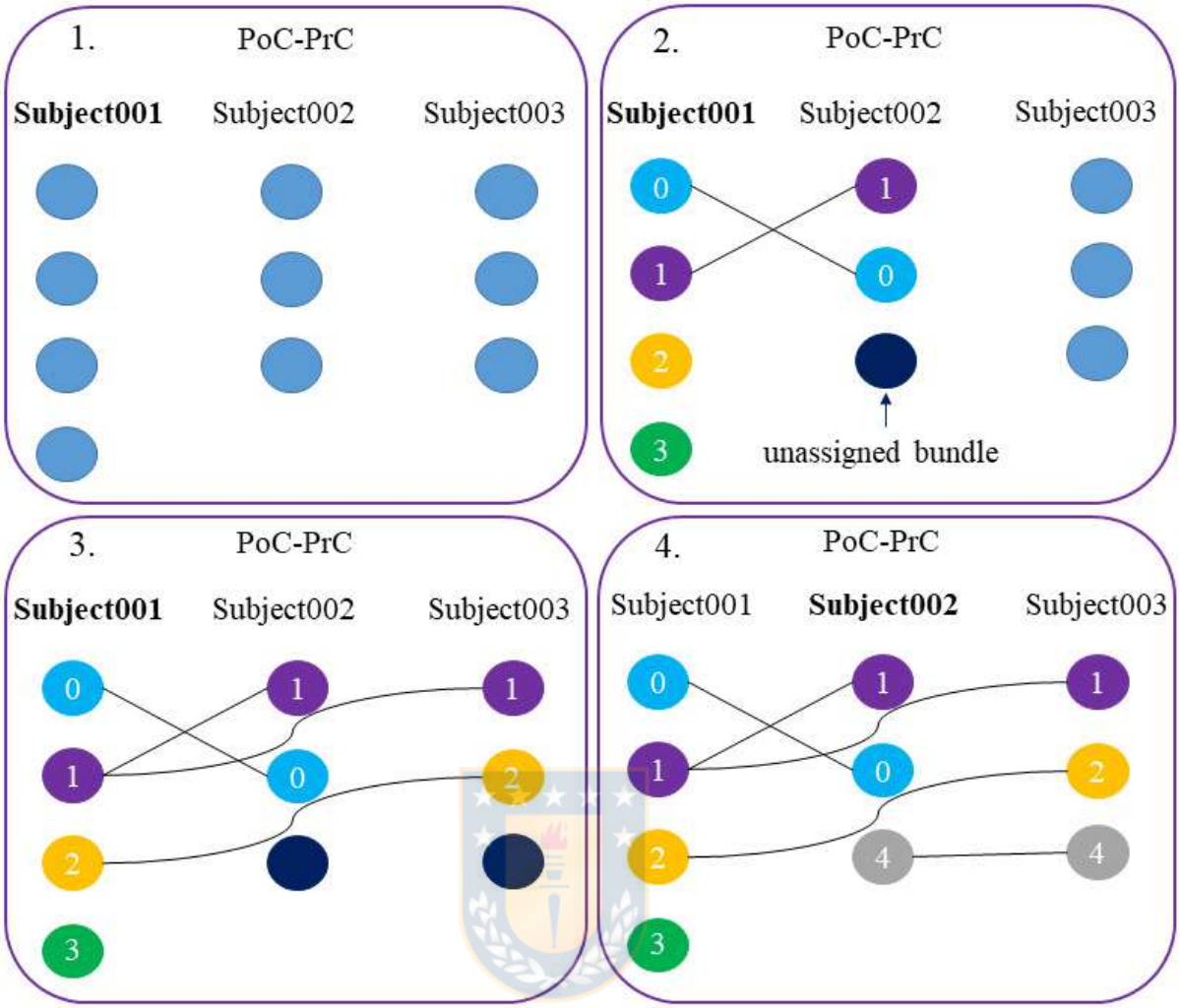


Figure 3.16: Schematics of the *Hungarian* algorithm for inter-subject labeling of bundles connecting PoC-PrC regions. First, the bundles are ordered from highest to lowest number of bundles. Second, the reference subject, *Subject001*, is compared to *Subject002*, leaving unassigned bundles. Third, it continues comparing to the rest of the subjects. Finally, the reference passes to the next subject with unassigned bundles, *Subject002* and these are compared with the rest of the subjects. This process is repeated until all subjects are analyzed.

the *PoC-PrC* regions. First, it starts with the computation of all the cluster centroids. Unlike matching, in this case, it is not necessary that the clusters are ordered (see Figure 3.17-(1.)). Second, the *QB* method is applied to all clusters, generating inter-subject clusters. Bundles within inter-subject clusters are labeled with the same name (see Figure 3.17-(2.)). Finally, the clusters of a subject that are in the same inter-subject

clusters are merged (see Figure 3.17-(3.)). The clusters with the highest reproducibility are 0 and 3 since they appear in all subjects, whose tags would be: *PoC-PrC_0* and *PoC-PrC_3*. In addition, there may be some loose cluster, which will be individually labeled with another index.

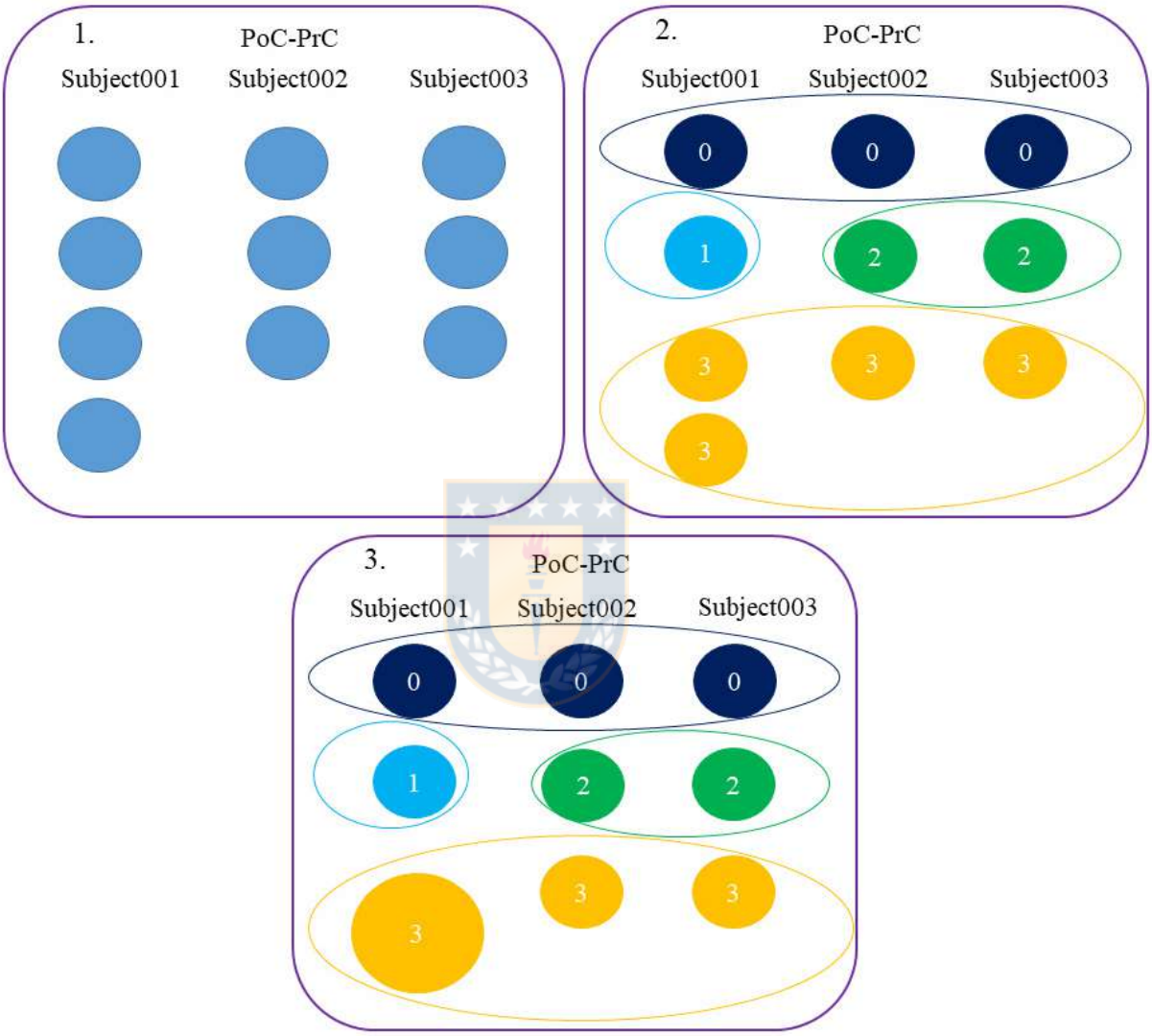


Figure 3.17: Schematics of the *QB* algorithm for labeling inter-subject bundles for PoC-PrC regions. First, the cluster centroids are computed. Second, *QB* is applied to all the intra-subject clusters, to obtain inter-subject clusters. Bundles belonging to an inter-subject cluster are labeled using the same name. Finally, clusters of the same subject that belong to the same inter-subject cluster are merged.

3.2.3. Experimental results

The experiments were executed on a computer with 4-core Intel Core i5-8250U CPU running at 1.60 GHz, 6MB of cache and 8GB of RAM, using Ubuntu 18.04.2 LTS with kernel 4.15.0-64 (64 bits). The programming language used to develop almost all stages is Python 3.6.

Twenty subjects were used for analysis, tractographies, cortical meshes and labels according to the *Desikan-Killiany* atlas. The intra-subject labeling was applied to all the subjects. First, the results for intra-subject labeling performed on four subjects are shown and analyzed. Next, the two inter-subject labeling methods, matching and clustering, were applied to the 20 subjects, using a set of distance thresholds, ranging from a conservative value of 10 mm to a moderate value of 21 mm. We use the minimum average direct-flip (MDF) distance (Equation 3.4). It is defined as the mean Euclidean distance between corresponding points of two centroids. We use centroids resampled with 21 equidistant points. Since centroids (or fibers) can be ordered in memory in opposite directions, the distance must be calculated with centroids in both directions (direct and flipped order), and then the minimum value must be selected, what will correspond to the correct order. The distance threshold defines the degree of similarity considered between bundle centroids. The smaller, the more restrictive the analysis will be.

The reproducibility of the bundles was evaluated by counting the number of subjects that had each inter-subject bundle. The quality of the labeling was evaluated using a distance measure between bundles across subjects (MDF). Also, heatmaps are shown to have an insight on the reproducibility and variability in terms of the number of fibers of the most reproducible bundles, for a restrictive distance threshold of 12 mm. Finally, some examples of bundles are displayed for a visual inspection of the results. *Hungarian* algorithm led to a high correspondence, but low reproducibility for all the thresholds, with 96 seconds of execution time. *QuickBundles* led to better correspondence and reproducibility with short execution time, of only 9 s. Hence, the whole processing for the inter-subject labeling over 20 subjects takes on average 1.17 h. In the following, we detail the results.

Intra-subject labeling

The intra-subject labeling was applied to the 20 subjects. The input data are the tractography datasets of each subject and a cortical mesh where each vertex is labeled with the corresponding gyri (according to *Desikan-Killiany* atlas). The method consists of four stages (see Figure 3.14). First, a fiber clustering *(stage 1)*, is applied to the whole-brain tractography dataset, which returns a set of clusters of similar fibers. Secondly, a cluster filtering *(stage 2)* is performed, where small and long fiber clusters are discarded, keeping only fibers on a reasonable range for short association fibers. Next, a fiber intersection *(stage 3)* is executed, to determine the intersection of the fibers with the cortical mesh. Finally, a cluster labeling *(stage 4)* is applied, that labels each cluster according to the two most connected cortical regions, and an index indicating its relative order in the region.

The fiber clustering *(stage 1)* led to about 43,000 clusters per subject. For fiber filtering *(stage 2)*, a filter with a minimum cluster size of $min_{nf} = 10$ fibers is used to discard small clusters. Also, a minimum cluster length of $min_{len} = 30$ mm and a maximum cluster length of $max_{len} = 80$ mm were employed to discard clusters that are too short or too long, leading to an average of 1100 clusters per subject. The filtering values are similar to those previously used [50, 106]. After applying fiber intersection and cluster labeling stages *(stage 3 and 4)*, the clusters of each subject were labeled according to the pair of anatomical regions connected by each bundle, and the position based on ascending order of y-axis on MNI space (default configuration), i.e., from the bottom-up.

An example of the relative ordering for intra-subject bundle labeling is presented in Figure 3.18. We can appreciate that bundles connecting *postcentral (PoC)* and *precentral (PrC)* regions are ordered according to y-axis in ascending order (from the bottom-up). Bundles are ordered according to the *PoC* parcel since it is indexed before in the *Desikan-Killiany* atlas.

Even though the method performs only an intra-subject analysis, a degree of correspondence between the four first subjects can be found, according to their relative

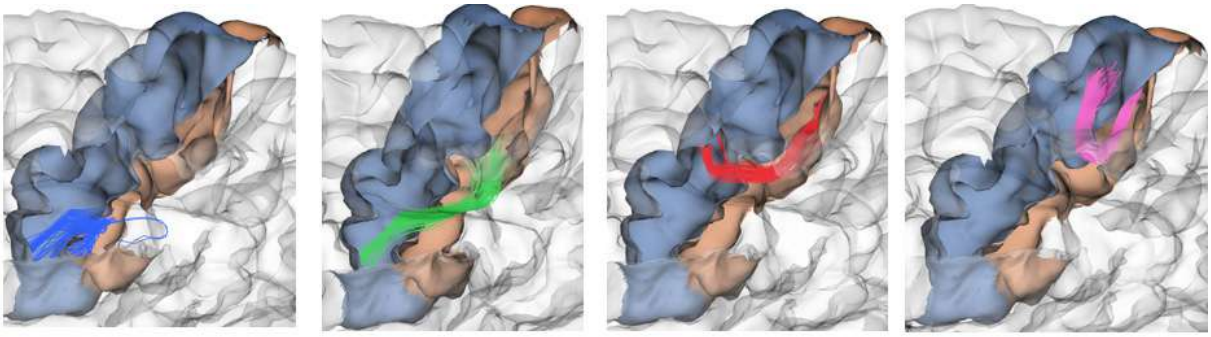


Figure 3.18: Bundles connecting right *PoC* and *PrC* regions. Example of bundle labeling according to the relative position of the bundles connecting *PoC* and *PrC* regions for Subject001.

position index. Because of inter-subject variability, the correspondence is not perfect, nor do they all have the same number of bundles. Figure 3.19 displays the first five bundles of the four subjects, which connect the left *PrC* gyri with the *supra-marginal (SM)* gyri.

A quantitative evaluation of the bundle correspondence among subjects is displayed in Figure 3.20, where the distance (MDF) between the bundle centroids of each pair of subjects for the five bundles connecting *PrC* and *SM* gyri (*PrC_SM_0* to *PrC_SM_4*) is represented with a color scale in mm. Bundles show a relatively good correspondence among them, with distances between centroids ranging from 7 to 36 mm, with an average of about 20 mm. Note that distances of 20-30 mm have been previously used for inter-subject analyses of superficial white matter [50, 106].

Finally, the average execution times for each stage of the intra-subject labeling are: 192 s for *stage 1*, 10.23 s for *stage 2*, 11 s for *stage 3* and 2.49 s for *stage 4*, taking on average a total time of 3.6 min.

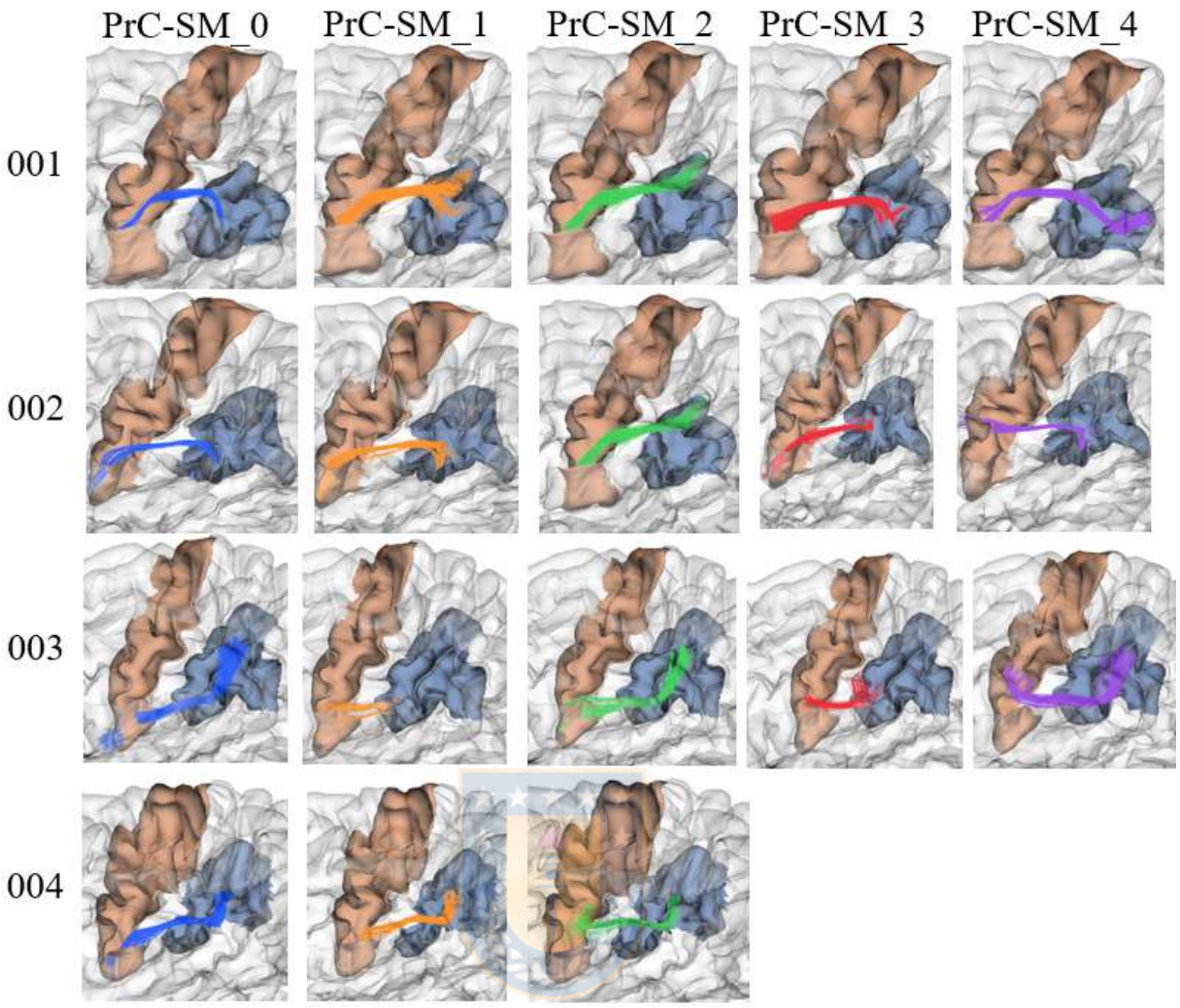


Figure 3.19: Correspondence of intra-subject bundle labels across subjects. Comparison of the first five bundles from four subjects (001-004), connecting *PrC* and *SM* gyri.

Inter-subject labeling

The inter-subject labeling was applied to 20 subjects from the ARCHI database [113]. A comparison has been made between the two implemented methods, matching with the *Hungarian* algorithm [40] and clustering with *QuickBundles (QB)* algorithm [42]. Both algorithms work with an input parameter, a distance threshold, using the minimum average direct-flip (MDF) distance [42] from one centroid to another. This distance threshold defines the minimum degree of similarity between bundles. In addition, tests have been carried out with four different distance thresholds. The first

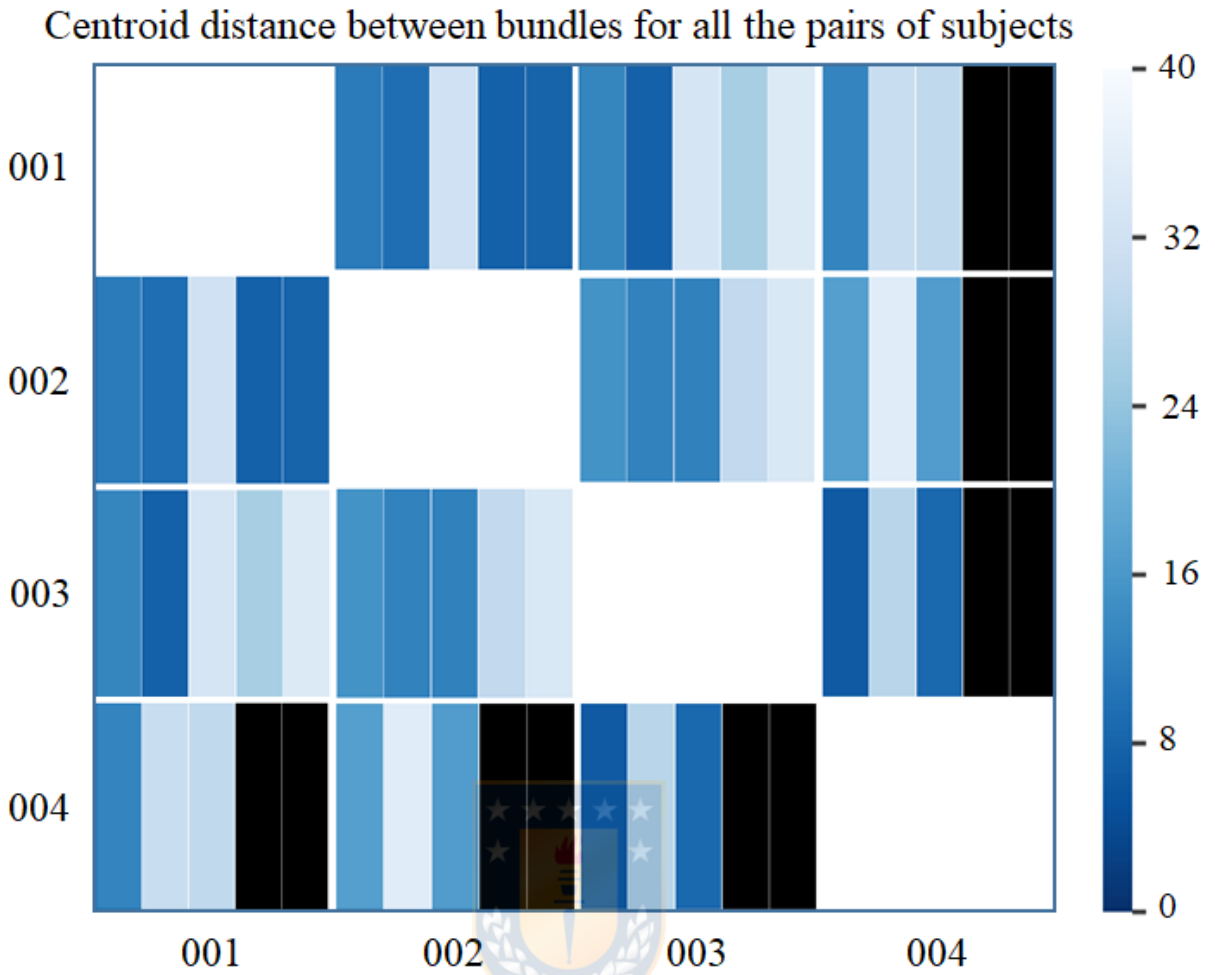


Figure 3.20: Bundle centroid distances between pairs of subjects for intra-subject labeling of four subjects. Each cell in the matrix contains five divisions. Each division represents a bundle connecting *PrC* and *SM* gyri (*PrC_SM_0* to *PrC_SM_4*). The color scale represents the distances (in mm) between bundle centroids for all the pairs of subjects. The black divisions represent the absence of bundles connecting the gyri.

threshold of 10 mm is very conservative, being the default threshold of *QB* for intra-subject clustering. We also used a 12 mm threshold, which is still conservative and aims to find similar bundles across subjects. Two other moderate thresholds were used: 18 mm and 21 mm, which are adequate considering that distances of 20-30 mm have been previously used for SWM inter-subject analyses [50, 106].

The reproducibility of the methods was evaluated by counting the number of subjects in which each bundle was found. Figure 3.21 shows the reproducibility of the

bundles for both methods and the four thresholds. Table 3.4 lists three reproducibility indices for the two inter-subject labeling methods, separated by hemisphere: the maximum number of subjects for the bundles within the 20 most reproducible bundles, and the number of bundles with reproducibility greater than or equal to 50% and 75%. As expected, for both algorithms, the higher the distance threshold, the greater the reproducibility. As can be seen, the method that shows the highest reproducibility is *QB*, presenting 94 bundles with more than 50% of reproducibility for a distance threshold of 21 mm, which is a good number, based on previous studies [50, 106]. On the other hand, the *Hungarian* algorithm only found 34 bundles with more than 50% of reproducibility for the same threshold. Furthermore, the *Hungarian* algorithm found no bundles present in all subjects, while *QB* found 19 for 21 mm threshold. As the *Hungarian* algorithm tries to match 1-1 the bundles, leads to less reproducibility than *QB*.

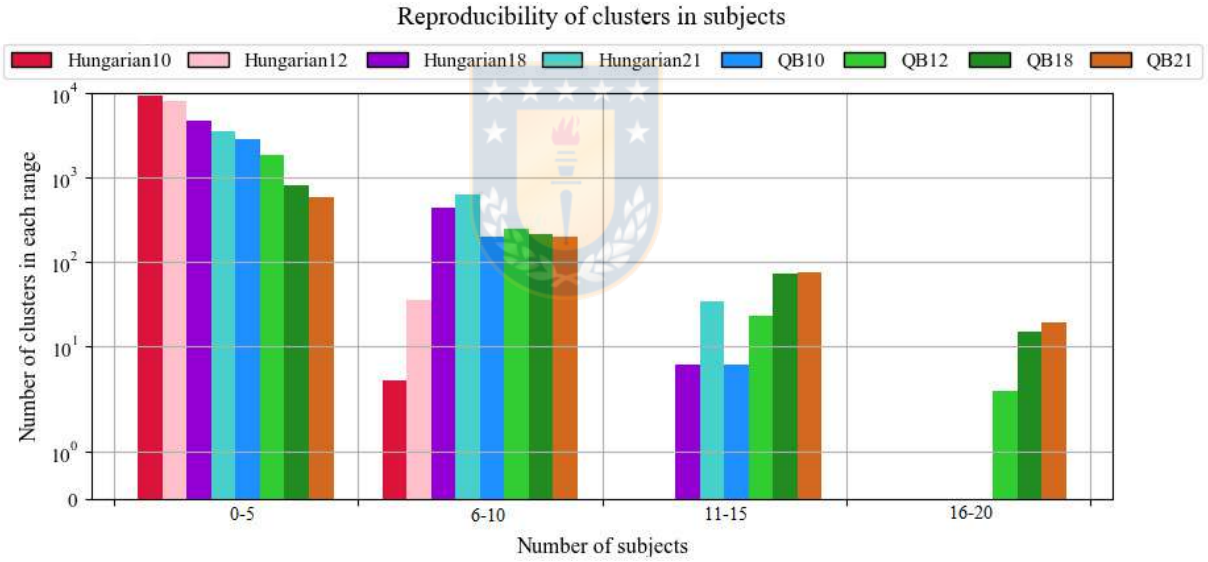


Figure 3.21: Reproducibility of bundles with inter-subject labeling for the two methods. The number of subjects is shown on the x-axis while the y-axis shows the number of clusters in each range.

Figure 3.22 shows inter-subject labeling quality for both methods with the four tested distance thresholds. The quality is evaluated using the inter-cluster distance (MDF), calculated per each bundle as the average distance between all the pairs of bundle

Method	Max	# Bundles $\geq$ 50%	# Bundles $\geq$ 75%
Hungarian12_left	7	0	0
Hungarian12_right	6	0	0
QB12_left	19	12	3
QB12_right	14	14	0
Hungarian18_left	13	3	0
Hungarian18_right	11	3	0
QB18_left	20	41	9
QB18_right	19	42	6
Hungarian21_left	13	22	0
Hungarian21_right	13	12	0
QB21_left	20	49	10
QB21_right	20	45	9

Table 3.4: Reproducibility values between for the two inter-subject labeling methods. The left column identifies the method (*Hungarian* or *QB*), hemisphere (left or right), and the thresholds (12 mm, 18 mm or 21 mm). The second column lists the maximum number of subjects for the bundles within the 20 most reproducible bundles. Columns three and four show the number of bundles with reproducibility greater than or equal to 50% and 75%, respectively.

centroids from the subjects where the bundle was labeled. Thus, the clusters classified with the same label are measured together, the closer the clusters are in all the subjects, the better the quality of the method. As expected, for both algorithms, the lower the distance threshold, the higher the quality. It can be seen that the most accurate algorithm is the *Hungarian* with a 10 mm threshold, at expenses of a low reproducibility, as shown above. The *QB* algorithm has a lower quality than *Hungarian* because it groups clusters of the same subject and merges them, thus increasing the inter-cluster distance. However, the merging of close clusters leads to a final better reproducibility, while keeping a moderate intra-cluster distance across subjects, with values inferior to 30 mm, and an average of about 15 mm.

To have an insight of the reproducibility and variability of the most reproducible bundles in all subjects for the two labeling methods, we created heatmaps (Figures 3.23, 3.24, 3.25 and 3.26). The heatmaps were created separately for the 20 bundles of the left and right hemispheres with the highest reproducibility in the 20 subjects, for

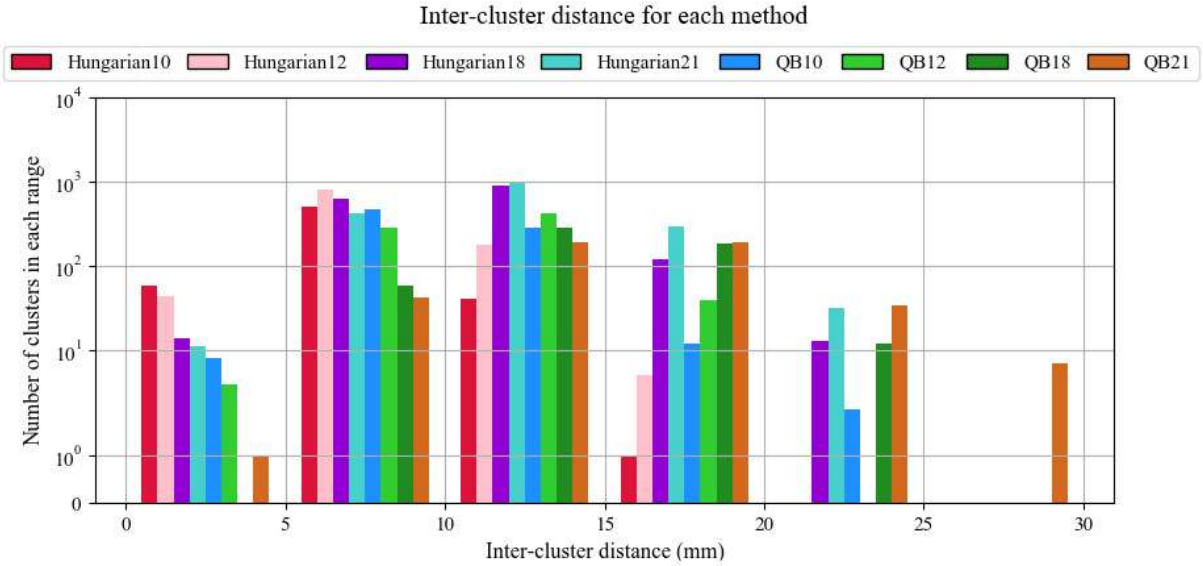


Figure 3.22: Inter-cluster bundle distance for both inter-subject labeling method. X-axis represents the inter-cluster distance measured in mm. Y-axis shows the number of clusters in each range.

a 12 mm threshold. Figures 3.23 and 3.24 display the heatmaps for the *Hungarian* algorithm, for left and right hemispheres, respectively, while Figures 3.25 and 3.26 show the heatmaps for the *QB* clustering algorithm. The bundles appear in descending order along the y-axis, according to the reproducibility between subjects, which appear along the x-axis. Empty (white) boxes indicate that a bundle does not exist in a certain subject. The colors represent the normalized number of fibers of each bundle (0-1), the darker, more fibers.

It can be seen that, as in Figure 3.21, the method with the highest reproducibility is *QB*. The number of fibers seems to be more homogeneous for *QB*, with a tendency of a low normalized number of fibers. This does not mean that the bundles have few fibers, but that their number is of a given value for most of the subjects, with very high values for a few subjects. The bundle with the highest reproducibility is *lh_PoC-SM_2* which was found in 19 subjects, followed by *lh_IP-SP_0* and *lh_Tr-RMF_0*, both found in 18 subjects. Reproducibility using matching is poorer, whose most reproducible bundle, *lh_IP-SP_69*, appears in only seven subjects.

Finally, some examples of bundles with high, medium, and low reproducibility are

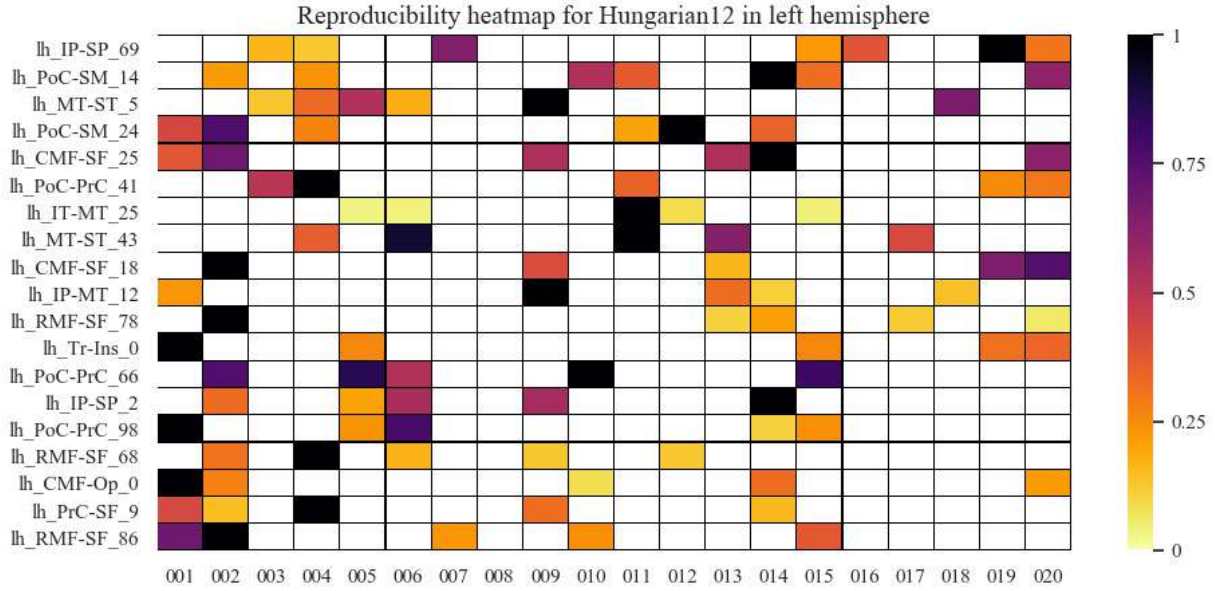


Figure 3.23: Reproducibility heatmap for *Hungarian* algorithm with threshold 12 mm, for the left hemisphere. On the x-axis are the subjects, on the y-axis are the 20 most reproducible bundles. The greater the number of fibers, the darker the color of the box on the heatmap that is normalized between 0 and 1.

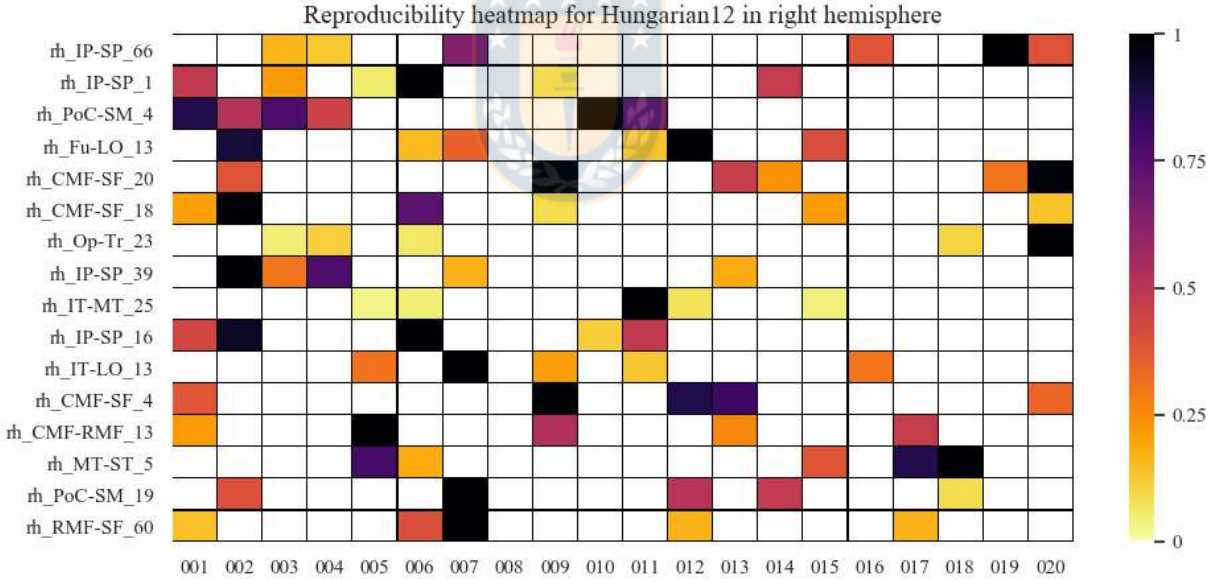


Figure 3.24: Reproducibility heatmap for *Hungarian* algorithm with threshold 12 mm, for the right hemisphere. X-axis displays the subjects used, the 20 most reproducible bundles are shown on the y-axis. The darker boxes indicate a higher concentration of fibers in the bundle. These values are normalized between 0 and 1.

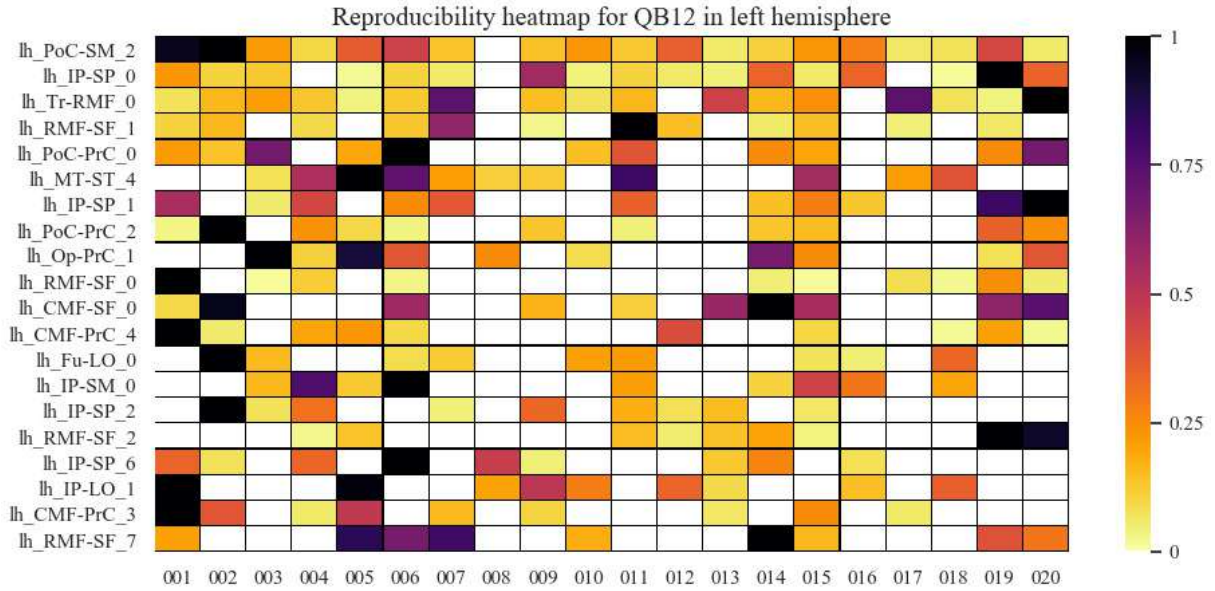


Figure 3.25: Reproducibility heatmap for *QB* with threshold 12 mm, for the left hemisphere. X-axis shows the subjects, while the y-axis shows the 20 most reproducible bundles among subjects in the left hemisphere. Darker boxes show bundles with more fibers in them. White boxes show the absence of the bundle in the determined subject. The heat bar shows the values of normalized fibers between 0 and 1.

displayed for a visual inspection of the results. Figure 3.27 shows the bundle *lh_PoC-SM_2*, belonging to the left hemisphere and classified by the *QB* clustering method with 12 mm threshold. This is the bundle with the highest reproducibility with this restrictive threshold, being present in 19 out of the 20 subjects, with the exception of *Subject008*, achieving a 95% of reproducibility. It can be seen how bundles connect approximately the same cortical regions in different subjects and have a similar main shape. Also, it can be seen that the number of fibers is very variable among subjects, which is usual in SWM bundles.

Figure 3.28 shows the bundle *lh_PoC-PrC_0*, of the left hemisphere and classified by the *QB* method with 12 mm of threshold. It appeared in 11 out of the 20 subjects, that is, a 55% of reproducibility. This is a small bundle connecting the *PoC* and *PrC* gyri.

Lastly, Figure 3.29 shows for the left hemisphere the cluster *lh_RMF-SF_7*, classified by the *QB* method using a threshold of 12 mm. This is the least reproducible

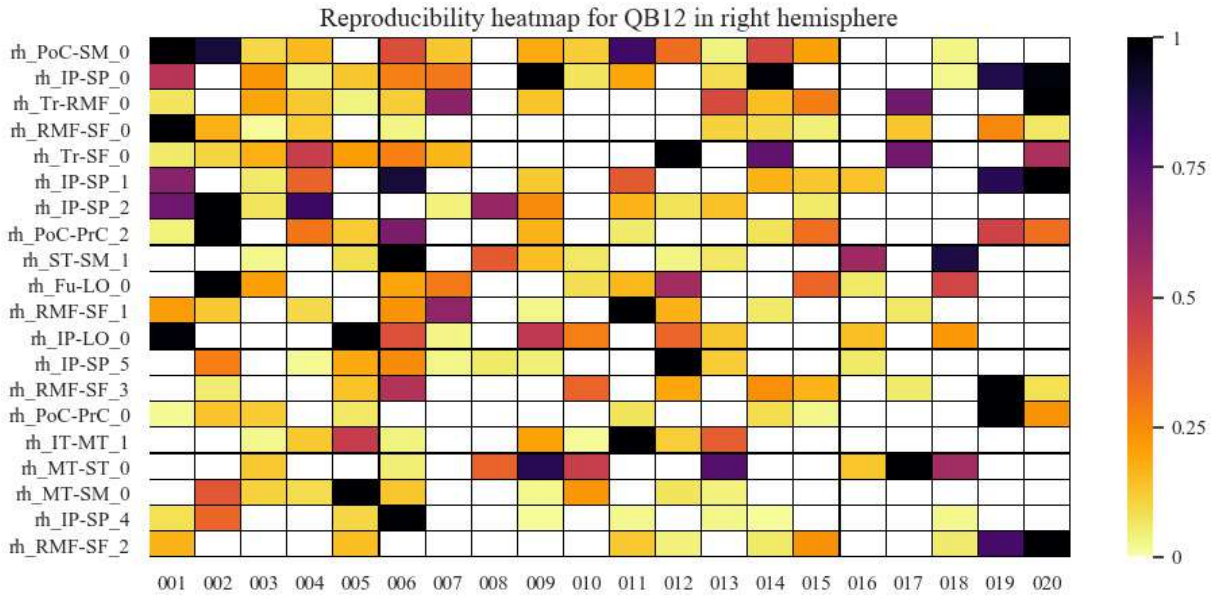


Figure 3.26: Reproducibility heatmap for *QB* with threshold 12 mm, for the right hemisphere. X-axis displays the subjects, and on the y-axis appears the 20 most reproducible bundles. The lighter the color of the box, the fewer fibers it contains. If the box is white, it indicates the absence of the bundle in the subject. The fiber values appear normalized between 0 and 1 in the heat bar.

cluster of the heatmap of Figure 3.25, appearing in 9 out of the 20 subjects, reaching only 45% of reproducibility. However, it can be seen that the bundles connect the same area in all subjects, slightly varying the position and the number of fibers.

3.2.4. Discussion

In the last two decades, a great number of methods have been proposed for the analysis of tractography datasets. Most of the works have been focused on the study of deep white matter bundles, such as the arcuate fasciculus or the inferior fronto-occipital fasciculus. These bundles are in general larger and more stable across subjects, and have been described by neuroanatomists several decades ago. The methods have been focused on the study of these bundles, the creation of WM bundle atlases and the segmentation of WM bundles. Most of the studies have been developed with a combination of ROI-based and clustering-based methods, and the important guidance of neuroanatomy experts. In general, the applications analyze the segmented bundles

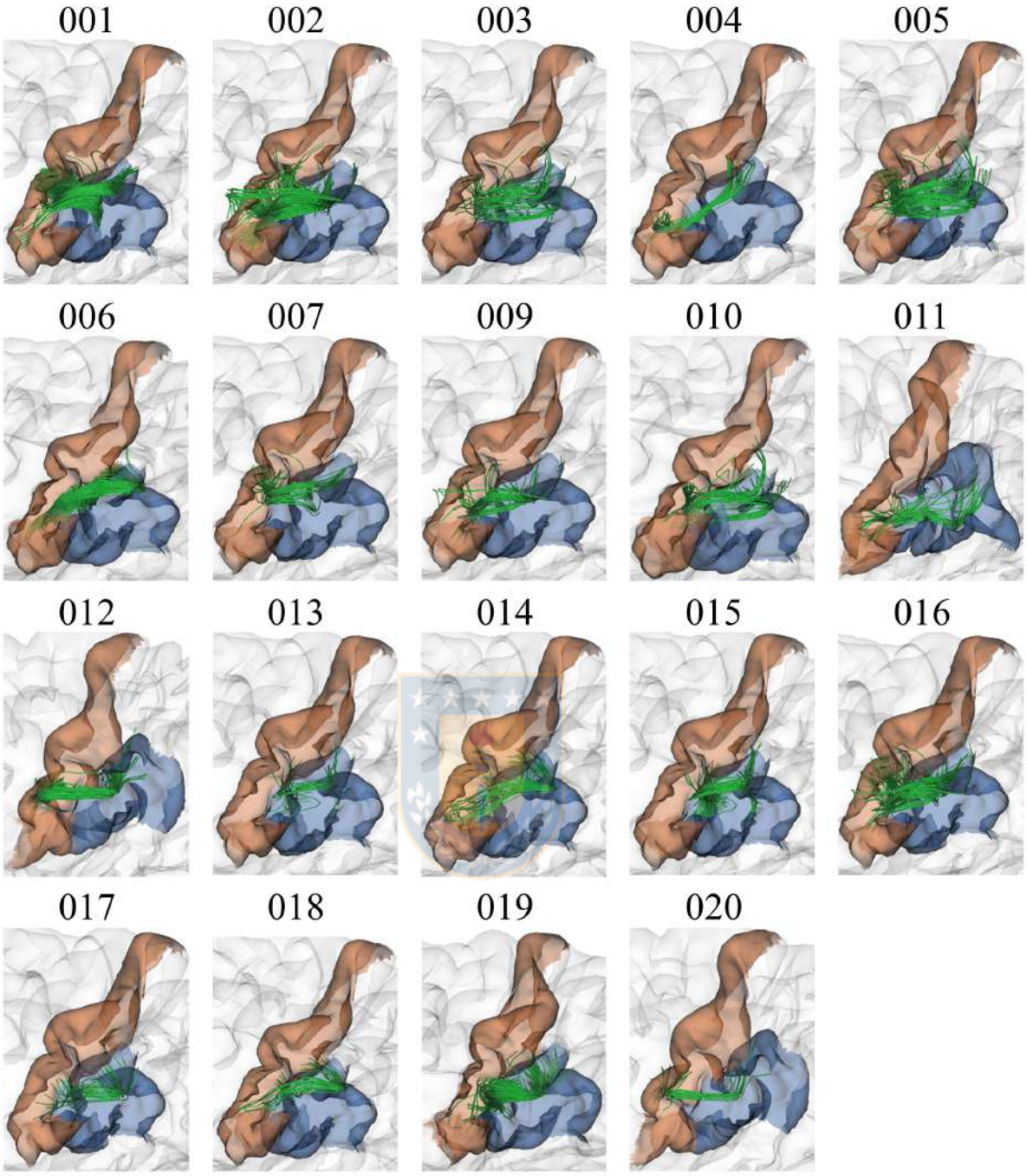


Figure 3.27: Bundle *lh_PoC-SM_2*, with the highest reproducibility in all subjects. The results show good reproducibility among subjects, appearing in 19 of the 20 subjects for the *QB* method with a 12 mm threshold.

across subjects and different populations of patients.

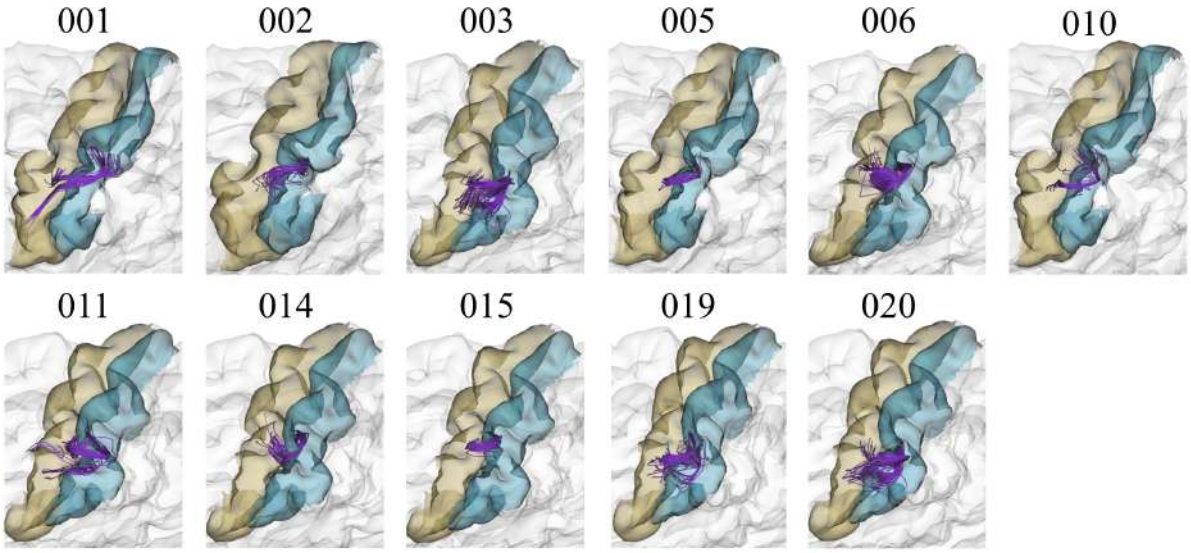


Figure 3.28: Bundle *PoC-PrC_0*, with medium reproducibility. The *PoC-PrC_0* bundle appears in 11 out of 20 subjects, achieving 55% of reproducibility, for the *QB* algorithm with a 12 mm threshold.

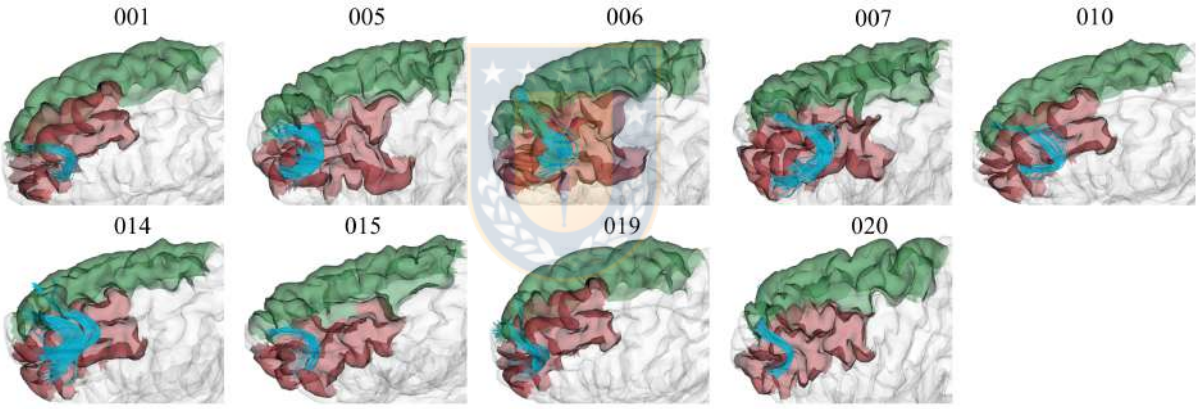


Figure 3.29: Bundle *lh_RMF-SF_7*, with low reproducibility. The bundle appears in nine out of 20 subjects using the *QB* clustering algorithm with a 12 mm threshold.

The methods have evolved with the increasing quality of the data. Tractography datasets have increased their size and complexity due to a higher resolution and better image quality, being able to provide a better representation of fiber crossing and small bundles. These advances are also associated with improved algorithms along all the processing pipeline, including artifacts and distortion corrections, diffusion local modeling, and fiber tracking. Furthermore, the use of more accurate tractography propagation masks (e.g., based on T1 images [52]) has helped to achieve a better

reconstruction of small and superficial white matter fibers.

Hence, in the last decade, due to the better quality of dMRI images and processing algorithms, it has been possible to start studying the short association WM bundles. A first whole-brain study used an atlas of gray and white matter to extract short fibers connecting adjacent gyri [144]. Other works combined a hierarchical fiber clustering and cortical parcellation information to extract reproducible short association bundles [50]. A recent study reported a great number of short association bundles, but without a labeling [142]. These works were mostly focused on the creation of SWM atlases. Hence, there is still a need for methods, open to the research community for the study of short association bundles in new databases.

The proposed methods provide an automatic labeling of SMW bundles. First, an efficient individual labeling was implemented. It generates compact clusters and labels them according to a cortical parcellation based on mesh information, for a high-quality ROI-based labeling. Furthermore, the bundles connecting each pair of anatomical regions (gyri) are ordered following an axis orientation. The resulting clusters could be used for fast and easy exploration of short association bundles in individual brains. Without a labeling, its exploration is very complex, since about one thousand clusters are produced for the whole-brain.

Subsequently, an inter-subject method has been added, to obtain a correspondence between the clusters (or bundles) across subjects. We tested two methods, a matching, based on the *Hungarian* algorithm, and a clustering method, based on the *QB* algorithm. Even though we used available implementations of both methods, we have adapted them to the processing of labeled intra-subject clusters from different subjects to generate automatically labeled inter-subject clusters.

Since there is a high inter-subject variability in tractography datasets, especially for short association bundles, the applied processing extracts the main fiber connections by using first an intra-subject clustering. It has also the advantage to remove noisy fibers, that are not grouped into the main clusters. Furthermore, the inter-subject analysis, along with a common labeling for clusters which are similar for most of the

subjects, performs an identification of non-reproducible fibers at the group level.

The results show a better reproducibility across subjects for the *QB* clustering method versus the *Hungarian* algorithm. *Hungarian* algorithm finds a good correspondence between subjects, with low inter-cluster distance, but at expenses of inferior reproducibility. Due to inter-subject variability, and the absence of bundles in some subjects, the one-to-one matching strategy seems not to be directly applicable to this kind of problem. On the other side, the clustering groups similar bundles on subjects and do not impose the existence of clusters in all the subjects. Indeed previous inter-subject analyses based on clustering have included a reproducibility constraint, e.g., a minimum number of subjects present in the clusters. Hence, an advantage of the proposed labeling is that this reproducibility information is easily extracted from the inter-subject labels, which is not the case for classic algorithms. Furthermore, even though the main goal of this work is not a study of the reproducibility of SWM bundles, the results of our inter-subject clustering strategy are competitive with state-of-the-art methods, with 94 reproducible bundles for a moderate MDF distance of 21 mm, compared to about 100 hundred bundles obtained for atlases proposed in [50, 106], created with a maximum Euclidean distance of 30 mm.

Note that several factors impact the results, including the quality of the tractography datasets, and the registration strategy. It has been shown that using non-linear registration increases the number of SWM identified [106]. In our experiments we used affine registration to Talairach space, however, other registration algorithms can be applied without problem. Furthermore, false positives are very likely to increase the variability across subjects and affect the results. Bundle variability may also be due to the inherent variability of the cortical folding patterns [83]. Since U-fibers are directly under the cortical sulci, present a smaller size, and connect small regions of the cortex, they are more sensitive to differences in the cortex anatomy than deep white matter bundles.

Finally, we highlight some advantages of the proposed methods. First, it is efficient, taking about 3.6 min for an intra-subject analysis and about 9 s to perform the inter-subject clustering. That is, for the whole inter-subject labeling processing, it takes about 1.17 h on average. This time is reasonable for an inter-subject analysis. Furthermore,

the algorithms are scalable and can be applied to larger tractography datasets and databases.

Inter-subject labeling can be used to discover patterns of connections in different groups of healthy subjects and patients. The inter-subject clusters can be used to create WM bundle atlases, which require the inspection of experts in anatomy. Note that the clusters from tractography can contain artifacts and false positives [82]. In fact, false-positives can be highly reproducible, hence these will be also labeled by the proposed methods. The inter-subject labeling was not conceived to identify bundles with an anatomical meaning or to discard false-positives. The objective is to identify and label all the reproducible bundles. These bundles can be used as input to other analyses that include anatomical knowledge, in order to validate the bundles. The proposed algorithm can also contribute to the analysis of tractography datasets for the improvement of tractography methods, through the incorporation of anatomical information and filtering. Finally, other applications include the study of brain connectomes and methods for diffusion-based cortical parcellations.

Chapter 4

Cortical surface parcellation

In this chapter, we present three methods to perform the cortical surface parcellation. The main parcellation method is presented in Section 4.1 and it generates the parcellation of the cortical surface by using an atlas of white matter bundles. It creates atlases of parcels with different granularities based on the *Desikan-Killiany* atlas as a delimiter of the anatomical parcels. The second parcellation method is presented in Section 4.2 and is a complementary method to perform the cortical surface parcellation of a subject made with clustering of fibers. Finally, Section 4.3 contains a complementary individual cerebral cortex parcellation based on the morphology of the cortical surface.

4.1. Cortical surface parcellation based on a fiber-bundle atlas

In this section, we present a hybrid method to create fine-grained parcellations of the cortical surface, from a coarse-grained parcellation according to an anatomical atlas, based on cortico-cortical connectivity. The connectivity information is obtained from segmented superficial and deep white matter bundles, according to bundle atlases, instead of the whole tractography. First, we present the state of the art of parcellations to date. The following is a description of the proposed method for the parcellation of the cerebral cortex. In the experimental section, we carry out reproducibility tests and a comparison with other parcellations based on different MRI modalities, such as macroanatomy, structural, functional, diffusion, and multimodal. Finally, we discuss on the cortical parcellation method and the obtained results.

4.1.1. Related work

The human connectome is of special interest to understand the brain structure and function [121]. The structural connectome is composed of two basic elements, the somas (nodes) and the axons (edges) that exist between them, formed by white matter (WM) tracts [118, 59, 13]. Magnetic Resonance Imaging (MRI) provides in-vivo techniques to study the human brain. MRI modalities include diffusion-weighted MRI (dMRI) that estimates the WM tracts of the brain [90],structural MRI (sMRI) which focuses on brain anatomy [11] and functional MRI (fMRI) that estimates brain function [124, 61]. dMRI allows researchers and clinicians to study non-invasively and in-vivo how white matter is organized in the brain giving details of its connectivity and structure [75]. It is based on measurements of the movement of hydrogen atoms present in water molecules of biological tissues.Tractography algorithms reconstruct an estimate of the main WM tracts of the entire brain based on dMRI information [7, 143]. The generated datasets represent an estimation of the main WM pathways, in the format of 3D polylines, also called *fibers*, even though they do not represent real neural fibers [92, 101]. This technique is indirect, and relies on models and inference, but allows a whole-brain exploration of WM structure in living humans, on large populations of subjects.

The structural networks of the human cerebral cortex have not yet been comprehensively mapped [118, 58, 121]. The brain's structural and functional systems have features of complex networks, such as "small-world" topology, highly connected hubs and modularity, at the whole human brain scale [13]. The study of brain connectivity, taking into account its function and structure, can be performed based on a cortex parcellation, which is the cortical division of the brain into macroscopic regions [28]. A parcellation may be based on resting-state fMRI (rs-fMRI) [111], anatomical structure [32], dMRI [76] or cytoarchitecture. Architectonic and other template-based atlases have been created [31, 123], but may not reflect the individual variations in regional functional boundaries. Data-driven parcellations can overcome this limitation by a better definition of individual cortical regions [117]. Parcellation atlases can be constructed using information from multiple modalities, and several scales. For example,

for a given population, information from cortical folding, myelin content, resting-state, and task-based fMRI was integrated to create a functionally relevant parcellation [46]. However, individual variability and the limitations of each modality make the application of those methods very difficult. Here, we focus on the development of a method for the tractography-based parcellation (TBP) of the cortical surface. The method could be posteriorly integrated to multimodal parcellation frameworks [99].

Connectivity-based methods use tractography information to find regions with common connectivity patterns between the cortical voxels, or cortical surface mesh vertices, that compose each region. All the methods have to deal with the high inter-subject variability, especially in the brain cortex and superficial white matter (SWM). Hence, to reduce the complexity of the problem, some methods have been focused or tested on a few brain regions, or have used an anatomical parcellation for initial regions [4, 69, 54, 105, 100, 77]. In general, the similarity between the connectivity profiles of the voxels (or vertices) is estimated using some similarity measure and then, a method is applied to regroup elements with common connectivity patterns. Some methods have been proposed to perform an analysis over the whole-brain cortex [76, 98, 89, 29, 96]. This kind of approach, in general, calculates the whole connectivity profile of each seed node (image voxels or mesh vertices) followed by the computation of a connectivity matrix and clustering of the nodes. A group of methods performs a tractography-based parcellation of the cortex using only connectivity information given by the fiber extremities [98, 76, 77], while other group embeds fiber shape information into the analysis [89, 29, 96]. Also, for inter-subject analysis, it is necessary to find the correspondence between subjects. One strategy is to create the parcellation taking into account the main connections present in the population of subjects [112]. Another approach is to detect individual connectivity patterns, or even parcels, from the tractography of each subject and then find consistent parcels among the population of subjects [89, 76, 77]. Furthermore, due to the high complexity and the huge size of connectivity data, all the methods use a dimension reduction criterion. The difficulties mentioned above, between others, make the parcellation of the human brain cortex a complicated and unachieved task. In the following, we briefly describe some methods to provide an insight into the complexity of the solution implementation.

An interesting approach of whole-brain TBP is based on hierarchical clustering [89]. The method selects GM/WM interface voxels as seeds and generates probabilistic tractography from them. For each seed voxel a tractogram is obtained (visitation map). Hierarchical clustering is applied over the tractograms using a noncentered variant of the Pearson's correlation coefficient as a similarity measure. The resulting dendrogram is post-processed to reduce the number of branchings. Next, a leaf-matching is iteratively applied to the two tractograms with the highest similarity, to find correspondence across subjects. Even though the method is promising, the different parameters were difficult to adjust and no perfect match was found. [76] apply a watershed to the connectivity profiles averaged from all the subjects of a gyrus (patch) in order to split the cortical surface into catchment basins [104]. A set of regions of interest strongly connected to the gyrus across subjects is then identified, and a joint patch connectivity matrix across subjects is calculated. Finally, to construct the final cortex parcellation, each gyrus is clustered using the classical k-medoids algorithm applied to the distance matrix. The method removes a large part of the connectivity data by filtering, however, a good reproducibility among subjects was obtained. Another interesting example is the work proposed by [96]. The method first calculates connectivity matrices from cortical vertices and subcortical voxels to the rest of the brain, based on probabilistic tractography. Then, creates an average matrix across the subjects and applies independent component analysis (ICA) to provide a group-average connectivity matrix. The dimensionality of this matrix is incrementally reduced in tractography space using principal component analysis (PCA) on subsets of the matrix. A post-processing is applied to obtain a hard parcellation of the cortex, without a straightforward mapping to tractography and gray matter, due to the high cortical and connectivity variability between subjects.

A different strategy for creating a parcellation is to use a hybrid method involving the use of bundles segmented from a bundle atlas. The first proof of concept used a subset of bundles manually selected from a multi-subject SWM bundles atlas [50]. This work processed 10 subjects, using fix parameters manually tuned for all the processing. It segmented the bundles for each subject and calculated the intersection regions of the bundles with the cortex. In case of overlapping between two regions, the parcel label of the smaller parcel prevailed over the bigger one. This very preliminary work showed the

potential advantage of using labeled bundles for the cortical parcellation, with relatively good correspondence in some regions of the brain. This method was next improved by the use of a graph representation of the overlapping between regions [115]. This first attempt tuned the parameters in one subject and subsequently applied them to four other subjects, giving some correspondence across the subjects. However, since no inter-subject analysis is performed for the merging of the connecting regions, the method is not applicable to a large group of subjects.

Hence, extending this idea, we propose a new hybrid method for the structural connectivity-based parcellation of the cortical surface based on segmented bundles. Unlike most of the methods proposed in the literature, which use full tractography, we use fibers labeled into bundles, according to short and long bundle atlases. The advantage is that the correspondence of connecting regions is given in advance for the different subjects in a database. Furthermore, the generation of parcels from segmented bundles could give a better representation of the main regions or nodes of the human brain connectome, since these were identified as the main short and long connections of the brain, represented in the atlases of bundles. Then, the resulting parcellation will represent a subdivision of the cortex into the regions that connect the most probable bundles. The method still has the difficulty to clearly define the nodes (cortex parcels), knowing that the bundles from tractography are very variable across subjects and may not exist in several subjects. This poses a big but interesting challenge. The key point of the proposed work is the automatic analysis of the density and variability of the connecting regions among subjects over the cortical mesh so that the most probable ones are selected, merged and homogenized. The overlapping is solved using a graph representation of the intersected regions taking into account the degree of overlapping of their density centers, across subjects.

The method was applied to a group of 79 subjects from a HARDI database. Several quantitative and qualitative evaluations were performed. Twenty parcellations were generated, based on different sets of the three parameters of the method, and compared to evaluate the similarity between them. Furthermore, a reproducibility analysis was also performed, based on the similarity of connectivity matrices across subjects,

constructed with the whole tractography.

A comparison with a macroanatomical parcellation using Dice's coefficient between subject's connectivity matrices was performed, showing a slightly better reproducibility in a resultant parcellation generated with the proposed method. Moreover, other comparisons are made with state-of-the-art parcellations based on different MRI modalities, finding a degree of similarity with dMRI, functional, anatomical, and multimodal atlases. The higher similarity was found for our parcellation composed of 185 sub-parcels with another parcellation containing 239 parcels, based on dMRI data from the same database, but created with a totally different approach. This comparison led to 130 parcels in common based on a Dice's coefficient ≥ 0.5 and 75 parcels in common with a Dice coefficient ≥ 0.6. Finally, complementary analyses were performed that are included in the appendix B.

4.1.2. Our approach

We propose a hybrid method for the creation of fine-grained parcellations of the cortical surface, from a coarse-grained parcellation according to an anatomical atlas, based on structural connectivity information, given by segmented bundles for a population of subjects. The bundle segmentation is based on atlas bundles from three different atlases. The parcellation method receives as input the tractography and labeled mesh of each subject, the fused bundle atlas with selected superficial and deep white matter bundles. The method returns an average parcellation atlas for the input dataset, which consists of a subdivision of the anatomical parcels (gyri) of *Desikan-Killiany* atlas, based on the most stable connectivity-based sub-parcels across the subjects. The data consists of the labels associated with each cortical mesh vertex. Note that the used cortical meshes, based on Freesurfer processing output, contain the same number of triangles and vertices in all the subjects. For all the subjects, corresponding triangles will represent the same anatomical region, but with local differences, according to the morphology of each subject. To create the final parcellation, the method uses the probability and density information of the sub-parcels from all the subjects. Hence,

an intermediate output of the method is the probabilistic representation of each sub-parcel. Figure 4.1 shows a schematic of the parcellation method. The selection of the final bundle atlas is performed as a pre-processing stage *(A)*. Next, the method is composed of six steps: *(B)* fiber bundle segmentation, *(C)* extraction of meshes and labels, *(D)* intersection of the fibers with the mesh, *(E)* fiber filtering, *(F)* cortex parcellation, and *(G)* sub-parcel post-processing.

Pre-processing: Fusion of atlases

This pre-processing aims to create a fused atlas of white matter bundles, containing the main WM connections across subjects and, consequently, to create a more complete parcellation of the cortex (see Figure 4.1-(A)). We used two atlases of superficial white matter (SWM) and one atlas of deep white matter (DWM). The first SWM atlas, *swm_atlas_1*, is composed of 50 bundles in both hemispheres, with a total of 7,857 fibers [50]. The second SWM atlas *swm_atlas_2*, has 44,345 fibers and is made up of 44 bundles in the left hemisphere and 49 bundles in the right hemisphere [106]. Finally, the DWM atlas contains 18 bundles per hemisphere, corresponding to 11,755 fibers [53]. Those atlases were created using the ARCHI database, representing the most reproducible bundles across subjects (see Figure B.2 of the appendix B).

The bundles from both SWM atlases are labeled following the same naming convention, based on the anatomical regions of *Desikan-Killiany* atlas [31]. The name contains *lh* or *rh* to denote the left or right hemisphere, followed by the name of the two regions connected by the bundle, according to the abbreviation of the region (see Table B.3). Finally, a correlative number is added to indicate the index, as many bundles can connect the same two anatomical regions in an atlas. For example, a bundle connecting the postcentral and precentral gyri of the left hemisphere is called: *lh_PoC-PrC_0*, where 0 is the index given by the atlas. *swm_atlas_1* contains only bundles connecting two different anatomical regions (gyri), while *swm_atlas_2* also contains bundles connecting different areas of an anatomical region. On the other side, the DWM atlas labels the bundles according to an abbreviation of their anatomical name, followed by *LEFT* or *RIGHT* to denote the hemisphere.

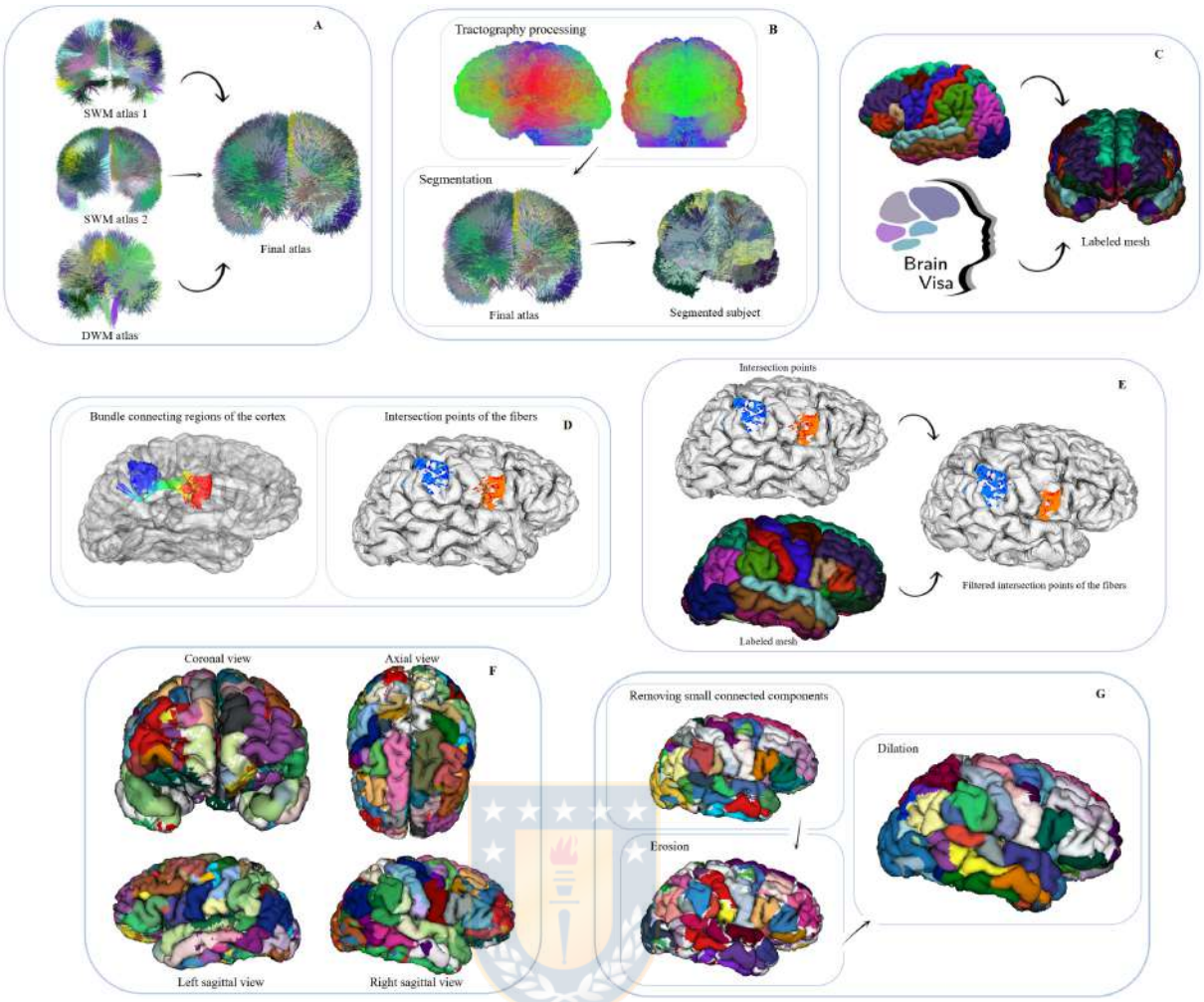


Figure 4.1: Schematic of the parcellation method. **(A) Pre-processing: Fusion of atlases**. The bundles of a long and two short WM bundle atlases are fused into a final fiber bundle atlas. **(B) Step 1: Fiber bundle segmentation**. Classifies a subject's fibers with respect to the bundle atlas. **(C) Step 2: Extraction of meshes and labels**. By using FreeSurfer and then BrainVISA software, the cortical meshes and their corresponding labels, according to *Desikan-Killiany* atlas, are obtained. **(D) Step 3: Intersection of the fibers with the mesh**. This step obtains the initial and final triangles intersected by each fiber bundle. **(E) Step 4: Fiber filtering**. This algorithm first obtains the label of each intersected triangle, then removes misclassified fibers and then performs a fiber alignment according to the corresponding atlas bundle. **(F) Step 5: Cortex parcellation**. The objective is to subdivide each region (anatomical parcel) into sub-parcels. **(G) Step 6: Sub-parcel post-processing**. To get more homogeneous parcels, the small connected components are eliminated, followed by a closing of the parcels over the cortex.

To fuse the atlases, we first analyzed the bundles that are very similar in both SWM atlases, connecting the same regions. In general, in the case of a big similarity between bundles from both atlases, we selected the most compact bundle. After a visual comparison of both atlases, the bundles of *swm_atlas_1* are better defined in their ends, and therefore are more suitable for performing a cortical parcellation. Some bundles with high similarity in both SWM atlases are shown in Figure B.1. Hence, all the bundles of the *swm_atlas_1* were selected (see the first row of Figure B.2). Next, 27 bundles in the left hemisphere and 34 in the right hemisphere for *swm_atlas_2* were selected, as shown in the second row of Figure B.2. Most of the selected bundles of *swm_atlas_2* connect different areas within an anatomical region.

Respecting the DWM atlas, we first discarded the Corticospinal Tract, Fornix and Thalamix Radiations, as those bundles do not represent cortico-cortical connections. Also, we discarded the Corpus Callosum as it is a very large bundle that would not be very informative for the definition of subdivisions of the anatomical regions. The selected bundles are: Arcuate fasciculus, with its anterior and posterior portions (*AR*, *AR_ANT*, *AR_POST*), Cingulum (*CG*), Inferior Fronto-Occipital (*IFO*), Inferior Longitudinal (*IL*) and Uncinate (*UN*) bundles (see the third row of Figure B.2). These bundles cover the cortical regions that the two SWM atlases do not cover, achieving a complete coverage of the cortex. The fused atlas is in MNI space and contains a total of 179 bundles, distributed in 86 bundles in the left hemisphere (see Table B.1) and 93 bundles in the right hemisphere (see Table B.2), as we see in Figure B.3 (first row). Finally, the centroid of each atlas bundle is calculated as the mean of the corresponding points of all the fibers in a bundle, to later align the segmented fibers.

STEP 1: Fiber bundle segmentation

This step performs the segmentation of white matter bundles for each subject (see 4.1-(B)). Segmenting the fibers gives direct correspondence of the bundles and the connected cortical regions across the subjects. The segmentation algorithm [128] is a parallel version of the algorithm proposed in [53]. It classifies the fibers of a subject's tractography based on a multi-subject WM bundle atlas. It calculates the maximum

Euclidean distance between corresponding points of each subject fiber and each atlas fiber. A subject's fiber is labeled with the closest bundle if the distance does not exceed the maximum threshold defined for the bundle. The algorithm returns the subject's fibers that were correctly classified, labeled with the corresponding bundle name. Figure B.3 shows the final atlas of white matter bundles as well as a subject segmented with the atlas.

To perform the segmentation, the tractography datasets are resampled with 21 equidistant points, since it is a sufficient number to perform an analysis of the similarity between fibers, as used in others works [55, 53]. Before the calculation, the tractographies are transformed to the MNI space.

This algorithm receives as input the tractography of a subject, resampled with 21 equidistant points in MNI space, the fused bundle atlas, and the distance thresholds to be used for each bundle, defined for each original atlas [50, 106, 53]. It returns the segmented bundles for each subject, according to the labeling of the atlas bundles.

STEP 2: Extraction of meshes and labels

This step aims to obtain the meshes of the cortical surface and the labels of the anatomical regions given by the *Desikan-Killiany* atlas, as in Figure 4.1-(C). First, FreeSurfer software [39] is employed to calculate the cortical surfaces for each subject. By using this software, a direct correspondence between the cortical surface mesh of the subjects is obtained, since the number of vertices is the same for all of them, changing only their 3D coordinates in the mesh according to the individual morphology. For the labeling of the cortical surface, FreeSurfer uses the *Desikan-Killiany* (DK) atlas, which consists of 35 regions per hemisphere [31]. Each region in the atlas (see Table B.3) has associated a label (integer number). Hence, the labeling consists in assigning to each vertex of the mesh, the label of the region that corresponds to it. Next, BrainVISA software [24] was used to apply the pipeline that converts the formats and transforms the mesh to the subject's T1 space. It provides the mesh file with 81,924 vertices per subject and a file with the vertex labels. This step receives as input the NIFTI T1 image of each subject. The output is the cortical mesh and the labels

according to DK atlas, associated with each subject.

STEP 3: Intersection of the fibers with the mesh

This step calculates the intersection of the fibers with the cortical mesh [115]. The Möller-Trumbore algorithm [87] is used to determine whether the end of a fiber intersects a mesh triangle. For each end of the fiber, the algorithm selects the nearest triangle. Finally, only those fibers whose intersections at both ends were correctly identified are used. The algorithm returns the set of initial and final intersection points for each fiber. Figure 4.1-(D) illustrates a bundle and its intersection points over the cortical mesh. In this step, the algorithm receives as input the cortical mesh and the segmented fiber bundles of each subject in T1 space. It returns for each subject and each bundle, the indices of the intersected triangles by each fiber of the bundle, at both bundle ends.

STEP 4: Fiber filtering

This step filters out the fibers that do not connect the anatomical regions that should be connected, following the definition of the bundle to which they belong (see Figure 4.1-(E)). The algorithm receives as input the fiber intersection information (intersected triangles) and the label (cortical region) of each mesh vertex. This step returns as output the filtered fibers, i.e. those that intersect exactly within the corresponding anatomical parcel (gyri) so that all fibers that do not belong to that anatomical parcel are removed. Specifically, the filtering algorithm consists of three sub-steps.

Sub-step 1: Obtaining the fiber labels. First, for each bundle, the labels (according to the *Desikan-Killiany* atlas) of the triangles intersected by the start and end fiber points of each fiber are obtained. Next, the names of the two regions connected by each bundle are extracted from the bundle names. For example, for bundle PoC-PrC, the initial region is PoC (postcentral) and the final region is PrC (precentral) (see Figure 4.2-(A)).

Sub-step 2: Removing of misclassified fibers. Each fiber is analyzed and those in which the intersected triangle label does not correspond to the initial or final bundle regions are removed (see Figure 4.2-(B)). This processing removes the fibers that were misclassified by the fiber bundle segmentation method.

Sub-step 3: Fiber alignment. The fibers on a whole-brain tractography dataset have different orientations and are stored in the direction they were tracked. Hence, on average, half of the fibers are stored in the inverse direction. In these cases, the fiber points are swapped to align the fibers according to the atlas bundles, using the bundle centroids (see Figure 4.2-(C)). Finally, the filtered fibers for each bundle are obtained (see Figure 4.2-(D)).

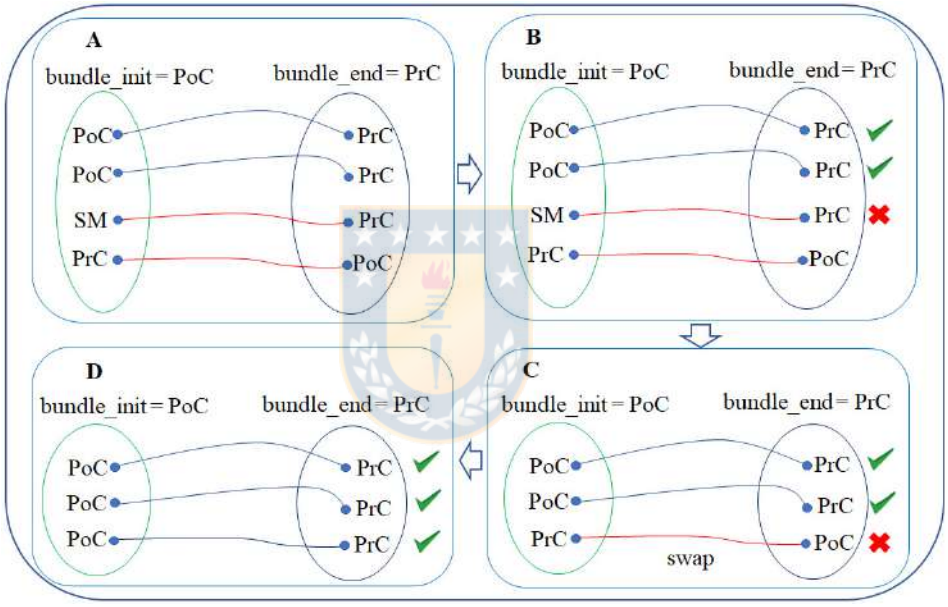


Figure 4.2: WM fiber filtering for a bundle. **(A) Sub-step 1: Obtaining the fiber labels**. The labels of the triangle vertices that are intersected by the start and end of each fiber are obtained. **(B) Sub-step 2: Removing of misclassified fibers**. Fibers that were misclassified by respect to the bundle anatomical definition are discarded. **(C) Sub-step 3: Fiber alignment**. The fibers that are stored in the inverse direction, by respect to the atlas bundle centroid, are inversed. **(D)** Filtered final fibers for a bundle.

STEP 5: Cortex parcellation

This is the main step of the method. It creates a fine-grained cortex parcellation, from a coarse-grained anatomical parcellation, based on the connectivity of segmented white matter bundles. The sub-parcels are probabilistic but a final hard parcellation is obtained with the most probable label for each triangle of the mesh. This algorithm receives as input for each subject the filtered fibers of the previous step (STEP 4), the intersection information (STEP 3), and the labeled cortical mesh (cortical mesh and vertex labels according to *DK* atlas). It returns as output the mesh vertex labels for the new parcel subdivisions. As mentioned above, it also generates the probabilistic representation of each sub-parcel, that is used to generate the final parcellation. The step can be subdivided into four sub-steps (see Figure 4.3). Next, we explain each one of the sub-steps.

Sub-step 5.1: Creating preliminary sub-parcels. This sub-step creates *preliminary sub-parcels* based on the fiber bundle intersection information of each triangle. Each bundle in the atlas will define two preliminary sub-parcels, corresponding to the two extremities of the bundle. Sub-parcel names were defined following the bundle names. Also, a label is internally associated to identify each sub-parcel. Figure 4.3-(A) shows an example of the two preliminary sub-parcels created for bundle *PrC-PoC_3* of a subject. Each anatomical parcel, given by *Desikan-Killiany* atlas, is formed by several preliminary sub-parcels, which overlap each other, representing all the bundles that connect the region.

Sub-step 5.2: Calculating probability maps. This sub-step computes the probability of each sub-parcel in each triangle across the subjects. With this information, the probability maps for all the sub-parcels over the mesh are inferred, and the most probable sub-parcels for each triangle are also obtained. Let us denote t_i with $i = 1, ..., n$ a triangle of the mesh, and consider the neighborhood of t_i as all the triangles that share a vertex or and edge with t_i. For each triangle, we count the number of times a sub-parcel appears in the neighborhood N_i, with $i = 1, ..., n$. To achieve this, for each triangle t_i, a list is created with the labels of the sub-parcels that intersect the triangle or

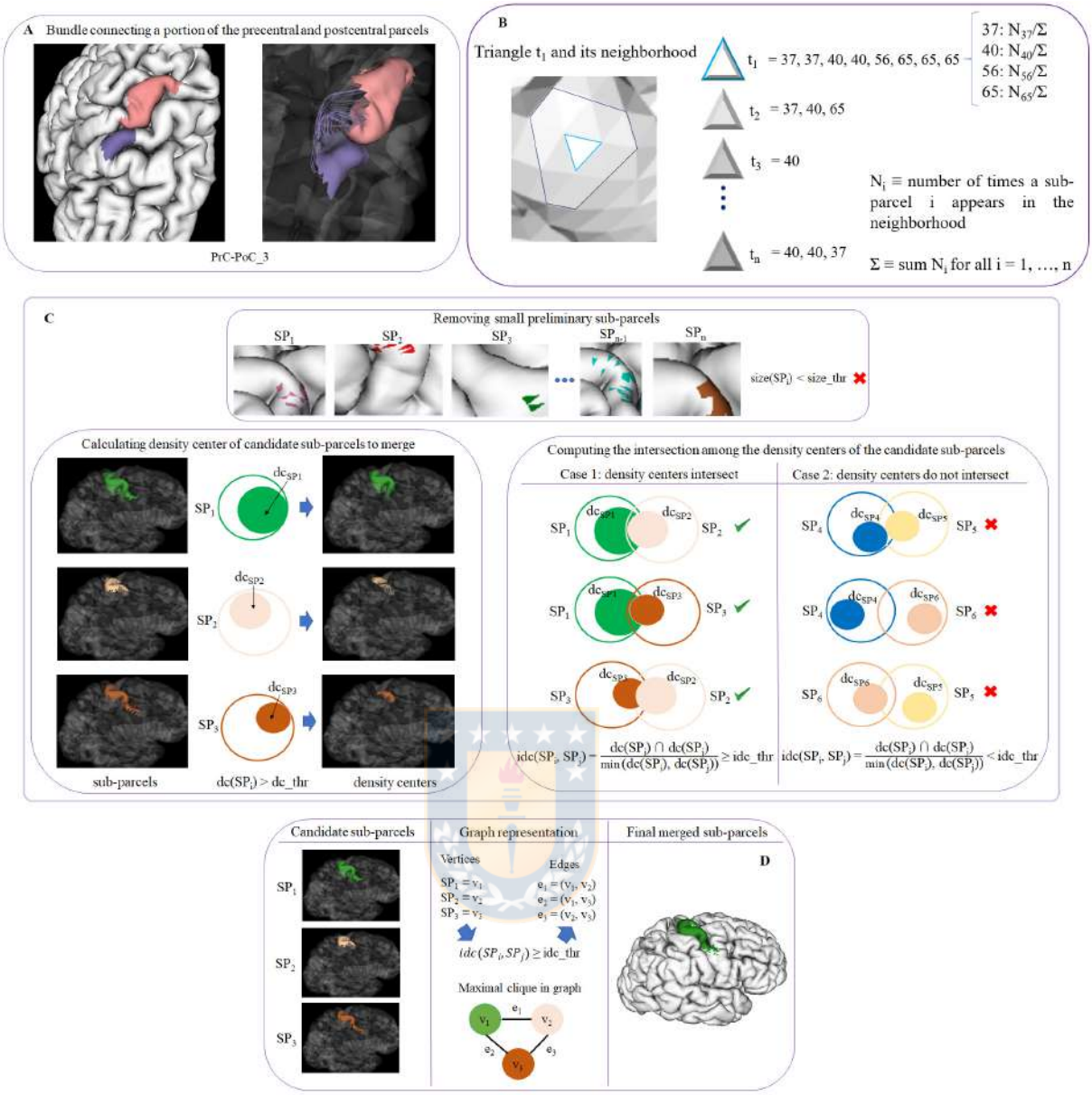


Figure 4.3: Schematics of cortex parcellation (STEP 5) sub-steps. **(A) Sub-step 5.1: Creating preliminary sub-parcels**. Preliminary sub-parcels are created based on the fiber bundle intersection and the labels of each triangle. **(B) Sub-step 5.2: Calculating probability maps**. The probability of each sub-parcel label in each triangle across the subjects is calculated. **(C) Sub-step 5.3: Processing preliminary sub-parcels**. This step deals with the preliminary sub-parcel overlapping. **(D) Sub-step 5.4: Merging of candidate sub-parcels**. This sub-step groups the sub-parcels to be merged.

its neighborhood. A label is added for each bundle fiber intersection, for each subject. Figure 4.3-(B) shows an example for a triangle t_1 and its neighborhood. For instance, t_1 has associated the list of sub-parcel labels that intersect the triangle and its neighbor triangles, for all the subjects. Each label has also associated the number of times the sub-parcel appears in the neighborhood (N_i). We also calculate Σ equal to the sum of all the counts N_i. To obtain the probability of each label in each triangle, the value N_i is divided by Σ. Finally, for each triangle, the list of probabilities is sorted in descending order.

Sub-step 5.3: Processing preliminary sub-parcels. The purpose of this step is to solve the overlapping that exists between the preliminary sub-parcels within each anatomical parcel of the cortex.

First, small preliminary sub-parcels are eliminated. We denote a preliminary sub-parcel as SP_i, with $i = 1, ..., n$ the label of the sub-parcel, and $size\_thr$ the threshold used to eliminate the smaller sub-parcels. The size is measured in terms of the number of triangles of the sub-parcel, as the areas of the mesh triangle are very homogeneous. The criterion of elimination $size\_thr$ is defined in terms of the percentage of triangles of the sub-parcel by respect to the corresponding anatomical region. After this processing, we get a set of *candidate sub-parcels* and recalculate the probability maps according to the reduced set of sub-parcels (see Figure 4.3-(C) top image).

The sub-parcels are highly overlapped, but if we look at their intersection density, we can observe that the overlapping can occur in regions with low density, or in a region of high density for one sub-parcel, but for low density for another sub-parcel. Indeed, the fiber intersection density is not homogeneous across the mesh surface for most of the sub-parcels. In fact, in most of the cases, only a portion of the sub-parcels presents a high density. Since a merging of the sub-parcels is required to obtain a hard parcellation, we calculate the density center of each sub-parcel to perform a better analysis. The density center is defined as the area where the highest concentration of fibers exists for each triangle. We denote the sub-parcel as SP_i with $i = 1, ..., n$ the label of each sub-parcel, and the density center as $dc(SP_i)$. Each triangle of a sub-parcel has associated a list with the probability of each sub-parcel present in the

triangle. Then, the density centers are defined using a minimum probability threshold of dc_thr (density center threshold) for $dc(SP_i)$. A sub-parcel can have several density centers spread over the sub-parcel. Some examples of parcel density centers are shown in Figure 4.3-(C) bottom left image. The left column displays three sub-parcels corresponding to the precentral anatomical parcel. In the middle column, we represent each sub-parcel SP_i as a circle of the same color and mark each density center with another filled circle inside. The third column shows the regions for $dc(SP_i) > dc\_thr$, corresponding to the density center of each sub-parcel.

Once the density centers have been calculated, we compute the intersection among them, for the candidate sub-parcels. The objective is to check if there is a significant overlap between the candidate sub-parcels to merge them. Given all the pairs of sub-parcels, we denote the density center of the first sub-parcel as $dc(SP_i)$ and the second sub-parcel as $dc(SP_j)$, being i and j the labels of each sub-parcel, with $i, j = 1, ..., n$ and $i <> j$. The intersection between each pair of sub-parcels is calculated based on the intersection of their density centers (triangles), following equation 4.1:

$$idc(SP_i, SP_j) = \frac{dc(SP_i) \cap dc(SP_j)}{min(dc(SP_i), dc(SP_j))} \tag{4.1}$$

To define a significant intersection, we use a threshold idc_thr (intersection of density centers threshold), where $idc \geq idc\_thr$ will define the sub-parcels that are candidates to merge. Figure 4.3-(C) bottom right image, illustrates an example of sub-parcel intersection analysis. The first column shows a case where three sub-parcels intersect between them, considering the intersection of the three pairs of sub-parcels. The second column shows the opposite case, where do not exist important overlaps between the sub-parcels, and thus they are not candidates to merge. Once all the candidate sub-parcels have been obtained, the next processing performs the merging of them.

Sub-step 5.4: Merging of candidate sub-parcels. In this step, the overlapping between sub-parcels is analyzed using a graph representation of the sub-parcels and their intersection, to merge the parcels that are significantly intersected with each other. More specifically, the objective is to find the groups of sub-parcels that are all intersected with each other within each group. This problem can be solved using a graph

representation of the sub-parcel intersections and a maximal clique algorithm.

Let $G = (V, E)$ be an undirected graph, where each vertex $v \in V$ represents a candidate sub-parcel. For each pair of candidate sub-parcels v_1 and v_2 in G, we create an edge between them $e = (v_1, v_2) \in E$ if the probability that the intersection of the density centers, $idc(SP_i, SP_j)$ (Eq. 4.1) is superior to the threshold idc_thr. Hence, the graph will contain only the relevant intersections between sub-parcels, i.e., where the density centers present a minimum percentage of overlapping. Once G is created, the idea is to obtain the vertices in the graph that are all connected with each other. Therefore, we use the graph algorithm called *clique* [67] that aims to find subsets of vertices that are adjacent (connected), and merge them in a single sub-parcel. We use a *clique* variant called *maximal clique*, which finds a clique with the largest possible number of vertices. The problem of finding maximal cliques is computational expensive (NP-hard) [135, 116], however, for sparse graphs the complexity is less [37]. After having calculated all the maximal cliques that are in G, these are sorted by size (number of vertices), in descending order. Following this order, the candidate sub-parcels of each maximal clique are merged to get the biggest number of fusions. This processing leads to the final sub-parcels, composed of merged candidate sub-parcels, candidate sub-parcels that were not merged, and the sub-parcels that were not candidates to merge (not included in the graph).

Figure 4.3-(D) shows an example of merging for three candidate sub-parcels of the precentral anatomical parcel. In the first column, the candidate sub-parcels, denoted by SP_i, are displayed. In the second column is included the graph representation of the intersections, in which each sub-parcel SP_i is a vertex v_i. If the idc (see Eq. 4.1) between a pair of sub-parcels is superior to the threshold idc_thr, an edge is created between both vertices. The graph G and the maximal clique are also graphically represented. The third column shows the final sub-parcel, resulting from the merging of the three sub-parcels.

Finally, the probability maps for each triangle are recalculated. Also, the most probable label is determined, with the purpose of obtaining a hard parcellation (see Figure 4.1-(F)).

STEP 6: Sub-parcel post-processing

The last step of the parcellation method deals with post-processing sub-parcels, to better define the final sub-parcels and the hard parcellation. It receives as input the mesh vertex labels of the parcel subdivision from the previous step. The post-processing is composed of three morphological operations performed over the cortical mesh.

Removing small connected components. The sub-parcels obtained in the previous step may be formed by more than one connected component. Some small components are in fact groups of a few triangles isolated from the main component. Hence, these small components are removed using a graph representation of each sub-parcel. The connected components of a graph can be easily calculated [119], and then ordered by size in descending order. Next, the largest connected component is kept. For each small connected component, the second most probable label in the list containing the probability map of the corresponding triangles is selected. The neighborhood of each connected component is then analyzed to verify if there exists a match between the second label of the triangle and its neighborhood. In most cases this value is appropriate, but if this is not the case, the label is removed. Figure B.5 shows an example of this processing for the supramarginal (SM) parcel, with three sub-parcels.

Sub-parcel opening. For each sub-parcel, the morphological operation called *opening* [60] is applied over the mesh, in order to eliminate isolated triangles that are scattered throughout the mesh. This operation is the result of the application of *erosion* + *dilation* operations. Hence, these two operations were sequentially applied.

Figure B.6 shows an example of results after applying the post-processing, with $size\_thr = 0.1$, $dc\_thr = 0.1$ and $idc\_thr = 0.1$. This hard parcellation or parcellation result, consists of 85 sub-parcels in the left hemisphere and 72 sub-parcels in the right hemisphere.

Parcellation method parameter settings

This section provides the parcellation method configuration parameters. The parcellation method has three configurable parameters for generating a hard parcellation: $size\_thr$, dc_thr and idc_thr. Note that all the parameters are adapted to the anatomical region and sub-parcel size, as are defined as percentages. The parameters are:

- **Minimum preliminary sub-parcel size threshold ($size\_thr$)**

  This parameter is used to eliminate small preliminary sub-parcels that do not exceed a certain size, concerning the average size of the sub-parcels of an anatomical parcel. We visually evaluated the results with different values of $size\_thr$, between 0.05 and 0.40. Big values, greater than 0.25, eliminate big preliminary sub-parcels, and therefore, leave uncovered some regions in the cortex. On the other hand, values inferior to 0.1 remove only very small preliminary sub-parcels. Thus, we selected a conservative value of $size\_thr$ = 0.10, which will only eliminate small sub-parcels, with a size inferior to the 10% of the average sub-parcel size on a region.

  Figure B.4 shows an example of *Removing of small preliminary sub-parcels* substep, belonging to Step 5 of the parcellation method for the precentral anatomical parcel (PrC), using $size\_thr$ = 0.10 and $size\_thr$ = 0.30.

- **Preliminary sub-parcel density center threshold (dc_thr)**

  This parameter determines the size of the density center (dc) of a preliminary sub-parcel. It defines the minimum percentage of probability of the sub-parcel in a triangle used to consider the triangle as part of the density center, and potentially be considered for the intersection analysis. We varied its value between 0.10 and 0.30. The higher the chosen value, the smaller the density centers are, and fewer intersections will be found when determining the intersection of the sub-parcels.

- **Intersection of density centers threshold (idc_thr)**

  This parameter defines the minimum intersection between the density centers of two sub-parcels to be considered overlapped, and hence, candidates to merge. This parameter is varied from 0.10 to 0.40. The lower the idc_thr, the more likely it is that sub-parcels will merge.

4.1.3. Experimental results

We implemented Steps 2 and 4 in C++11, which are also parallelized with OpenMP. The rest of the steps were performed in the Python programming language version 3.6. We executed our experiments on a computer with 12-core Intel Core i7-8700K CPU 3.70GHz, 12MB of shared L3 cache and 32GB of RAM, using Ubuntu 18.04.2 LTS with kernel 4.15.0-51 (64 bits). For the tests, we used the tractography datasets of the 79 subjects of the ARCHI database. We used the same database for the atlas creation and the parcellation creation to guarantee the best results.

Table 4.1 shows 20 parcellations (atlases) generated with the proposed method, resulting from different sets of parameters, with varying density center threshold ($dc\_thr = 0.10, 0.15, 0.20, 0.25, 0.30$) and intersection of density centers threshold ($idc\_thr = 0.10, 0.20, 0.30, 0.40$), and a fixed value of $size\_thr = 0.10$. The atlases are identified by a number. Also, the table lists the number of sub-parcels in the left hemisphere (# SP lh) and right hemisphere (# SP rh) obtained for each atlas. As dc_thr and idc_thr increase, the number of sub-parcels in both hemispheres increases, as more restrictive values are used to define a significant sub-parcel intersection, leading to an inferior number of merges.

A first observation is the asymmetry in the number of sub-parcels between both hemispheres, with more sub-parcels in the left hemisphere for all the parcellations. A similar result was found in the parcellation created by [76], based on the same database, with 126 parcels for the left hemisphere and 113 parcels for the right hemisphere. This could be due to a higher variability found in the left hemisphere, in the number of fibers across subjects. To have an insight on how much variability there is

atlas name	dc_thr	idc_thr	# SP lh	# SP rh
atlas 1	0.10	0.10	85	72
atlas 2	0.10	0.20	92	80
atlas 3	0.10	0.30	96	79
atlas 4	0.10	0.40	108	82
atlas 5	0.15	0.10	86	74
atlas 6	0.15	0.20	96	80
atlas 7	0.15	0.30	110	83
atlas 8	0.15	0.40	119	94
atlas 9	0.20	0.10	90	79
atlas 10	0.20	0.20	105	83
atlas 11	0.20	0.30	119	96
atlas 12	0.20	0.40	126	102
atlas 13	0.25	0.10	101	84
atlas 14	0.25	0.20	117	93
atlas 15	0.25	0.30	127	106
atlas 16	0.25	0.40	128	111
atlas 17	0.30	0.10	115	91
atlas 18	0.30	0.20	120	107
atlas 19	0.30	0.30	128	111
atlas 20	0.30	0.40	130	111

Table 4.1: Parameters used and the number of sub-parcels obtained per hemisphere for each configuration of parameters for the cortex parcellation method. The first column shows the name given to each atlas (parcellation result), based on a correlative number. The second and third columns list the different values for the density center (dc_thr) and the intersection of the density center (idc_thr) thresholds, for each generated atlas. Columns four and five list the number of sub-parcels obtained for the left (#SP lh) and right (# SP rh) hemispheres, respectively.

with respect to the population average, we calculated the coefficient of variation (CV) of the number of fibers in both hemispheres across subjects. CV measures the ratio between the standard deviation (σ) and the mean (μ), i.e., $CV = \sigma/\mu$. The resulting CV for the left hemisphere is 0.23, while for the right hemisphere is 0.21. One possible cause of the asymmetry in the number of sub-parcels may be that the higher variability in the number of fibers in the left hemisphere could produce a higher variability of

connections in the cortex, resulting in smaller parcels.

In the following subsections, we first perform a reproducibility analysis of the connectivity for the generated parcellations across subjects. Then, a comparison of the generated atlases is performed based on the similarity of their sub-parcels. Finally, a comparison is carried out with some state-of-the-art parcellations based on different modalities.

Reproducibility analysis

To test the consistency of the generated parcellations across the subjects, a reproducibilty analysis was performed [5]. For this purpose, for each subject, its tractography is taken and intersected with its mesh by means of the obtained parcellation. Afterwards, a binary connectivity matrix of size $n * n$ is created, where n is the total number of sub-parcels belonging to the resulting parcellation. If there is a connection between two sub-parcels, it is indicated with a one, otherwise it is indicated with a zero. All this procedure is shown in Figure 4.4. Finally, the Dice coefficient [34] is used to measure the similarity of the binary connectivity matrices across subjects for each obtained parcellation. The Dice coefficient measures the similarity between two sets (Eq. 4.2):

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{4.2}$$

where |A| and |B| are the number of elements of sets A and B, respectively, and DSC is the Dice coefficient. DSC ranges between 0 and 1, the closer to 1, the greater the similarity between the two sets.

To compute the Dice coefficient between two connectivity matrices, the *bctpy*[1] Python library was used, which is an adaptation of the Matlab Brain Connectivity Toolbox for Python [108]. For each generated parcellation, DSC was calculated between the connectivity matrices of each pair of subjects and then averaged (see Figure 4.5).

[1] https://github.com/aestrivex/bctpy

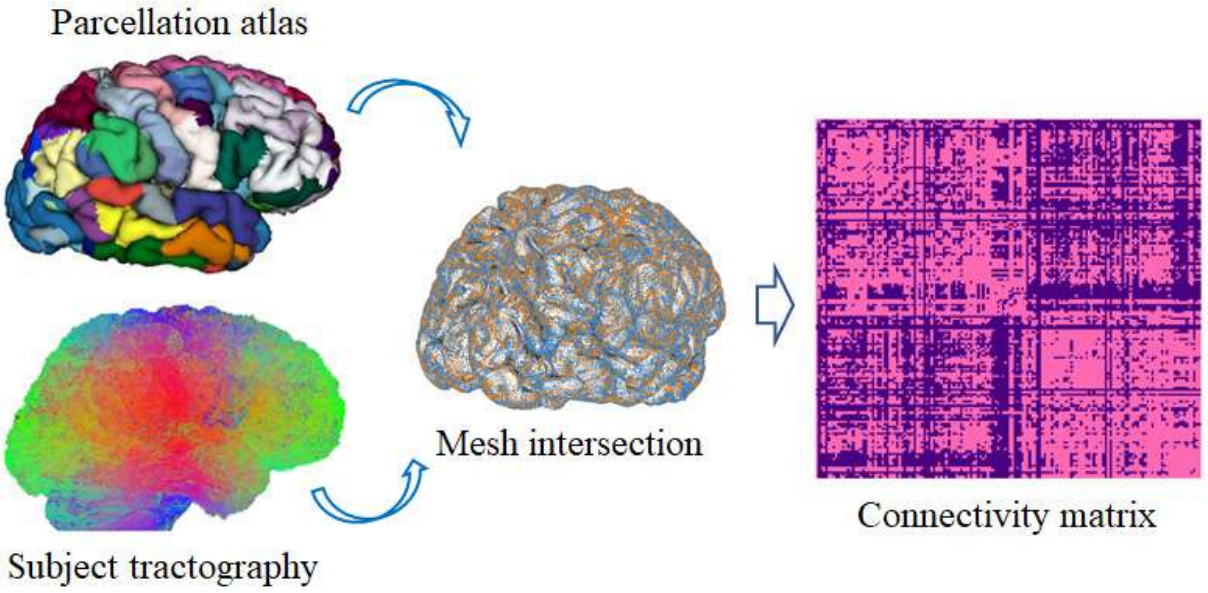


Figure 4.4: Brain connectivity analysis. First, the tractography of each subject is intersected with the subject's cortical mesh, using the generated cortical parcellation. Then, a square connectivity matrix $n*n$ is created, where n is the total number of sub-parcels in the parcellation. The matrix contains a 1 where there exists a connection between the pair of corresponding sub-parcels and zero in other case.

A cross-validation analysis was also performed an included in section B.7 of the appendix B.

The results show that there is no great variability between the generated atlases in terms of the similarity between the connectivity matrices obtained for the different subjects. In general, the similarity decreases with the number of sub-parcels, which is expected due to the relatively bigger effect of noise and inter-subject variability but is still good for a high number of parcels. Hence, the atlases with the least number of sub-parcels have the highest similarity between the subjects, which are *atlas 1* and *atlas 5*.

Additionally, we performed some tests based on network analysis [23, 13], provided in the appendix B. Section B.5 describes the graph construction, while section B.6 contains the network metrics calculation. These metrics are highly sensitive to the number of sub-parcels. For example, the results show a better small-world ω coefficient

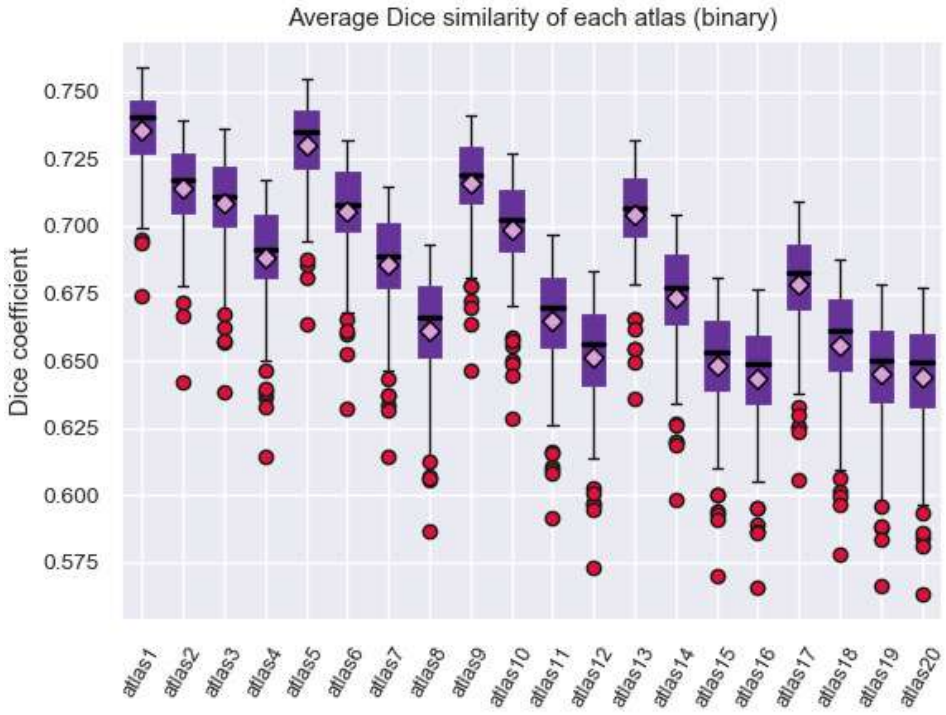


Figure 4.5: Average Dice coefficient for each parcellation (atlas) configuration, given by Table 4.1. The Dice coefficient was calculated between the connectivity matrices of each pair of subjects and then averaged. The results show a slight variability between the generated atlases. Due to the inter-subject variability, the similarity is smaller with a larger number of parcels.

for *atlas 1* and *atlas 5*. We selected the *atlas 5* as a candidate for comparison with other state-of-the-art methods, due to its high reproducibility, and as an example of atlas with a small number of sub-parcels. Table B.4 contains a complete description of the number of sub-parcels of *atlas 5*.

We have applied another criterion to select a parcellation, based on the atlas that is most similar to the remaining generated atlases. This atlas is in some way the most homogeneous atlas among all the atlases generated by the parcellation method. To select the atlas, we compared between them all the sub-parcels from the 20 generated atlases, by the construction of a similarity matrix. For each pair of atlases, we used the Dice coefficient to compare all the sub-parcels between the two atlases. To obtain the final result between two atlases, the Dice coefficient of all sub-parcels was

averaged, representing the degree of similarity between the two atlases. The closer the coefficient is to one, the more similar are the two atlases. Figure 4.7-left illustrates the similarity matrix for the 20 atlases. Finally, we selected the most reproducible atlas among the others, presenting the higher Dice coefficient on average, resulting to be the *atlas 13*. Table B.4 (fourth and fifth columns) illustrates the number of sub-parcels per hemisphere in *atlas 13*. Table B.5 shows a detailed description of the number of sub-parcels per hemisphere that have in common the generated atlases. Furthermore, Figure 4.7-right illustrates the sub-parcels that *atlas 13* has in common with the others atlases, based on a Dice coefficient $>= 0.6$ (47 sub-parcels in the left hemisphere and 41 sub-parcels in the right hemisphere).

As mentioned above, we selected *atlas 5* for comparisons with state-of-the-art atlases because it is an example of an atlas with a low number of sub-parcels, and high reproducibility. On the other hand, *atlas 13* was selected as the most similar atlas to the remaining generated atlases. *Atlas 5* contains 160 sub-parcels, while *atlas 13* contains 185 sub-parcels. Table B.4 of the appendix B lists the differences between *atlas 5* and *atlas 13* in terms of the number of sub-parcels per each *DK* atlas region. About 70% of the sub-parcels are similar. The differences, in general, refer to a subdivision of the sub-parcels.

Finally, we have selected an example to illustrate the biological significance of a pair of sub-parcels obtained by our parcellations. For this purpose, we have taken the sub-parcels of *atlas 13* that better match the most common definitions of Broca's [2] and Wernicke's [45] areas, related to language processing. As illustrated in Figure 4.6, these regions seem to correspond to the Broca's (in red) and Wernicke's (in green) areas. Besides, we illustrate the fibers connecting both sub-parcels, which correspond to the arcuate fasciculus, in agreement with the literature [20]. In fact, this fascicle is present in the fused bundle atlas used to create the parcellations, and the segmentation of this bundle is very stable across subjects. Hence, its connections define the described sub-parcels, also present in *atlas 5*. Further studies are required to validate in detail the biological significance of diffusion-based parcellations.

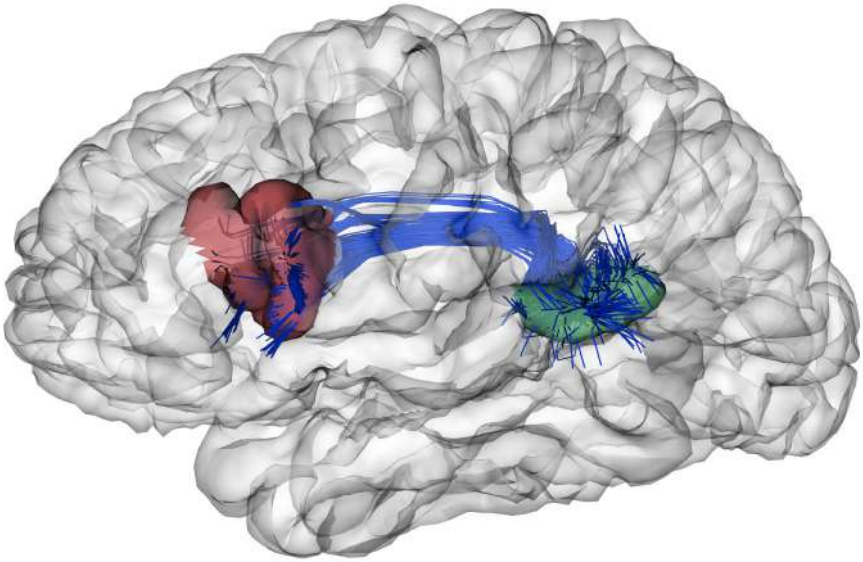


Figure 4.6: Example of two sub-parcels of *atlas 13* with biological relevance. These sub-parcels follow the most common definitions of the Broca's [2] (in red) and Wernicke's [45] (in green) areas, related to language processing. Also, the fibers connecting these sub-parcels are illustrated in blue, which correspond to the arcuate fasciculus [20].

Comparison with state-of-the-art parcellations

This section provides a comparison between *atlas 5* and *atlas 13*, generated by our parcellation, with other state-of-the-art parcellations based on different modalities. For the comparisons we used *Destrieux* atlas [32], based on macroanatomy with 150 parcels, and *Lefranc* parcellation [76], based on dMRI, containing 239 parcels. Based on a multimodal approach, we used *Brainnetome* atlas [38], which is composed of 210 cortical parcels and 36 sub-cortical parcels and *Glasser*'s atlas [46], which contains 360 parcels, 180 per hemisphere. In addition, based on functional MRI, we used *PrAGMATiC* atlas [64], which has 320 parcels, *Schaefer*'s atlases [111], consisting of several parcellations varying from 100 to 1,000 parcels, and *Yeo*'s atlas [139], which is formed by 7 or 17 networks. All the atlases are in MNI space, and are available in image format, by the exception of *Lefranc* that is available as a labeled mesh.

Hence, to be able to compare the parcellations in image format with our atlases, we first performed a mapping of the atlases to a cortical mesh. For each atlas, we labeled

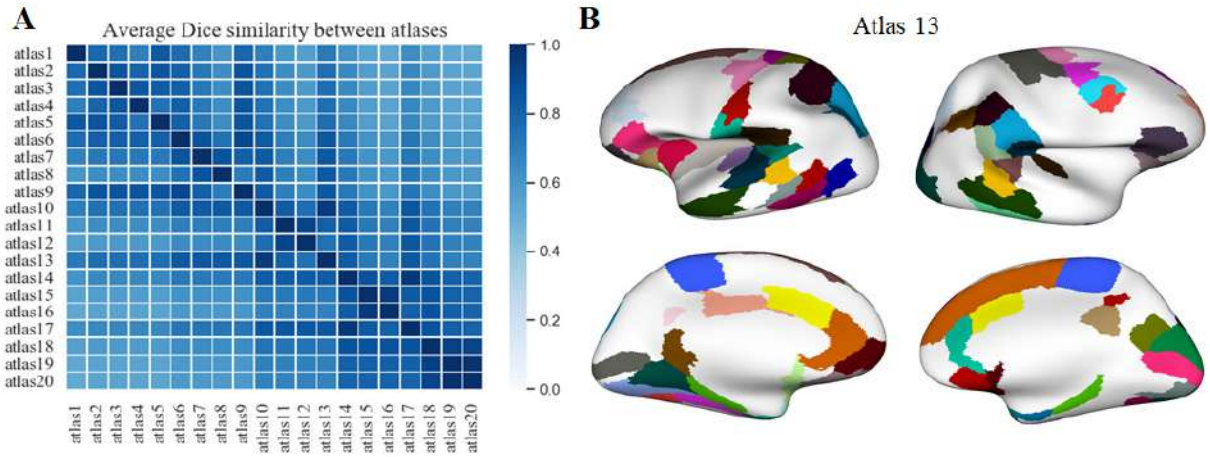


Figure 4.7: A comparison between the twenty atlases generated with the proposed method, based on different sets of the three parameters of the method. **(A)** Matrix obtained from the pairwise comparison of the sub-parcels of the different atlases, based on the average Dice coefficient (the closer to one, the more similar). The atlas most similar to the other atlases, i.e., the atlas with the higher mean Dice coefficient, is *atlas 13*. **(B)** A visualization of the sub-parcels of *atlas 13* that are in common with all the other parcellations, based on a Dice coefficient ≥ 0.6. The first column shows the left hemisphere with 47 sub-parcels in common, while the second column illustrates the 41 sub-parcels in common for the right hemisphere.

a cortical mesh in MNI space, by assigning to each mesh vertex the label of the closest voxel in the image atlas. Then, we compared our atlases with the other parcellations, by evaluating the similarity of each sub-parcel of *atlas 5* and *atlas 13* with each parcel of the other atlases, using the Dice coefficient over the mesh triangle labels. In this case, Dice's coefficient will evaluate the degree of overlap between a pair of parcels, ranging from 0, for a total dissimilarity, to 1, for a complete similarity. Moreover, to identify the sub-parcels in *atlas 5* and *atlas 13*, we named them based on the anatomical regions connected by these sub-parcels, based on *Desikan-Killiany* atlas. For example, the parcel lh_RMF-CMF-SF_0 connects RMF with CMF and SF regions. The number (_0) denotes the index of the sub-parcel, which depends on the number of sub-parcels in an atlas with the same connections. Prefixes *lh* or *rh* refer to the left or right hemisphere. For more details, Table B.3 in the appendix B lists the names of anatomical regions of *Desikan-Killiany* atlas. To perform the tests, both atlases, *atlas 5* and *atlas 13*, were applied to one subject in MNI space as a representative subject (Subject 001 from the

ARCHI database). The same subject was used for all the tests. Any other subject transformed to MNI space could be used, because there is a correspondence between the mesh triangles of all the subjects, which have the same triangle indexes.

Tables 4.2 and 4.3 contain the results of the comparison between *atlas 5* and *atlas 13* with the other atlases. For each state-of-the-art parcellation, the total number of parcels or networks that it contains, as well as the number of parcels similar to *atlas 5* and *atlas 13*, between different ranges of Dice coefficient, are indicated. In the table, the minimum Dice value considered is 0.5, while similarity values superior to 0.9 were not found.

Parcellation name	≥ 0.5 and < 0.6	≥ 0.6 and < 0.7	≥ 0.7 and < 0.8	≥ 0.8 and < 0.9
Brainnetome (210 parcels)	37 parcels	21 parcels	4 parcels	0 parcels
Destrieux (150 parcels)	26 parcels	8 parcels	2 parcels	1 parcel
Glasser (360 parcels)	31 parcels	7 parcels	2 parcels	0 parcels
Lefranc (239 parcels)	47 parcels	32 parcels	26 parcels	16 parcels
PrAGMATiC (320 parcels)	33 parcels	13 parcels	2 parcels	0 parcels
Schaefer (100 parcels)	26 parcels	9 parcels	2 parcels	0 parcels
Schaefer (200 parcels)	37 parcels	15 parcels	4 parcels	0 parcels
Yeo (7 networks)	0 parcels	0 parcels	0 parcels	0 parcels
Yeo (17 networks)	4 parcels	0 parcels	1 parcel	0 parcels

Table 4.2: Number of similar parcels found between *atlas 5* and parcellations based on different MRI modalities. For each parcellation, the total number of parcels or networks that compose it is listed, as well as the number of similar parcels based on Dice's coefficient. This coefficient ranges from 0 to 1, with 1 being total similarity. The number of similar parcels is divided into four groups, according to Dice's coefficient value.

Destrieux atlas was generated using 12 datasets and algorithms that classified each vertex in a computer-assisted manual manner and divided the cerebral cortex into 75 parcels per hemisphere, giving a total of 150 parcels. The comparison of *Destrieux* and *atlas 5*, obtained 37 similar parcels with Dice $>= 0.5$ in the two hemispheres. The most similar parcel is in the right hemisphere, corresponding to *G_cuneus*, located in the occipital lobe (cuneus) and corresponding to *rh_Cu-Li-MOF_0* of *atlas 5*. In the left hemisphere we have the parcels *S_suborbital* and *G&S_subcentral* corresponding to the frontal (suborbital sulcus) and parietal (subcentral gyrus) lobes. In *atlas 5* these

Parcellation name	≥ 0.5 and < 0.6	≥ 0.6 and < 0.7	≥ 0.7 and < 0.8	≥ 0.8 and < 0.9
Brainnetome (210 parcels)	40 parcels	26 parcels	5 parcels	0 parcels
Destrieux (150 parcels)	33 parcels	8 parcels	4 parcels	0 parcels
Glasser (360 parcels)	40 parcels	14 parcels	0 parcels	0 parcels
Lefranc (239 parcels)	55 parcels	35 parcels	27 parcels	13 parcels
PrAGMATiC (320 parcels)	43 parcels	19 parcels	3 parcels	0 parcels
Schaefer (100 parcels)	24 parcels	10 parcels	2 parcels	0 parcels
Schaefer (200 parcels)	31 parcels	22 parcels	4 parcels	0 parcels
Yeo (7 networks)	0 parcels	0 parcels	0 parcels	0 parcels
Yeo (17 networks)	4 parcels	0 parcels	0 parcels	0 parcels

Table 4.3: Amount of parcels in common between *atlas 13* and other parcellations based on the Dice coefficient. The first column lists the names for each parcellation and the number of parcels or networks that compose them. The other columns detail the total number of parcels in common for each Dice interval. This coefficient ranges between 0 and 1, being 1 the biggest similarity achieved.

parcels correspond to the sub-parcels *lh_MOF-LOF-LO_0* and *PoC-Ins-SM_0* respectively. The comparison between *Destrieux* and *atlas 13* obtains 45 similar parcels with Dice $>= 0.5$ between both hemispheres. We highlight from the left hemisphere, the parcel *S_suborbital* of *Destrieux*, which has its equivalent in *atlas 13* of the sub-parcel *lh_MOF-IC-PrCu_0*, located in the frontal lobe (suborbital sulcus). In the right hemisphere, the parcels *G_cuneus* and *G&S_subcentral* of *Destrieux* are similar to the sub-parcels *rh_Cu-Li_0* and *rh_PoC-SM_0* of *atlas 13*, and are located in the occipital (cuneus) and parietal (subcentral gyrus) lobes respectively.

With respect to the comparative with *Lefranc*, this parcellation has in common with our method, that was based on the same database (ARCHI) and uses the regions of *Desikan-Killiany* atlas as input. Furthermore, the method uses the whole dMRI tractography as input. *Lefranc* algorithm compresses the connectivity profiles of each gyrus, taking into account the inter-subject variability, and considering inter-subject high-density connectivity areas extracted using a surface-based watershed algorithm. Finally, it applies a clustering algorithm over the reduced connectivity profiles to obtain a group-wise parcellation, which consists of 239 parcels (126 in the left hemisphere and 113 in the right hemisphere). In the comparison made between *Lefranc* and *atlas*

5, we found 121 similar parcels with Dice $>= 0.5$. The most relevant parcel in the left hemisphere is *lh.caudalmiddlefrontal.1* which corresponds to the *CMF-PrC-RMF_1* sub-parcel in *atlas 5*, located in the frontal lobe (caudal middle frontal gyrus). The most similar parcels in the right hemisphere are *rh.inferiorparietal.3* and *rh.precuneus.2*, both belonging to the parietal lobe, specifically, the inferior parietal cortex and precuneus cortex. They correspond in *atlas 5* to the sub-parcels *IP-IT-MT_0* and *PrCu-CAC-PoCi-SF_0* respectively. Respecting *Lefranc* and *atlas 13*, they have 130 parcels in common. The most similar parcels in the left hemisphere are *lh.supramarginal.2* and *lh.postcentral.3*. Both parcels are located in the parietal lobe, namely in the supramarginal and postcentral gyri. The equivalent sub-parcels in *atlas 13* are *lh_SM-PrC_0* and *lh_PoC-Ins-SM_0*. In the right hemisphere, we found *rh.inferiorparietal.2* parcel, which is located in the parietal lobe (inferior parietal cortex) and is similar to sub-parcel *rh_IP-SM-PrC_0* in *atlas 13*.

Brainnetome atlas relies on a multimodal approach. Multimodal information consists of diffusion MRI, functional MRI, and structural MRI data. This parcellation was based on 80 subjects of the Human Connectome Project (HCP) database and contains 210 cortical parcels (105 per hemisphere) and 36 subcortical parcels. This atlas has in common with *atlas 5* and *atlas 13* that uses dMRI, but employs probabilistic rather than deterministic tractography. Furthermore, both methods use the *Desikan-Killiany* atlas as input information. *Atlas 5* and *Brainnetome* have 62 parcels in common. We can highlight *parcel A9_46d_L* (left hemisphere) located in the frontal lobe (middle frontal gyrus), which is linked to *atlas 5* sub-parcel called *lh_RMF-CMF-SF_0*, and is related to inhibition, social cognition, and word generation. In the right hemisphere, parcel *msOccG_R* of *Brainnetome*, located in the occipital lobe (lateral occipital cortex), is similar to sub-parcel *rh_SP-LO-MT_0* of *atlas 5*, related to spatial ability, shape vision, motion vision and inhibition. Also, we have parcel *A23c_R* that is similar to sub-parcel *rh_PoCi-CAC-PrCu-RAC_0*, located in the limbic lobe (cingulate gyrus), which is related to emotions, reward, and pain. Moreover, between *atlas 13* and *Brainnetome* there are 71 parcels in common. The three most similar parcels correspond to the left hemisphere. *Brainnetome* parcel *A9_46d_L* corresponds in *atlas 13* to the sub-parcel *lh_RMF-CMF-SF_0* (middle frontal gyrus) and is related to inhibition, social cognition

and word generation. Parcel *A40rv_L* is linked to sub-parcel *lh_SM-PrC_0* in *Atlas 13* and is located in the parietal lobe (inferior parietal) and has functions related to audition, pain, grasping and discrimination. Finally, parcel *A8dl_L* corresponds to sub-parcel *lh_CMF-Op_0* of *atlas 13*, is located in the frontal lobe (superior frontal gyrus) and is related to emotion, cognition and memory.

Glasser atlas is also based on a multimodal approach. This parcellation is based on functional connectivity (resting-state), microstructural architecture, functional specialization (task-fMRI), and topography information. In addition, it uses data from 449 subjects of the HCP database and generates a final parcellation which consists of 360 parcels (180 per hemisphere). We found 40 parcels in common with *atlas 5*. In the left hemisphere, sub-parcel *lh_Or-LOF-Ins-LO_0* of *atlas 5* corresponds to parcel *L_a47r_ROI* located in the frontal lobe (Orbital and Polar Frontal Cortex) of *Glasser* and is linked to relational-match, gambling, working memory, language (story and math) and faces-shapes recognition. Continuing in the same hemisphere, another relevant *Glasser* parcel is *L_POS1_ROI*, located in the parietal lobe (Posterior Cingulate Cortex), which is similar to sub-parcel *lh_PrCu-PH-En_0* and is related to language (story and math) and scene selection. As for the right hemisphere, *R_V3A_ROI* parcel which is located in the occipital lobe (Dorsal Stream Visual Cortex) is equivalent to *rh_SP-LO-MT_0* sub-parcel of *atlas 5*, which is related to retinotopic areas, gambling, and emotion. The comparison between *Glasser* and *atlas 13* led to 54 parcels in common. The most similar parcels in the left hemisphere for *Glasser* are *L_TPOJ1_ROI* and *L_11l_ROI*, which are linked to the sub-parcels *lh_Ban-MT_0* and *LOF-LO_0* of *atlas 13*. The former is located in the Temporo-parieto-occipital junction, area associated with faces-shapes recognition, language (story and math), audition, visual concepts, and gambling. *LOF-LO_0* parcel is in the frontal lobe (Orbital and Polar Frontal Cortex) and has the functionalities of memory and faces-shapes recognition. Regarding the right hemisphere, *R_V3A_ROI* parcel of *Glasser* has its equivalent in the sub-parcel *rh_IP-IT_0* of *atlas 13*, located in the occipital lobe (Dorsal Stream Visual Cortex) and is related to gambling, emotion and retinotopic areas.

PrAGMATiC atlas is the result of using fMRI and ROI-based methods in a probabilistic and Bayesian generative model approach. The model was applied using 12 subjects, and the resulting atlas, containing 320 parcels, represents the distribution of semantically selective functional areas in the human cerebral cortex. The comparison made between *PrAGMATiC* and *atlas 5* found 48 common parcels. We highlight in the left hemisphere parcel *IPFC_L8* of *PrAGMATiC* located in the frontal lobe (inferior prefrontal cortex), which is similar to sub-parcel *lh_Tr-Ins-SF-IT_0* of *atlas 5* and is related to the violence, emotions, social, and abstract skills. Continuing in the same hemisphere, we have the parcel *IPFC_L12* located in the frontal lobe, similar to sub-parcel *lh_LOF-ST-TEM-LO_0* of *atlas 5*, related to abstract, tactile and numeric skills. On the other hand, in the right hemisphere, parcel *LTC_R3* of *PrAGMATiC* is equivalent to sub-parcel *rh_Ban-IP_0* of *atlas 5* located in the temporal lobe (lateral temporal cortex) and is related to violence, social and emotion skills. *PrAGMATiC* has 65 parcels in common with *atlas 13*. The *LTC* parcel in the left hemisphere is the most similar, located in the temporal lobe (lateral temporal) and equivalent to sub-parcel *lh_Ban-MT_0* of *atlas 13*, with associated functionalities such as violence, emotional and communal skills. In the right hemisphere we highlight the parcels *LTC_R3* and *SPFC_R15*, located in the temporal (lateral temporal) and frontal (superior prefrontal) lobes. The former has the functionalities of violence, social, and emotional concepts, while the latter is associated with mental, emotional, and violent concepts.

Schaefer is a parcellation based on resting-state fMRI from a database of 1,489 subjects, which uses a gradient-weighted Markov Random Field (gwMRF) model to generate the final parcellations, with 100, 200, 400, 600, 800 and 1,000 parcels. To compare *Schaefer* atlas with our results, we chose the parcellations of 100 and 200 parcels because they are the most similar in number of parcels to *atlas 5* (160 sub-parcels) and *atlas 13* (185 sub-parcels). Therefore, starting with *Schaefer* parcellation with 100 parcels, we found 37 parcels in common with *atlas 5*. In the left hemisphere, parcel *DefaultB_IPL_1* of *Schaefer* is similar to sub-parcel *lh_IP-IT-MT_0* of our atlas and is located in the temporal lobe, while in the right hemisphere, parcel *ContB_IPL_1* is equivalent to sub-parcel *rh_SM-PrC-SP_0* and is located in the frontal lobe. Both

parcels are associated to language skills (story and math). We also highlight sub-parcel *rh_RMF-LOF-SF-LO_0* of *atlas 5*, which corresponds to parcel *ContB_PFCl_1* of *Schaefer* located in the parietal lobe, and is related to the working memory.

On the other hand, the *Schaefer* parcellation of 200 parcels, has in common a total of 56 parcels with *atlas 5*. In the left hemisphere, parcel *Temp_Par_1* (temporal lobe) from *Schaefer* is similar to sub-parcel *lh_Ban-IT-MT_0* of our atlas, which has associated language functionality (story and math). For the right hemisphere, parcel *TempPar_3* (temporal lobe) of *Schaefer* is similar to *rh_Ban-IP_0* sub-parcel of *atlas 5*, which is related to language (story and math), and *VisCent_ExStr_3* parcel is similar to *rh_LO-Or-MT-RMF_0* sub-parcel, which is located in the occipital lobe and is associated to visual areas and relational skills (matching and fixation). Moreover, the comparison between *Schaefer* parcellation *atlas 13* found 57 similar parcels. In the left hemisphere the most similar is the parcel *LimbicA_TempPole_1* located in the temporal lobe that is related to the sub-parcel *lh_TEM-LOF-MOF-LO_0* of the *atlas 13* and has as associated functionality the language (story and math). For the right hemisphere, *TempPar_3* (temporal lobe) parcels is equivalent to sub-parcel *rh_Ban-IP_0* of *atlas 13*, asociated to language functionalities (story and math), and parcel *VisCent_ExStr_3* (occipital lobe) is similar to *LO1-LO0-MT_1* sub-parcel, associated to relational skills functionalities (matching and fixation).

Yeo atlas was based on fMRI data from 1,000 subjects. The comparison with this atlas leads to less similar parcels, since the sub-parcels of *atlas 5* and *atlas 13* are generally smaller than the 17 networks of *Yeo* atlas. Five parcels were found in common between *Yeo* atlas and *atlas 5*, and four parcels between *Yeo* atlas and *atlas 13*. We highlight sub-parcel *lh_LO1-LO0-ST-MT_1* from *atlas 5* located in the left occipital lobe, which is related to $V1_c$ region of *Yeo* atlas and is associated to the visual area and central vision.

Figure 4.8 illustrates the parcels found similar between *atlas 5* and other parcellations, mainly based on functional information (*Brainnetome*, *Schaefer* and *Glasser*), considering a Dice coefficient ≥ 0.6. Figure 4.9 shows the common parcels between *atlas 13* and *Destrieux*, *Lefranc* and *Brainnetome* parcellations, with a Dice coefficient

≥ 0.6. For more information on comparisons see section B.9 of the appendix B.

The evaluation of the differences between the atlases based on dMRI (*Lefranc*, *Brainnetome*, and our parcellations) is a difficult task. The coarse anatomical regions in which the atlases were based is the main difference. *Lefranc* and our parcellations present a high dependency on the anatomical regions of *Desikan-Killiany* atlas (35 per hemisphere), while *Brainnetome* uses *DK* atlas but with several regions combined (20 cortical and 4 subcortical regions per hemisphere). Hence, we can compare the granularity of the *DK* regions for the different atlases, where a higher difference exists for *Brainnetome* in the combined regions. Table B.6 of the appendix B lists the number of subdivisions of *DK* anatomical regions for all the atlases. In some cases, *Brainnetome* parcels cannot be matched with DK standard regions. In the table, an asterisk is used to indicate the *DK* anatomical parcels where *Brainnetome* performs a different subdivision of the regions and only an approximate number of subdivisions is provided. Another big difference is that *Brainnetome* has equivalent parcels in both hemispheres, while *Lefranc* and our parcellations are asymmetric, presenting more sub-parcels in the left hemisphere. This is due to the different approaches used, where the method that created *Brainnetome* used, in addition to stability across the population, the interhemispheric anatomical homology. Another difference is the number of total parcels, where *Lefranc* has 239, *Brainnetome* 210, *atlas 13* 185, and *atlas 5* 160 cortical sub-parcels. Hence, depending on the application, the total granularity could be determinant for the atlas selection. Furthermore, *Lefranc* presents subdivisions in almost all the *DK* atlas regions, which is not the case for the other atlases. This could be produced by the watershed algorithm applied to the cortical surface, which may be more sensitive to local connectivity density variations. Also, this atlas presents a high granularity in some regions, such as the Fusiform, Lateral occipital (left), Middle temporal, Pars orbitalis, Pericalcarine, and Transverse temporal. *Brainnetome*, on his side, presents more subdivisions for the Inferior temporal, Superior frontal, and Insula regions. Finally, the four atlases present higher subdivisions for the Superior temporal, Superior frontal, Precentral, Postcentral, and Inferior temporal gyri. Furthermore, *Brainnetome* is the only atlas that has subcortical parcels (18 per hemisphere).

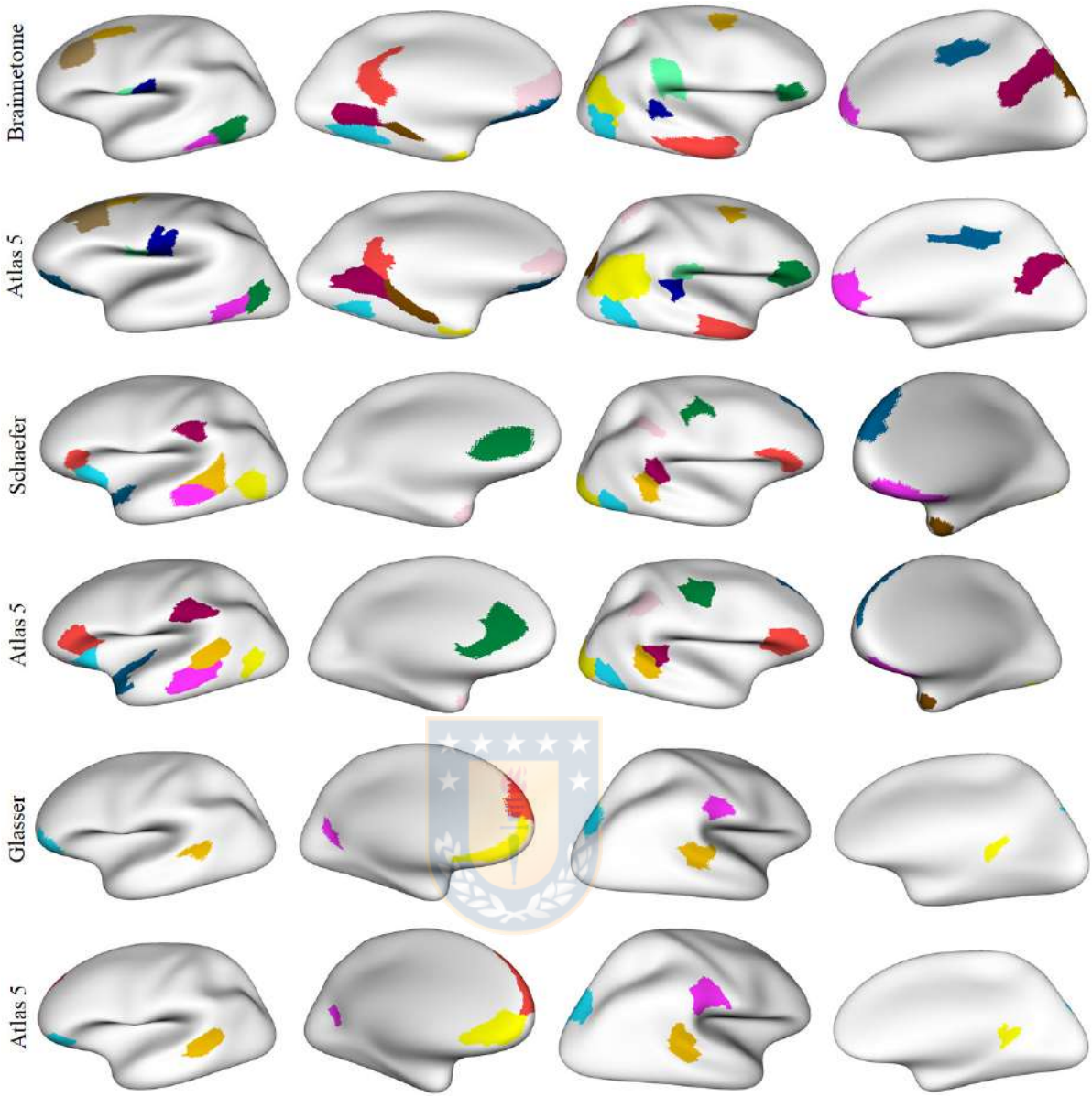


Figure 4.8: Parcels in common between *atlas 5* parcellation and some parcellations based on different MRI modalities, with Dice coefficient ≥ 0.6. Both hemispheres are shown for each parcellation with the inflated mesh. First and second rows: comparison with *Brainnetome* (210 cortical parcels) [38], 13 similar parcels were found in the left hemisphere and 12 in the right hemisphere. Third and fourth rows: comparison with *Schaefer* parcellation (200 parcels) [111], with 9 similar parcels in the left hemisphere and 10 in the right hemisphere. Fifth and sixth rows: comparison with *Glasser* parcellation (360 parcels) [46], with 5 similar parcels in the left hemisphere and 4 in the right hemisphere. This gives a total of 25, 19, and 9 parcels in common, respectively.

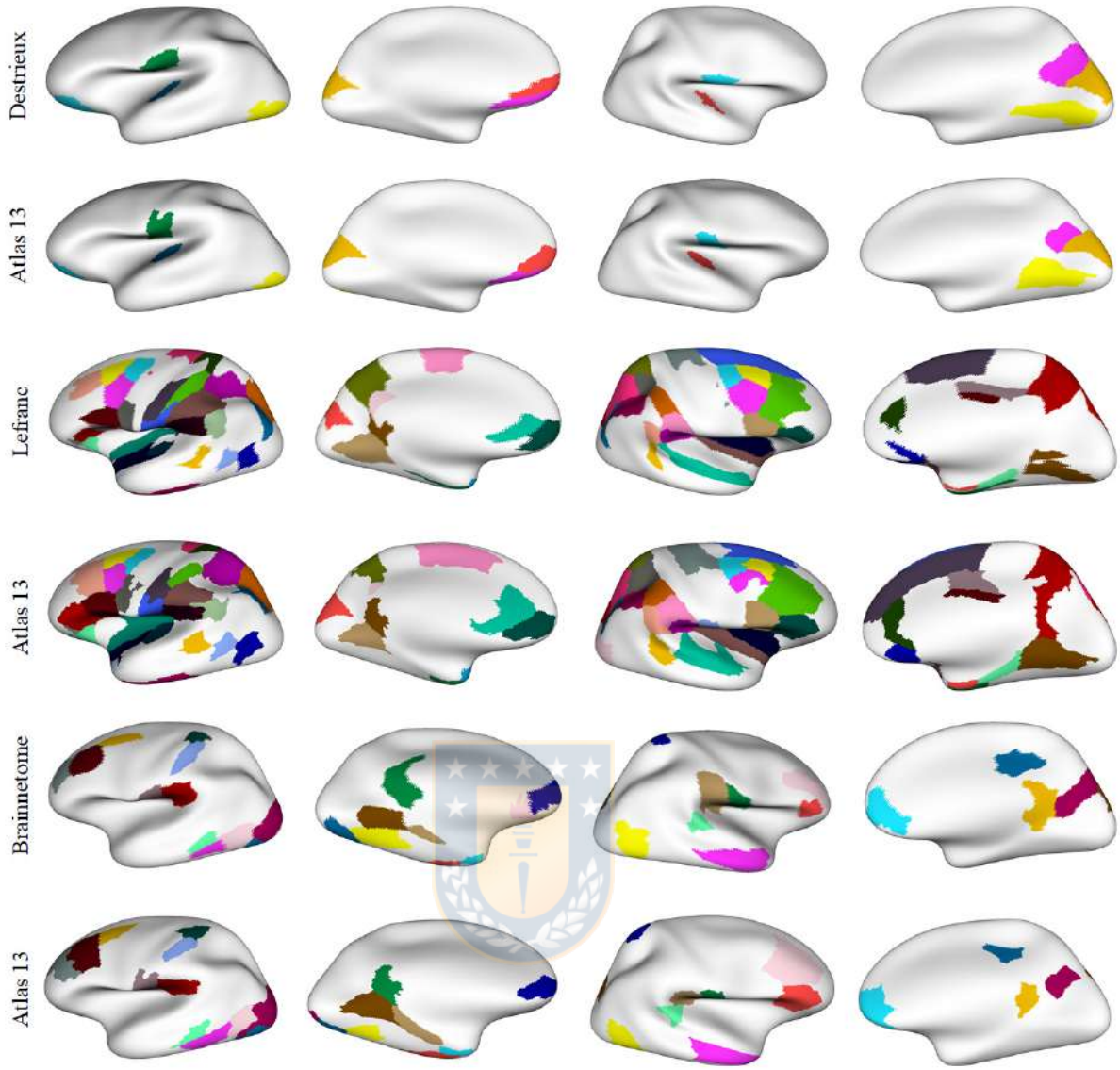


Figure 4.9: Common parcels between *atlas 13* parcellation and some parcellations from the state of the art with Dice coefficient ≥ 0.6. Both hemispheres are shown for each parcellation with the inflated mesh. First and second rows: comparison with *Destrieux* atlas (150 parcels) [32], with 7 parcels in common in the left hemisphere and 5 in the right hemisphere. Third and fourth rows: comparison with *Lefranc* atlas (239 parcels) [76], with 40 common parcels in the left hemisphere and the 35 parcels in the right hemisphere. Fifth and sixth rows: comparison with *Brainnetome* (210 cortical parcels) [38], with 19 parcels in common in the left hemisphere and 12 parcels in the right hemisphere. This gives a total of 12, 75, and 31 similar parcels, respectively.

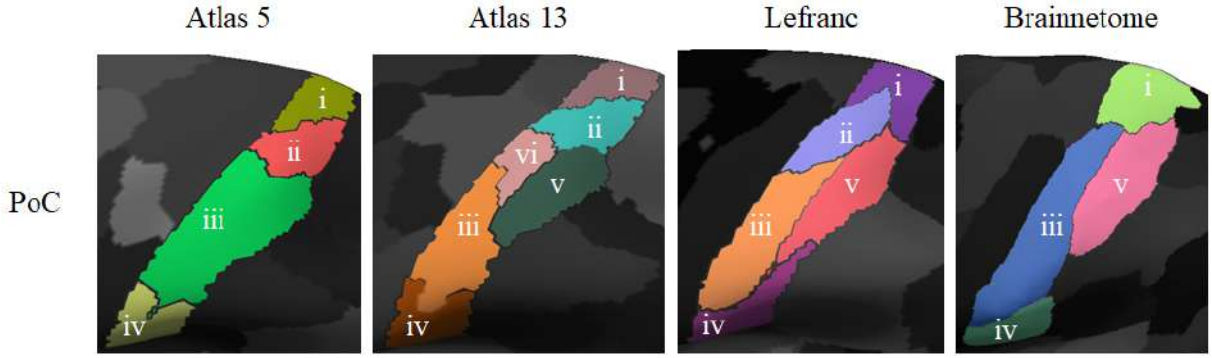


Figure 4.10: Comparison of the sub-parcels obtained by the different atlases based on dMRI for the postcentral (PoC) anatomical parcel. From left to right the atlases are: *atlas 5* (four subdivisions), *atlas 13* (six subdivisions), *Lefranc* (five subdivisions) and *Brainnetome* (four subdivisions). The sub-parcels were enumerated following the best correspondence between atlases. It can be observed that sub-parcels *iii*, *v*, and *vi* of *atlas 13* are a subdivision of sub-parcel iii of *atlas 5*. Also sub-parcels *i* and *iv* are similar in all the atlases. Furthermore, sub-parcels *v* in *atlas 13*, *Lefranc* and *Brainnetome* are very similar.

Making a comparison between the atlases based on dMRI modality, we found similarities and differences in the number of subdivisions per anatomical parcel and per hemisphere. Figure 4.10 shows a visual comparison between *atlas 5*, *atlas 13*, *Lefranc*, and *Brainnetome* for the postcentral anatomical parcel. The sub-parcels found were enumerated according to their correspondence between the different atlases. Sub-parcels *i* and *iv* are similar for all the atlases. In addition, sub-parcel *iii* of *atlas 5* is divided into sub-parcels *iii*, *v*, and *vi* of *atlas 13*. On the other hand, there is a high similarity between sub-parcels *v* in *atlas 13*, *Lefranc*, and *Brainnetome*.

Finally, we compared the connectivity matrices obtained for *atlas 5* (160 sub-parcels) and *Destrieux* atlas (150 parcels), which is based on macroanatomy. First, the connectivity matrices of each subject for both atlases were calculated (79 subjects). Then, the matrices were binarized and the Dice coefficient was calculated between each pair of subjects and posteriorly averaged for each atlas, to compare the reproducibility of the connectivity matrices generated by both atlases. Figure 4.11 shows the results of the Dice coefficient for both parcellations. As shown, *atlas 5* parcellation is a little more reproducible (≈ 0.01) than *Destrieux* atlas, despite having 10 more parcels. In general,

the higher the number of parcels, the lower the value of inter-subject reproducibility using Dice's coefficient, as subdividing the cortex into a larger number of sub-parcels will lead to more variable connectivity, due to inter-subject variability. With the obtained result, it seems that the boundaries of the sub-parcels produce a better agreement with the underlying connections.

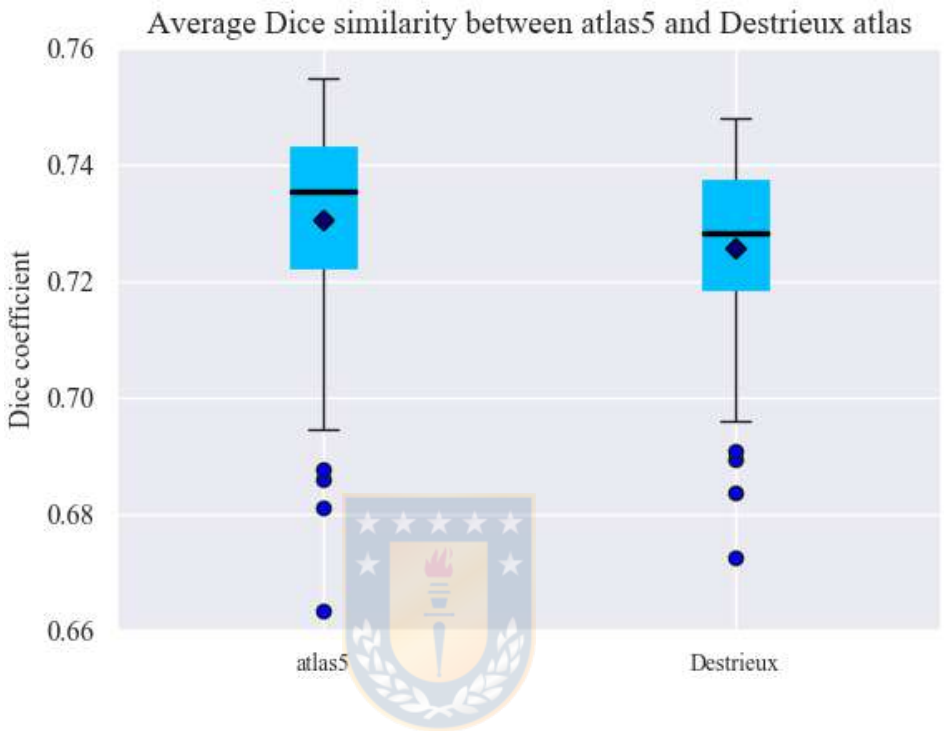


Figure 4.11: Dice coefficient for connectivity reproducibility for *atlas 5* parcellation and *Destrieux* atlas. Dice's coefficient is in the range 0 to 1, the closer to 1 the more reproducibility between subjects. *Atlas 5* composed of 160 sub-parcels is slightly better in terms of reproducibility than the *Destrieux* parcellation consisting of 150 parcels.

4.1.4. Discussion

We propose a new hybrid method for the creation of fine-grained parcellations of the cortical surface from a coarse-grained anatomical parcellation, based on the connectivity given by a fiber-bundle atlas. Since the bundles have a correspondence between subjects, a direct match is obtained between the regions intersected by the extremities of the bundles across subjects. However, due to the overlap of cortical bundle intersections, inter-subject variability, and tractography limitations, several processing steps

are applied to find consistent parcels among subjects.

The main analysis uses the probability of each sub-parcel along with the fiber density over the cortex to detect reproducible regions. Also, the overlapping between regions is solved using a graph representation of sub-parcel density center intersections. The method has the advantage of being conceptually simple, despite the complexity of its implementation, with few parameters that represent characteristics that are also easy to understand. Therefore, parameter variation has an understandable effect on the final parcellation, in particular, in the number of sub-parcels. Results are very promising, showing an expected behavior of the method for a wide range of parameters and a high similarity between the generated atlases. Even though the final number of sub-parcels per hemisphere depends on the parameter configuration, there is a high dependence on the maximum number of sub-parcels with the used bundle atlas. This is the reason why the atlas *swm_atlas_1* [50] was chosen first, as it contained much more compact bundles at its extremities, leading to more candidate sub-parcels. Furthermore, the optimal number of sub-parcels and the method itself will depend on the subsequent analysis to be performed. If the objective is to analyze and compare structural connectomes, it seems convenient to use connectivity-based parcellations created from tractography data.

In general, our method leads to good inter-subject correspondence in all the created parcellations, given by a relatively high average Dice coefficient for connectivity matrices. The comparison with *Destrieux* atlas showed a slightly better reproducibility for *atlas 5*, despite having 10 more sub-parcels. In any case, the contribution of this work is a method for the creation of a fine-grained parcellation from an anatomical coarse parcellation, based on a bundle-atlas that can be used for further analyses.

The comparisons between *atlas 5* and *atlas 13* with some atlases based on different MRI modalities found a set of similar parcels, with a Dice coefficient greater than or equal to 0.5. The comparison with parcellations based on fMRI gave some insights on the functions related to some sub-parcels obtained for our atlases, which in turn are associated with specific structural connections. Even though the objective of the present work is to propose a diffusion-based parcellation, the comparison shows that

good correspondence is found between several sub-parcels of our atlases and parcels from other modalities. For comparison, the size of the parcels is crucial. For example, for *Yeo* atlas with 7 networks, no similar parcels were found due to the higher size of the networks. On the other hand, $\approx$ 51% of the parcels in *Lefranc* atlas (with 239 parcels), based on dMRI, have a similar sub-parcel in *atlas 5* (with 160 sub-parcels). The comparison between *Lefranc* atlas and *atlas 13* achieves $\approx$ 54% of common parcels (130 sub-parcels) being the most similar parcellation. This result is interesting since the proposed method is based on the same database than *Lefranc* atlas, but with a totally different approach. Despite the different number of parcels, two other parcellations have more than 30% of similarity with our atlases, which are *Brainnetome* and *Schaefer* with 100 parcels. Of course, we are not considering a perfect match between the parcels. Further analyses could be performed, in particular, to compare the different state-of-the-art parcellations, but are out of the scope of this work.

A limitation of the method may be the use of the fused fiber atlas, composed of SWM and DWM bundles, to generate the input data, instead of using the whole-brain tractography. However, all the diffusion-based methods perform at some stage a filtering of the data, since it is necessary to keep only the reproducible regions or connections. A concrete limitation is the maximum number of sub-parcels that could be created, which depends on the final atlas bundles. We found a total number of sub-parcels ranging from 157 to 241, which is around the number of parcels obtained by the state-of-the-art methods based on tractography, with 15-250 parcels for [89], 239 parcels for [76] and 50-300 parcels for [96]. To reach a higher number of parcels it would be necessary to add more bundles or subdivide the current bundles. Another limitation of the method is the use of *Desikan-Killiany* atlas to define the coarse granularity of the sub-parcels, instead of generating a parcellation without such limits. Also, the whole method is difficult to reproduce due to the use of different platforms and methods, which is not infrequent in this kind of analysis. For that reason, we have created a code repository[2] with the necessary codes and files to apply all the processing, including the fused bundle atlas, and the segmentation, intersection and parcellation codes, among others. Furthermore, the resulting data will be available to the public in a data repository.

[2]https://github.com/andvazva/Parcellation

On the other hand, the use of the fused fiber bundle atlas is an advantage, since it allows a direct correspondence between subjects, avoiding the search for such correspondence at the end of the process employing clustering algorithms. Other positive aspect is the low execution time, where the segmentation algorithm is capable of segmenting a subject of 1,500,000 fibers in less than 20 seconds and the cortical parcellation algorithm performs the subdivision of the anatomical DK parcels in approximately 10 minutes. Furthermore, this algorithm has only three configurable parameters that allow the generation of parcellations with a smaller or larger number of sub-parcels.

In any case, the limitations of diffusion MRI should always be considered when analyzing results based on dMRI tractography. This technique is used to non invasivelly reconstruct the major white matter tracts of the brain. Tractography algorithms are able to generate valid bundles, however, due to the limited spatial resolution of the voxels and the large amounts of fiber pathways that can pass through them, false positives and false negatives are also generated. In fact a non negligible number of false positive bundles is produced, some of them reproducible across subjects. One of the next challenges of tractography will be to control these false positives and improve the spatial reconstruction of existing WM tracts [82]. Therefore, special care must be taken when interpreting the results given by tractography algorithms in a study [15]. The differences in connectivity profiles can be produced due to artifacts in the tractography. Hence, the parcellations based on diffusion tractography are only valid when the differences in connectivity profiles reflect true anatomical differences [15]. We have shown one example of biological significance, for the Broca's and Wernicke's areas, but further studies need to be performed to validate the diffusion-based parcellations.

4.2. Cortical surface parcellation based on intra-subject white matter fiber clustering

In this section, we present a hybrid method to perform the cortical parcellation of an individual based on the information provided by the connectivity of the white matter fibers. The first section shows the work related to the parcellation of the cortex. The proposed method consists of five steps, which perform the complete labeling of cortical mesh vertices, representing the different cortex sub-parcels, with strong connections to other sub-parcels. In the experimental results, we evaluated the method with brain connectivity metrics. Finally, we present the discussion of the method and the results obtained. This is a preliminary implementation and requires future improvements and validations.

4.2.1. Related work

Advances in brain imaging have allowed the study of the structure and connectivity of white matter (WM), a research area that is constantly growing. One of the most used techniques to understand the anatomical connectivity of the brain is the diffusion-weighted Magnetic Resonance Imaging (dMRI). It is a non-invasive and in-vivo technique, based on measurements of the movement of hydrogen molecules present in water [74]. Tractography algorithms use dMRI information to estimate the main trajectories of the WM tracts [92]. When applied to the whole-brain, resulting datasets contain a large amount of 3D polylines, called fibers, that represent the main brain WM connectivity.

Understanding how the brain works requires a detailed description of the network of connections that form it [118]. A cortical parcellation represents a way to divide the brain cortex into macroscopic regions, according to their structure or functioning, in order to study brain connectivity [28]. The best known parcellation is Brodmann's atlas, based on post-mortem cytoarchitecture study, focused on the size, density, shape and distribution of cell bodies in cortical layers [3]. In contrast, in-vivo techniques based on MRI, enable the development of other parcellations, based on anatomical structures [32], functional MRI (fMRI) [111] or a multi-modal approach [46]. Performing a cortical

parcellation is a difficult task due to the high variability that exists between subjects in terms of white matter and gray matter, as well as the disadvantages of each imaging modality.

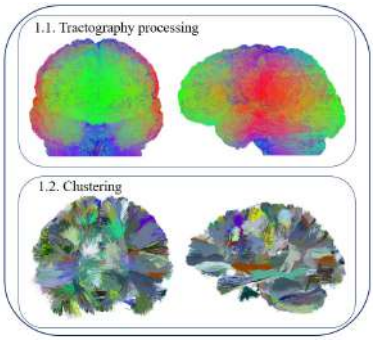
The most common approaches to estimate brain connectivity are diffusion tractography, structural covariance, resting-state functional connectivity and meta-analytic connectivity modeling [36]. Diffusion tractography provides information about structural connectivity, but has the limitations of not being able to delimit the beginnings and terminations of fiber bundles [35] and produce false positives due to the large number of fibers that cross between the WM tracts. In addition, short association fiber connections can be lost due to the limited resolution of the tractographic methods [82]. Two strategies can be used to perform diffusion-based cortical parcellations. One approach first determines corresponding connections across subjects and then creates a parcellation according to the main connections in all the subjects. For example, a fiber bundle atlas of superficial WM connections was used to segment bundles in a group of subjects and get some consistent parcels for the 10 analyzed subjects [51]. The difficulty here is to detect a representative set of the common connections for a population of subjects and create the final parcels. The second approach detects robust individual parcels from the whole tractography dataset, and then manages to find and delineate consistent parcels across subjects [76].

In this work, we propose a new hybrid method of individual cortical parcellation based on WM connectivity and intra-subject fiber clustering, with automatic parcel labeling. Our goal is to perform a good quality individual cortical parcellation to be used for a group-wise parcellation in the future. We applied the method to a group of subjects and evaluated several measures of brain connectivity. We demonstrate that the resulting networks for each subject comply with the integration and functional segregation as well as with the small-world definition.

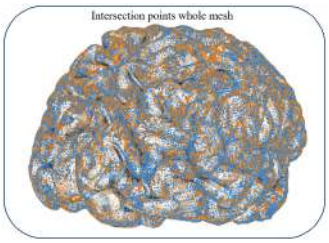
4.2.2. Our approach

We perform the parcellation of the cortex with a hybrid method based on WM connectivity given by fiber clusters, leading to an automatic labeling of cortical regions. In

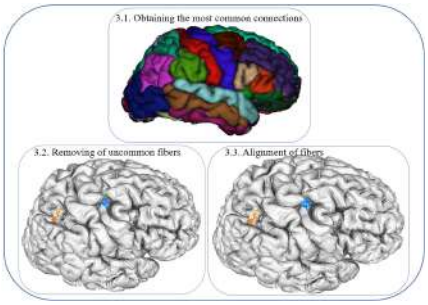
the following, we explain the whole method (see Figure 4.12), which is composed of five steps: *(1)* fiber clustering, *(2)* intersection with the mesh, *(3)* WM fiber filtering, *(4)* parcellation of the cortex and *(5)* sub-parcel post-processing.



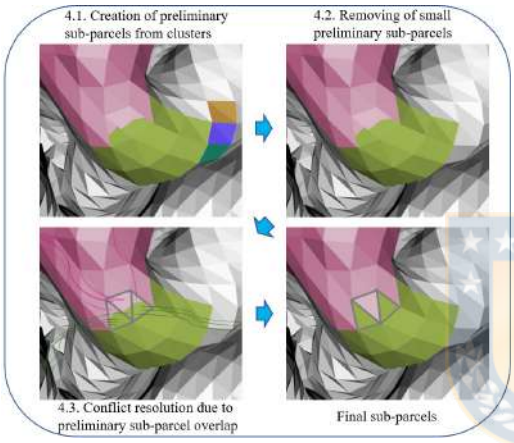
(a) Step 1: Fiber clustering.



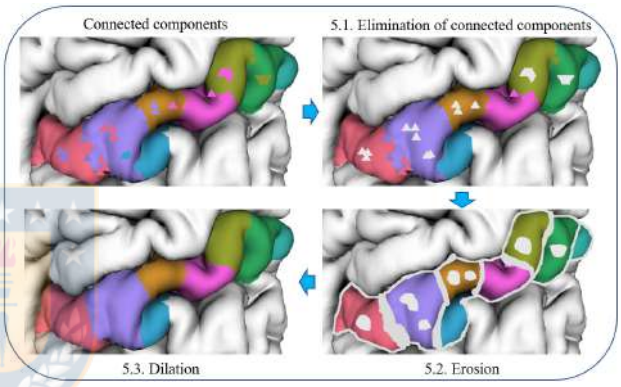
(b) Step 2: Intersection with the mesh.



(c) Step 3: WM fiber filtering.



(d) Step 4: Parcellation of the cortex.



(e) Step 5: Sub-parcel post-processing.

Figure 4.12: Parcellation method. **Step 1: Fiber clustering**. First, the whole tractography is resampled with 21 points and transformed to T1 space. Next, a fiber clustering is applied to obtain compact clusters. **Step 2: Intersection with the mesh**. The intersection of the fiber clusters with the cortical mesh is calculated. **Step 3: WM fiber filtering**. The fibers of each cluster are labeled according to an anatomical parcellation. Fibers that do not correspond to the most common connections are filtered out. Also, inverted fibers are realigned. **Step 4: Parcellation of the cortex**. Preliminary sub-parcels, from each cluster extremity are created. Next, small preliminary sub-parcels are removed. Finally, the overlap among sub-parcels is solved, by assigning each triangle to the most connected sub-parcel. **Step 5: Sub-parcel post-processing**. The main connected component of each sub-parcel is kept, in order to remove small isolated areas. Next, an erosion followed by a dilation (opening operation) are applied to eliminate some imperfections in the perimeter of the sub-parcels.

STEP 1: Fiber clustering

We apply an intra-subject clustering to all the fibers of the whole-brain tractography dataset. The objective is to create clusters with similar fibers, according to their position and shape, which we call fiber bundles.

First, the tractography obtained from the database is preprocessed (see Figure 4.12(a)$_{1.1}$). Fibers are resampled with 21 equidistant points, a number of points big enough to represent all the brain fibers. Next, the tractography datasets are transformed from T2 to T1 space. Finally, an intra-subject clustering algorithm is applied, which is composed of three main steps [109]. Figure 4.12(a)$_{1.2}$ shows an example of the results after applying the clustering to a tractography dataset.

STEP 2: Intersection with the mesh

This step determines the intersection of the fiber clusters with the cortical mesh [115], and was modified to perform the intersection over the whole tractography. Figure 4.12(b) shows an example of intersection points of the fibers with the mesh.

STEP 3: WM fiber filtering

This step aims to label the clusters that intersect with the mesh and filter them to delimit anatomical cortical regions (brain circonvolutions). To perform the filtering, we carry out three sub-steps (see Figure 4.12(c)):

1. **Obtaining the most common connections**: First, the fibers of each cluster are labeled according to the *Desikan-Killiany* cortical atlas labels [31]. For that, we use the mesh vertex labeling information given by FreeSurfer (see Figure 4.12(c)$_{3.1}$). Next, the most common connection, at the beginning and the end of each cluster, are determined. Figure 4.13$_{3.1}$ shows an example for different clusters connecting the Superior-Parietal (SP) and PreCentral (PrC) parcels.

2. **Removing of uncommon fibers**: The next sub-step aims to remove the fibers that do not correspond to the most common connections, in order to keep only

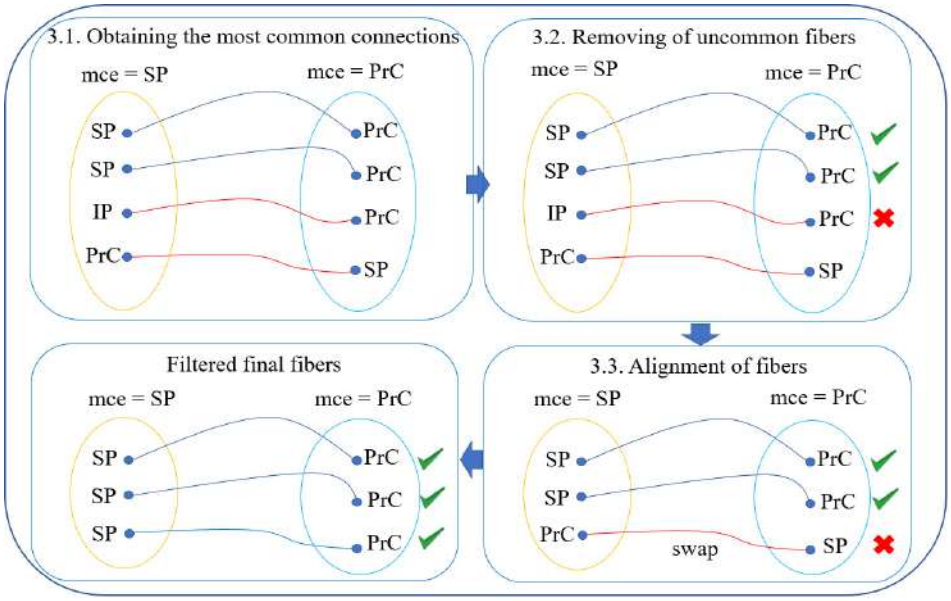


Figure 4.13: Example of WM Fiber filtering for a cluster. **Sub-step 3.1**: Obtaining the most common connections. In this case, the most common connections are SP and PrC. **Sub-step 3.2**: Removing of uncommon fibers. The fiber labeled with the IP parcel is removed since it does not belong to SP. **Sub-step 3.3**: Alignment of fibers. Fibers that are inverted according to the most common connections, are swapped.

the fibers corresponding to the anatomical parcels given by the *Desikan-Killiany* cortical atlas. See Figure $4.13_{3.2}$ for an example.

3. **Alignment of fibers**: Finally, a fiber alignment is applied, where fibers that are inverted by respect to the most common connections are swapped, as shown in Figure $4.13_{3.3}$.

Once the aforementioned sub-steps have been carried out, white matter fiber clusters are filtered, getting a better delimitation between bundles.

STEP 4: Parcellation of the cortex

This step creates sub-parcels from the preliminary sub-parcels defined by the intersection of each cluster. It solves the conflicts between the overlaps of the preliminary sub-parcels. The creation of the sub-parcels is done in three sub-steps:

1. **Creation of preliminary sub-parcels from clusters**: Each fiber cluster will define two preliminary sub-parcels, one from each extremity of the cluster. The set

of preliminary sub-parcels is created by constructing a list of the triangles intersecting each cluster extremity. For the example in Figure 4.12(d)$_{4.1}$, if the cluster goes from PrC parcel to PostCentral (PoC) parcel, in the PrC anatomical parcel a sub-parcel is created with the start triangles of the cluster, that is, the triangles that are in PrC and a sub-parcel is also created for the end triangles that belong to the PoC anatomical parcel. Note that a triangle can be intersected by several cluster extremities.

2. **Removing of small preliminary sub-parcels**: Preliminary sub-parcels, that are too small, with a size 10% smaller than the average size of the preliminary sub-parcels within an anatomical parcel, are removed (see Figure 4.12(d)$_{4.2}$). We denote as SP_i a sub-parcel, where $i = 1, ..., n$ is the index of the sub-parcel, and $size_{thr}$ is the threshold for the elimination of small sub-parcels. Then, if $size_{thr}(SP_i) < 0.10$ the sub-parcel is removed. With this step a big amount of isolated triangles and noisy preliminary sub-parcels are removed.

3. **Conflict resolution due to preliminary sub-parcel overlap**: Some preliminary sub-parcels present an overlapping within an anatomical region. To solve this problem, the conflicting triangles (belonging to several preliminary sub-parcels) are analyzed, and assigned to the sub-parcel with the higher number of intersecting fibers (see Figure 4.12(d)$_{4.3}$).

At the end of this step, the existing conflicts between preliminary sub-parcels disappear, thus obtaining a set of sub-parcels corresponding to each anatomical parcel (circonvolution).

STEP 5: Sub-parcel post-processing: This is the last step of the cortical parcellation method, that aims to perform a refinement of the sub-parcels, in order to get more uniform areas. It consists of three sub-steps:

1. **Elimination of connected components**: Within an anatomical parcel, a graph is created for each sub-parcel and then, the connected components of each graph

are calculated. The largest connected component of each sub-parcel is kept, leading to the removal of small isolated areas (see Figure 4.12(e)$_{5.1}$).

2. **Erosion**: The edges of each sub-parcel are eroded over the mesh, to finish the removal of some peaks or protrusions that deform the sub-parcel perimeter (see Figure 4.12(e)$_{5.2}$).

3. **Dilation**: Finally, the sub-parcels are expanded, filling the gaps left by erosion, resulting in smooth, uniform and well-defined sub-parcels (see Figure 4.12(e)$_{5.3}$). The morphological operation of *erosion + dilation* corresponds to an *opening*, which is an operation for noise elimination.

After carrying out the post-processing, we obtain the complete parcellation of the cortical mesh. It is defined by the subdivision of the cortex into sub-parcels, given by a label for each mesh vertex.

4.2.3. Experimental results

We calculated the cortical parcellation of five subjects from the ARCHI database, using their complete clustered tractographies. To evaluate the quality of the connections among parcels, we generated a connectivity map for each subject, from the resulting parcellation, and evaluated it using network graph metrics.

A connectivity map is built up by the tractography of each subject and the mesh parcellated into sub-parcels, performing the following steps:

1. The intersection of the complete tractography with the parcellated mesh is calculated.

2. A square matrix $n * n$ is created with n equal to the total number of sub-parcels, initialized to 0.

3. For each fiber in the tractography, connecting two sub-parcels, a 1 is added to the cells corresponding to both sub-parcels in the connectivity matrix.

Hence, a connectivity map was calculated for each subject, from its individual parcellation and tractography. In the following subsections we use these connectivity maps with brain connectivity metrics and also perform a qualitative analysis of the parcellations obtained from the five subjects.

Measures of brain connectivity

There are many metrics for the evaluation of the characteristics of brain networks [23]. The properties we choose to analyze them are functional segregation (Clustering Coefficient), functional integration (Characteristic Path Length) and Small-Worldness [108]. These properties have been shown to be present in the brains of the higher vertebrates [122]:

1. Functional segregation: It is the presence of strongly interconnected groups or clusters in the brain. The metric used to measure this property is the Clustering Coefficient [133]. A value closer to 0 denotes a random network, however, a complex-network shows higher clustering coefficient values. Equation 4.3 defines the clustering coefficient for undirected graphs:

$$C_i = \frac{2N_i}{k_i(k_i - 1)} \tag{4.3}$$

   where N_i is the amount of links in the neighborhood of i, k_i is the degree of a particular node i and C_i is the clustering coefficient for node i.

   The equation 4.4 measures the average clustering coefficient for the entire network:

$$C = \frac{1}{n} \sum_{i \in G} C_i \tag{4.4}$$

   where G is the graph of the undirected network, n is the total number of nodes in the network and C is the average clustering coefficient.

2. Functional integration: It is the ability to easily distribute information across the different specialized regions of the brain. The better the information is distributed,

the higher is the functional integration. The measure used to measure this property is the Path Length, specifically the Characteristic Path Length [133] that averages the shortest path length between each pair of nodes (sub-parcels) in the network. Equation 4.5 describes the characteristic path length for an undirected graph:

$$L = \frac{1}{n(n-1)} \sum_{i,j \in G, i \neq j} d_{ij} \tag{4.5}$$

where G is the graph of the whole undirected network, n is the total number of nodes in the network, d_{ij} between nodes i and j is the smallest distance between them and L is the characteristic path length.

3. Small-Worldness: This metric is very relevant, since it combines the two previous ones. A brain network must have good functional segregation, keeping functional integration a little lower, that is, strongly interconnected internal regions, and in turn, a good amount of links to other regions [108]. The σ coefficient is used to measure the small-world property, which is the ratio between the clustering coefficient and its equivalent random network divided by the path length and its corresponding random network. Equation 4.6 details the sigma coefficient:

$$\sigma = \frac{\frac{C}{C_r}}{\frac{L}{L_r}} \tag{4.6}$$

where C is the clustering coefficient, C_r is the clustering coefficient for the equivalent random network, L is the path length, L_r is the path length of its equivalent random network, and σ is the coefficient that measures the small-worldness. A network is considered small-world if $C \gg C_r$ and $L \approx L_r$, then $\sigma > 1$.

To evaluate the network we used the *bctpy* toolbox[3] for Python, that provides functions to calculate the Clustering Coefficient and the Characteristic Path Length metrics. Moreover, to calculate the Small-World metric, we transformed our matrices into graphs and used the *networkx* library for Python [63]. Figure 4.14 displays the metrics for the

[3]https://github.com/aestrivex/bctpy

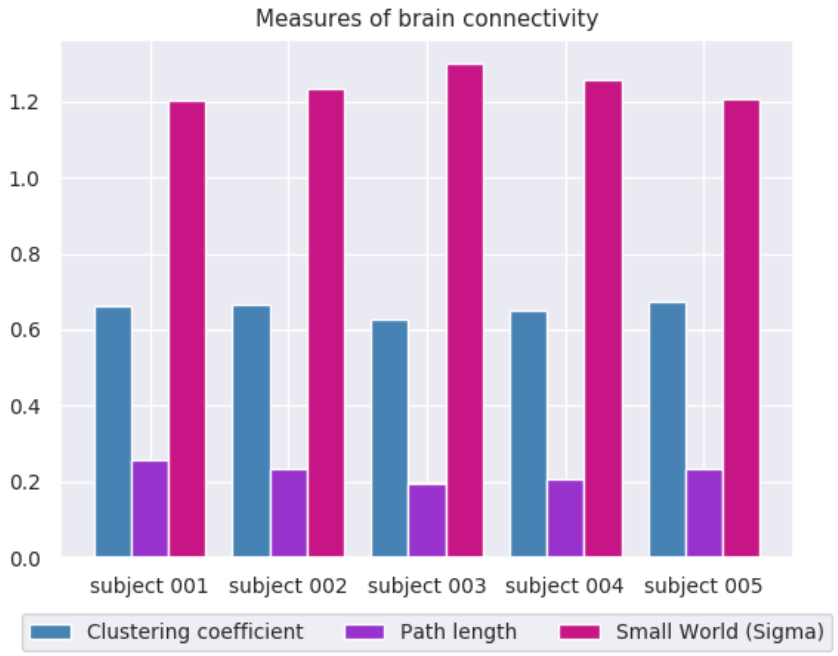


Figure 4.14: Measures of brain connectivity obtained for the five subjects: Clustering Coefficient, Characteristic Path Length and Small-Worldness. On the x-axis the five subjects are shown and on the y-axis the coefficients for the different metrics appear.

five subjects. A high Clustering Coefficient, while maintaining a lower Characteristic Path Length, and therefore a Small-Worldness > 1 was obtained for all the subjects.

These results demonstrate that the connectivity maps obtained for each subject, created by our parcellation method, are considered small-world networks, and therefore, maintain the properties of segregation and functional integration of the brain. In addition, as seen in Figure 4.14, the results are very similar among the five subjects.

Qualitative Analysis

In this subsection we show the different views of the cerebral cortex of a subject, as well as the different individual parcellations for five subjects in the database.

Figure 4.15 shows the coronal, axial, right and left sagittal views, resulting from the individual parcellation of subject 001, consisting of 430 sub-parcels for the whole brain, with 209 in the left hemisphere and 221 in the right hemisphere. Finally, Figure 4.16

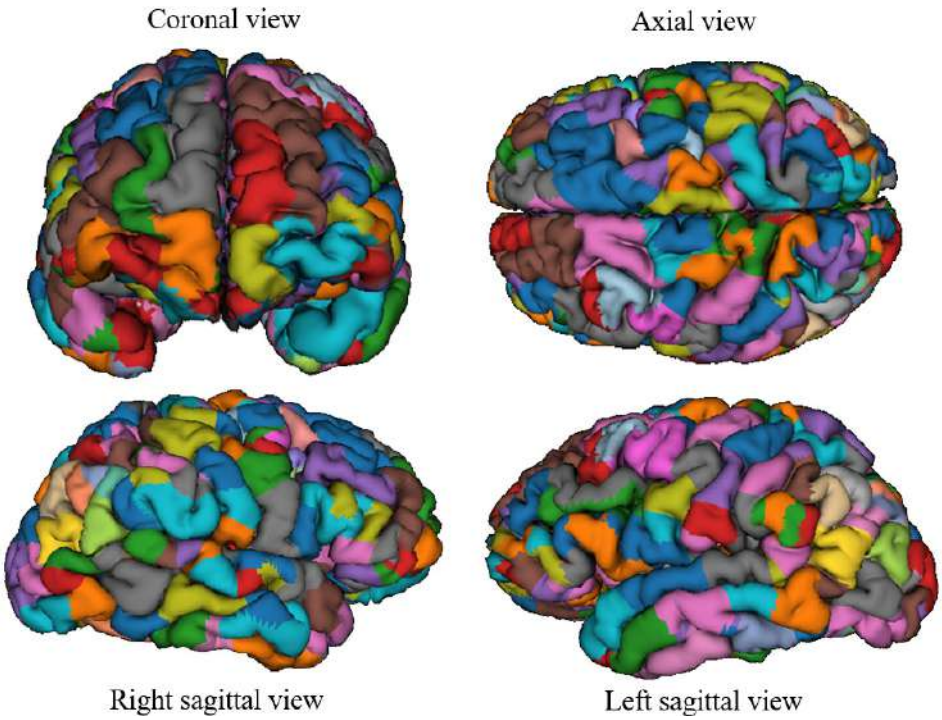


Figure 4.15: Individual parcellation for subject 001. Coronal, axial, right and left sagittal views are displayed. The parcellation subdivides the cortex into 430 sub-parcels, 209 in the left hemisphere and 221 in the right hemisphere.

displays the parcellation results for the five subjects. An average of 400 sub-parcels was obtained for the whole cortex, with approximately 200 sub-parcels in average per hemisphere.

4.2.4. Discussion

We developed a hybrid method that performs the complete parcellation of the cerebral cortex of an individual, based on the connectivity information of the white matter fibers from a whole-brain tractography dataset. The fiber clustering helps to define compact connections and filter out outliers. The method provides good quality results in the connectivity maps of the five analyzed subjects, evaluated by network graph metrics. Resulting networks show a high Clustering Coefficient, low Characteristic Path Length and Small-World property. These properties indicate good integration and functional segregation of the brain [108].

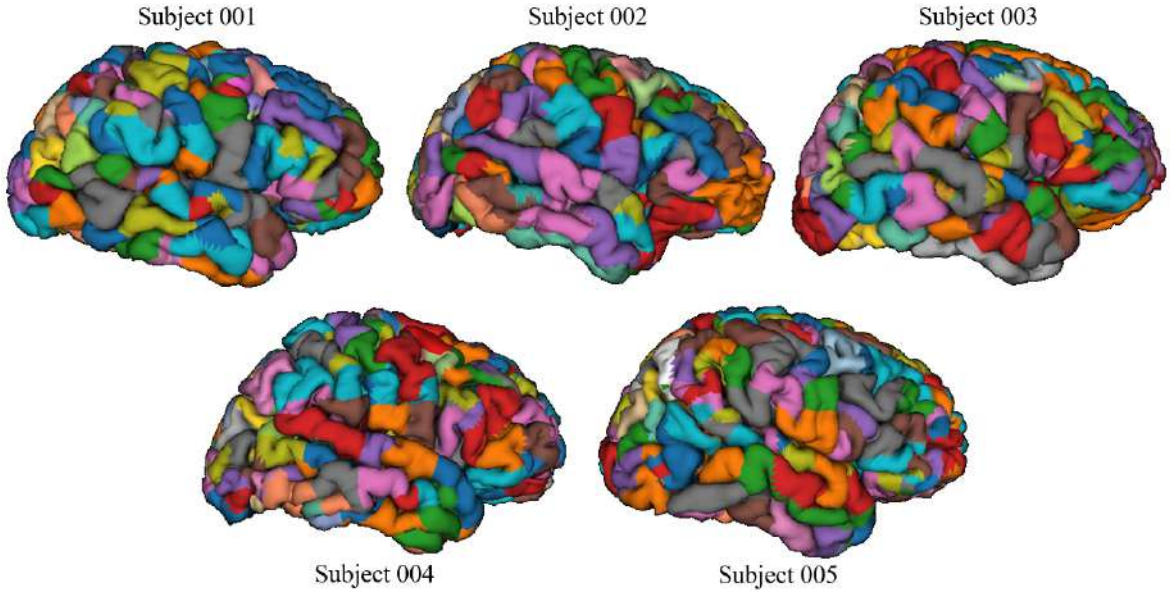


Figure 4.16: Individual cortex parcellation for five subjects (right sagittal views). The average of sub-parcels obtained for a hemisphere is 200, with 400 sub-parcels on average of the entire cerebral cortex.

4.3. Cortical surface parcellation based on geodesic distance

In this section, we present a method that performs the parcellation of the cerebral cortex for an individual based on the information of the brain's topology, that is, taking into account the geodesic distance over the cortical surface. In the first section, we describe the works related to this types of parcellations, based on the structure. After that, we present the developed method. In the experimental results section, we carried out the comparison with two state-of-the-art atlases based on macroanatomy. Finally, we include a discussion of the proposed method and the obtained results.

4.3.1. Related work

Magnetic Resonance Imaging (MRI) allows the study of the brain in a non-invasive and in-vivo way. In particular, structural MRI (sMRI) gives an anatomical differentiation of main brain tissues, enabling the automatic segmentation of them. The cortical surface can be extracted by softwares like FreeSurfer[4] [27, 47] or BrainVISA[5] [12].

A cortex parcellation, i.e., a subdivision of the cortex into several parcels or regions [141], can be performed based on different criteria, mostly based on anatomical, functional or diffusion-based information. This is a very complicated task due to the restrictions of each modality and the high inter-subject variability that exists, in particular, in white matter (WM) and gray matter (GM).

When studying the human connectome, brain region definition takes an important role in the study of brain connectivity and function [118]. Anatomical parcellation methods take into account the macroscopical anatomy, like the gyri and sulci [32, 14]. For example, Cachia et al. used a geodesic distance to label the cortex mesh vertices, using two nested Voronoï diagrams and labeled sulci [14]. Other method uses a statistical surface-based atlas, which includes information of the cortex curvature and the manual labeling of 35 regions of interest (ROIs) per hemisphere [31].

In order to evaluate diffusion-based [117] or functional-based [26] parcellations of

[4] https://surfer.nmr.mgh.harvard.edu/fswiki
[5] http://brainvisa.info/web/index.html

the cortical surface, these can be compared to a geodesic parcellation which is based on the geodesical properties of the mesh. However, this calculation can be time-consuming. Therefore, in this work, we propose a parallel method for the complete parcellation of the cortical surface, based on the geodesic distance. The goal is to create a fast individual cortical parcellation, available to the community for parcellation comparisons. The algorithm can be applied to subdivide each anatomical parcel given by *Desikan-Killiany* (*DK*) atlas, or to perform the cortical parcellation of the entire brain, depending on the method to be evaluated.

4.3.2. Our approach

The method implemented is called *GeoSP*, and performs the cortical parcellation based on a geodesic distance over the surface. The algorithm has two different modes. The default mode is based on the *DK* atlas to delimit the anatomical parcels and performs a geodesic subdivision of each anatomical parcel. Note that other atlases could also be used. The second mode creates a cortical parcellation for the entire brain. For the first mode, the method receives for each anatomical parcel (35 in total) a value k, used to divide each anatomical parcel into the specified k sub-parcels (for both hemispheres), i.e. an anatomical parcel with $k = 2$ will be divided into two sub-parcels. On the other hand, the second mode receives a unique k value, which will be used to divide each brain hemisphere into k sub-parcels, based on a geodesic distance, without using any other cortical parcellation. The method can be subdivided into two main steps: *(1)* a pre-processing that creates a graph representation of the mesh, and *(2)* K-means clustering based on geodesic distance over the mesh [41].

1) Pre-processing

Each anatomical parcel (for default mode) or each hemisphere (for the second mode) is represented with an undirected graph. The graph $G = (V, E)$ directly represents the mesh structure, formed by the vertices V and the edges E that join them. For the default mode, that performs the subdivision of each anatomical parcel given by the *DK* atlas, the labels of each region are used to identify each parcel. Finally, Euclidean distance (d_E) is calculated between each pair of vertices to create weighted

Algorithm 1 Parallel_kmeans

$groups \leftarrow []$ {list of sub-parcels containing the indexes of the vertices}
if $k > 1$ **then** {number of clusters}
  $centers \leftarrow initialize()$ {initializing centroids}
  $centers\_old \leftarrow centers$
  $converge \leftarrow FALSE$
  $i \leftarrow 0$
  while $i \leq nIter$ **and** $!converge$ {iterations and criterion for convergence}
    $groups \leftarrow calc\_groups()$ {calculating groups}
    $centers \leftarrow comp\_centroids()$ {computing centroids}
    $converge \leftarrow stop\_critery()$
    $centers\_old \leftarrow centers$
    $i \leftarrow i + 1$
  end while
else
  $groups \leftarrow [all\_indices]$ {all the indices for one anatomic parcel}
end if
return $groups$ {returns the list whose elements are the sub-parcel groups(indices))}

graphs.

2) K-means clustering

To subdivide an anatomical parcel or a hemisphere into k sub-parcels, a K-means clustering is applied. The algorithm consists of the following sub-steps: *(a)* initializing centroids, *(b)* (re)calculating groups and *(c)* (re)computing centroids. The algorithm uses a parallel implementation and its pseudocode is shown in Algorithm 1. For default mode, the method launches the K-means algorithm in parallel for each anatomical parcel given by *DK* atlas, while for the second mode, it launches a single thread per hemisphere. To exploit the capabilities of parallelism, it is launched in the sub-step of *(re)computing centroids*.

1. **Initializing centroids**. To perform this sub-step, K-means++ algorithm [73] is used to select the initial centroids. It has a low time complexity of $\mathcal{O}(logk)$. First, the method receives k, the number of sub-parcels (clusters) to divide each anatomical parcel (or each hemisphere). For each anatomical parcel, there may be different k. Also, k can be randomly set. Although the selection of starting centroids takes additional time, by using K-means++ the convergence of K-means

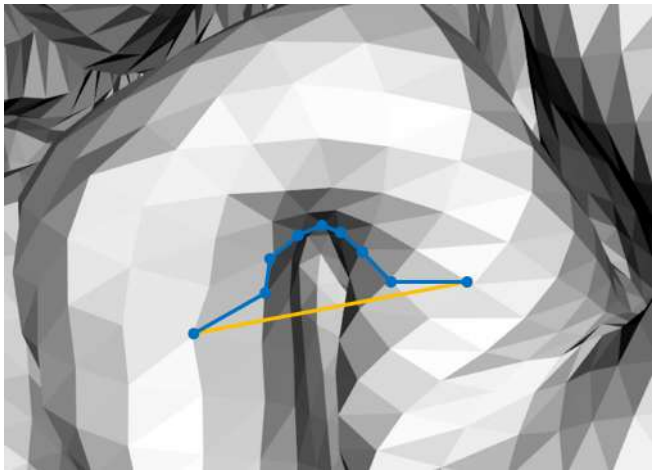


Figure 4.17: Euclidean distance versus geodesic distance. By using the Euclidean distance between two points, a straight line is obtained (orange path). While the geodesic distance considers the shortest path across the gyri and sulci of the mesh (blue path).

occurs quickly with reduced computation time. This leads to initial centroids better distributed than random selection across the anatomical parcels.

2. **(Re)Calculating groups**. In this sub-step, clusters are (re)calculated by assigning each vertex to the closest centroid. To achieve this, the *single-source shortest path problem* ($SSSP$) is used, which looks for the shortest path from a vertex v (centroid) to the rest of the vertices of the graph G. To calculate the distance between vertices, instead of the Euclidean distance, the geodesic distance is used. Then, based on an implementation of the *Dijkstra algorithm with Fibonacci heaps* [22], the $SSSP$ is calculated for all the centroids, that is, the shortest path from each centroid to all other vertices. This algorithm runs with low complexity ($\mathcal{O}(|E| + |V|log|V|)$). Finally, for each graph vertex, the distances obtained to the different centroids are compared, and each vertex is assigned to the centroid with the smallest geodesic distance. Figure 4.17 illustrates the Euclidean and geodesic distances for two vertices over the mesh. The path between two points for Euclidean distance is a straight line while the path for the geodesic distance is a route along the surface of the mesh, taking into account the gyri and the sulci.

3. **(Re)Computing centroids**. This is the last sub-step of the algorithm, in which the centroids of the clusters must be (re)calculated. First, the *all-pairs shortest*

paths problem has to be solved, that is, for each pair of vertices, the shortest path has to be calculated. This is done with the *Floyd–Warshall algorithm* [102], which runs in $\mathcal{O}(|V|^3)$. Although the temporal complexity of this step is high, it is still a polynomial running time (cubic) depending on the size of the input. The result obtained is a new centroid, which is the vertex closest to all other vertices in the cluster.

Sub-steps (b) and (c) are executed until the convergence criterion is reached. For this, the centroids of the current iteration are compared with the previous one. The algorithm stops if the distance is less than 2 mm or a maximum of 20 iterations is reached.

4.3.3. Experimental results

First, Figure 4.18 displays the results for one subject with 140 sub-parcels and 350 sub-parcels, for both modes of the method. To obtain 140 sub-parcels using the *DK* atlas, we divide each anatomical parcel into $k = 2$ sub-parcels. Since *DK* atlas has 35 anatomical parcels per hemisphere, with $k = 2$, we obtain 70 sub-parcels per hemisphere, leading to a total of 140 sub-parcels for the whole brain. Following the same procedure, to obtain 350 sub-parcels we divide each anatomical parcel into 5 sub-parcels, which generates 175 sub-parcels per hemisphere. It can be seen that the method generates homogeneous parcels both for the entire cortex and for the *DK* atlas-based parcellation.

Then, to illustrate an example of use, we calculated the reproducibility of structural connectivity across subjects for three different parcellations: *GeoSP*, *DK* and *Destrieux*. Figure 4.19 displays the scheme of processing performed to obtain the reproducibility analysis for a parcellation. For each subject, we used the tractography dataset in T1 space to calculate the structural connectivity matrix, based on each parcellation. To construct a matrix, the intersection of the fibers with the cortical mesh is determined and the labels of the pair of parcels connected by each fiber are used to add a count in the corresponding cell of the matrix. Next, the matrx is binarized and

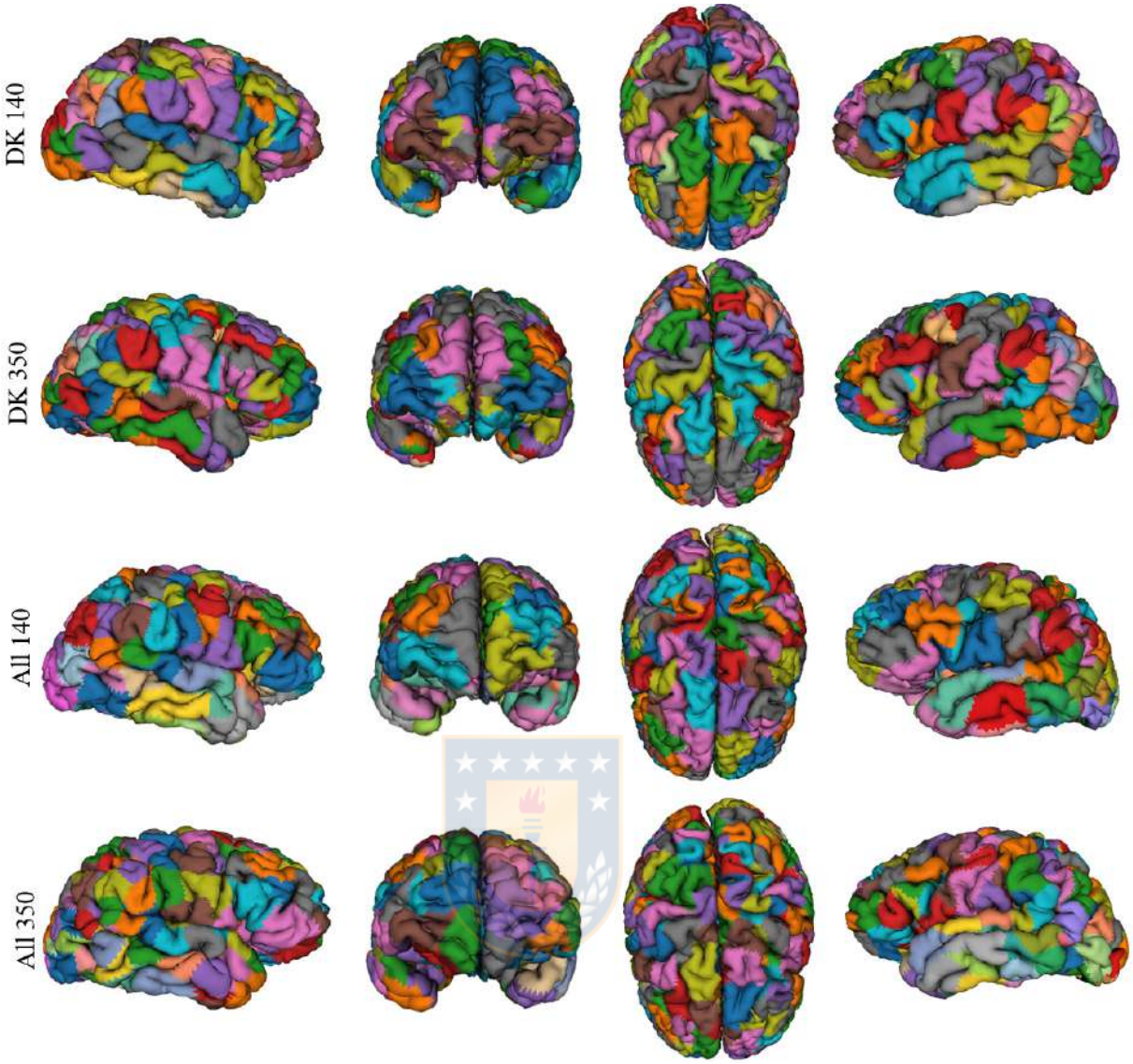


Figure 4.18: Parcellation of the cortical mesh obtained in one subject for modes based on *DK* atlas and for the entire cortex. Right sagittal, coronal, axial and left sagittal views. First and second rows: parcellation based on *DK* atlas subdivided into 140 (first row) and 350 (second row) sub-parcels, with execution times of 42.9s and 18.1s respectively. Third and fourth rows: parcellation for the entire cortex into 140 (first row) and 350 (second row) sub-parcels, with execution times of 75.4s and 82.25s, respectively.

converted into a graph to use network metrics. One of these metrics is the Dice coefficient and was calculated between each pair of subjects, for each method. Figure 4.20 shows a boxplot of the reproducibility among the 50 subjects between *GeoSP* and the

other anatomical atlases. The reproducibility is slightly higher for *GeoSP* in both cases, with a difference of 0.024 between *GeoSP* and *DK* (70 parcels) and of 0.043 for *GeoSP* and *Destrieux* (150 parcels).

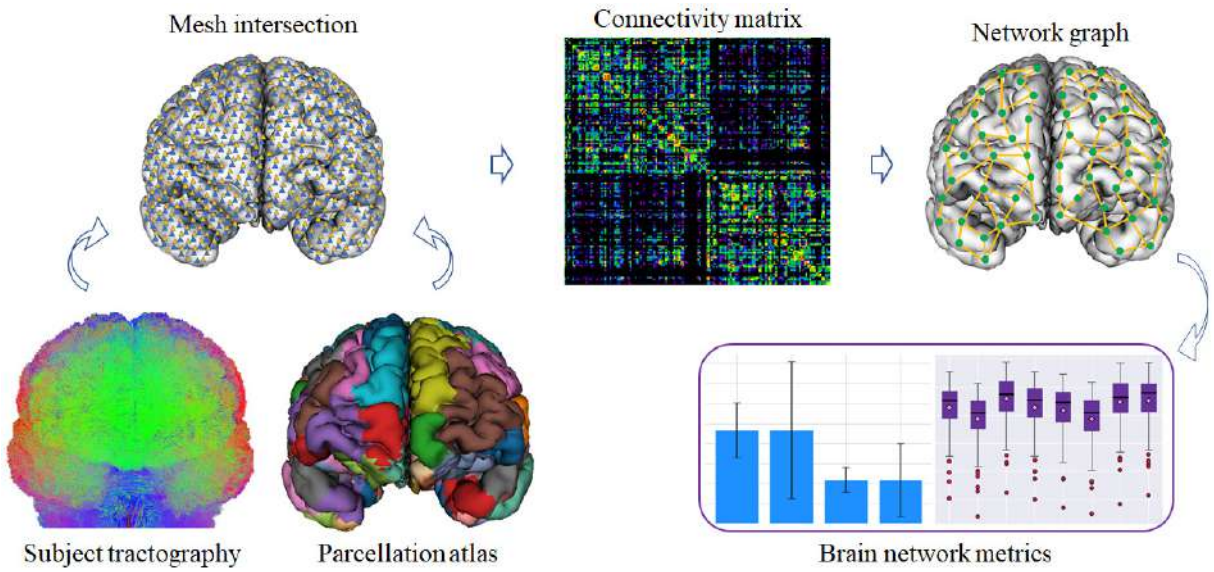


Figure 4.19: Brain connectivity analysis scheme. First, the tractography of each subject in T1 space is intersected with the cortical mesh, which is parcellated based on an atlas. Then, a connectivity matrix is created containing for each cell, corresponding to a pair of sub-parcels, the total number of connections between them. The matrix is then binarized, indicating with a 1 if the sub-parcels are connected and with a 0 if there is no connectivity between them. The connectivity matrix is finally converted to a network graph to analyze network metrics as the Dice coefficient.

Finally, the execution time for both modes was compared. Figure 4.21 displays the execution times according to the number of sub-parcels in which the cortex is subdivided. For mode one, based on *DK* parcellation, the execution time decreases with the number of parcels. This is because the greater the number of sub-parcels, and being delimited by the anatomical parcels of the atlas, the algorithm has to perform fewer computations when recomputing the centroids. On the other hand, for the entire cortex, with a greater number of sub-parcels, more time is needed to subdivide the cortex. This is due to the size of the graphs (one for each hemisphere), where the recalculation of centroids becomes very expensive since it is necessary to recalculate all the shortest paths between all the pairs.

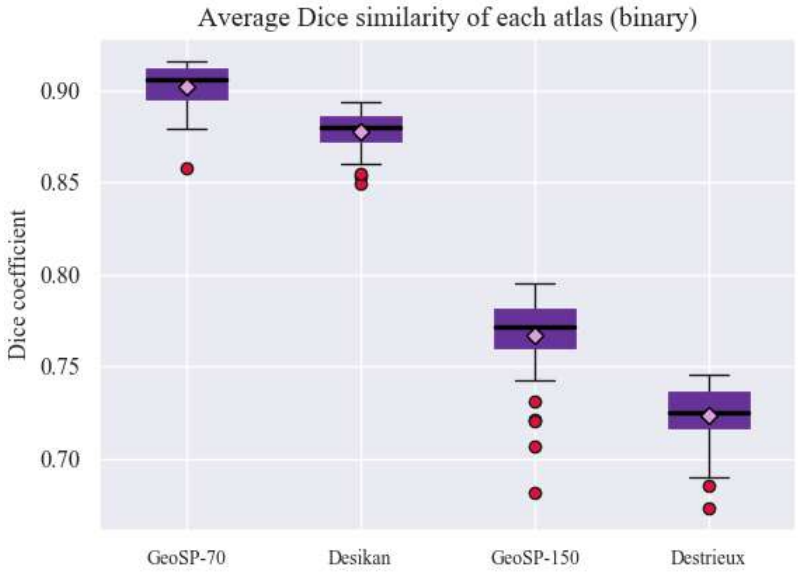


Figure 4.20: Comparison of the structural connectivity reproducibility between *GeoSP* and the two atlases, with equal number of parcels. X-axis shows the different atlases used. Y-axis contains the Dice coefficient, the closer to one, the greater the reproducibility. The rhombus indicates the mean and the black line the median for each atlas. Results show a difference of 0.024 between *GeoSP* and *DK* atlas, and 0.043 between *GeoSP* and *Destrieux* atlas.

4.3.4. Discussion

We propose a parallel method to perform a parcellation of the cortical surface mesh based on geodesic distance, in order to consider gyri and sulci topology. This method is called *GeoSP*, represents the mesh with a graph and performs a K-means clustering in parallel. It has two modes of use, by default, it performs the geodesic cortical parcellation based on the boundaries of the anatomical parcels provided by the *Desikan-Killiany* atlas. The other mode performs the complete parcellation of the cortex.

The algorithm was tested in 50 subjects. Results show homogeneous sub-parcels for both modes and different number of sub-parcels. Structural connectivity reproducibility between *GeoSP* and two anatomical atlases is very similar and slightly higher for *GeoSP*. This may be due to the higher homogeneity of the parcels with *GeoSP*. Moreover, the greater the number of parcels, the less reproducibility will be obtained. Hence, this test shows that special attention should be given to the indices to be used in comparisons between parcellations. In any case, we provide a fast and configurable

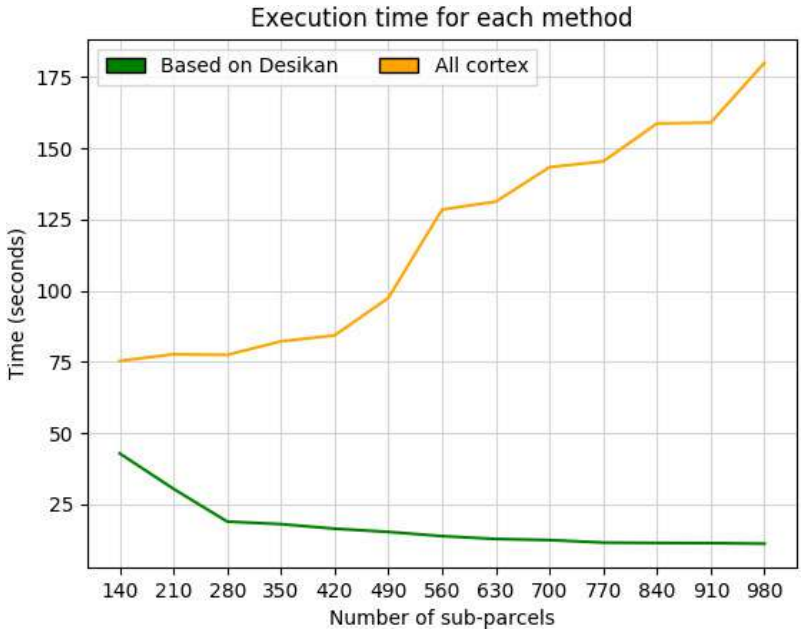


Figure 4.21: Execution time (seconds) for each mode, depending on the number of sub-parcels. As expected, the subdivision into sub-parcels according to the delimitation given by the *Desikan-Killiany* atlas is less expensive than subdividing the entire cortex.

parcellation method based on geodesic distance, available to the community, to perform the comparison and evaluation of data-driven parcellations, like those based on diffusion or functional MRI.

Chapter 5

Conclusions and future work

In this chapter, we present the conclusions of all the work carried out in the thesis. In addition, we show the future lines of research that would be interesting to explore.

5.1. Conclusions

The developments of this thesis are related to the computational neuroscience area, specifically to algorithms and methods for the study of brain connectivity based on dMRI. We have developed efficient algorithms that can deal with the noise present in the diffusion tractography. These artifacts are produced by tractography algorithms, generating valid bundles that are taken into account when using the white matter fibers. In addition, the high variability of fibers (more than a million) among subjects generates a high complexity when dealing with these massive 3D datasets. Therefore, when using diffusion tractography, additional validation is needed, such as contrasting with the anatomy and obtaining a biological meaning, using post-mortem studies or applying other MRI techniques like functional MRI.

We hope that this research will be of benefit to neuroscientists in the area since modern studies can be developed by analyzing new databases and contributing to the decoding of the human connectome.

The conclusions of the works presented are set out below:

- **Clustering and labeling of white matter fibers**. We present *FFClust* which is an efficient algorithm for clustering of white matter fibers. It is at least an order of magnitude faster than *QuickBundles*, which is one of the most widely used algorithms for streamline clustering. For example, a subject with approximately 1 million fibers is processed by the sequential version of *FFClust* in 1.99 min, and its

parallel implementation in 45 s, while *QB* takes 2.2 h using its best configuration. However, using the Davies-Bouldin (DB) index to measure cluster quality, *QB* achieves the best result, followed by *FFClust*, while *QBX* does not improve the quality of *QB*.

Thanks to this method, new brain connectivity studies can be carried out, in particular on fibers of the superficial white matter (SWM). Furthermore, *FFClust* and the *Guevara*'s method achieve similar Davies-Bouldin index values and can identify all bundles when applying segmentation. They also have the lowest percentage errors when identifying selected bundles. However, *FFClust* is faster and simpler than the *Guevara*'s method.

On the other hand, we implemented a fast method for the automatic labeling of white matter fiber bundles, specifically for the SWM, based on an intra-subject clustering and the connectivity of the clusters with the cortical mesh, based on an anatomical ROI atlas. The algorithm also adds a label associated with the relative position of the bundles. Results for intra-subject labeling show a degree of correspondence between subjects, which is further improved with inter-subject labeling. A complete intra-subject labeling is executed on an average time of 3.5 minutes for a tractography dataset of about one million fibers. This enables a fast and easy exploration, visualization, and analysis of labeled short association bundles in individuals, which is very difficult without any additional information.

Besides, we developed an inter-subject labeling by using two methods. One approach is matching, in particular, the *Hungarian* algorithm, and the other is clustering, employing *QuickBundles* algorithm. The results show a better reproducibility across subjects for the clustering method versus the matching algorithm while keeping a moderate inter-cluster distance, indicating a good quality of the clusters. Furthermore, the algorithm is scalable and the whole processing for the inter-subject labeling executes at a reasonable time, of about 1.17 h for 20 subjects. The obtained clusters could be used to perform group-wise connectivity studies, such as the creation of WM bundle atlases, and the development of new

methods for the analysis of brain connectome.

- **Cortical surface parcellation**. The proposed method creates a fine-grained parcellation of the cortical surface, consisting of the subdivision of coarse anatomical parcels, from a diffusion-based fiber bundle atlas. The generated parcellation depends on three configurable parameters that generate a parcellation with a smaller or larger number of sub-parcels. Furthermore, an intermediate output of the method is the probabilistic representation of the preliminary sub-parcels, associated to the two connections of each bundle.

  As a final conclusion, this method can create a fine-grained cortical parcellation based on structural connectivity, from coarse anatomical parcels, leading to sub-parcels with high consistency in connectivity profiles across a population of subjects, and a degree of correspondence with state-of-the-art parcellations based on different MRI modalities.

  Furthermore, we have developed a method to perform individual cortical parcellations, based on the information of the connectivity of the white matter fibers. Fiber clustering helps to define more compact connections and therefore regions, in addition to filtering outliers. The results show good quality in the analyzed subjects, in particular, in the connectivity maps. These were evaluated with brain network metrics, such as clustering coefficient, path length, and small-worldness. These metrics indicate good functional integration and segregation of the brain. However, there is no consensus in the community about the validity of these metrics. Further improvements and validations are required.

  Finally, we present a parallel method, called *GeoSP*, which creates an individual parcellation of the cortical mesh based on geodesic distance, thus considering the brain topology, i.e., the gyri and sulci. The method represents the mesh by means of a network and has two modes of use. By default, it performs the parcellation based on the boundaries of the *Desikan-Killiany* atlas. The other mode performs the division for the complete cortex. The results show homogeneous sub-parcels. Execution times for the atlas-based mode are 18 s and 82 s for

the full cortex mode. Furthermore, compared to two macroanatomical atlases, *GeoSP* has fewer differences in structural connectivity reproducibility among subjects, probably due to a higher homogeneity of the sub-parcels. However, this does not mean that the parcellation has biological meaning or that is better than the others. Further analyses need to be performed. The method is available for making comparisons between parcellations, whether based on dMRI or fMRI.

5.2. Future work

In this section, we present the future lines of research that could be carried out based on the work of this thesis. We present these ideas for each of the contributions made:

- **Clustering and labeling of white matter fibers**. Since *FFClust* has four steps, as future work we propose to improve the execution times for the slower stages. This could be useful for integrating the clustering algorithm with visualization applications to enable the quick exploration and other post-processing analyses of the structure of the white matter for one or multiple subjects. Then, we suggest that *FFClust* can be used in similar applications where the *Guevara* has been successfully used [53, 50].

  Regarding the labeling of SWM bundles, future work could be focused on the application of the method in high-quality databases, such as the *Human Connectome Project (HCP)* database, for the creation of a SWM atlas and diffusion-based cortical parcellations. In addition, other algorithms could be used to generate better labeling of the SWM bundles.

- **Cortical surface parcellation**. The cortical parcellation based on a fiber-bundle atlas could use a new bundle atlas, based on a larger database, like the HCP database and probabilistic tractography. Also, the bundles could be obtained from an inter-subject fiber clustering from the same database, which could lead to a better representation of WM bundle connections of the population of subjects. However, the post-processing of candidate sub-parcels would probably be

more complicated due to a larger amount of bundles and a higher sub-parcel overlapping. Also, functional information could be used to create a parcellation by using multimodal parcellation frameworks [99]. Furthermore, another line to explore is the inclusion of some atlas bundles based on known functional areas. Finally, the probabilistic information of the preliminary sub-parcels could be used, in combination with individual segmented bundles, to create individual parcellations, adapted to each subject. Its effect will be small changes on the boundaries of the sub-parcels of each subject, due to individual differences in the segmented bundles. Adapted parcellations should lead to increased consistency in structural connectome across subjects.

Regarding cortical surface parcellation based on clustering, as future work, we could explore the implementation of a multi-subject version of this parcellation method and test it in different databases, such as the Human Connectome Project. Hence, we could obtain an atlas (or model) of cortical parcels with similar connectivity profiles across a population of healthy subjects. Also, other information could be integrated, such as fibers segmented with a bundle atlas [128], or data from other modalities, like fMRI.

Finally, as future lines of research with respect to the geodesic parcellation, the execution time of the proposed method for both modes of operation could be improved. On the other hand, other anatomical atlases could be used as a basis for drawing the boundaries between anatomical parcels instead of *Desikan-Killiany* atlas.

Appendix A

Publications related to this thesis

A.1. Publications

In this appendix we present the list of publications related to the thesis.

A.1.1. Journals

- N. López-López, A. Vázquez, J. Houenou, C. Poupon, J.-F. Mangin, S. Ladra, and P. Guevara. From coarse to fine-grained parcellation of the cortical surface using a fiber-bundle atlas. Frontiers in Neuroinformatics, 14:32, 2020.
- A. Vázquez, N. López-López, A. Sánchez, J. Houenou, C. Poupon, J.-F. Mangin, C. Hernández, and P. Guevara. FFClust: Fast fiber clustering for large tractography datasets for a detailed study of brain connectivity. NeuroImage, 220:117070, 2020[1].
- A. Vázquez, N. López-López, J. Houenou, C. Poupon, J.-F. Mangin, S. Ladra, and P. Guevara. Automatic group-wise whole-brain short association fiber bundle labeling based on clustering and cortical surface information. BioMedical Engineering OnLine, 19(1):1–24, 2020.

A.1.2. International conferences

- N. López-López, A. Vázquez, C. Poupon, J.-F. Mangin, S. Ladra, and P. Guevara. GeoSP: A parallel method for a cortical surface parcellation based on geodesic distance. In 2020 42th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 44-49, 2020.

[1]This publication is directly related to the Master's thesis of A. Vázquez [125].

- I. Huerta, A. Vázquez, N. López-López, J. Houenou, C. Poupon, J.-F. Mangin, P. Guevara, and C. Hernández. Inter-subject clustering of brain fibers from whole-brain tractography. In 2020 42th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 124-128, 2020.
- I. Osorio, D. Bonometti, D. Carrasco, N. López-López, A. Vázquez, C. Román, C. Poupon, J. Mangin, and P. Guevara. FiberVis: a tool for a fast fiber tractography visualization and segmentation. In 25th Annual Meeting of the Organization for Human Brain Mapping (HBM), 2019.
- A. Vázquez, N. López-López, N. Labra, M. Figueroa, C. Poupon, J.-F. Mangin, C. Hernández, and P. Guevara. Parallel optimization of fiber bundle segmentation for massive tractography datasets. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 178–181, 2019.

A.1.3. National conferences

- N. López-López, A. Vázquez, C. Poupon, J.-F. Mangin, and P. Guevara. Cortical surface parcellation based on intra-subject white matter fiber clustering. In 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), pp. 1–6, 2019.
- I. Huerta, A. Vázquez, N. López-López, P. Guevara, and C. Hernández. Inter-subject clustering of brain fibers from whole-brain tractography. In 2019 XII Congreso Anual de Ingeniería Biomédica (CAIB XII), 2019.
- A. Vázquez, N. López-López, N. Labra, M. Figueroa, C. Poupon, J.-F. Mangin, C. Hernández, and P. Guevara. Optimización paralela para la segmentación de fascículos en conjuntos masivos de datos de tractografía. In 2018 IEEE XXVII INGELECTRA. IEEE, 2018. This work received the first national prize in the postgraduate category.

Appendix B

Supplementary material for parcellation

B.1. Comparison of superficial white matter bundles for both SWM atlases

Figure B.1 shows similar bundles between the two superficial white matter (SWM) atlases, *swm_atlas_1* [50] and *swm_atlas_2* [106]. Three bundles of the left hemisphere (CMF-Op_0, CMF-PrC_0 and RMF-SF_0) and three bundles of the right hemisphere (IP-LO_0, PrC-SM_0, Tr-Ins_0) with high similarity, were taken as examples for comparison between both atlases.

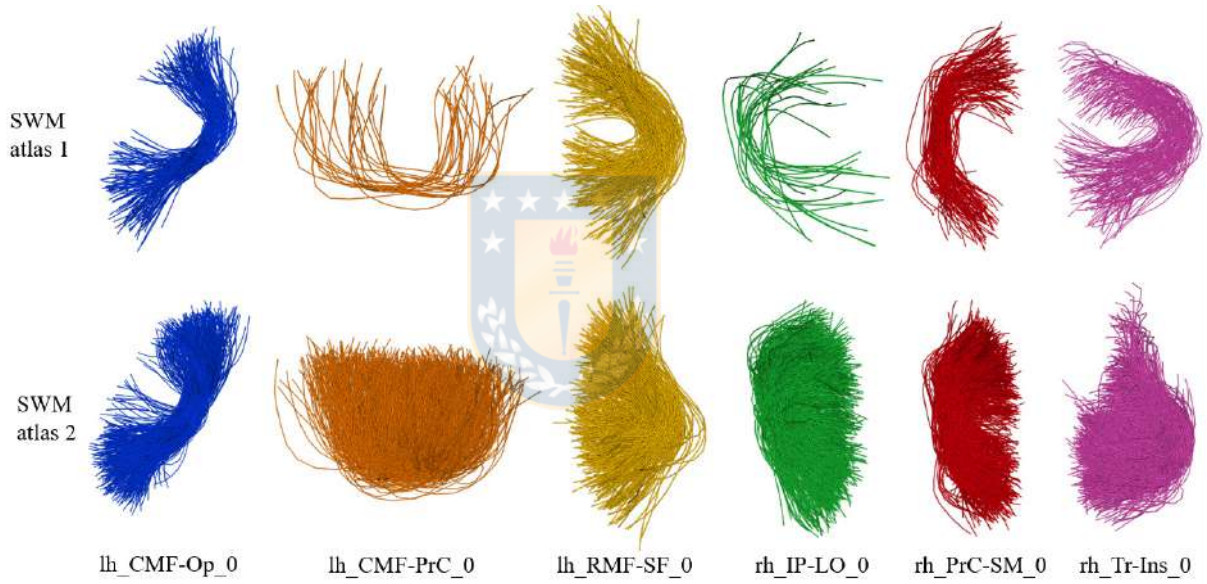


Figure B.1: Some examples of bundles with high similarity between both SWM atlases. First row: *swm_atlas_1*, second row: *swm_atlas_2*.

Figure B.2 displays the superficial white matter and deep white matter (DWM) bundles used to create the final fused white matter atlas.

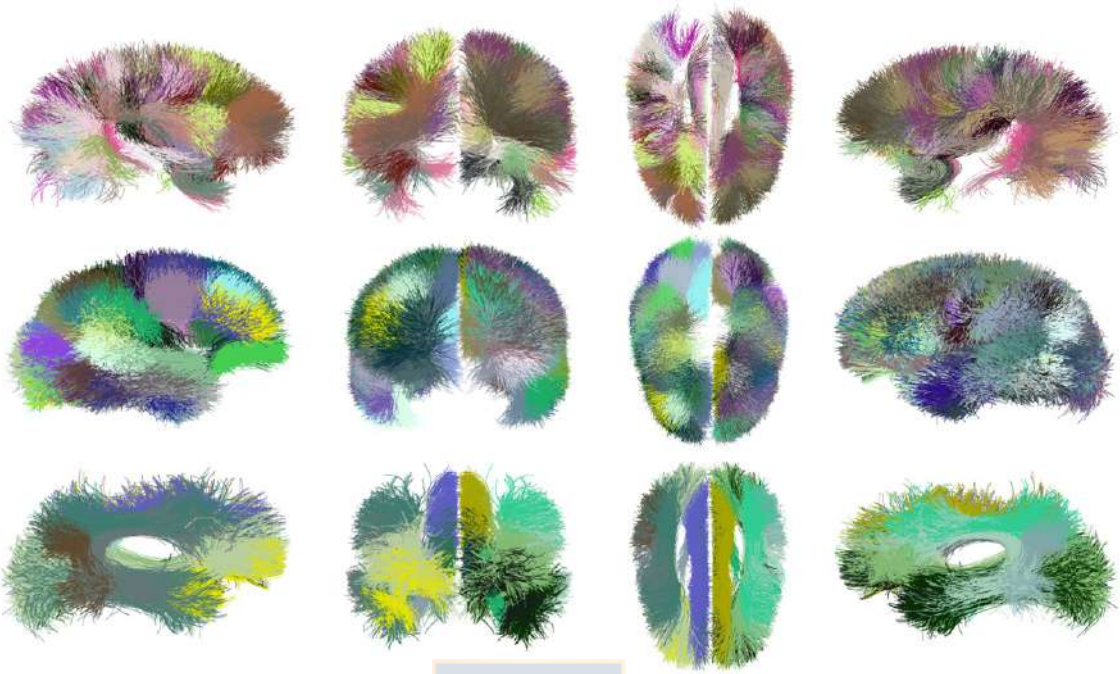


Figure B.2: Bundles of the SWM and DWM atlases used for the creation of the fused final atlas. First row: *swm_atlas_1* [50], composed of 50 bundles in each hemisphere. Second row: *swm_atlas_2* [106], with 27 bundles in the left hemisphere and 34 in the right hemisphere, after selecting the bundles that complement *swm_atlas_1*. Third row: *DWM atlas* [53], composed of 9 bundles per hemisphere. Right sagittal, coronal, axial and left sagittal views.

B.2. Fiber bundle segmentation

Figure B.3 shows in the first row the final fused atlas of white matter bundles. The bottom row contains a segmented subject with the final atlas.

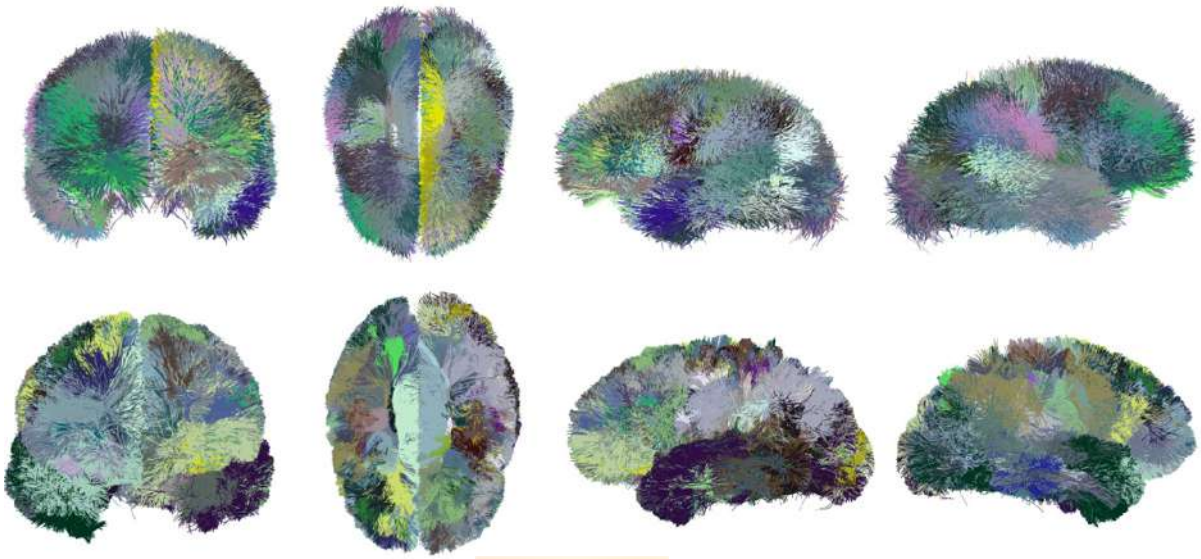


Figure B.3: Example of bundle segmentation. The first row shows the final fused atlas of white matter bundles used by the proposed method. The second row shows a segmented subject, based on the atlas. Coronal, axial, left sagittal and right sagittal views.

B.3. Parameter configuration for parcellation creation

In this section, an example of the parameter $size\_thr$ is shown. Figure B.4 illustrates the *Removing of small preliminary sub-parcels* sub-step, belonging to Step 5 of the parcellation method for the precentral anatomical parcel (PrC), by using $size\_thr$ = 0.10 and $size\_thr$ = 0.30.

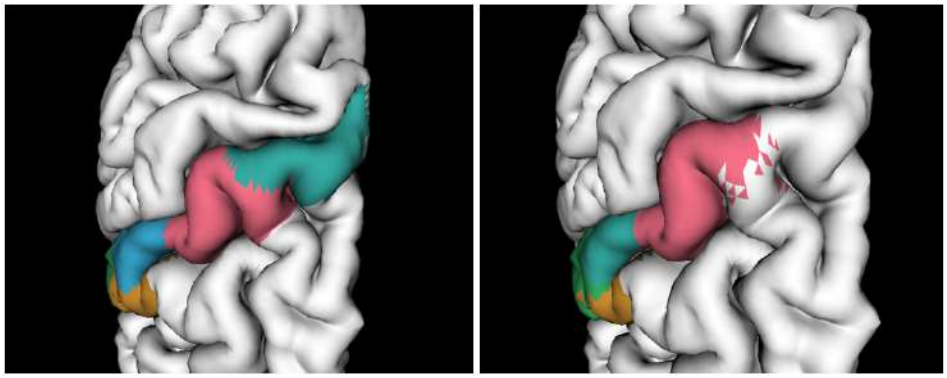


Figure B.4: Example of size threshold for the removing of small preliminary sub-parcels ($size\_thr$), for the precentral (PrC) anatomical parcel. Left: $size\_thr$ = 0.1, Right: $size\_thr$ = 0.3. A small $size\_thr$ will prevent the removal of relatively big sub-parcels, which could leave uncovered regions in the cortex.

B.4. Parcel post-processing

This section shows the results of parcel post-processing (Step 6). Figure B.5 gives an example of *Removing small connected components* sub-step, belonging to Step 6 of the proposed method (Sub-parcel post-processing). Finally, Figure B.6 illustrates the final hard parcellation from *atlas 1* after applying *removing small connected components, erosion and dilation*.

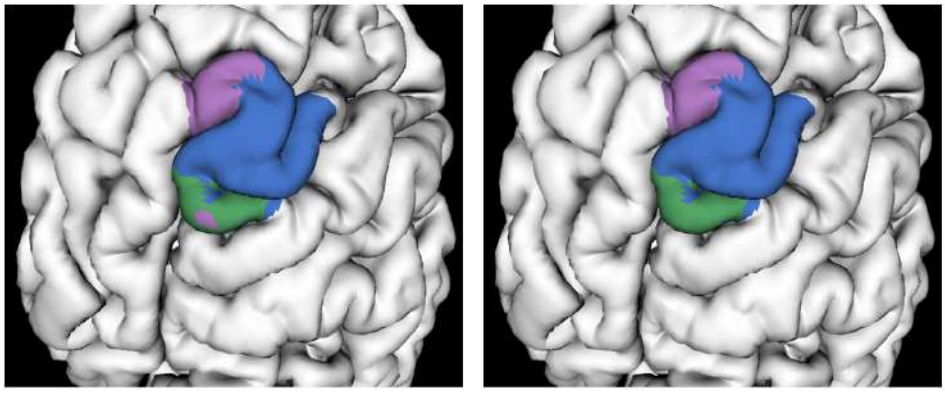


Figure B.5: Removing small connected components (Step 6). The image on the left shows the SM anatomical parcel, divided into three sub-parcels with small connected components among them. Each connected component is represented with a color but internally is modeled as a graph. The image on the right shows the SM anatomical parcel after applying the removal of the small connected components, resulting in more uniform sub-parcels.

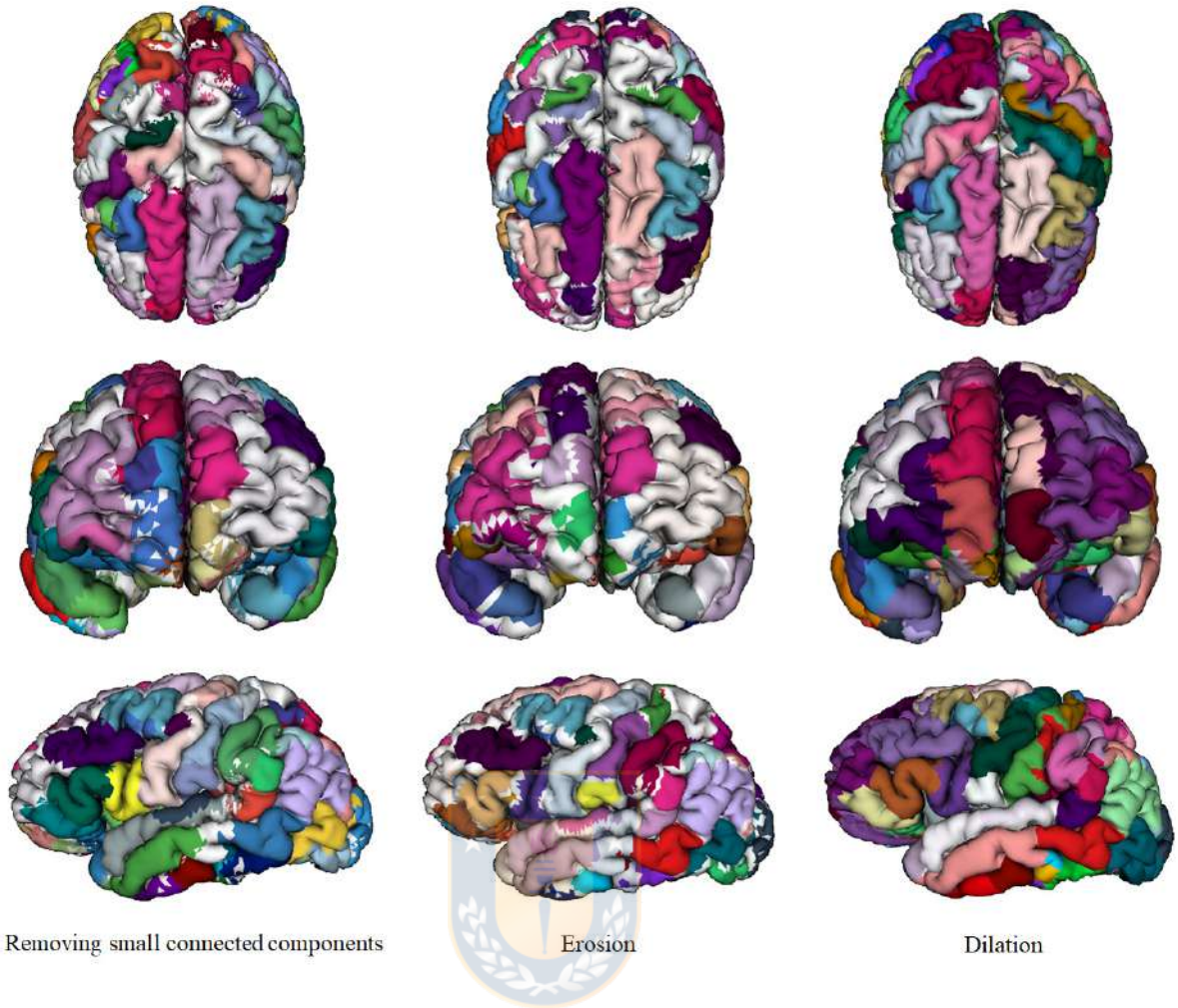


Figure B.6: Results for Parcel post-processing (Step 6) stages. The first column shows the hard parcellation after applying the removal of the small connected components. The second column shows the result after applying erosion to the sub-parcels and the third column is the result of applying dilation. This is the final hard parcellation obtained, for $size\_thr = 0.1$, $dc\_thr = 0.1$ and $idc\_thr = 0.1$, consisting of 85 sub-parcels in the left hemisphere and 72 sub-parcels in the right hemisphere. For all the columns the axial, coronal, and left sagittal views are displayed.

B.5. Brain graph construction

We performed some tests, based on graph network analysis. For that, the connectivity matrix of each of the 79 subjects was calculated for each cortical parcellation result. Each matrix was obtained using the generated parcellation, given by the sub-parcel labels, applied to the subject's cortical mesh, and the whole tractography of the subject. The constructed connectivity matrices are binary, denoting the existence or absence of a connection between the pair of sub-parcels, given by at least one fiber.

Complex networks in graph theory are modeled as a graph G formed by nodes $v \in V$ and linked by edges $e \in E$, such that $G = (V, E)$. Hence, each matrix was converted to a graph, for analyzing it using network graph metrics. Although using binary undirected graphs is a simplification of reality in terms of brain networks [8], in neuroimaging it is an accepted technique because the signal-to-noise ratio is limited in the data [1].

Figure B.7 shows the connectivity matrices for three subjects using *atlas 5*. Three matrices are displayed for each subject: binary, count (not binarized), and the logarithm count. Note that different values were tested to set the minimum number of fibers used to define the existence of a connection, leading to non-significant differences in the comparisons detailed below. As an example, Figure B.8 shows the sub-parcels and connections obtained in *atlas 5*, for the precentral (PrC) and postcentral (PoC) gyri, associated with the main motor area and the primary somatosensory cortex, respectively [16].

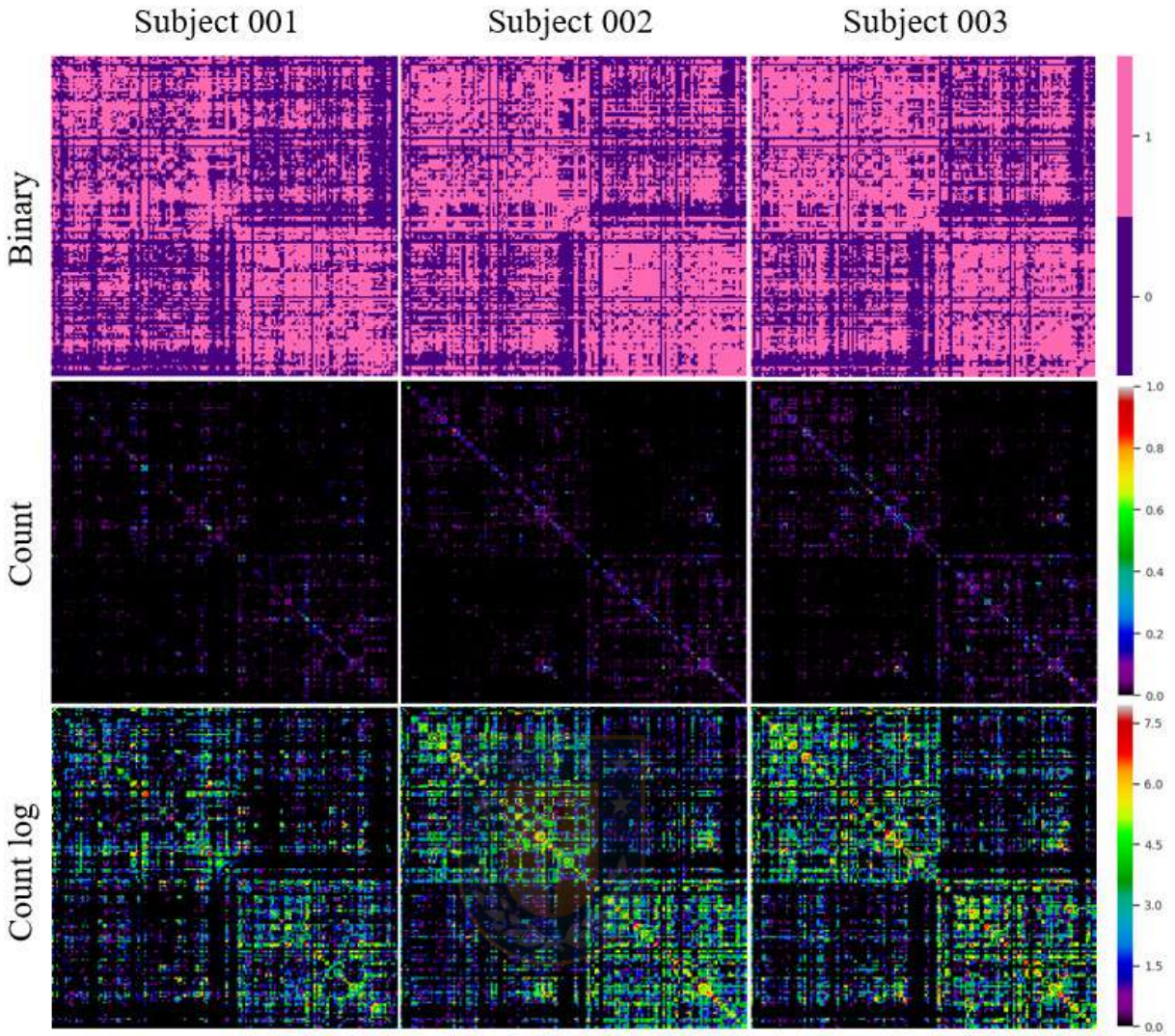


Figure B.7: Connectivity matrices for three subjects, using *atlas 5*. The matrices are squared and have a dimension equal to the number of sub-parcels of the atlas. The first row represents the binary matrices, which indicate the presence or absence of a connection between the sub-parcels. The second row displays the count matrices, which contain the number of fibers connecting each pair of sub-parcels, normalized between 0 and 1. Finally, the third row displays the logarithmic count matrices, which contain the logarithm of the non-normalized count matrix. A high similarity can be seen between the different subjects.

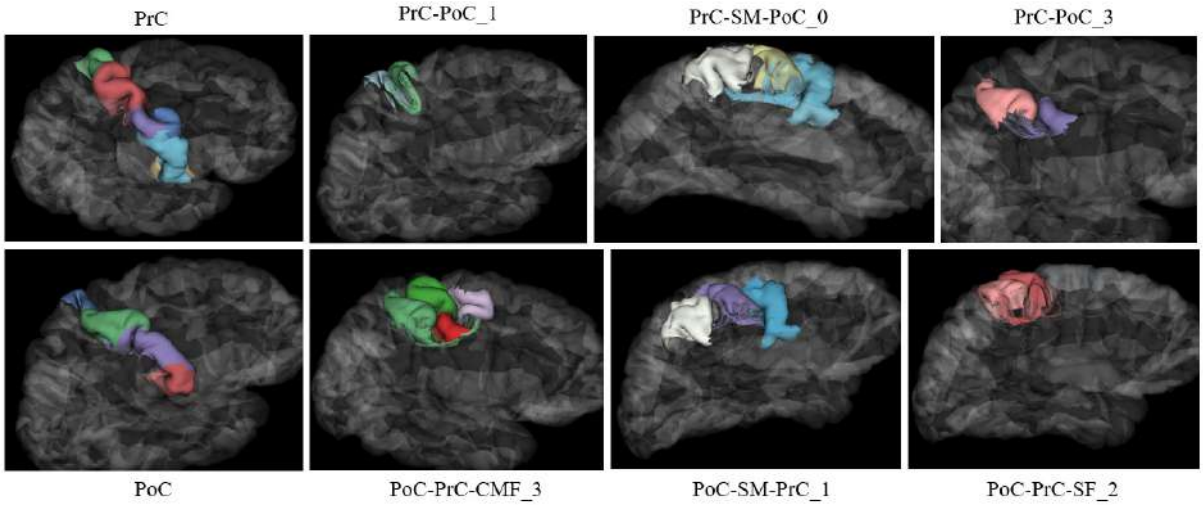


Figure B.8: Example of structural connectivity between different cortical regions of the precentral (PrC) and postcentral (PoC) sub-parcels, according to *atlas 5*, based on parameter configuration given by Table 4.1. Upper row: subdivision into sub-parcels of the precentral anatomical parcel, and the connections between these sub-parcels with other sub-parcels of PoC and SM (supramarginal) regions. Lower row: subdivision into sub-parcels of the postcentral anatomical parcel, and the connections between these sub-parcels with other sub-parcels of the CMF (caudal middle frontal), SM and SF (superior frontal) regions.

B.6. Brain network metrics

To conduct the study on the characteristics of a brain network, there is a wide range of metrics [23, 13]. Three features, representatives of the parcellation problem were selected: functional segregation, functional integration, and small-worldness [108]. To perform the calculations, we used the *bctpy* library[1], which is an adaptation of the Matlab Brain Connectivity Toolbox for Python [108]. Specifically, *bctpy* was used to calculate the clustering coefficient C (functional segregation) and path length L (functional integration) metrics. In addition, to evaluate the metric of small-world (ω coefficient), the *networkx* Python library [120, 63, 62] was employed. Supplementary material contains a description of each one of these metrics.

Clustering coefficient

This measure of functional segregation in the brain is defined as the existence of large highly connected groups of nodes or clusters in brain regions [133]. It is a measure of the completeness of a node's neighborhood. The metric can be calculated for the entire network, as the average clustering coefficient. This metric ranges from 0 to 1. A low clustering coefficient defines a random network, while a complex network has a high clustering coefficient [13]. For undirected graphs, the clustering coefficient is defined by Equation 4.3.

Equation 4.4 describes the average clustering coefficient that measures the whole network.

Figure B.9 (top left) shows a box plot for the average clustering coefficient C, obtained for each parcellation configuration (atlas), given by Table 4.1. Each box represents a generated parcellation. The median is displayed as a line inside the box, and the mean appears as a rhombus. The lines known as *mustaches* indicate the variability outside the lower quartile and the upper quartile. Finally, the red dots indicate the outliers in the data. Following the definition of the clustering coefficient, a complex network must have a high clustering coefficient, so in our chart, the generated parcellations that best meet the definition are *atlas 1* and *atlas 5*. Besides, they are the ones that show

[1] https://github.com/aestrivex/bctpy

less variability. Note that a larger number of sub-parcels will give a lower value of C, which is congruent with the literature.

Path length

It is the most used measure for functional integration and measures the number of edges that exist to get from one node to another in the network. In other words, it is the easiness to distribute information among the different regions of the brain. The path length, also called *average shortest path length* or *characteristic path length* [133], is calculated based on the shortest path length of each pair of nodes that make up the network. Lattice networks have long average path length, unlike complex networks or random networks that have short average path length [13]. For undirected graphs, the characteristic path length is defined by Equation 4.5.

Figure B.9 (top right) shows the box plots obtained for the characteristic path length L, for each parcellation configuration (atlas). We can observe that the characteristic path length is proportionally lower than the respective average clustering coefficient, which complies with the characteristic that a complex network must have a small average shortest path length.

Small-worldness

All natural networks comply with the small-world topology showing their specific functionality by moving away from randomness. The small-world metric combines a high clustering coefficient (functional segregation) while keeping a low characteristic path length (functional integration), which links all the nodes in the network [13]. In other words, all nodes are locally strongly interconnected, and they are linked to other regions through a few links.

Small-worldness can be quantified with the coefficient called ω [120]. It is calculated comparing the clustering coefficient of the analyzed network (C) to an equivalent lattice network (C_l), and its path length (L) to an equivalent random network (L_r), as described in Equation B.1.

$$\omega = \frac{L_r}{L} - \frac{C}{C_l} \tag{B.1}$$

When using the clustering coefficient of a lattice network, this metric is less sensitive to the fluctuations than the clustering coefficient of a random network. The ω coefficient ranges from -1 to 1 regardless of the size of the network. The closer the value of ω coefficient is to zero, the network is considered closer to small-world property, that is, $L \approx L_r$ and $C \approx C_l$ [120].

Figure B.9 (bottom) shows the ω coefficient for the different parcellation configurations (atlas). This coefficient has to be as close as possible to zero, to present the property of a small-world network. All parcellations comply with this property.

Best parcellation configuration according to brain network metrics

Taking into account the previous results, the configuration of parameters that achieve the best results, according to the small-world ω coefficient, are *atlas 1* and *atlas 5*. As shown, *atlas 1* and *atlas 5* have the value of ω coefficient closest to zero, presenting the best small-world network property. This is mainly due to the parameters dc_thr (density center threshold) and idc_thr (intersection of density centers threshold) since as we increase these two thresholds, fewer intersections are considered significant and fewer sub-parcels are merged. As known, fewer sub-parcels will lead to smaller path lengths, better complying with functional integration and a small shortest path length. Given that the size of the network (number of sub-parcels) is extremely determinant in small-world values, the previous results only allow us to conclude that all networks tested have properties characteristic to small-world networks, being better for networks with fewer nodes. However, obtaining a large number of sub-parcels could be a desired feature, for example, for the analysis of small functional areas.

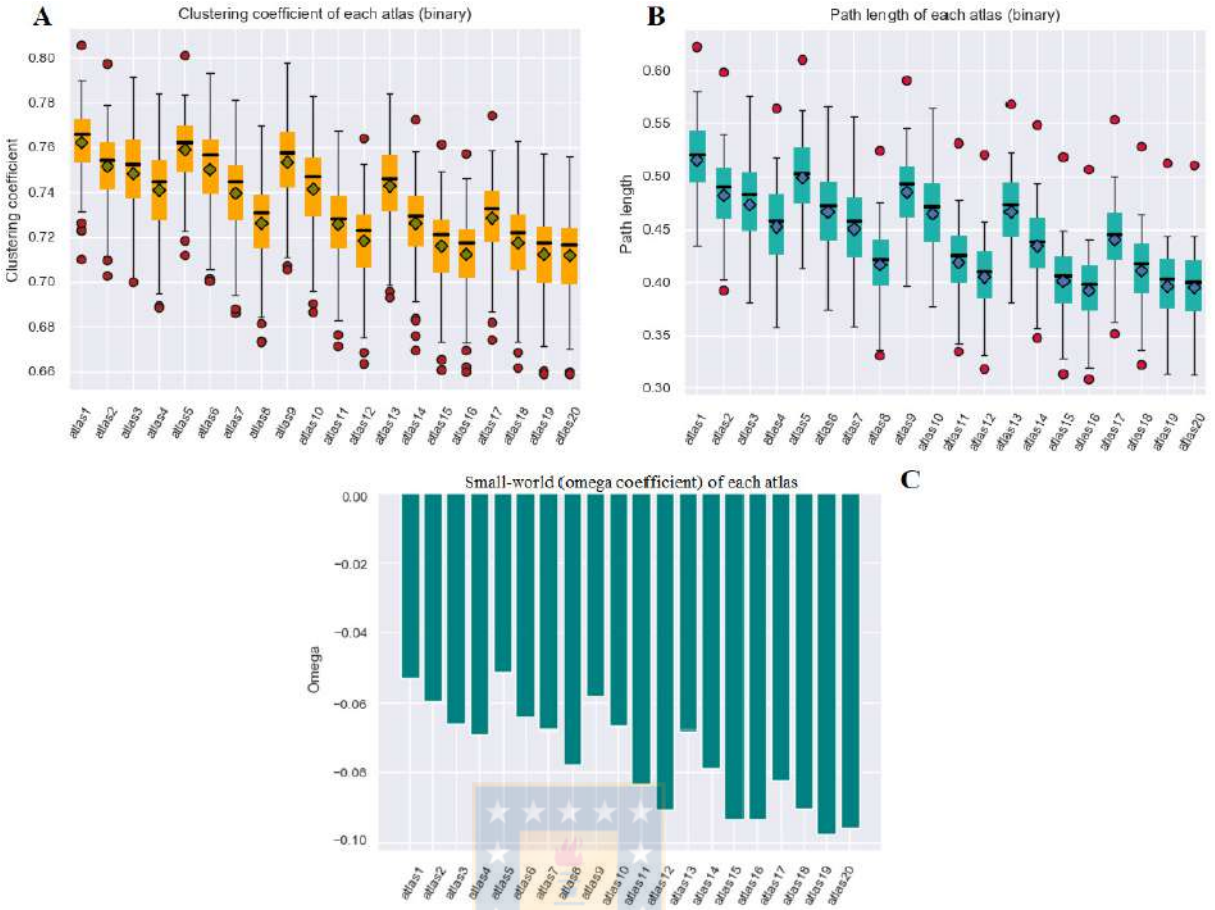


Figure B.9: Metrics of brain networks. **(A)** Average clustering coefficient C for all parcellation (atlas) configurations, given by Table 4.1. *Atlas 1* and *atlas 5* achieve the best results for the clustering coefficient, being also those that have less variability. **(B)** Path length L for all parcellation (atlas) configurations, given by Table 4.1. For all atlases, the average shortest path length is small, complying with the complex network definition. **(C)** Small-world ω coefficient for all parcellation (atlas) configurations, given by Table 4.1. *Atlas 5* appears to be slightly better, since the closer to zero, the more it complies with the property of small-world.

B.7. Cross-validation analysis

To test the dependence of the proposed method on the input data, we applied the method to the ARCHI database using k-fold cross-validation [71]. We created eight different datasets, composed of 69 subjects for the parcellation creation (atlas datasets), and the 10 remaining subjects for testing (testing datasets). The parameters used are those for *atlas 5*, i.e., $dc\_thr = 0.15$ and $idc\_thr = 0.10$. For more details, Table B.7 lists the subjects used for each dataset.

The cortical parcellation method was applied to the eight atlas datasets. Next, each created parcellation was applied to the 10 subjects of the corresponding testing dataset and the connectivity matrix of each subject was calculated according to the corresponding parcellation. Finally, the different graph network metrics were calculated for each subject (average Dice coefficient, average clustering coefficient, path length, small-world ω coefficient), and averaged for each testing dataset. In general, a high similarity between the metrics for the different parcellations was found. The average Dice coefficient presented an average between 0.70 and 0.76, which indicates a good similarity of the connectivity matrix between subjects for all the parcellations, even though some testing datasets lead to better results.

To evaluate the influence of the testing dataset on the average Dice coefficient, we calculated the connectivity matrix and Dice coefficient for the whole dataset (79 subjects), separately for each one of the eight atlases generated. We found a higher similarity between subjects, with an average Dice coefficient varying between 0.72 and 0.74. Finally, the other metrics also showed high similarity, with an average clustering coefficient between 0.76 and 0.77, an average path length between 0.50 and 0.54, and small-world ω coefficient between -0.04 and -0.08. Hence, all networks comply with the small-world property.

Figures B.10 and B.11 show the metrics of Dice similarity, clustering coefficient, path length and small-world ω coefficient applied to the testing datasets (Figure B.10) and the whole dataset (B.11) for the eight atlases created, according to Table B.7.

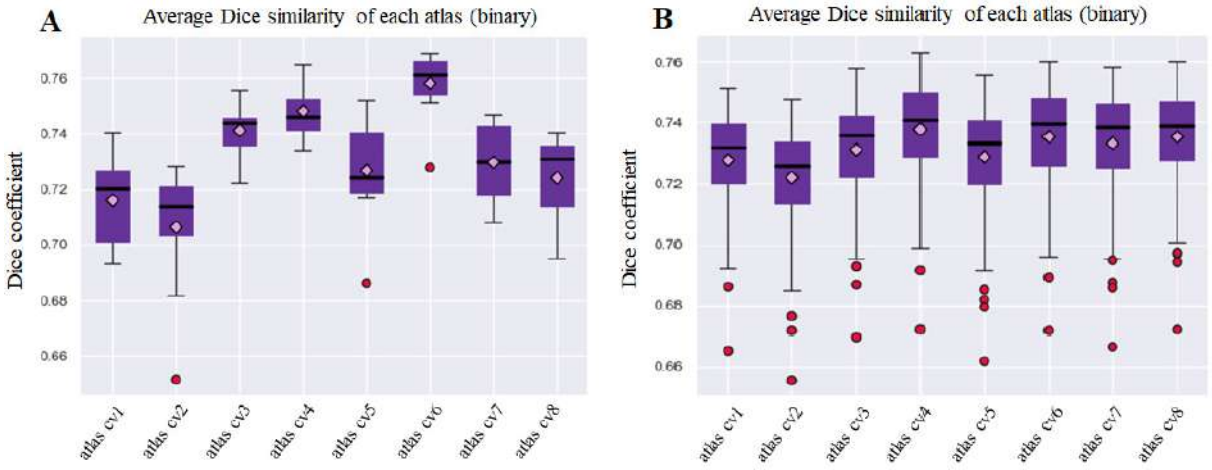


Figure B.10: **(A)** Cross-validation analysis: Dice coefficient for the 10 subjects of the testing datasets, for each generated atlas, according to Table B.7. Results show high similarity between subjects for all the atlases, even though some variability is found between the testing datasets. **(B)** Cross-validation analysis: Dice coefficient for all subjects using each generated atlas, according to Table B.7. A lower variability between subjects was found between the different atlases when using the 79 subjects.

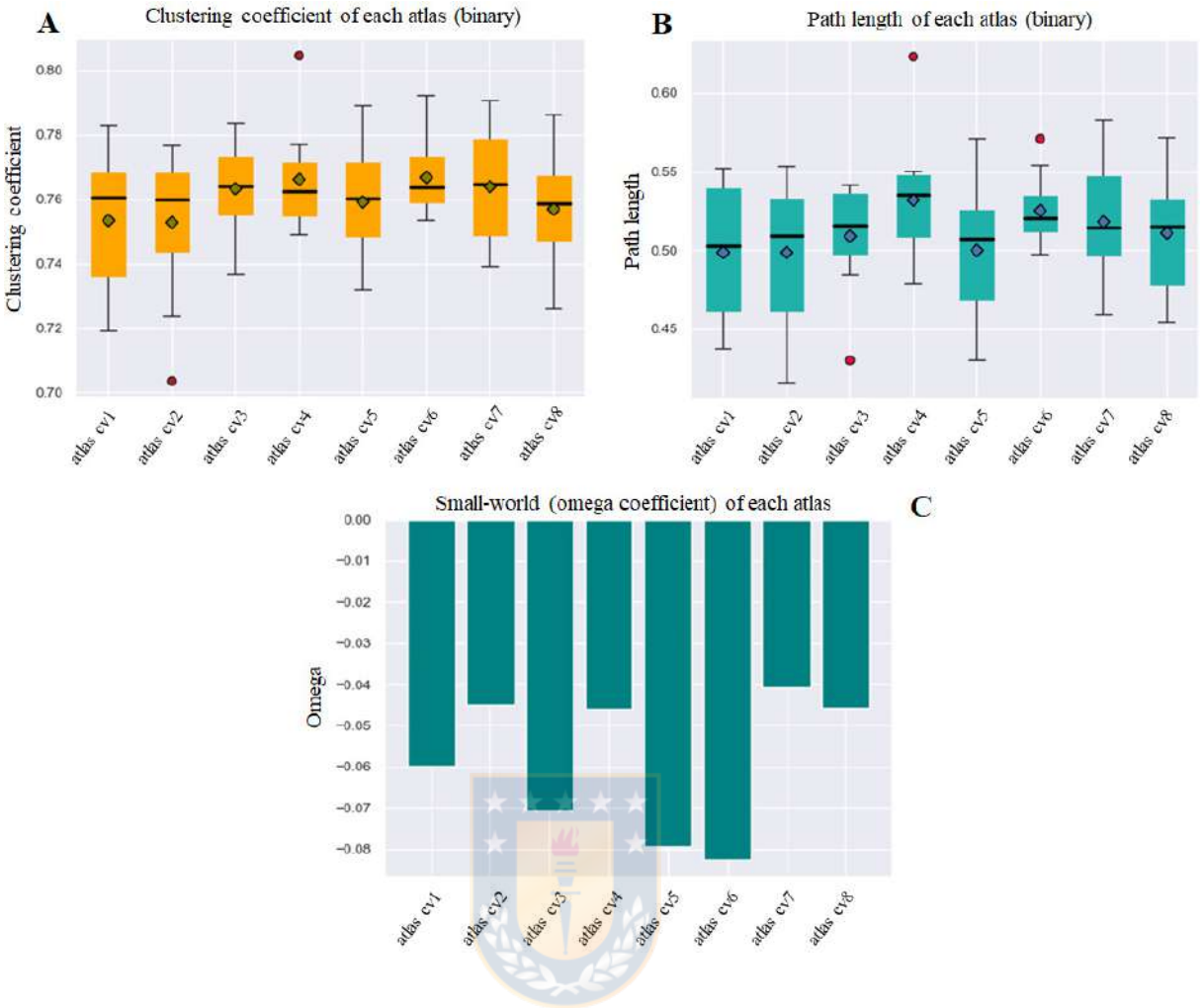


Figure B.11: Cross-validation analysis: network metrics obtained for the 10 subjects of the testing datasets, for each generated atlas, according to Table B.7. **(A)** Clustering coefficient. **(B)** Path length. **(C)** Small-world ω coefficient. In all the cases, results show a low variability between the different testing datasets. Also, an ω coefficient very close to zero was found for all the datasets, complying with the small-world property.

B.8. From Desikan-Killiany atlas to a finer granularity, example for *atlas 5*

Figure B.12 illustrates the result of the parcellation obtained for *atlas 5* as well as *Desikan-Killiany* atlas. The first two rows contain the *Desikan-Killiany* parcellation, composed of 35 parcels per hemisphere. Rows three and four show the resulting parcellation for *atlas 5*, that subdivides the *Desikan-Killiany* atlas, based on white matter fiber connectivity. The parcellation contains 160 sub-parcels, 86 sub-parcels in the left hemisphere and 74 sub-parcels in the right hemisphere. For more details, see Table B.4, which illustrates the number of sub-parcels obtained for each anatomical region in *atlas 5*, for each hemisphere.

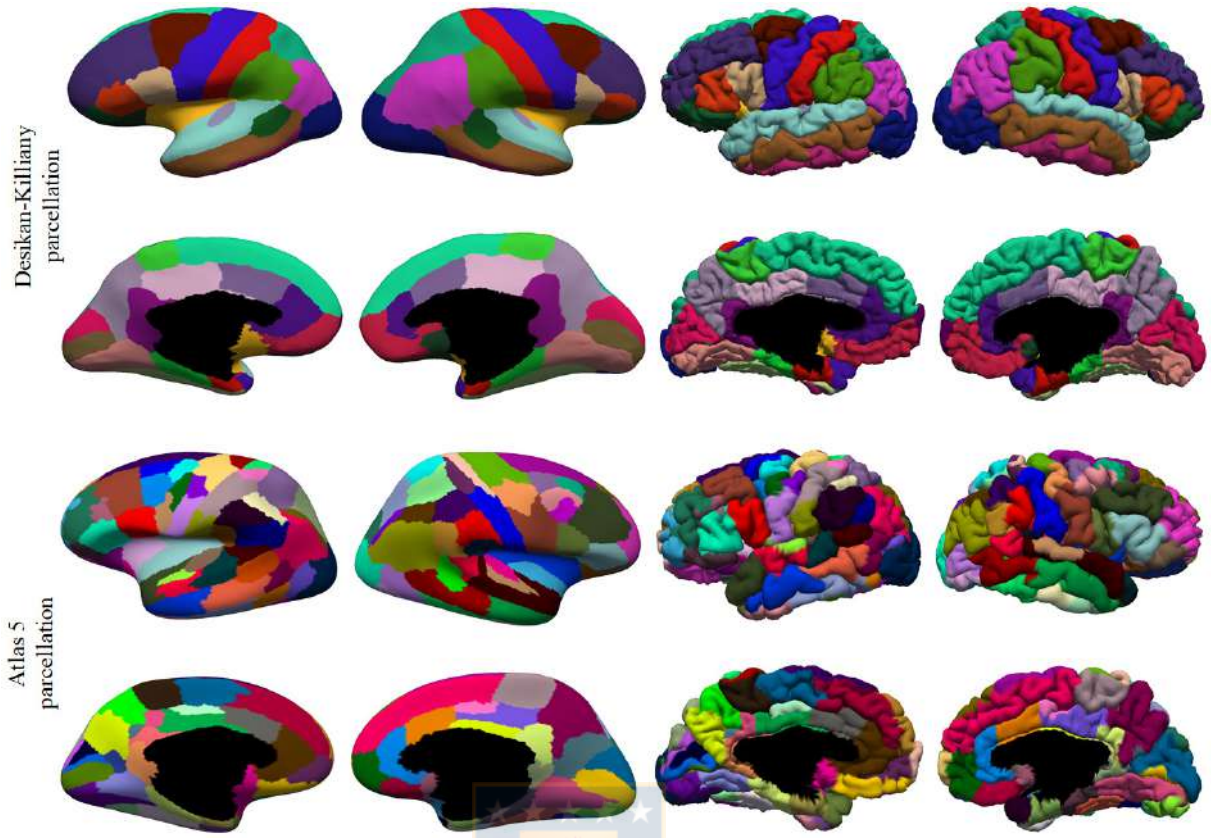


Figure B.12: *Desikan-Killiany* atlas and *atlas 5*, obtained using the parameter configuration given by Table 4.1. The first and second rows show the *Desikan-Killiany* atlas, which is formed by 70 parcels, 35 in each hemisphere. Rows three and four illustrate *atlas 5* parcellation which is composed of 160 sub-parcels, 86 sub-parcels in the left hemisphere and 74 sub-parcels in the right hemisphere. Columns one and two show both parcellations with the inflated surface, while columns three and four show the pial surface.

B.9. Comparisons with state-of-the-art parcellations based on different MRI modalities

Figure B.13 shows the comparisons performed between *atlas 5* and *Schaefer* (100 parcels), *PrAGMATiC* (320 parcels) and *Yeo* (7 or 17 networks) parcellations. Moreover, Figure B.14 illustrates the comparisons between *atlas 13* and *Glasser* (360 parcels), *PrAGMATiC* and *Schaefer* (200 parcels) parcellations. Only parcels with a Dice's coefficient ≥ 0.6 are shown.

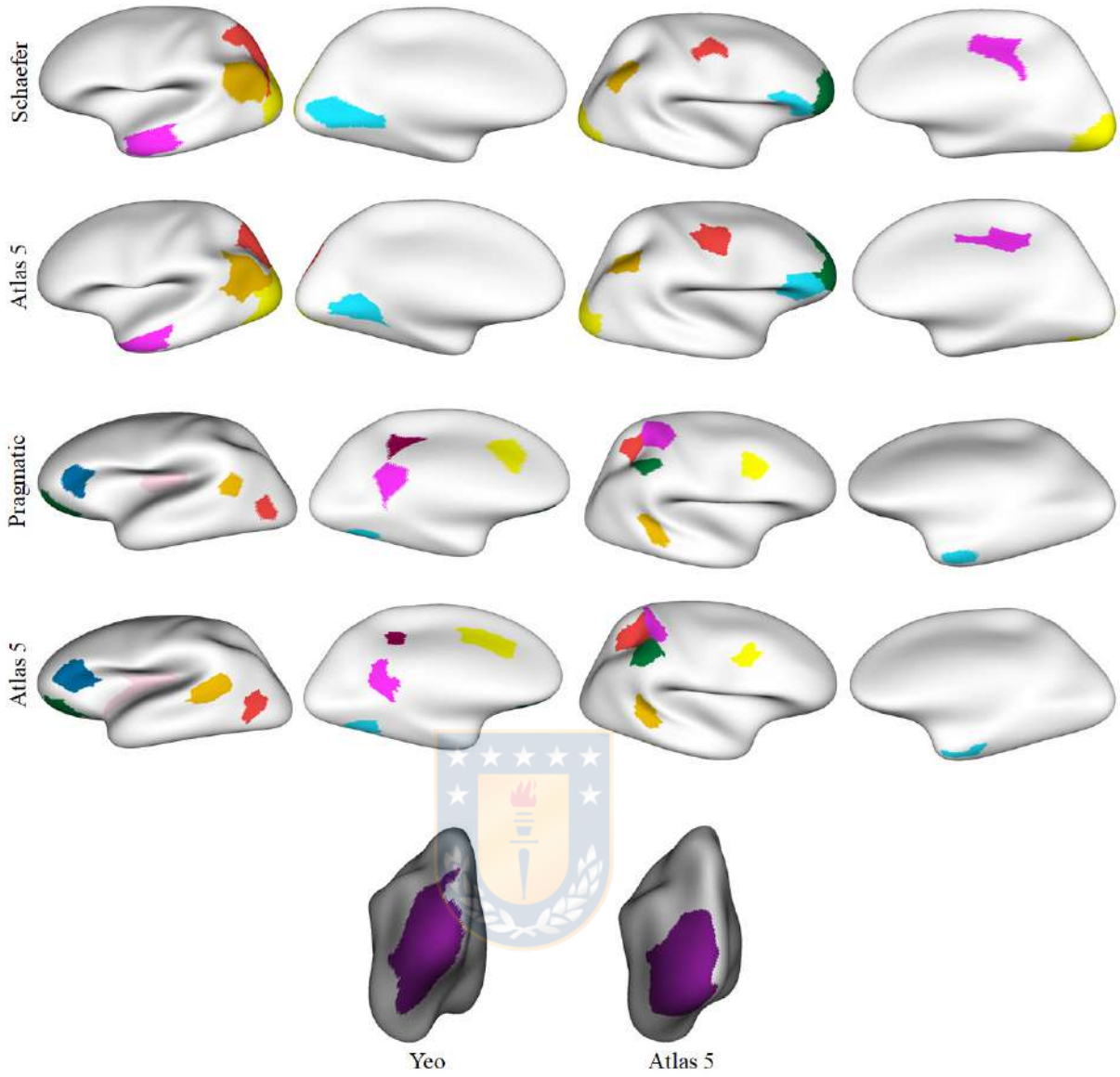


Figure B.13: Parcels in common between *atlas 5* and state-of-the-art parcellations based on different MRI modalities, with a Dice coefficient ≥ 0.6. Comparisons are shown for *Schaefer* (100 parcels), *PrAGMATiC* and *Yeo* atlases. All the meshes are inflated. *Schaefer* has 11 parcels in common with *atlas 5*, 5 in the left hemisphere and 6 in the right hemisphere. *PrAGMATiC* has 15 similar parcels with *atlas 5*, 9 in the left hemisphere and 6 in the right hemisphere. Finally, *Yeo* atlas with 17 networks has one parcel in common with *atlas 5* in the left hemisphere.

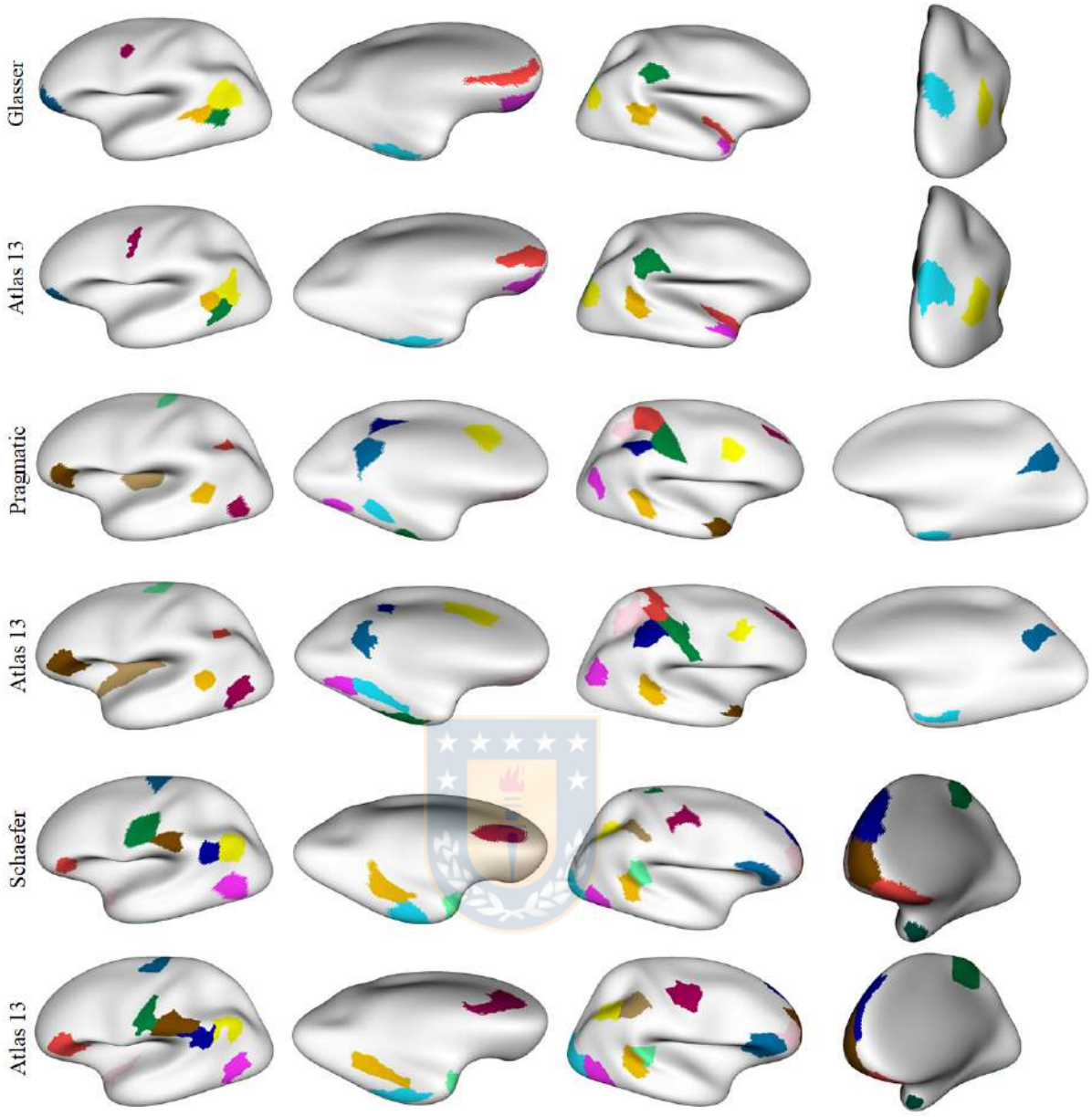


Figure B.14: Common parcels between *atlas 13* and state-of-the-art parcellations based on different MRI modalities, with a Dice coefficient ≥ 0.6. Comparisons are shown for *Glasser*, *PrAGMATiC* and *Schaefer* atlases. All the meshes are inflated. *Glasser* has 14 parcels in common with *atlas 13*, 8 in the left hemisphere and 6 in the right hemisphere. On the other hand, *PrAGMATiC* has 22 similar parcels with *atlas 13*, 12 parcels in the left hemisphere and 10 parcels in the right hemisphere. Finally, *Schaefer* parcellation (200 parcels) has 26 common parcels, 12 the left hemisphere and 14 in the right hemisphere.

B.10. Complementary Tables

Tables B.1 and B.2 list the bundles of the final fused atlas for the left and right hemispheres, respectively. The left hemisphere is composed of 50 bundles of *swm_atlas_1*, 27 bundles of *swm_atlas_2* and nine of DWM atlas. On the other hand, the right hemisphere contains 50 bundles of *swm_atlas_1*, 34 bundles of *swm_atlas_2* and nine of DWM atlas.

Table B.3 contains the 35 regions (gyri) and abbreviations of the *Desikan-Killiany* atlas [31].

Table B.4 lists for each hemisphere the number of sub-parcels obtained for *atlas 5* and *atlas 13*, respectively, for each anatomical region (gyrus) after applying the parcellation method. On the other hand, Table B.5 shows the sub-parcels that *atlas 13* has in common with the other atlases generated by the method, for a Dice coefficient ≥ 0.6.

Table B.6 illustrates the comparison between *atlas 5*, *atlas 13*, *Lefranc* and *Brainnetome* on the number of parcels per anatomical parcel, based on *Desikan-Killiany (DK)* atlas. All the atlases use *DK* atlas as a coarse anatomical parcellation, but *Braintomme* is based on some regions of this atlas that are combined. Therefore in some cases, *Braintomme* parcels cannot be perfectly matched as subdivisions of *DK* standard regions. The * is used to indicate the *DK* anatomical parcels where *Brainnetome* performs a different subdivision of the regions and only an approximate number of subdivisions is provided. In the following, we briefly describe these cases.

The Pars Opercularis and the Pars Triangularis *DK* atlas regions correspond to the Inferior Frontal Gyrus (IFG) in *Brainnetome*, that is divided into six parcels. Three parcels correspond approximately to the Pars Opercularis and the other three parcels to the Pars Triangularis.

The Orbital Gyrus (OrG) in *Brainnetome* is subdivided into six parcels, that correspond to the Pars Orbitalis (one parcel), Frontal Pole (one parcel), Lateral Orbito Frontal (one parcel), and Medial Orbito Frontal(three parcels).

The Superior Temporal, Transverse Temporal, and Temporal Pole *DK* anatomical

parcels correspond to the Superior Temporal Gyrus (STG) in *Brainnetome*. This region is subdivided into six parcels, having a good correspondence with the Superior Temporal (four parcels), the Temporal Pole (one parcel), and the Transverse Temporal (one parcel) regions.

The Parahippocampal Gyrus (PhG) in the *Brainnetome* atlas is subdivided into six parcels, corresponding to the Entorhinal (two parcels) and Parahippocampal (four parcels) *DK* anatomical regions.

The *DK* Inferior Parietal and Supramarginal anatomical parcels correspond to the Inferior Parietal Lobule (IPL) in *Brainnetome*, where each *DK* region is subdivided into three parcels.

The MedioVentral Occipital Cortex (MVOcC) of *Brainnetome* atlas is subdivided into five parcels. These correspond to two parcels for the Cuneus, one parcel for the Pericalcarine, and one parcel for the Lingual region.

Finally, the Cingulate Gyrus (CG) in *Brainnetome* is composed of 7 parcels. These parcels correspond in *DK* atlas to two parcels for the Posterior Cingulate, two parcels for the Isthmus, two parcels for the Anterior Cingulate and one parcel for the Rostral Anterior Cingulate region.

See the supplementary material[2] of “The Human Brainnetome Atlas” for more information on equivalences between atlases.

Finally, Table B.7 lists the datasets used for cross-validation analysis from the 79 subjects of the ARCHI database.

[2]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4961028/

Bundle	Atlas	Bundle	Atlas	Bundle	Atlas
AR_ANT_LEFT	DWM	lh_IT-IT_1l	*swm_atlas_2*	lh_PoC-SM_0	*swm_atlas_1*
AR_LEFT	DWM	lh_IT-MT_0	*swm_atlas_1*	lh_PoC-SM_1	*swm_atlas_1*
AR_POST_LEFT	DWM	lh_LOF-LOF_0i	*swm_atlas_2*	lh_PrC-Ins_0	*swm_atlas_1*
CG_LEFT	DWM	lh_LOF-Or_0	*swm_atlas_1*	lh_PrC-PrC_0l	*swm_atlas_2*
CG2_LEFT	DWM	lh_LOF-RMF_0	*swm_atlas_1*	lh_PrC-SF_0	*swm_atlas_1*
CG3_LEFT	DWM	lh_LOF-RMF_1	*swm_atlas_1*	lh_PrC-SM_0	*swm_atlas_1*
IFO_LEFT	DWM	lh_LOF-ST_0	*swm_atlas_1*	lh_PrC-SM_1i	*swm_atlas_2*
IL_LEFT	DWM	lh_LO-LO_0i	*swm_atlas_2*	lh_PrCu-PrCu_0l	*swm_atlas_2*
lh_CAC-PrCu_0	*swm_atlas_1*	lh_LO-LO_1i	*swm_atlas_2*	lh_RAC-SF_1	*swm_atlas_1*
lh_CMF-CMF_0i	*swm_atlas_2*	lh_LO-LO_2l	*swm_atlas_2*	lh_RMF-RMF_0i	*swm_atlas_2*
lh_CMF-Op_0	*swm_atlas_1*	lh_MOF-ST_0	*swm_atlas_1*	lh_RMF-RMF_1i	*swm_atlas_2*
lh_CMF-PoC_0	*swm_atlas_1*	lh_MT-MT_0i	*swm_atlas_2*	lh_RMF-SF_0	*swm_atlas_1*
lh_CMF-PrC_0	*swm_atlas_1*	lh_MT-MT_1i	*swm_atlas_2*	lh_RMF-SF_1	*swm_atlas_1*
lh_CMF-PrC_1	*swm_atlas_1*	lh_MT-MT_1l	*swm_atlas_2*	lh_SF-SF_0l	*swm_atlas_2*
lh_CMF-RMF_0	*swm_atlas_1*	lh_MT-SM_0	*swm_atlas_1*	lh_SM-Ins_0	*swm_atlas_1*
lh_CMF-SF_0	*swm_atlas_1*	lh_MT-ST_0	*swm_atlas_1*	lh_SM-SM_0i	*swm_atlas_2*
lh_Cu-Li_0l	*swm_atlas_2*	lh_Op-Ins_0	*swm_atlas_1*	lh_SM-SM_1i	*swm_atlas_2*
lh_Fu-Fu_0i	*swm_atlas_2*	lh_Op-PrC_0	*swm_atlas_1*	lh_SM-SM_2i	*swm_atlas_2*
lh_Fu-Fu_1l	*swm_atlas_2*	lh_Op-SF_0	*swm_atlas_1*	lh_SP-SM_0	*swm_atlas_1*
lh_Fu-IT_0i	*swm_atlas_2*	lh_Or-Ins_0	*swm_atlas_1*	lh_SP-SP_0i	*swm_atlas_2*
lh_Fu-LO_0	*swm_atlas_1*	lh_PoC-Ins_0	*swm_atlas_1*	lh_ST-Ins_0	*swm_atlas_1*
lh_IC-PrCu_0	*swm_atlas_1*	lh_PoCi-PrCu_0	*swm_atlas_1*	lh_ST-ST_0i	*swm_atlas_2*
lh_IP-IT_0	*swm_atlas_1*	lh_PoCi-PrCu_1	*swm_atlas_1*	lh_ST-ST_1l	*swm_atlas_2*
lh_IP-LO_1	*swm_atlas_1*	lh_PoCi-RAC_0	*swm_atlas_1*	lh_ST-TT_0	*swm_atlas_1*
lh_IP-MT_0	*swm_atlas_1*	lh_PoCi-SF_0	*swm_atlas_1*	lh_Tr-Ins_0	*swm_atlas_1*
lh_IP-SM_0	*swm_atlas_1*	lh_PoC-PrC_0	*swm_atlas_1*	lh_Tr-RMF_0i	*swm_atlas_2*
lh_IP-SP_0	*swm_atlas_1*	lh_PoC-PrC_1	*swm_atlas_1*	lh_Tr-SF_0	*swm_atlas_1*
lh_IP-SP_1	*swm_atlas_1*	lh_PoC-PrC_2	*swm_atlas_1*	UN_LEFT	DWM
lh_IT-IT_0l	*swm_atlas_2*	lh_PoC-PrC_3	*swm_atlas_1*		

Table B.1: Bundles of the left hemisphere of the final fused bundle atlas. The name of the bundles in the original SWM and DWM atlases and the name of the atlas to which the bundles belong are listed. Nine bundles are from DWM atlas, 50 bundles are from *swm_atlas_1* and 27 bundles are from *swm_atlas_2*.

Bundle	Atlas	Bundle	Atlas	Bundle	Atlas
AR_ANT_RIGHT	DWM	rh_LOF-LOF_1r	*swm_atlas_2*	rh_PrC-Ins_0	*swm_atlas_1*
AR_POST_RIGHT	DWM	rh_LOF-MOF_0	*swm_atlas_1*	rh_PrC-SF_0i	*swm_atlas_2*
AR_RIGHT	DWM	rh_LOF-RMF_0	*swm_atlas_1*	rh_PrC-SM_0	*swm_atlas_1*
CG_RIGHT	DWM	rh_LOF-RMF_1	*swm_atlas_1*	rh_PrC-SM_1i	*swm_atlas_2*
CG2_RIGHT	DWM	rh_LOF-ST_0	*swm_atlas_1*	rh_PrC-SP_0	*swm_atlas_1*
CG3_RIGHT	DWM	rh_LO-LO_0i	*swm_atlas_2*	rh_PrCu-PrCu_0r	*swm_atlas_2*
IFO_RIGHT	DWM	rh_LO-LO_1i	*swm_atlas_2*	rh_RAC-SF_0	*swm_atlas_1*
IL_RIGHT	DWM	rh_LO-SP_0	*swm_atlas_1*	rh_RMF-RMF_0i	*swm_atlas_2*
rh_CAC-PoCi_0	*swm_atlas_1*	rh_MOF-ST_0	*swm_atlas_1*	rh_RMF-RMF_0r	*swm_atlas_2*
rh_CAC-PrCu_0	*swm_atlas_1*	rh_MT-MT_0i	*swm_atlas_2*	rh_RMF-RMF_1i	*swm_atlas_2*
rh_CMF-CMF_0i	*swm_atlas_2*	rh_MT-MT_0r	*swm_atlas_2*	rh_RMF-RMF_1r	*swm_atlas_2*
rh_CMF-Op_0i	*swm_atlas_2*	rh_MT-MT_1i	*swm_atlas_2*	rh_RMF-SF_0	*swm_atlas_1*
rh_CMF-PrC_0	*swm_atlas_1*	rh_MT-SM_0	*swm_atlas_1*	rh_RMF-SF_0r	*swm_atlas_2*
rh_CMF-PrC_1	*swm_atlas_1*	rh_MT-ST_0	*swm_atlas_1*	rh_RMF-SF_1	*swm_atlas_1*
rh_CMF-RMF_0	*swm_atlas_1*	rh_Op-Ins_0	*swm_atlas_1*	rh_SF-SF_1r	*swm_atlas_2*
rh_CMF-SF_0	*swm_atlas_1*	rh_Op-PrC_0	*swm_atlas_1*	rh_SF-SF_2r	*swm_atlas_2*
rh_CMF-SF_1	*swm_atlas_1*	rh_Op-SF_0	*swm_atlas_1*	rh_SM-Ins_0	*swm_atlas_1*
rh_Cu-Li_0	*swm_atlas_1*	rh_Op-Tr_0	*swm_atlas_1*	rh_SM-SM_0i	*swm_atlas_2*
rh_Fu-Fu_0i	*swm_atlas_2*	rh_Or-Ins_0	*swm_atlas_1*	rh_SM-SM_1i	*swm_atlas_2*
rh_Fu-IT_0i	*swm_atlas_2*	rh_PoCi-PrCu_1	*swm_atlas_1*	rh_SM-SM_2i	*swm_atlas_2*
rh_Fu-LO_1	*swm_atlas_1*	rh_PoCi-PrCu_2	*swm_atlas_1*	rh_SP-SM_0	*swm_atlas_1*
rh_IC-PrCu_0	*swm_atlas_1*	rh_PoCi-RAC_0	*swm_atlas_1*	rh_SP-SP_0i	*swm_atlas_2*
rh_IP-IP_0r	*swm_atlas_2*	rh_PoC-PoC_1r	*swm_atlas_2*	rh_SP-SP_0r	*swm_atlas_2*
rh_IP-IT_0	*swm_atlas_1*	rh_PoC-PrC_0	*swm_atlas_1*	rh_ST-ST_0i	*swm_atlas_2*
rh_IP-LO_0	*swm_atlas_1*	rh_PoC-PrC_1	*swm_atlas_1*	rh_ST-TT_0	*swm_atlas_1*
rh_IP-MT_0	*swm_atlas_1*	rh_PoC-PrC_1r	*swm_atlas_2*	rh_Tr-Ins_0	*swm_atlas_1*
rh_IP-SM_0	*swm_atlas_1*	rh_PoC-PrC_2	*swm_atlas_1*	rh_Tr-RMF_0i	*swm_atlas_2*
rh_IP-SP_0	*swm_atlas_1*	rh_PoC-PrC_3i	*swm_atlas_2*	rh_Tr-SF_0	*swm_atlas_1*
rh_IT-MT_1	*swm_atlas_1*	rh_PoC-SM_0	*swm_atlas_1*	rh_Tr-SF_1r	*swm_atlas_2*
rh_IT-MT_2	*swm_atlas_1*	rh_PoC-SP_0	*swm_atlas_1*	rh_Tr-Tr_0r	*swm_atlas_2*
rh_LOF-LOF_0i	*swm_atlas_2*	rh_PoC-SP_1	*swm_atlas_1*	UN_RIGHT	DWM

Table B.2: Bundles of the right hemisphere of the final fused bundle atlas. The name of the bundles in the original SWM and DWM atlases and the name of the atlas to which the bundles belong are listed. Nine bundles are from DWM atlas, 50 bundles are from *swm_atlas_1* and 34 bundles are from *swm_atlas_2*.

Region (gyrus)	Abbreviation
Bankssts	Ban
Caudal anterior cingulate	CAC
Caudal middle frontal	CMF
Corpus callosum	COC
Cuneus	Cu
Entorhinal	En
Fusiform	Fu
Inferior parietal	IP
Inferior temporal	IT
Isthmus cingulate	IC
Lateral occipital	LO
Lateral orbito frontal	LOF
Lingual	Li
Medial orbito frontal	MOF
Middle temporal	MT
Parahippocampal	PH
Paracentral	PC
Pars opercularis	Op
Pars orbitalis	Or
Pars triangularis	Tr
Pericalcarine	PeCa
Postcentral	PoC
Posterior cingulate	PoCi
Precentral	PrC
Precuneus	PrCu
Rostral anterior cingulate	RAC
Rostral middle frontal	RMF
Superior frontal	SF
Superior parietal	SP
Superior temporal	ST
Supramarginal	SM
Frontal pole	FRP
Temporal pole	TEM
Transverse temporal	TT
Insula	Ins

Table B.3: Regions and abbreviations of *Desikan-Killiany* atlas [31]. The regions are present in both hemispheres.

Region (gyrus)	# SP lh atlas 5	# SP rh atlas 5	# SP lh atlas 13	# SP rh atlas 13
Bankssts	1	1	**2**	1
Caudal anterior cingulate	1	1	1	1
Caudal middle frontal	3	4	3	4
Cuneus	3	2	**1**	**1**
Entorhinal	2	2	2	**1**
Fusiform	4	3	**5**	3
Inferior parietal	2	3	**6**	**5**
Inferior temporal	4	3	**3**	3
Isthmus cingulate	3	2	3	2
Lateral occipital	2	3	**3**	**2**
Lateral orbito frontal	3	3	**4**	3
Lingual	2	2	**3**	2
Medial orbito frontal	2	2	**3**	2
Middle temporal	3	2	**4**	2
Parahippocampal	1	1	1	1
Paracentral	1	1	1	1
Pars opercularis	1	1	1	1
Pars orbitalis	1	1	**2**	1
Pars triangularis	1	1	1	1
Pericalcarine	1	1	1	1
Postcentral	4	3	**6**	**6**
Posterior cingulate	3	3	3	3
Precentral	7	3	**9**	3
Precuneus	3	2	**4**	**5**
Rostral anterior cingulate	1	1	1	1
Rostral middle frontal	4	2	4	**4**
Superior frontal	4	3	4	**4**
Superior parietal	3	4	**2**	4
Superior temporal	5	4	**6**	4
Supramarginal	4	4	4	4
Frontal pole	1	1	1	1
Temporal pole	1	1	1	1
Transverse temporal	1	1	1	1
Insula	3	2	3	2

Table B.4: Number of sub-parcels for each anatomical region (gyrus) for *atlas 5* and *atlas 13*. First column: anatomical parcels from *Desikan-Killiany* atlas. Second and third columns: number of sub-parcels per hemisphere after applying the parcellation method for *atlas 5*. Fourth and fifth columns: number of sub-parcels per hemisphere for *atlas 13*. The sub-parcels of the *atlas 13* that differ from the *atlas 5* are marked in bold.

Region (gyrus)	# SP lh	# SP rh
Bankssts	1	1
Caudal anterior cingulate	1	1
Caudal middle frontal	0	3
Corpus callosum	0	0
Cuneus	0	1
Entorhinal	1	1
Fusiform	3	1
Inferior parietal	1	2
Inferior temporal	1	2
Isthmus cingulate	2	1
Lateral occipital	1	1
Lateral orbito frontal	2	3
Lingual	2	1
Medial orbito frontal	2	1
Middle temporal	3	1
Parahippocampal	1	1
Paracentral	1	1
Pars opercularis	0	0
Pars orbitalis	1	0
Pars triangularis	1	1
Pericalcarine	1	1
Postcentral	3	0
Posterior cingulate	1	0
Precentral	2	1
Precuneus	0	2
Rostral anterior cingulate	1	1
Rostral middle frontal	2	1
Superior frontal	2	3
Superior parietal	2	2
Superior temporal	2	1
Supramarginal	1	4
Frontal pole	1	0
Temporal pole	1	1
Transverse temporal	1	1
Insula	3	0

Table B.5: Number of sub-parcels in common between *atlas 13* and the other atlases generated by the method of parcellation for each anatomical region (gyrus) for a Dice coefficient ≥ 0.6. First column: anatomical parcels from *Desikan-Killiany* atlas. Second and third columns: number of sub-parcels per hemisphere after applying the parcellation method.

Region (gyrus)	lh_atlas 5	rh_atlas 5	lh_atlas 13	rh_atlas 13	lh_Lefranc	rh_Lefranc	lh_BN	rh_BN
Bankssts	1	1	2	1	4	2	2	2
Caudal anterior cingulate	1	1	1	1	4	7	**2***	**2***
Caudal middle frontal	3	4	3	4	3	3	**2***	**2***
Cuneus	3	2	1	1	2	3	**2***	**2***
Entorhinal	2	2	2	1	3	2	**2***	**2***
Fusiform	4	3	5	3	8	7	3	3
Inferior parietal	2	3	6	5	3	3	**3***	**3***
Inferior temporal	4	3	3	3	2	2	7	7
Isthmus cingulate	3	2	3	2	3	6	**2***	**2***
Lateral occipital	2	3	3	2	9	2	6	6
Lateral orbito frontal	3	3	4	3	4	2	**1***	**1***
Lingual	2	2	3	2	2	3	**2***	**2***
Medial orbito frontal	2	2	3	2	2	4	**3***	**3***
Middle temporal	3	2	4	2	6	7	4	4
Parahippocampal	1	1	1	1	3	1	**4***	**4***
Paracentral	1	1	1	1	4	3	2	2
Pars opercularis	1	1	1	1	2	2	**3***	**3***
Pars orbitalis	1	1	2	1	4	3	**1***	**1***
Pars triangularis	1	1	1	1	2	2	**3***	**3***
Pericalcarine	1	1	1	1	3	4	**1***	**1***
Postcentral	4	3	6	6	5	5	4	4
Posterior cingulate	3	3	3	3	3	3	**2***	**2***
Precentral	7	3	9	3	5	5	6	6
Precuneus	3	2	4	5	2	2	4	4
Rostral anterior cingulate	1	1	1	1	2	2	**1***	**1***
Rostral middle frontal	4	2	4	4	5	2	**5***	**5***
Superior frontal	4	3	4	4	5	3	7	7
Superior parietal	3	4	2	4	3	4	5	5
Superior temporal	5	4	6	4	8	3	**4***	**4***
Supramarginal	4	4	4	4	3	5	**3***	**3***
Frontal pole	1	1	1	1	2	3	**1***	**1***
Temporal pole	1	1	1	1	2	2	**1***	**1***
Transverse temporal	1	1	1	1	3	3	**1***	**1***
Insula	3	2	3	2	4	2	6	6

Table B.6: Comparison of the number of subdivisions of *Desikan-Killiany (DK)* atlas regions, for the different atlases based on dMRI. First column: the *DK* anatomical parcel. The second and third columns correspond to *atlas 5*, while the fourth and fifth columns correspond to *atlas 13*. The sixth and seventh columns refer to *Lefranc* atlas and the last two columns to the *Brainnetome* atlas. For each atlas, the left and right hemisphere are indicated by *lh* and *rh*, respectively. The numbers that appear in bold with an asterisk in *Brainnetome* represent the parcels that do not perfectly match with *Desikan-Killiany* atlas regions.

# atlas	testing subjects	subjects for the atlas
cv1	1-10	11-79
cv2	11-20	1-10, 21-79
cv3	21-30	1-20, 31-79
cv4	31-40	1-30, 41-79
cv5	41-50	1-40, 51-79
cv6	51-60	1-50, 61-79
cv7	61-70	1-60, 71-79
cv8	70-79	1-69

Table B.7: Subject datasets used for cross-validation analysis from the 79 subjects of the ARCHI database. Left column: the eight groups of atlas generation and testing datasets. The atlases are named from $cv1$ to $cv8$. Second column: the range of subjects used for the testing datasets. Note that in $cv8$, subject 70 is repeated, to complete 10 subjects. Third column: the range of subjects used for the atlas creation datasets.

Appendix C

Resumen del trabajo realizado

C.1. Introducción

El cerebro humano es altamente complejo y su funcionamiento completo no ha sido descifrado en la actualidad. Su estudio se remonta al antiguo Egipto, y aunque siempre ha existido el deseo de comprender su funcionamiento y su estructura anatómica, su anatomía no se ha conocido hasta hace poco, especialmente en lo que respecta a la conectividad de la materia blanca y su estructura [103]. Gracias al estudio de las diferentes áreas cerebrales en base a diferentes modalidades es posible comprender mejor su funcionamiento, así como las múltiples patologías asociadas. La estructura y la función del cerebro están fuertemente ligadas.

Para entender cómo funciona el cerebro, es necesario conocer la red de conexiones del cerebro o el conectoma del cerebro humano. Los elementos básicos del conectoma son una composición de cuerpos de neuronas cerebrales (nodos) y las conexiones (aristas) que existen entre ellos, formados por las fibras neuronales de materia blanca [118, 59, 13].

Actualmente, se conocen diferentes técnicas para poder observar el cerebro de manera no invasiva. Una de ellas es la Resonancia Magnética (MRI), a través de la cual se han podido obtener imágenes de alta resolución para estudiar las diferentes áreas del cerebro. Esta técnica tiene diferentes modalidades, una de las cuales se conoce como Imagen de Resonancia Magnética de difusión (dMRI) [90]. Esto nos permite caracterizar el proceso de difusión de las moléculas de agua [6, 75]. Gracias a ello, se pueden obtener las trayectorias de las fibras que forman los principales fascículos que conectan las diferentes áreas del cerebro [101] y, mediante algoritmos de tractografía, se pueden reconstruir en 3D [92]. Así, la estructura del cerebro puede

ser estudiada in vivo y de manera no invasiva [91, 129, 53].

Sin embargo, los resultados basados en la tractografía de dMRI deben ser analizados cuidadosamente debido a las limitaciones de esta modalidad. El análisis de los resultados de los algoritmos de tractografía contienen ruido debido a los artefactos de la tractografía que pueden conducir a diferencias en los perfiles de conectividad en los estudios con poblaciones de sujetos [15]. A pesar de que los algoritmos de tractografía generan fascículos válidos para el estudio de la conectividad cerebral, el gran número de líneas o fibras que pasan por los vóxeles tienen la limitación espacial del tamaño de los mismos, generando así una serie de falsos positivos y negativos. Uno de los desafíos a conseguir en los próximos años será reducir estos falsos positivos y mejorar la resolución de los tractos de materia blanca [82]. Además, la alta variabilidad de las fibras de materia blanca entre los sujetos dificulta el estudio de la conectividad cerebral, ya que los resultados de la tractografía generan complejos conjuntos de datos de al menos 1 millón de fibras, especialmente si se tienen en cuenta las fibras de materia blanca superficiales.

El estudio del conectoma humano es un área clave y en crecimiento de investigación [121]. También es la intersección de los siguientes campos: Biología, Ingeniería Electrónica, Ciencias de la Computación, Física y Neurociencia. En particular, ciencias de la computación es una disciplina muy importante porque proporciona métodos eficientes y robustos que pueden trabajar con grandes conjuntos de datos en un tiempo razonable. Además, estos métodos son capaces de tratar el ruido producido en las tractografías, así como manejar la alta variabilidad que existe entre los sujetos. El manejo de los datos de las tractografías no es una tarea trivial, ya que es más compleja de lo que parece a priori, pues requiere la manipulación de muchos formatos, los datos contienen ruido y es necesario aplicar transformaciones. Desde ese punto de vista, se necesita una mayor validación de los resultados al utilizar la tractografía de difusión, que debe contrastarse con la anatomía, la resonancia magnética funcional y los estudios post mortem. Esta labor se centra en los métodos para analizar los datos de la tractografía a fin de obtener la mejor información sobre las conexiones y contribuir a la comprensión del conectoma humano.

En esta tesis, proponemos algoritmos y métodos eficientes para el estudio de la conectividad del cerebro capaces de trabajar con grandes conjuntos de fibras generadas por algoritmos de tractografía. Por lo tanto, contribuimos al desarrollo de un método de clustering de fibras de materia blanca para tratar los datos de la tractografía. Actualmente, hay muchos métodos de clustering de fibras [42, 43, 55, 106]. Sin embargo, no todos ellos realizan una validación tan minuciosa como la que proponemos en esta tesis. El clustering de fibras no valida si las fibras son verdaderas o no, ya que busca representar los fascículos de materia blanca presentes en los sujetos para posteriormente realizar estudios y análisis de conectividad cerebral. Además, desarrollamos un método que realiza el etiquetado de los fascículos de materia blanca superficial. Existe un amplio conocimiento sobre las fibras de larga asociación y proyección, utilizadas para identificar regiones funcionales a través de su conectividad, pero los fascículos de materia blanca superficial son menos estudiados en el estado del arte, debido a que los fascículos de fibras tienen una gran variabilidad entre los sujetos [49, 106] y tienen un tamaño pequeño, lo que hace que etiquetarlas e identificarlas sean tareas complejas. Por lo tanto, mediante los métodos intra e intersujeto desarrollados, hemos podido identificar los fascículos intersujeto con un alto grado de reproducibilidad y darles un nombre automático ya que no están descritos en detalle en la anatomía. Finalmente, en esta tesis, se proponen tres métodos de parcelación, el principal basado en un atlas de fascículos de fibras de materia blanca, y dos métodos secundarios basados en clustering y la topología del cerebro. Una parcelación cortical es un método para subdividir la corteza cerebral en parcelas [141]. Existen parcelaciones basadas en diferentes modalidades, como la anatómica, la información basada en la difusión, la resonancia magnética funcional y la multimodal. Como actualmente hay muchos tipos diferentes de parcelaciones, no hay consenso sobre qué parcelación es mejor que otra, o cuál es el mejor método para llevar a cabo una parcelación. Por esta razón, hemos desarrollado un método de parcelación basado en un atlas de fibras de materia blanca que puede ayudar en el estudio y descifrado del conectoma humano.

Además, como prueba de concepto, hemos llevado a cabo una parcelación individual basada en el clustering de fibras para hacer frente a la limitación que se encuentra cuando se utiliza un atlas de fascículos de fibras. Al utilizar la tractografía completa, el

número de parcelas obtenidas es mayor ya que se utilizan todas las fibras del cerebro y no sólo un conjunto de fibras que forman el atlas. Finalmente, hemos desarrollado un método de parcelación individual basado en la topología de la malla cerebral. Este método es capaz de llevar a cabo la parcelación cortical en un tiempo eficiente y con una buena homogeneidad en las parcelas obtenidas, siendo altamente competitivo con los métodos del estado del arte basados en la macroanatomía.

Para concluir, es probable que esta investigación sea de gran interés para los neurocientíficos, los neuroanatomistas y los neurólogos, ya que se pueden utilizar métodos de preprocesamiento y programas informáticos de alta calidad para estudiar la conectividad del cerebro. Gracias a este trabajo será posible desarrollar estudios clínicos modernos analizando nuevas bases de datos y así poder obtener respuestas sobre las funciones y la estructura del cerebro.

C.2. Metodología utilizada

C.2.1. Objetivo principal

Esta tesis se centra en el desarrollo de algoritmos y métodos para analizar y estudiar la conectividad anatómica del cerebro utilizando datos de dMRI, y así cooperar a la decodificación del conectoma del cerebro humano.

C.2.2. Objetivos específicos

Se proponen los siguientes objetivos específicos para abordar el trabajo de tesis:

- **OBJ1**: Mejorar un método eficiente para el clustering de fibras.
- **OBJ2**: Evaluar el clustering de fibras mediante la realización de análisis cualitativos y cuantitativos y comparaciones con otros algoritmos existentes.
- **OBJ3**: Diseñar e implementar un método para el etiquetado automático de los fascículos de materia blanca superficial.
- **OBJ4**: Realizar la evaluación del etiquetado por clustering intra e intersujeto.
- **OBJ5**: Crear un método para la generación de atlas de parcelas para la población de sujetos estudiados.
- **OBJ6**: Evaluar el método de parcelación con métricas de reproducibilidad y hacer la comparación con otros atlas de parcelas del estado del arte.
- **OBJ7**: Generar un método de parcelación cortical individual basado en la información de conectividad de la materia blanca a partir de un conjunto de datos de tractografía y realizar la evaluación del método de parcelación individual con métricas de conectividad de redes cerebrales.
- **OBJ8**: Crear un método para realizar la parcelación individual de la corteza teniendo en cuenta la topología. Evaluar la reproducibilidad del método generado con otros atlas basados en la macroanatomía.

C.2.3. Metodología

Para lograr los objetivos específicos de esta tesis, deben completarse las siguientes tareas:

- **T1**: Revisión y estudio del estado del arte de los métodos de clustering de fibras de materia blanca, etiquetado automático de las fibras de materia blanca superficial y parcelación de la superficie cortical basada en diferentes modalidades (**OBJ1, OBJ3, OBJ5, OBJ7, OBJ8**).

- **T2**: Colaboración en la implementación de un método rápido y eficiente de clustering de fibras. Para lograrlo, se utilizará un algoritmo de clustering para reducir el número de elementos de entrada de las fibras de materia blanca. Por otro lado, se adaptará un algoritmo de segmentación de fibras para reasignar pequeños clústeres a clústeres más grandes. Finalmente, los clústeres se mezclarán, reduciendo así el número de clústeres generados (**OBJ1**).

- **T3**: Evaluación cualitativa del método de clustering de fibras con algoritmos del estado del arte. Analizaremos los clústeres con diferentes configuraciones. Además, se extraerán imágenes de clústeres anatómicamente significativos, comprobando que el método presenta una mejor calidad y delimitación de los clústeres (**OBJ2**).

- **T4**: Evaluación cuantitativa para medir el tiempo de ejecución del método de clustering de fibras, así como las distancias intra e inter-clúster (**OBJ2**).

- **T5**: Diseño e implementación del método de etiquetado de fascículos de materia blanca superficial intrasujeto. Primero se aplicará el clustering a la tractografía del sujeto, y luego se filtrará para los fascículos cortos. Las fibras se intersecarán con la malla cerebral y, finalmente, los clústeres serán etiquetados basados en un atlas (**OBJ3**).

- **T6**: Diseño e implementación del método de etiquetado de la materia blanca superficial entre sujetos. Más específicamente, se etiquetará a un grupo de sujetos, manteniendo la correspondencia entre ellos. Para ello, se utilizarán un algoritmo

de matching y un algoritmo de clustering para buscar la correspondencia de los clústeres de diferentes sujetos (**OBJ3**).

- **T7**: Evaluación del método de etiquetado intrasujeto. En primer lugar, la correspondencia de los fascículos se encontrará a través de los sujetos. Además, se medirán las distancias de los centroides de los fascículos entre los pares de sujetos (**OBJ4**).

- **T8**: Evaluación de los métodos de etiquetado entre sujetos para medir la reproducibilidad de los clústeres con diferentes umbrales entre los sujetos (**OBJ4**).

- **T9**: Fusión de un atlas de materia blanca por medio de dos atlas de fibras cortas y uno de fibras largas. Además, la información anatómica se generará a partir de la segmentación cortical para descartar fibras mal segmentadas para cada uno de los sujetos de la base de datos ARCHI. Luego, se utilizará el pipeline del software BrainVisa para realizar la conversión entre los espacios de la malla cerebral obtenida (**OBJ5**).

- **T10**: Diseño, implementación y evaluación del algoritmo de segmentación de fibras de materia blanca optimizado para obtener fascículos de fibras estables presentes en la mayoría de los sujetos (**OBJ5**).

- **T11**: Diseño e implementación de un método para generar atlas de parcelas en diferentes granularidades y obtener así una parcelación representativa para un grupo de sujetos. Este método tendrá en cuenta las superposiciones que se producen entre las diferentes subparcelas de los distintos sujetos. Además, obtendrá las probabilidades de las conexiones subyacentes para generar las subparcelas. Finalmente, se aplicará un posprocesamiento de las subparcelas para obtener una mejor definición de las mismas (**OBJ5**).

- **T12**: Evaluación del método de parcelación cortical mediante el análisis de la conectividad cerebral y la reproducibilidad de los diferentes atlas generados. A continuación, se compararán también dos atlas del método con otros atlas del estado del arte basados en modalidades de resonancia magnética, buscando subparcelas y funcionalidades comunes (siempre que sea posible). Por otra parte,

se realizará un análisis de validación cruzada y se utilizarán métricas de cálculo de redes para el estudio de la conectividad del cerebro (**OBJ6**).

- **T13**: Diseño e implementación de un método para llevar a cabo la parcelación individual de la corteza cerebral. Posteriormente, el clustering de fibras se aplicará a toda la tractografía. Finalmente, se utilizará el algoritmo de parcelación individual y se realizará un posprocesamiento de las subparcelas. Además, se utilizarán las métricas de conectividad cerebral para la evaluación. Finalmente, se realizará un análisis cualitativo con el resultado de la parcelación para los sujetos (**OBJ7**).

- **T14**: Diseño e implementación de un método de parcelación cortical paralela basado en la topología del cerebro que tendrá en cuenta las circunvoluciones y los surcos. Además, se realizará un análisis cualitativo y se medirá el tiempo de ejecución del método de parcelación cortical. Por último, se evaluará la reproducibilidad con otros métodos del estado del arte (**OBJ8**).

C.3. Resultados más relevantes

En esta sección se presentan las principales contribuciones de la tesis. Todas ellas tratan sobre el desarrollo de métodos para el estudio de la conectividad cerebral. Estas contribuciones se listan a continuación:

- Los algoritmos de clustering de fibras en la investigación de la neuroanatomía generan datos que permiten el estudio de la estructura de la materia blanca. La primera contribución de esta tesis es la colaboración en el desarrollo de un método automatizado de clustering de fibras llamado *FFClust*, que identifica fascículos de materia blanca a partir de grandes conjuntos de datos de tractografía. La primera versión de este trabajo se desarrolló en otra tesis de máster [125]. El objetivo principal es desarrollar un clustering eficiente para agrupar las fibras en clústeres compactos y regulares, que representen la estructura de la materia blanca de todo el cerebro. Los clústeres resultantes describen el conjunto de los principales fascículos de materia blanca presentes en un cerebro individual. En los individuos, los clústeres pueden utilizarse para estudiar la conectividad local en cerebros patológicos, mientras que a nivel de población, el procesamiento y análisis de los fascículos reproducibles y otros algoritmos de posprocesamiento pueden llevarse a cabo para estudiar la conectividad del cerebro y crear nuevos atlas de fascículos de materia blanca. Un interés especial es su utilización para el estudio de los fascículos de asociación cortos y su segmentación, así como de las subdivisiones de los fascículos anatómicos largos. El método propuesto es unas 8,6 veces más rápido que el método del estado del arte, y permite un rápido procesamiento y visualización de los principales conjuntos de fibras de materia blanca.

  Mis contribuciones a este trabajo son la colaboración en las siguientes tareas: redacción del primer borrador, revisión de todas las versiones y borradores posteriores, estudio y revisión del estado del arte en el clustering de fibras, creación de esquemas y de la mayoría de las imágenes del paper, cambio de la concurrencia por paralelismo en los pasos 1 y 4, paralelización de la etapa 2 de mapeo

que no tenía paralelismo, implementación de todo el código, dando una mejora de al menos 2,5 veces más rápido que su versión anterior, optimización de entrada/salida y cambio de formato, prueba y búsqueda de errores e implementación de los dos códigos del cálculo de centroides, las pruebas se hicieron de nuevo con 50 sujetos (antes se hacían con un solo sujeto), uso de la media y la desviación estándar en todas las gráficas, agregados a la comparación con el estado del arte los métodos de *QuickBundlesX* (*QBX*) [43] y *Guevara* [55], implementación de la distancia máxima en *QuickBundles* (*QB*) [42] y *QBX* (que requiere tiempos de ejecución superiores a un día) para las pruebas de distancia con todos los métodos, pruebas de tiempo de ejecución de *QB* y *QBX* con distancia máxima (horas de pruebas), la implementación de las distancias entre los clústeres (no existentes), adición del índice de Davies-Bouldin (DB) [9] y la comparación con todos los métodos, el análisis cualitativo mediante la segmentación para demostrar que los clústeres obtenidos tienen significado biológico, adición del método de *Guevara* a todos los análisis cualitativos, la comparación de los clústeres más similares y el tiempo de ejecución entre los métodos. Este paper fue aceptado en la revista NeuroImage [127].

Las siguientes contribuciones que se presentan son resultado directo de este trabajo de tesis.

- Se han utilizado métodos de clustering de fibras para agrupar automáticamente fibras similares en clústeres. Sin embargo, debido a la variabilidad entre los sujetos y a los artefactos, los clústeres resultantes son difíciles de procesar para encontrar conexiones comunes entre los sujetos, especialmente para la materia blanca superficial. La segunda contribución de esta tesis es un método de etiquetado automático de fascículos de asociación cortos en un grupo de sujetos. El método se basa en un clustering de fibras intrasujeto que genera clústeres de fibras compactos. Después, tomando como referencia el atlas de *Desikan-Killiany* [31], los clústeres son etiquetados en base a la conectividad cortical de las fibras, y nombrados según su posición relativa a lo largo de un eje. Las etiquetas proporcionan información útil para la visualización y el análisis de las conexiones

individuales, lo cual es muy difícil sin ninguna información adicional. Finalmente, comparamos y aplicamos dos estrategias diferentes para el etiquetado de los fascículos entre sujetos: una de ellas basada en un famoso algoritmo de clustering del estado del arte, y la otra basada en un algoritmo de matching. El rendimiento de ambas implementaciones se compara en términos de reproducibilidad y distancia de los fascículos entre sujetos. Los clústeres obtenidos podrían utilizarse para realizar análisis de conectividad manual o automático en individuos o entre sujetos. El trabajo resultante está publicado en la revista Biomedical Engineering Online [126].

- La tercera contribución de esta tesis, y también la más importante, es un método para crear parcelaciones de la superficie cortical a partir de un atlas anatómico. La información de conectividad se obtiene de un atlas de fascículos de materia blanca, en lugar de la tractografía completa. Este atlas está compuesto por un atlas de fibras de materia blanca profunda y dos atlas de materia blanca superficial. Así, se obtiene una correspondencia directa entre los fascículos y las regiones corticales entre los sujetos, pudiendo obtener una buena representación del conectoma del cerebro humano, ya que los fascículos obtenidos se basan en los atlas de fascículos que contienen las conexiones cortas y largas más reproducibles que se encuentran en una población de sujetos. Además, el método produce otro resultado que contiene la representación probabilística de las subparcelas preliminares. De esta manera, la información podría utilizarse con fascículos segmentados de cada individuo y crear así parcelaciones individuales adaptadas a cada sujeto que deberían de conducir a una mayor consistencia en la conexión estructural entre los sujetos. Este trabajo ha sido publicado en la revista Frontiers in Neuroinformatics [79].

- La cuarta contribución es un método complementario al método de parcelación y realiza la parcelación cortical completa de un individuo teniendo en cuenta la información de conectividad de las fibras de la materia blanca. Nuestro objetivo es realizar una parcelación cortical individual de buena calidad que se utilizará

para una parcelación en grupo en el futuro. El resultado es el etiquetado completo de los vértices de la malla cortical, que representan las diferentes subparcelas de la corteza, con fuertes conexiones a otras subparcelas. Utilizamos las métricas de redes del cerebro para evaluar el método en un conjunto de sujetos. Estas métricas cumplen con la segregación e integración funcional, así como con la definición de *small-world*. El resultado de este trabajo fue publicado en la IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) [80].

- Finalmente, la última contribución de esta tesis es un método paralelo para la parcelación completa de la superficie cortical, basado en la distancia geodésica. El método tiene dos modos de uso, el primero subdivide cada parcela anatómica dada por el atlas de *Desikan-Killiany*. El segundo modo de uso realiza la división cortical completa de un sujeto. El método propuesto estará a disposición de la comunidad para llevar a cabo la evaluación de las parcelaciones corticales basadas en datos. Como ejemplo, comparamos la parcelación *GeoSP* con los atlas de *Desikan-Killiany* y *Destrieux* en 50 sujetos, obteniendo parcelas más homogéneas para *GeoSP* y diferencias menores en la conectividad estructural entre los sujetos. El paper resultante fue publicado en la 42 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [81].

C.4. Conclusiones y trabajo futuro

En esta sección presentamos las conclusiones de todo el trabajo realizado en la tesis. Además, mostramos las futuras líneas de investigación que sería interesante explorar.

C.4.1. Conclusiones

Los desarrollos de esta tesis están relacionados con el área de la neurociencia computacional, específicamente con algoritmos y métodos para el estudio de la conectividad cerebral basada en dMRI. Hemos desarrollado algoritmos eficientes que pueden tratar el ruido presente en la tractografía de difusión. Estos artefactos son producidos por los algoritmos de la tractografía, generando fascículos válidos que se tienen en cuenta cuando se utilizan las fibras de materia blanca. Además, la alta variabilidad de las fibras (más de un millón) entre los sujetos genera una gran complejidad al tratar estos masivos conjuntos de datos 3D. Por lo tanto, cuando se utiliza la tractografía de difusión, se necesita una validación adicional, como contrastar con la anatomía y obtener un significado biológico, utilizando estudios post mortem o aplicando otras técnicas de resonancia magnética como la resonancia magnética funcional.

Esperamos que esta investigación sea beneficiosa para los neurocientíficos del área, ya que pueden desarrollarse estudios modernos analizando nuevas bases de datos y contribuyendo a la decodificación del conectoma humano.

Las conclusiones de los trabajos presentados se exponen a continuación:

- **Clustering y etiquetado de fibras de la materia blanca**. Presentamos *FFClust* que es un algoritmo eficiente de clustering de fibras de la materia blanca. Es por lo menos un orden de magnitud más rápido que *QuickBundles*, que es uno de los algoritmos más utilizados para clustering de fibras. Por ejemplo, un sujeto con aproximadamente 1 millón de fibras es procesado por la versión secuencial de *FFClust* en 1,99 min, y su implementación paralela en 45 s, mientras que *QB* tarda 2,2 h usando su mejor configuración. Sin embargo, usando el índice de

Davies-Bouldin (DB) para medir la calidad de los clústeres, *QB* logra el mejor resultado, seguido de *FFClust*, mientras que *QBX* no mejora la calidad de *QB*.

Gracias a este método se pueden realizar nuevos estudios de conectividad cerebral, en particular sobre las fibras de la materia blanca superficial. Además, el método de *FFClust* y el método de *Guevara* consiguen un índice Davies-Bouldin similar y pueden identificar todos los fascículos al aplicar la segmentación. También tienen los menores porcentajes de error al identificar los fascículos seleccionados. Sin embargo, *FFClust* es más rápido y más simple que el método *Guevara*.

Por otro lado, implementamos un método rápido para el etiquetado automático de los fascículos de fibras de materia blanca, específicamente para la materia blanca superficial, basado en un clustering intrasujeto y la conectividad de los clústeres con la malla cortical, basado en un atlas anatómico de regiones de interés. El algoritmo también añade una etiqueta asociada a la posición relativa de los fascículos. Los resultados del etiquetado intrasujeto muestran un grado de correspondencia entre los sujetos, que se mejora aún más con el etiquetado intersujeto. Un etiquetado completo entre sujetos se ejecuta en un tiempo medio de 3,5 minutos para un conjunto de datos de tractografía de alrededor de un millón de fibras. Esto permite una exploración, visualización y análisis rápidos y fáciles de los fascículos de asociación corta etiquetados en los individuos, lo cual es muy difícil sin ninguna información adicional.

Además, desarrollamos un etiquetado entre sujetos usando dos métodos. Uno de ellos consiste en matching, en particular, el algoritmo *Húngaro*, y el otro en clustering, empleando el algoritmo *QuickBundles*. Los resultados muestran una mejor reproducibilidad entre los sujetos para el método de clustering frente al algoritmo de matching, manteniendo una distancia moderada entre los clústeres, lo que indica una buena calidad de los clústeres. Además, el algoritmo es escalable y todo el procesamiento para el etiquetado entre sujetos se ejecuta en un

tiempo razonable, de aproximadamente 1,17 h para 20 sujetos. Los clústeres obtenidos podrían utilizarse para realizar estudios de conectividad en grupo, como la creación de atlas de fascículos de materia blanca, y el desarrollo de nuevos métodos para el análisis de la conectividad del cerebro.

- **Parcelación de la superficie cortical**. El método propuesto crea una parcelación de la superficie cortical, que consiste en la subdivisión de parcelas anatómicas, a partir de un atlas de fascículos de fibras basado en la difusión. La parcelación generada depende de tres parámetros configurables que generan una parcelación con un número menor o mayor de subparcelas. Además, un resultado intermedio del método es la representación probabilística de las subparcelas preliminares, asociadas a las dos conexiones de cada fascículo.

  Como conclusión final, este método puede crear una parcelación cortical basada en la conectividad estructural, a partir de parcelas anatómicas, lo que da lugar a subparcelas con perfiles de conectividad muy consistentes en toda una población de sujetos, y un grado de correspondencia con parcelas del estado del arte basadas en modalidades de resonancia magnética.

  Además, hemos desarrollado un método para realizar parcelaciones corticales individuales, basado en la información de la conectividad de las fibras de la materia blanca. El clustering de fibras ayuda a definir conexiones más compactas y por lo tanto regiones, además de filtrar los valores atípicos. Los resultados muestran una buena calidad en los sujetos analizados, en particular, en los mapas de conectividad. Estos fueron evaluados con métricas de redes cerebrales, como el *coeficiente de clustering*, el *path length* y *small-worldness*. Estas métricas indican una buena integración funcional y segregación del cerebro. Sin embargo, no hay consenso en la comunidad sobre la validez de estas métricas. Se necesitan más mejoras y validaciones.

  Finalmente, presentamos un método paralelo, llamado *GeoSP*, que crea una parcelación individual de la malla cortical basada en la distancia geodésica, considerando así la topología del cerebro, es decir, las circunvoluciones y los surcos. El

método representa la malla por medio de una red y tiene dos modos de uso. Por defecto, realiza la parcelación basada en los límites del atlas de *Desikan-Killiany*. El otro modo realiza la división para la corteza completa. Los resultados muestran subparcelas homogéneas. Los tiempos de ejecución para el modo basado en el atlas es de 18 s y de 82 s para el modo de corteza completa. Además, en comparación con dos atlas macroanatómicos, *GeoSP* tiene menos diferencias en la reproducibilidad de la conectividad estructural entre los sujetos, probablemente debido a una mayor homogeneidad de los subparcelas. Sin embargo, esto no significa que la parcelación tenga significado biológico o que sea mejor que las otras. Es necesario realizar más análisis. El método está disponible para realizar comparaciones entre parcelaciones, ya sean basadas en MRI de difusión o funcional.

C.4.2. Trabajo futuro

En esta sección, presentamos las líneas futuras de investigación que podrían llevarse a cabo en base al trabajo de esta tesis. Presentamos estas ideas para cada una de las contribuciones realizadas:

- **Clustering y etiquetado de fibras de la materia blanca**. Ya que *FFClust* tiene cuatro pasos, como trabajo futuro proponemos mejorar los tiempos de ejecución para las etapas más lentas. Esto podría ser útil para integrar el algoritmo de clustering con aplicaciones de visualización para permitir la exploración rápida y otros análisis de posprocesamiento de la estructura de la materia blanca para uno o varios sujetos. Entonces, sugerimos que *FFClust* puede utilizarse en aplicaciones similares en las que el método de *Guevara* se ha utilizado con éxito [53, 50].

  En lo que respecta al etiquetado de los fascículos de la materia blanca superficial, la labor futura podría centrarse en la aplicación del método en bases de datos de alta calidad, como la base de datos del *Human Connectome Project (HCP)*, para la creación de un atlas de materia blanca superficial y parcelaciones corticales

basadas en la difusión. Además, se podrían utilizar otros algoritmos para generar un mejor etiquetado de los fascículos de la materia blanca superficial.

- **Parcelación de la superficie cortical**. La parcelación cortical basada en un atlas de fascículos de fibras podría utilizar un nuevo atlas de fascículos, basado en una base de datos más grande, como la base de datos HCP y la tractografía probabilística. Además, los fascículos podrían obtenerse de un clustering de fibras entre sujetos de la misma base de datos, lo que podría dar lugar a una mejor representación de las conexiones de los fascículos de materia blanca de la población de sujetos. Sin embargo, el posprocesamiento de las subparcelas candidatas sería probablemente más complicado debido a una mayor cantidad de fascículos y una mayor superposición de las subparcelas. Además, la información funcional podría utilizarse para crear una parcelación mediante el uso de frameworks de parcelación multimodal [99]. Además, otra línea a explorar es la inclusión de algunos fascículos de atlas basados en áreas funcionales conocidas. Por último, la información probabilística de las subparcelas preliminares podría utilizarse, en combinación con los fascículos individuales segmentados, para crear parcelaciones individuales adaptadas a cada sujeto. Su efecto será el de pequeños cambios en los límites de las subparcelas de cada sujeto, debido a las diferencias individuales de los fascículos segmentados. Las parcelaciones adaptadas deberían conducir a una mayor consistencia en la conexión estructural entre los sujetos.

  En cuanto a la parcelación de la superficie cortical basada en clustering, como trabajo futuro podríamos explorar la aplicación de una versión multisujeto de este método de parcelación y probarlo en diferentes bases de datos, como la del Human Connectome Project. De este modo, podríamos obtener un atlas (o modelo) de parcelas corticales con perfiles de conectividad similares en una población de sujetos sanos. También se podría integrar otra información, como fibras segmentadas con un atlas de fascículos [128], o datos de otras modalidades, como la MRI funcional.

Por último, como futuras líneas de investigación con respecto a la parcelación geodésica, se podría mejorar el tiempo de ejecución del método propuesto para ambos modos de funcionamiento. Por otra parte, otros atlas anatómicos podrían utilizarse como base para trazar los límites entre las parcelas anatómicas en lugar del atlas de *Desikan-Killiany*.

Bibliography

[1] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72, 2006.

[2] K. Amunts and K. Zilles. Architecture and organizational principles of Broca's region. *Trends in Cognitive Sciences*, 16(8):418–426, 2012.

[3] K. Amunts and K. Zilles. Architectonic mapping of the human brain beyond Brodmann. *Neuron*, 88(6):1086–1107, 2015.

[4] A. Anwander, M. Tittgemeyer, D. Y. von Cramon, A. D. Friederici, and T. R. Knösche. Connectivity-Based Parcellation of Broca's Area. *Cerebral Cortex*, 17(4):816–825, 2006.

[5] S. Arslan, S. I. Ktena, A. Makropoulos, E. C. Robinson, D. Rueckert, and S. Parisot. Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage*, 170:5–30, 2018.

[6] P. J. Basser, J. Mattiello, and D. LeBihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 1994.

[7] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44(4):625–632, 2000.

[8] D. S. Bassett, B. G. Nelson, B. A. Mueller, J. Camchong, and K. O. Lim. Altered resting state complexity in schizophrenia. *NeuroImage*, 59(3):2196–2207, 2012.

[9] J. C. Bezdek and N. R. Pal. Some new indexes of cluster validity. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 28(3):301–315, 1998.

[10] M. Brant-Zawadzki, G. D. Gillan, and W. R. Nitz. MP RAGE: a three-dimensional, t1-weighted, gradient-echo sequence–initial experience in the brain. *Radiology*, 182(3):769–775, 1992.

[11] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.

[12] L. Brun, A. Pron, J. Sein, C. Deruelle, and O. Coulon. Diffusion MRI: Assessment of the Impact of Acquisition and Preprocessing Methods Using the BrainVISA-Diffuse Toolbox. *Frontiers in Neuroscience*, 13:536, 2019.

[13] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186, 2009.

[14] A. Cachia, J.-F. Mangin, D. Rivière, D. Papadopoulos-Orfanos, F. Kherif, I. Bloch, and J. Régis. A generic framework for the parcellation of the cortical surface into gyri using geodesic Voronoï diagrams. *Medical Image Analysis*, 7(4):403–416, 2003.

[15] J. S. Campbell and G. B. Pike. Potential and limitations of diffusion MRI tractography for the study of language. *Brain and Language*, 131:65–73, 2014.

[16] M. Catani. A little man of some importance. *Brain*, 140(11):3055–3061, 2017.

[17] M. Catani and M. T. De Schotten. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 44(8):1105–1132, 2008.

[18] M. Catani, F. Dell'Acqua, F. Vergani, F. Malik, H. Hodge, P. Roy, R. Valabregue, and M. T. De Schotten. Short frontal lobe connections of the human brain. *Cortex*, 48(2):273–291, 2012.

[19] M. Catani, R. J. Howard, S. Pajevic, and D. K. Jones. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage*, 17(1):77–94, 2002.

[20] M. Catani and M. Mesulam. The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex*, 44(8):953–961, 2008.

[21] A. Chekir, M. Descoteaux, E. Garyfallidis, M. Côté, and F. O. Boumghar. A hybrid approach for optimal automatic segmentation of White Matter tracts in HARDI. In *2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 177–180, 2014.

[22] R. A. Chowdhury and V. Ramachandran. Cache-oblivious buffer heap and cache-efficient computation of shortest paths in graphs. *ACM Transactions on Algorithms (TALG)*, 14(1):1–33, 2018.

[23] J. R. Cohen and M. D'Esposito. The segregation and integration of distinct brain networks and their relationship to cognition. *Journal of Neuroscience*, 36(48):12083–12094, 2016.

[24] Y. Cointepas, D. Geffroy, N. Souedet, I. Denghien, D. Rivière, and F. Roses. The BrainVISA project: a shared software development infrastructure for biomedical imaging research. In *16th Annual Meeting of the Organization for Human Brain Mapping (HBM)*, 2010.

[25] M. Cousineau, P.-M. Jodoin, E. Garyfallidis, M.-A. Côté, F. C. Morency, V. Rozanski, M. Grand'Maison, B. J. Bedell, and M. Descoteaux. A test-retest study on parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical*, 16:222–233, 2017.

[26] K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, A. D. N. Initiative, et al. Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*, 192:115–134, 2019.

[27] A. M. Dale, B. Fischl, and M. I. Sereno. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194, 1999.

[28] M. A. de Reus and M. P. Van den Heuvel. The parcellation-based connectome: limitations and extensions. *NeuroImage*, 80:397–404, 2013.

[29] M. T. de Schotten, M. Urbanski, R. Valabrègue, D. J. Bayle, and E. Volle. Subdivision of the occipital lobes: an anatomical and functional MRI connectivity study. *Cortex*, 56:121–37, 2014.

[30] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine*, 58(3):497–510, 2007.

[31] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006.

[32] C. Destrieux, B. Fischl, A. Dale, and E. Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1):1–15, 2010.

[33] D. Duclap, B. Schmitt, A. Lebois, O. Riff, P. Guevara, and L. Marrakchi-Kacem. Connectomist-2.0: a novel diffusion analysis toolbox for BrainVISA. In *29th European Society for Magnetic Resonance in Medicine and Biology (ESMRMB)*, 2012.

[34] J. T. Duda, P. A. Cook, and J. C. Gee. Reproducibility of graph metrics of human brain structural networks. *Frontiers in Neuroinformatics*, 8:46, 2014.

[35] S. B. Eickhoff, B. Thirion, G. Varoquaux, and D. Bzdok. Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*, 36(12):4771–4792, 2015.

[36] S. B. Eickhoff, B. T. Yeo, and S. Genon. Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, 19(11):672–686, 2018.

[37] D. Eppstein and D. Strash. Listing all maximal cliques in large sparse real-world graphs. In *Experimental Algorithms*, pp. 364–375, 2011.

[38] L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral Cortex*, 26(8):3508–3526, 2016.

[39] B. Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, 2012.

[40] A. Frank. On Kuhn's Hungarian method—a tribute from Hungary. *Naval Research Logistics (NRL)*, 52(1):2–5, 2005.

[41] P. Fränti and S. Sieranoja. How much can k-means be improved by using better initialization and repeats? *Pattern Recognition*, 93:95–112, 2019.

[42] E. Garyfallidis, M. Brett, M. M. Correia, G. B. Williams, and I. Nimmo-Smith. Quickbundles, a method for tractography simplification. *Frontiers in Neuroscience*, 6:175, 2012.

[43] E. Garyfallidis, M.-A. Côté, F. Rheault, and M. Descoteaux. QuickBundlesX: sequential clustering of millions of streamlines in multiple levels of detail at record execution time. In *24th International Society of Magnetic Resonance in Medicine (ISMRM)*, 2016.

[44] E. Garyfallidis, M.-A. Côté, F. Rheault, J. Sidhu, J. Hau, L. Petit, D. Fortin, S. Cunanne, and M. Descoteaux. Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage*, 170:283–295, 2018.

[45] N. Geschwind. The organization of language and the brain. *Science*, 170(3961):940–944, 1970.

[46] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171, 2016.

[47] K. Gopinath, C. Desrosiers, and H. Lombaert. Graph convolutions on spectral embeddings for cortical surface parcellation. *Medical Image Analysis*, 54:297–305, 2019.

[48] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

[49] M. Guevara, P. Guevara, C. Román, and J.-F. Mangin. Superficial white matter: A review on the dMRI analysis methods and applications. *NeuroImage*, 212:116673, 2020.

[50] M. Guevara, C. Román, J. Houenou, D. Duclap, C. Poupon, J.-F. Mangin, and P. Guevara. Reproducibility of superficial white matter tracts using diffusion-weighted imaging tractography. *NeuroImage*, 147:703–725, 2017.

[51] M. Guevara, C. Román, P. Silva, D. Rivière, C. Poupon, J.-F. Mangin, and P. Guevara. Cortex surface parcellation based on short association white matter bundles. In *23th Annual Meeting of the Organization for Human Brain Mapping (HBM)*, 2017.

[52] P. Guevara, D. Duclap, L. Marrakchi-Kacem, D. Rivière, Y. Cointepas, C. Poupon, and J. Mangin. Accurate tractography propagation mask using T1-weighted data rather than FA. In *19th International Society of Magnetic Resonance in Medicine (ISMRM)*, 2011.

[53] P. Guevara, D. Duclap, C. Poupon, L. Marrakchi-Kacem, P. Fillard, D. Le Bihan, M. Leboyer, J. Houenou, and J.-F. Mangin. Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *NeuroImage*, 61(4):1083–1099, 2012.

[54] P. Guevara, M. Perrin, P. Cathier, Y. Cointepas, D. Riviere, C. Poupon, and J. Mangin. Connectivity-based parcellation of the cortical surface using q-ball imaging. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 903–906, 2008.

[55] P. Guevara, C. Poupon, D. Rivière, Y. Cointepas, M. Descoteaux, B. Thirion, and J.-F. Mangin. Robust clustering of massive tractography datasets. *NeuroImage*, 54(3):1975–1993, 2011.

[56] V. Gupta, S. I. Thomopoulos, C. K. Corbin, F. Rashid, and P. M. Thompson. FIBERNET 2.0: An automatic neural network based tool for clustering white matter fibers in the brain. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 708–711, 2018.

[57] V. Gupta, S. I. Thomopoulos, F. M. Rashid, and P. M. Thompson. FiberNET: An ensemble deep learning framework for clustering white matter fibers. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, pp. 548–555. Springer International Publishing, 2017.

[58] P. Hagmann, L. Cammoun, X. Gigandet, S. Gerhard, P. E. Grant, V. Wedeen, R. Meuli, J.-P. Thiran, C. J. Honey, and O. Sporns. MR connectomics: principles and challenges. *Journal of Neuroscience Methods*, 194(1):34–45, 2010.

[59] P. Hagmann, M. Kurant, X. Gigandet, P. Thiran, V. J. Wedeen, R. Meuli, and J.-P. Thiran. Mapping human whole-brain structural networks with diffusion MRI. *PLOS ONE*, 2(7):e597, 2007.

[60] H. J. Heijmans. *Morphological Image Operators*. Academic Press Boston, 1994.

[61] S. A. Huettel, A. W. Song, G. McCarthy, et al. *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, MA, 2004.

[62] M. D. Humphries and K. Gurney. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PLOS ONE*, 3(4):e0002051, 2008.

[63] M. D. Humphries, K. Gurney, and T. J. Prescott. The brainstem reticular formation is a small-world, not scale-free, network. *Proceedings of the Royal Society B: Biological Sciences*, 273(1585):503–511, 2005.

[64] A. G. Huth, W. A. De Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458, 2016.

[65] Y. Jin, Y. Shi, L. Zhan, B. A. Gutman, G. I. de Zubicaray, K. L. McMahon, M. J. Wright, A. W. Toga, and P. M. Thompson. Automatic clustering of white matter fibers in brain diffusion MRI with an application to genetics. *NeuroImage*, 100:75–90, 2014.

[66] T. Jitsuishi, S. Hirono, T. Yamamoto, K. Kitajo, Y. Iwadate, and A. Yamaguchi. White matter dissection and structural connectivity of the human vertical occipital fasciculus to link vision-associated brain cortex. *Scientific Reports*, 10:820, 2020.

[67] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pp. 85–103. Springer, 1972.

[68] J. Katz, M.-A. d'Albis, J. Boisgontier, C. Poupon, J.-F. Mangin, P. Guevara, D. Duclap, N. Hamdani, J. Petit, D. Monnet, P. L. Corvoisier, M. Leboyer, R. Delorme, and J. Houenou. Similar white matter but opposite grey matter changes in schizophrenia and high-functioning autism. *Acta Psychiatrica Scandinavica*, 134(1):31–39, 2016.

[69] J. C. Klein, T. E. J. Behrens, M. D. Robson, C. E. Mackay, D. J. Higham, and H. Johansen-Berg. Connectivity-based parcellation of human cortex using diffusion MRI: Establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *NeuroImage*, 34(1):204–211, 2007.

[70] T. M. Kodinariya and P. R. Makwana. Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6):90–95, 2013.

[71] R. Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, pp. 1137–1145, 1995.

[72] N. Labra, P. Guevara, D. Duclap, J. Houenou, C. Poupon, J.-F. Mangin, and M. Figueroa. Fast automatic segmentation of white matter streamlines based on a multi-subject bundle atlas. *Neuroinformatics*, 15(1):71–86, 2017.

[73] S. Lattanzi and C. Sohler. A better k-means++ algorithm via local search. In *International Conference on Machine Learning*, pp. 3662–3671, 2019.

[74] D. Le Bihan and M. Iima. Diffusion magnetic resonance imaging: what water tells us about biological tissues. *PLOS Biology*, 13(7):e1002203, 2015.

[75] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–546, 2001.

[76] S. Lefranc, P. Roca, M. Perrot, C. Poupon, D. Le Bihan, J.-F. Mangin, and D. Rivière. Groupwise connectivity-based parcellation of the whole human cortical surface using watershed-driven dimension reduction. *Medical Image Analysis*, 30:11–29, 2016.

[77] H. Li, L. Fan, J. Zhuo, J. Wang, Y. Zhang, Z. Yang, and T. Jiang. ATPP: A Pipeline for Automatic Tractography-Based Brain Parcellation. *Frontiers in Neuroinformatics*, 11:35, 2017.

[78] H. Li, Z. Xue, L. Guo, T. Liu, J. Hunter, and S. T. Wong. A hybrid approach to automatic clustering of white matter fibers. *NeuroImage*, 49(2):1249–1258, 2010.

[79] N. López-López, A. Vázquez, J. Houenou, C. Poupon, J.-F. Mangin, S. Ladra, and P. Guevara. From coarse to fine-grained parcellation of the cortical surface using a fiber-bundle atlas. *Frontiers in Neuroinformatics*, 14:32, 2020.

[80] N. López-López, A. Vázquez, C. Poupon, J.-F. Mangin, and P. Guevara. Cortical surface parcellation based on intra-subject white matter fiber clustering. In *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, pp. 1–6, 2019.

[81] N. López-López, A. Vázquez, C. Poupon, J.-F. Mangin, S. Ladra, and P. Guevara. GeoSP: A parallel method for a cortical surface parcellation based on geodesic distance. In *2020 42th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 44–49, 2020.

[82] K. H. Maier-Hein, P. F. Neher, J.-C. Houde, M.-A. Côté, E. Garyfallidis, J. Zhong, M. Chamberland, F.-C. Yeh, Y.-C. Lin, Q. Ji, et al. The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, 8(1):1–13, 2017.

[83] J.-F. Mangin, J. Lebenberg, S. Lefranc, N. Labra, G. Auzias, M. Labit, M. Guevara, H. Mohlberg, P. Roca, P. Guevara, J. Dubois, F. Leroy, G. Dehaene-Lambertz, A. Cachia, T. Dickscheid, O. Coulon, C. Poupon, D. Rivière, K. Amunts, and Z. Y. Sun. Spatial normalization of brain images and beyond. *Medical Image Analysis*, 33:127–133, 2016.

[84] P. Mansfield. Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55–L58, 1977.

[85] J. Martino, P. C. De Witt Hamer, F. Vergani, C. Brogna, E. M. de Lucas, A. Vázquez-Barquero, J. A. García-Porrero, and H. Duffau. Cortex-sparing fiber dissection: an improved method for the study of white matter anatomy in the human brain. *Journal of Anatomy*, 219(4):531–541, 2011.

[86] T. Meynert. Psychiatry, a Clinical Treatise on Disease of the Fore-Brain, based upon a Study of its Structure, Functions, and Nutrition Part I.- The Anatomy, Physiology, and Chemistry of the Brain. *Journal of Mental Science*, 32(139):400–405, 1886.

[87] T. Möller and B. Trumbore. Fast, Minimum Storage Ray-Triangle Intersection. *J. Graph. Tools*, 2(1):21–28, 1997.

[88] A. E. Monge and C. P. Elkan. The Field Matching Problem: Algorithms and Applications. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pp. 267–270. AAAI Press, 1996.

[89] D. Moreno-Dominguez, A. Anwander, and T. R. Knösche. A hierarchical method for whole-brain connectivity-based parcellation. *Human Brain Mapping*, 35(10):5000–5025, 2014.

[90] S. Mori and P. B. Barker. Diffusion magnetic resonance imaging: its principle and applications. *The Anatomical Record: An Official Publication of the American Association of Anatomists*, 257(3):102–109, 1999.

[91] S. Mori, 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*, 45(2):265–269, 1999.

[92] S. Mori and P. van Zijl. Fiber tracking: principles and strategies–a technical review. *NMR in Biomedicine*, 15(7-8):468–480, 2002.

[93] L. J. O'Donnell, A. J. Golby, and C.-F. Westin. Fiber clustering versus the parcellation-based connectome. *NeuroImage*, 80:283–289, 2013.

[94] L. J. O'Donnell, Y. Suter, L. Rigolo, P. Kahali, F. Zhang, I. Norton, A. Albi, O. Olubiyi, A. Meola, W. I. Essayed, P. Unadkat, P. A. Ciris, W. M. Wells, Y. Rathi, C.-F. Westin, and A. J. Golby. Automated white matter fiber tract identification in patients with brain tumors. *NeuroImage: Clinical*, 13:138–153, 2017.

[95] L. J. O'Donnell and C.-F. Westin. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Transactions on Medical Imaging*, 26(11):1562–1575, 2007.

[96] J. O'Muircheartaigh and S. Jbabdi. Concurrent white matter bundles and grey matter networks using independent component analysis. *NeuroImage*, 170:296–306, 2017.

[97] L. O'Donnell, M. Kubicki, M. E. Shenton, M. H. Dreusicke, W. E. L. Grimson, and C.-F. Westin. A method for clustering white matter fiber tracts. *American Journal of Neuroradiology*, 27(5):1032–1036, 2006.

[98] S. Parisot, S. Arslan, J. Passerat-Palmbach, W. M. Wells, and D. Rueckert. Tractography-driven groupwise multi-scale parcellation of the cortex. In *International Conference on Information Processing in Medical Imaging*, pp. 600–612. Springer, 2015.

[99] S. Parisot, B. Glocker, S. I. Ktena, S. Arslan, M. D. Schirmer, and D. Rueckert. A flexible graphical model for multi-modal parcellation of the cortex. *NeuroImage*, 162:226–248, 2017.

[100] M. Perrin, Y. Cointepas, A. Cachia, C. Poupon, B. Thirion, D. Riviere, P. Cathier, V. El Kouby, A. Constantinesco, D. Le Bihan, et al. Connectivity-based parcellation of the cortical mantle using q-ball diffusion imaging. *Journal of Biomedical Imaging*, 2008:368406, 2008.

[101] M. Perrin, C. Poupon, Y. Cointepas, B. Rieul, N. Golestani, C. Pallier, D. Rivière, A. Constantinesco, D. Le Bihan, and J.-F. Mangin. Fiber tracking in q-ball fields using regularized particle trajectories. In *Biennial International Conference on Information Processing in Medical Imaging*, pp. 52–63. Springer, 2005.

[102] Z. Ramadhan, A. Putera Utama Siahaan, and M. Mesran. Prim and Floyd-Warshall Comparative Algorithms in Shortest Path Problem. In *Proceedings of the Joint Workshop KO2PI and the 1st International Conference on Advance & Scientific Innovation*, ICASI'18, pp. 47–58. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2018.

[103] G. C. Ribas. The cerebral sulci and gyri. *Neurosurgical Focus*, 28(2):E2, 2010.

[104] P. Roca, D. Riviere, P. Guevara, C. Poupon, and J.-F. Mangin. Tractography-Based Parcellation of the Cortex using a Spatially-Informed Dimension Reduction of the Connectivity Matrix. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, pp. 935–942. Springer, 2009.

[105] P. Roca, A. Tucholka, D. Rivière, P. Guevara, C. Poupon, and J.-F. Mangin. Inter-subject connectivity-based parcellation of a patch of cerebral cortex. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, pp. 347–354. Springer, 2010.

[106] C. Román, M. Guevara, R. Valenzuela, M. Figueroa, J. Houenou, D. Duclap, C. Poupon, J.-F. Mangin, and P. Guevara. Clustering of Whole-Brain White Matter Short Association Bundles Using HARDI Data. *Frontiers in Neuroinformatics*, 11:73, 2017.

[107] C. Ros, D. Güllmar, M. Stenzel, H.-J. Mentzel, and J. R. Reichenbach. Atlas-guided cluster analysis of large tractography datasets. *PLOS ONE*, 8(12):e83847, 2013.

[108] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

[109] A. Sanchez, C. Hernández, C. Poupon, J.-F. Mangin, and P. Guevara. Clustering of tractography datasets based on streamline point distribution. In *27th International Society of Magnetic Resonance in Medicine (ISMRM)*, 2018.

[110] S. Sarrazin, C. Poupon, J. Linke, M. Wessa, M. Phillips, M. Delavest, A. Versace, J. Almeida, P. Guevara, D. Duclap, E. Duchesnay, J.-F. Mangin, K. L. Dudal, C. Daban, N. Hamdani, M.-A. D'Albis, M. Leboyer, and J. Houenou. A Multicenter Tractography Study of Deep White Matter Tracts in Bipolar I Disorder. *JAMA Psychiatry*, 71(4):388, 2014.

[111] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X. Zuo, A. J. Holmes, S. B. Eickhoff, and B.-T. T. Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9):3095–3114, 2017.

[112] P. Schiffler, J.-G. Tenberge, H. Wiendl, and S. G. Meuth. Cortex Parcellation Associated Whole White Matter Parcellation in Individual Subjects. *Frontiers in Human Neuroscience*, 11:352, 2017.

[113] B. Schmitt, A. Lebois, D. Duclap, P. Guevara, F. Poupon, D. Rivière, Y. Cointepas, D. LeBihan, J. Mangin, and C. Poupon. CONNECT/ARCHI: an open database to infer atlases of the human brain connectivity. In *29th European Society for Magnetic Resonance in Medicine and Biology (ESMRMB)*, 2012.

[114] D. Sculley. Web-scale k-means clustering. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pp. 1177–1178, New York, NY, USA, 2010. Association for Computing Machinery.

[115] F. Silva, M. Guevara, C. Poupon, J.-F. Mangin, C. Hernández, and P. Guevara. Cortical Surface Parcellation Based on Graph Representation of Short Fiber Bundle Connections. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1479–1482, 2019.

[116] M. Sipser. *Introduction to the Theory of Computation*. International Thomson Publishing, 1st edition, 1996.

[117] S. N. Sotiropoulos and A. Zalesky. Building connectomes using diffusion MRI: why, how and but. *NMR in Biomedicine*, 32(4):e3752, 2019.

[118] O. Sporns, G. Tononi, and R. Kötter. The human connectome: a structural description of the human brain. *PLOS Computational Biology*, 1(4):e42, 2005.

[119] R. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.

[120] Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, and P. J. Laurienti. The ubiquity of small-world networks. *Brain Connectivity*, 1(5):367–375, 2011.

[121] A. W. Toga, K. A. Clark, P. M. Thompson, D. W. Shattuck, and J. D. Van Horn. Mapping the human connectome. *Neurosurgery*, 71(1):1–5, 2012.

[122] G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11):5033–5037, 1994.

[123] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, 2002.

[124] M. P. Van Den Heuvel and H. E. H. Pol. Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8):519–534, 2010.

[125] A. Vázquez. Método eficiente de clustering de fibras cerebrales basado en distribución de puntos. Master's thesis in Computer Science, Universidad de Concepción, Concepción, 2019.

[126] A. Vázquez, N. López-López, J. Houenou, C. Poupon, J.-F. Mangin, S. Ladra, and P. Guevara. Automatic group-wise whole-brain short association fiber bundle labeling based on clustering and cortical surface information. *BioMedical Engineering OnLine*, 19(1):1–24, 2020.

[127] A. Vázquez, N. López-López, A. Sánchez, J. Houenou, C. Poupon, J.-F. Mangin, C. Hernández, and P. Guevara. FFClust: Fast fiber clustering for large tractography datasets for a detailed study of brain connectivity. *NeuroImage*, 220:117070, 2020.

[128] A. Vázquez, N. López-López, N. Labra, M. Figueroa, C. Poupon, J. Mangin, C. Hernández, and P. Guevara. Parallel optimization of fiber bundle segmentation for massive tractography datasets. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 178–181, 2019.

[129] S. Wakana, H. Jiang, L. M. Nagae-Poetscher, P. C. Van Zijl, and S. Mori. Fiber tract–based atlas of human white matter anatomy. *Radiology*, 230(1):77–87, 2004.

[130] D. Wassermann, L. Bloy, E. Kanterakis, R. Verma, and R. Deriche. Unsupervised white matter fiber clustering and tract probability map generation: Applications of a Gaussian process framework for white matter fibers. *NeuroImage*, 51(1):228–241, 2010.

[131] D. Wassermann, N. Makris, Y. Rathi, M. Shenton, R. Kikinis, M. Kubicki, and C.-F. Westin. The white matter query language: a novel approach for describing human white matter anatomy. *Brain Structure and Function*, 221(9):4705–4721, 2016.

[132] J. Wasserthal, P. Neher, and K. H. Maier-Hein. TractSeg - fast and accurate white matter tract segmentation. *NeuroImage*, 183:239–253, 2018.

[133] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440, 1998.

[134] R. J. Wilson. *Introduction to Graph Theory*. John Wiley & Sons, Inc., New York, NY, USA, 1986.

[135] G. J. Woeginger. Exact Algorithms for NP-Hard Problems: A Survey. In *Combinatorial Optimization - Eureka, You Shrink!*, pp. 185–207. Springer, 2003.

[136] Y. Wu, F. Zhang, N. Makris, Y. Ning, I. Norton, S. She, H. Peng, Y. Rathi, Y. Feng, H. Wu, and L. J. O'Donnell. Investigation into local white matter abnormality in emotional processing and sensorimotor areas using an automatically annotated fiber clustering in major depressive disorder. *NeuroImage*, 181:16–29, 2018.

[137] D. Xu and Y. Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, 2015.

[138] R. Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645 – 678, 2005.

[139] B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125, 2011.

[140] S. W. Yoo, P. Guevara, Y. Jeong, K. Yoo, J. S. Shin, J.-F. Mangin, and J.-K. Seong. An Example-Based Multi-Atlas Approach to Automatic Labeling of White Matter Tracts. *PLOS ONE*, 10(7):e0133337, 2015.

[141] F. Zhang, Y. Wu, I. Norton, Y. Rathi, A. J. Golby, and L. J. O'Donnell. Test–retest reproducibility of white matter parcellation using diffusion MRI tractography fiber clustering. *Human Brain Mapping*, 40(10):3041–3057, 2019.

[142] F. Zhang, Y. Wu, I. Norton, L. Rigolo, Y. Rathi, N. Makris, and L. J. O'Donnell. An anatomically curated fiber clustering white matter atlas for consistent white matter tract parcellation across the lifespan. *NeuroImage*, 179:429–447, 2018.

[143] T. Zhang, H. Chen, L. Guo, K. Li, L. Li, S. Zhang, D. Shen, X. Hu, and T. Liu. Characterization of U-shape streamline fibers: Methods and applications. *Medical Image Analysis*, 18(5):795–807, 2014.

[144] Y. Zhang, J. Zhang, K. Oishi, A. V. Faria, H. Jiang, X. Li, K. Akhter, P. Rosa-Neto, G. B. Pike, A. Evans, et al. Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *NeuroImage*, 52(4):1289–1301, 2010.

[145] Z. Zhang, M. Descoteaux, J. Zhang, G. Girard, M. Chamberland, D. Dunson, A. Srivastava, and H. Zhu. Mapping population-based structural connectomes. *NeuroImage*, 172:130–145, 2018.