

Brain networks: classification and controllability

Dissertation presented by
Benjamin CHIÊM

for obtaining the Master's degree in
Mathematical Engineering

Supervisor(s)
Jean-Charles DELVENNE, Frédéric CREVECOEUR

Reader(s)
John LEE

Academic year 2016-2017

“Which statement seems more true: (1) I have a brain. (2) I am a brain.”

Douglas R. Hofstadter & Daniel C. Dennett (1982).

Université Catholique de Louvain

Abstract

Ecole Polytechnique de Louvain
Applied Mathematics

Master in Mathematical Engineering

**Brain networks:
classification and controllability**

by Benjamin CHIÊM

The human brain is one of the most complex and fascinating objects we could imagine. Composed of a large number of interconnected neural regions, its structure can be represented as a graph, called a connectome. This mathematical representation of human brain connectivity offers new tools to analyze and understand our brain, and opens the door to perspectives in health and disease. In this work, we begin by explaining how to generate structural connectomes, starting from Magnetic Resonance Imaging data. We build our own dataset composed of 204 structural connectomes and show that their topology significantly differ from that of random networks. Afterwards, motivated by potential applications such as the early diagnosis of neurological disorders, we adopt a machine learning framework to classify connectomes with respect to some label. We propose four classification strategies, along with two state-of-the-art methods, to try to discriminate connectomes based on the gender, the handedness, the alcohol consumption or the smoking habit of the corresponding individuals. Results based on our dataset show that (i) connectomes of men and women are significantly different, (ii) connectomes of smokers and non-smokers are slightly different and (iii) our strategies are often better than the considered state-of-the-art methods. Results related to handedness and alcohol consumption are inconclusive and highlight some limitations of our dataset. Finally, we introduce some notions about the controllability of connectomes, a concept that may impact future treatments of neuropathologies. In particular, we show that cognitive systems exhibit different behaviors in terms of controllability, depending on their function. This result highlights the important relationship between brain structure and function.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Mr Jean-Charles Delvenne and Mr Frédéric Crevecoeur, who helped me and motivated me throughout the realization of the present work.

I would also like to thank Michael Schaub, Leonardo Gutierrez Gomez and Mauro Faccin for their ideas and suggestions.

A special thank to Mr Etienne Huens, whose technical help was crucial for the realization of this thesis.

I would like to sincerely thank Brent McLean for his patience and his invaluable help during the proofreading of this manuscript.

Thanks also to my parents, my sisters Nathalie and Aurore, and my brother Jean-Christophe for their moral support throughout this year.

I am grateful to all my friends for supporting me on any occasion. These five years spent with you were incredible; thank you!

Finally, I want to sincerely thank the one who is always there when I need her... Because she is essential to me. Thank you Sarah, my love.

Human Connectome Project Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Motivations	2
1.2 Historical overview and state of the art	2
1.3 Organization of the thesis and contributions	4
2 Representation of human brain connectivity	5
2.1 Methods	5
2.1.1 Structural MRI processing	5
2.1.2 Diffusion MRI processing	6
Tractography	7
2.1.3 Dataset construction	7
2.2 Results	8
2.2.1 Null models	9
2.2.2 Characteristic path length	9
2.2.3 Clustering coefficient	10
2.2.4 Small-world organization	11
2.2.5 Modularity	12
2.3 Discussion	12
3 Classification of structural brain networks	17
3.1 Methods	17
3.1.1 Classification and Support Vector Machines (from [51])	17
Hard and soft margins	19
The kernel trick and Mercer's condition	20
Dealing with unbalanced classes	21
3.1.2 Classical kernels	22
Topological and spectral feature vector	22
Bag-of-edges and the curse of dimensionality	23
3.1.3 Graph kernels	24
Random walk kernel	24
Weisfeiler-Lehman kernel	25
3.1.4 Proposed kernels	26
DeltaCon kernel	27
Kernel based on the Regularized Laplacian	29
3.2 Results	29
3.2.1 Preliminary analysis, testing procedure and performance evaluation	30
Kernel positive semidefiniteness	30
Label redundancy and class imbalance	30
Training and test phases	31

	Performance metrics	32
3.2.2	Discrimination with respect to gender	34
3.2.3	Discrimination with respect to handedness	34
3.2.4	Discrimination with respect to alcohol abuse	34
3.2.5	Discrimination with respect to smoking	35
3.3	Discussion	36
3.3.1	Limitations of the dataset	36
3.3.2	Brain differences related to gender	36
3.3.3	Choice of the classifier	36
3.3.4	Choice of the performance metric	38
3.3.5	Choice of the kernel	39
3.3.6	Interpretation of a classifier decision	41
4	Controllability of structural brain networks	43
4.1	Methods	43
4.1.1	The connectome as a linear system	43
4.1.2	Global controllability	44
4.1.3	Regional controllability	45
4.2	Results	46
4.2.1	Preliminaries	46
4.2.2	Global controllability of the structural connectome	47
4.2.3	Correlation between controllability and strength	47
4.2.4	Controllability-by-system interaction	48
4.3	Discussion	48
5	Conclusion	53
A	Topological properties of connectomes	55
A.1	Z-scores histograms	55
B	Topological and spectral features for graph classification	59
B.1	Mathematical definition of the coefficients	59
B.2	Most discriminative features	63
C	Controllability-by-system	65
C.1	Correspondence between ROIs and cognitive systems	65

List of Figures

2.1	Schematic representation of the pipeline used to generate macroscale structural connectomes.	6
2.2	Parcellation of the human brain into Regions of Interest. Left-center: 34 cortical ROIs are defined per hemisphere. Right: 7 subcortical ROIs are defined per hemisphere. Note: the cerebellum is not depicted in this Figure, but was taken into account in our final parcellation scheme. Reproduced from [10] and publicly available at http://www.jneurosci.org/content/31/44/15775/tab-figures-data	6
2.3	Schematic representation of the reconstruction of a streamline (red) via deterministic tractography, in two dimensions. The grid represents the pixels of the image, and the arrows represent the preferred direction of diffusivity. . .	7
2.4	Illustration of the definition of an edge weight. A and B are ROIs. Dashed lines represent reconstructed axonal beams. The solid line represents a weighted edge in the connectome.	8
2.5	Structural connectome: abstract representation of the structure of the human brain as a complex network. Gray spheres represent the ROIs. The color of the edges corresponds to the orientation. Left: axial view. Right: sagittal view.	8
2.6	Illustration of one iteration of the Maslov-Sneppen rewiring algorithm.	9
2.7	Characteristic path length computed for 20 connectomes of the dataset (blue) an their respective rewired null model (green).	10
2.8	Illustration of the clustering coefficient of a node. The degree of the red node is 3. Left: No triangle is attached to the red node; $C_{red} = 0$. Center: One triangle is attached to the red node; $C_{red} = 1/3$. Right: Three triangles are attached to the red node; $C_{red} = 1$	10
2.9	Weighted clustering coefficient computed for 20 connectomes of the dataset (blue) and their respective rewired null model (green).	11
2.10	Illustration of community (or module) detection by modularity maximization.	12
2.11	Modularity optimized following the method introduced by Newman in [43], for 20 connectomes of the dataset (blue) an their respective rewired null model (green).	13
3.1	Illustration of the margin maximization, when classifying blue and red points. Highlighted points are called "support vectors".	18
3.2	Illustration of the soft margin principle for two non-linearly separable classes, with slack variable ξ^i	19
3.3	Illustration of the construction of an induced feature space in which classes are linearly separable.	20
3.4	Illustration of the Weisfeiler-Lehman principle. Aggregation of the labels in a neighborhood of size 1 to create new labels.	26
3.5	Splitting of the dataset into a training set and a test set.	32
3.6	Illustration of the 5-fold cross-validation procedure. Each line represents an independent test.	33

3.7	Confusion matrix.	34
3.8	Example of a confusion matrix for a naive classifier.	34
3.9	Left: average female connectome. Right: average male connectome. The 400 most discriminative edges obtained by RFE are depicted in the connectome where their weight is the highest. Inter-hemispheric edges are traced in green, intra-hemispheric edges are traced in red.	37
3.10	Scatter plot of the inter-hemispheric ratio for male and female connectomes. Threshold line: $\zeta^* = 0.111$. The expansion along the Y-axis is just for visualization.	37
3.11	Local explanation of a classification decision using an interpretable classifier. To explain the classifier decision for the red cross, we can train a linear SVM locally, considering neighboring points in an understandable feature space.	42
4.1	Percentage of average controllability by cognitive system, when considering the 25 strongest average controllability hubs and the correspondence in Appendix C. The proportions were normalized by the number of ROIs by system.	49
4.2	Percentage of modal controllability by cognitive system, when considering the 25 strongest modal controllability hubs and the correspondence in Appendix C. The proportions were normalized by the number of ROIs by system.	49
A.1	Z-score histogram for the characteristic path length, on the whole dataset. (Z-score computed with respect to null models.)	56
A.2	Z-score histogram for the clustering coefficient, on the whole dataset. (Z-score computed with respect to null models.)	56
A.3	Z-score histogram for the modularity on the whole dataset. (Z-score computed with respect to null models.)	57

List of Tables

3.1	Spectral range of the proposed kernel matrices, on our dataset composed of 204 connectomes.	30
3.2	Labels redundancy LR , with respect to our dataset composed of 204 connectomes. (Equation (3.38))	31
3.3	Class imbalance for each tested label, with respect to our dataset composed of 204 connectomes.	31
3.4	Parameters of the methods.	32
3.5	One interpretation of Cohen's kappa. (From [67])	34
3.6	Classification results for the label Gender , and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.	35
3.7	Classification results for the label Handedness , and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.	35
3.8	Classification results for the label Alcohol abuse , and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.	35
3.9	Classification results for the label Smoking , and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.	36
4.1	Pearson correlation coefficient between node strength and corresponding average and modal controllability measures, averaged across all 204 connectomes of our dataset. (Mean $\pm$ standard deviation)	48

List of Abbreviations

ROI	Region Of Interest
MRI	Magnetic Resonance Imaging
fMRI	functional Magnetic Resonance Imaging
EEG	Electro Encephalo Graphy
MEG	Magneto Encephalo Graphy
SVM	Support Vector Machine
RFE	Recursive Features Elimination
FaBP	Fast Belief Propagation
RW	Random Walk
WL	Weisfeiler-Lehman
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
MKL	Multiple Kernel Learning
PBH	Popov-Belevitch-Hautus

Chapter 1

Introduction

The human brain is one of the most complex and fascinating objects we could imagine. Composed of approximately one hundred billion neurons, each of them forming several thousands synapses, it governs our everyday life and allows us to speak, to move, to think. For this reason, it has attracted the attention of numerous scientists around the world, especially since the last century. This interest is far from pointless: our brain is the result of millions of years of evolution, making it able to process information from multiple sources, to take important decisions and to coordinate the different organs composing our body, all in parallel. Unravelling the mysteries of the human brain constitutes the purpose of neurosciences.

Neurosciences gather a wide variety of research fields ranging from neurobiology to neuropsychology, all of them seeking to explore a specific aspect of the human brain. In particular, the convoluted architecture shaping our brain has long fascinated neuroscientists, despite its complexity. Recently, much interest has been granted to an innovative vision of human brain connectivity, thereby opening a new domain called connectomics. A connectome is a mathematical model of the interactions between brain regions or elements. These interactions are either structural or functional and create a large, complex network. Therefore, modern network theory, a field of engineering sciences, naturally became essential in the study of brain connectivity. The meeting between applied mathematics and neurosciences through connectomics marked the starting point of our interest in this emergent field and motivated the realization of the present work.

Alongside with the design of supercomputers whose impressive computing power continues to increase, a growing tendency to compare performances of the human brain with that of supercomputers has emerged. Unfortunately, a strict comparison would be meaningless: computers and the human brain were not designed to perform the same tasks. While almost no human is able to compute the first fifty decimals of π in a reasonable amount of time, machines are still not able to develop or to understand emotions. Actually, we can think about computers and the human brain as being part of a common loop. While the field of machine learning was originally inspired by the complex infrastructure of the cerebral cortex, thereby giving birth to the so-called artificial neural networks, machine learning algorithms are nowadays increasingly used to study the fascinating abilities of our brain.

In this context, integrating connectomics in a machine learning framework reveals an immense potentiality for understanding the organ which, paradoxically, allows us to learn, to understand, to create.

1.1 Motivations

In substance, a network (or a graph) consists of a set of vertices interconnected by edges. Thinking about neurons and axons as being the vertices and edges composing a large, complex network thus appears natural. Using graph theory to analyze the human brain offers exciting opportunities. It is a wide, versatile field of research supported by a large community of mathematicians and it provides an abstract but intelligible framework to understand complex interactions taking place inside a network. From this observation, it seems fair to consider using graph theoretic tools to study the relationships between cerebral regions and their impact on the functioning of our brain.

Meanwhile, neuroimaging techniques have recently made impressive progress in terms of reliability and resolution. This results in a colossal amount of precious data to process and analyze. Following this observation, it becomes imperative to develop new methods and algorithms in order to process these data. The fields of machine learning and data science thus appear unavoidable in the future of neurosciences. Conveniently, these two fields also deal with networked data of all kind. In fact, in the context of brain connectivity studies, dealing with networked data is more interesting and computationally less demanding than directly processing raw neuroimages. Moreover, in a connectome, interactions between brain regions or elements are explicitly taken into account through edges, whereas the analysis of raw images often considers local areas without caring about their relation with each other.

Eventually, as most of biomedical fields of research, one of the final goals of connectomics is to develop new techniques aiming at understanding, diagnosing and perhaps treating various neuropathologies such as Alzheimer's disease, Parkinson's disease, depression, epilepsy, schizophrenia and many others. Particularly, with the help of increasingly accurate machine learning methods, it is a safe bet that connectomics will take advantage of the multiplication of neuroimaging data emanating from both healthy and pathological subjects, in order to detect abnormalities in pathological brains and to find their origin. Although the path is still long before reaching this goal, the health and disease perspective is probably what persuaded us to start this master thesis.

1.2 Historical overview and state of the art

The paradigm of neurons being the fundamental and discrete units of the nervous system comes from the seminal work of the spanish neuroscientist Ramón y Cajal (1852-1934), co-winner of the Nobel Prize in Physiology or Medicine in 1906. This innovative view of the circuitry shaping our brain constitutes the base of modern neurosciences. Unfortunately, due to the microscopic nature of neurons, methods allowing to explore the brain were rather slow to come up.

In the last years however, new neuroimaging technologies have emerged, thereby paving the way to new perspectives in our understanding of the brain. The first brain networks were based on an invasive technique measuring the propagation of a tracer inside the brain of the cat or the macaque monkey. This technique, called tract-tracing, provides excellent results but is restricted to small areas, hence requiring many repeated experiments at different cortical locations. The next key step was the realization of the first complete mapping of a nervous system. The work of White et al. [1] concerned the nematode *Caenorhabditis Elegans* (abbreviated *C. Elegans*), a small worm of approximately 1 mm in length. White et al.

identified 302 individual neurons and about 5000 synaptic connections in the nervous system of *C. Elegans*. In [2], Watts and Strogatz mathematically modeled for the first time the nervous system of *C. Elegans* as a binary graph, and showed that this graph exhibits what they called a small-world topology. Although we cannot trace back the dawn of connectomics to a precise date, this was probably the first bridge between modern network theory and neurosciences.

Since then, non-invasive structural and functional neuroimaging modalities made it possible to explore the human brain more deeply. Nevertheless, a complete brain map similar to that of *C. Elegans* remains unavailable for the human. Therefore, as explained by Fornito, Zalesky and Bullmore in [3], connectomics is divided in three branches: macro-, meso- and microscale connectomics. While microscale refers to the level of single neurons and relies inherently on invasive imaging modalities applied to microscopic subparts of the brain, macroscale connectomics subdivides the entire cortex in a given number of Regions Of Interest (ROIs), thereby allowing *in vivo* non-invasive studies thanks to modern technologies. Meanwhile, mesoscale connectomics lies at the boundary of both its micro- and macroscale counterparts. In the present work, we focus our attention on macroscale human connectomes since their generation is easier thanks to efficient and non-invasive modalities such as Magnetic Resonance Imaging (MRI).

To build a macroscale connectome, several approaches are available in the literature. As reviewed by Fornito et al. in [4], these approaches can be grouped in two main classes, giving rise either to structural or to functional connectomes. Structural brain networks are based on the reconstruction of axonal pathways between predefined neural regions and mainly rely on diffusion-weighted imaging data. As discussed in [3] and [4], the validity of such connectomes essentially depends on the algorithms used for the reconstruction of the axonal beams, leading to the definition of edges. Besides, in functional brain networks, edges indicate the statistical dependence existing between activity patterns of two neural regions. There is no consensus on which dependence measure is the best suited, although the most commonly used is the Pearson correlation coefficient between time series of neural activity. Finally, let us mention a third, less popular kind of connectivity presented in [3], [4] and [5]. While both structural and functional connectivities give rise to undirected connectomes, effective connectivity aims at explicitly uncover the causal influences between neural regions, thereby allowing the definition of directed networks. Involving computationally intensive models of neuronal activity, the generation of effective connectomes is still hard at the macroscale, but should attract more attention in years to come.

Much work has already been done in the topological characterization of macroscale connectomes. The authors of [3, 5, 6, 7, 8] provide extensive reviews of the most important topological coefficients used to describe connectomes. In the same way as Watts and Strogatz established the small-world topology of the nervous system of *C. Elegans*, several studies confirmed this result for the human connectome. In [9], Bassett and Bullmore discuss the implications of small-world topology in both structural and functional brain networks. Another topological aspect that captivated the connectomic community is the rich-club organization of the human connectome, highlighted by Van Den Heuvel and Sporns in [10]. Interestingly, it turns out that the brain shares many topological properties with other kinds of networks such as social networks or electric power grids.

One of the most important topic in neurosciences is the interdependence between structure and function within the brain. As already discussed by several authors in [3, 5, 11, 12],

while the functionality of particular neural regions is constrained by anatomical connections, it is also true that structural links slowly evolve according to the functions required by the brain. For example, we can consider the evolution of the brain from childhood to adulthood. Although little attention has been paid until now to the relation between dynamics and structure in brain networks, some authors such as Sporns [7] and Stam et al. [8] suggest to give more importance to this topic in the future.

Along with the multiplication of neuroimaging databases, connectomics is already considered as a data science, as discussed by Kong and Yu in [13]. Indeed, while early connectomic studies focused their attention on single connectomes, the accessibility of new, large neuroimaging datasets allows us to compare connectomes across a large number of individuals. Furthermore, by combining neuroimaging data with behavioral and demographic aspects, new prospects are made possible thanks to machine learning and data mining, as explained by Glasser et al. in [14].

Finally, a large part of connectomic studies share one common goal: enhancing our understanding of neurological disorders. Many articles [5, 7, 15, 16] emphasize the future challenges of connectomics in health and disease. Although at the early stage, connectomic-based analysis of neuropathologies have already shown promising results in the context of depression [17], autism [18], schizophrenia [19, 20], epilepsy [21] and mild cognitive impairment [22, 23] among others. Clearly, connectomics will play a key role in the future of clinical neurosciences.

1.3 Organization of the thesis and contributions

In Chapter 2, we address the question of structural connectome extraction from MRI data provided by the Human Connectome Project. In particular, we present the methodological pipeline we adopted according to the literature, and we test several topological properties of the extracted connectomes in order to assess their consistency. Our contribution is the generation of our own dataset composed of 204 connectomes.

Chapter 3 is the core of the present work. We focus on a machine learning related topic: the classification of structural brain networks. We begin by introducing some fundamentals of supervised learning. Then, we present several approaches aiming at classifying the previously generated connectomes with respect to the gender, the handedness, the alcohol consumption or the smoking habit of the corresponding individuals. Along with two state-of-the-art methods, we propose four strategies based on the literature but not yet applied in this context. These strategies constitute our contribution to this chapter and are highlighted in blue boxes in the text. Results show that our strategies provide in general better results than state-of-the-art methods. We also discuss several aspects related to the class imbalance problem and the performances evaluation.

The aim of Chapter 4 is to introduce the rising field of controllability applied to structural connectomes. We base our approach on one specific paper, present concepts underlying the notion of controllability and reproduce some experiments conducted in the reference paper. We emphasize some encouraging results and discuss future prospects for this growing field.

To conclude, we come back in Chapter 5 on our main results and we mention some topics that aroused our interest but were not part of the present work.

Chapter 2

Representation of human brain connectivity

As discussed in the introduction, representing the human brain as a complex network appears to be natural. However, starting from neuroimaging data, several approaches allow us to construct connectomes having different physical meanings. Depending on the kind of study conducted, the choice of suitable modalities and algorithms is crucial. In this chapter, we describe the methodological pipeline used for the construction of our dataset, we evaluate some topological measures across our dataset to assess the consistency of our generation process and we justify our choices.

2.1 Methods

Intuitively, vertices and edges can be seen as analogies of neurons and axons respectively, forming a wide complex network. However, mapping the billions of neurons and trillions of axons composing the human brain remains impossible. By defining some Regions Of Interest (ROIs) over the entire cerebral cortex and measuring the relationship between them, macroscale connectomics aims at studying the connectivity of the human brain thanks to graph theoretic tools. In particular, we focus our attention on structural connectomes, which reflect the anatomical links lying inside the brain. To build a structural connectome, information from both structural and diffusion Magnetic Resonance Imaging (MRI) can be combined. While structural MRI will help to define the ROIs (i.e. the vertices of the connectome), diffusion MRI will allow us to reconstruct axonal beams lying in the white matter, thus defining the weights of the edges of the resulting brain network. This methodological pipeline (although not unique) is widespread in the literature, as illustrated by the work of Lufriu et al. [24], Gu et al. [25] or Hagmann [26, 27], among others. It is summarized in Figure 2.1 and described in the rest of this section.

2.1.1 Structural MRI processing

To feed our pipeline, we use T1-weighted¹ MRI images to provide good contrast between white matter and gray matter. The former consists of the inner part of the brain and is mainly composed of axons (also called tracts), whereas the latter contains most of the neuronal cells. Hence, we first need to perform image segmentation to separate white and gray matter in order to build anatomical information that will constrain further processing steps.

Next, 34 cortical ROIs per hemisphere were defined according to the automated atlas introduced by Desikan et al. [28], and 8 subcortical ROIs per hemisphere (including cerebellum) were segmented following the procedure described by Patenaude et al. in [29]. It is

¹The weighting schemes of MRI images refer to a particular setting of the acquisition parameters resulting in different levels of contrast between the white matter, the gray matter and the cerebrospinal fluid.

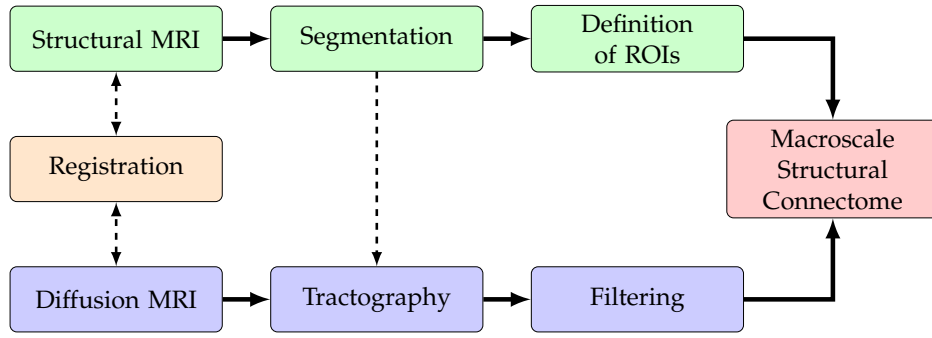


FIGURE 2.1: Schematic representation of the pipeline used to generate macroscale structural connectomes.

II

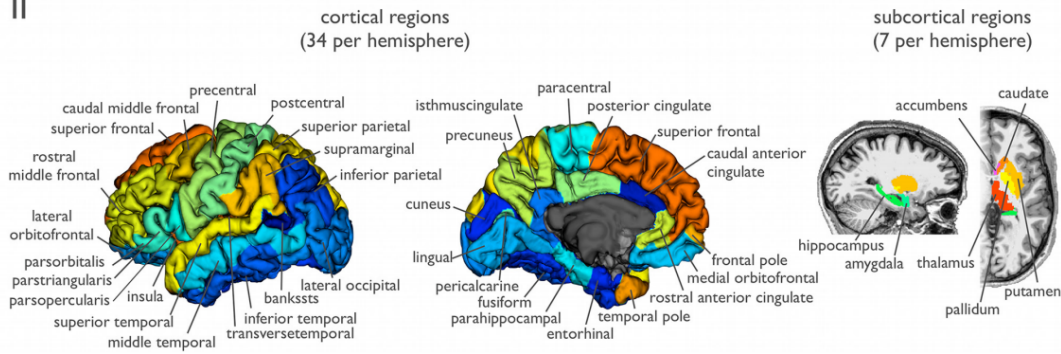


FIGURE 2.2: Parcellation of the human brain into Regions of Interest. Left-center: 34 cortical ROIs are defined per hemisphere. Right: 7 subcortical ROIs are defined per hemisphere. Note: the cerebellum is not depicted in this Figure, but was taken into account in our final parcellation scheme. Reproduced from [10] and publicly available at <http://www.jneurosci.org/content/31/44/15775/tab-figures-data>

interesting to notice that, following this procedure, the whole set of resulting connectomes will share the same 84 vertices. The global parcellation is depicted in Figure 2.2.

2.1.2 Diffusion MRI processing

Diffusion imaging allows us to non-invasively identify pathways inside the cerebral cortex by measuring the diffusivity of water molecules inside the brain. In white matter, this diffusivity is highly constrained by the direction of nerve fibers. Therefore, the measurement of the diffusivity along several directions allows us to deduce a privileged direction at a given voxel of the image.

It is important to specify that the voxel size in diffusion imaging ranges from 1 to 2 mm , while axons have a diameter of approximatively 1 μm . Thus, the reconstructed pathways correspond more to axonal beams than to single fibers. However, at the macroscale, these axonal beams provide sufficient information about the connectivity between ROIs.

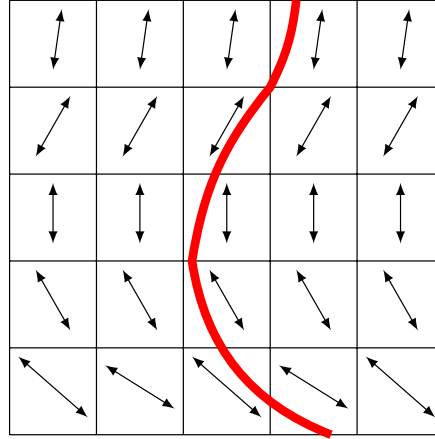


FIGURE 2.3: Schematic representation of the reconstruction of a streamline (red) via deterministic tractography, in two dimensions. The grid represents the pixels of the image, and the arrows represent the preferred direction of diffusivity.

Tractography

Tractography is the process allowing the reconstruction of axonal beams (also known as streamlines) from voxel to voxel. The simplest algorithms choose a random seed voxel, follow its privileged diffusion direction to reach the next voxel and the process is repeated until a given stopping criterion is met (see Figure 2.3). These algorithms, called deterministic, have the advantage of being simple but they are sensitive to noise and may lead to reconstruction errors, especially at voxels where several fibers cross each other [30]. To solve this issue, probabilistic tractography algorithms were developed. Although more computationally demanding (but easily parallelizable), they allow us to reduce reconstruction errors by evaluating, at each step, the probability density of a connection between two voxels. By exploiting the uncertainty of the direction of a fiber inside a voxel, probabilistic algorithms provide, in general, better results than deterministic algorithms [30].

Tractography algorithms are known to introduce some biases that could affect the final result, as explained by Smith et al. in [31]. Therefore, a filtering step is often needed after the tractography process in order to remove these biases and provide more biologically meaningful results.

2.1.3 Dataset construction

Preprocessed² structural and diffusion MRI data of 91 healthy men and 113 healthy women, aged from 22 to 37 years, have been retrieved from the Human Connectome Project database [34]. These data were then processed using the MRtrix3 toolbox developed by Tournier et al. [35] and following the guidelines available in the online³ documentation.

First of all, both anatomical and diffusion images were registered in order to get voxel correspondence. Then, we performed the whole-brain anatomically-constrained tractography introduced by Smith et al. in [36] by combining an improved probabilistic algorithm

²Preprocessing pipelines are described by Glasser et al. in [32]. In particular, data related to the automated cortical parcellation were produced by the FreeSurfer software [33].

³http://mrtrix.readthedocs.io/en/latest/quantitative_structural_connectivity/ismrm_hcp_tutorial.html



FIGURE 2.4: Illustration of the definition of an edge weight. A and B are ROIs. Dashed lines represent reconstructed axonal beams. The solid line represents a weighted edge in the connectome.

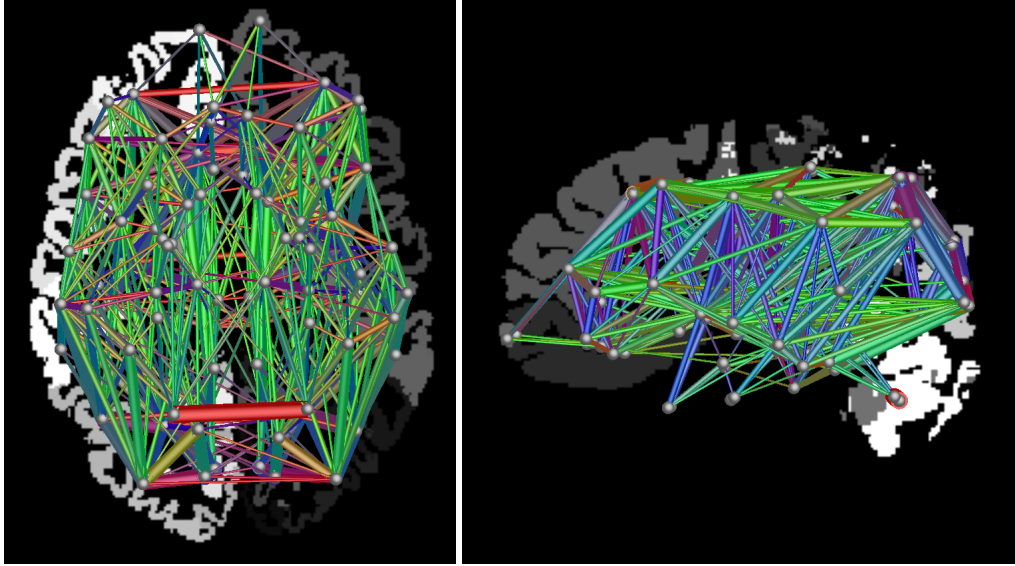


FIGURE 2.5: Structural connectome: abstract representation of the structure of the human brain as a complex network. Gray spheres represent the ROIs. The color of the edges corresponds to the orientation. Left: axial view. Right: sagittal view.

based on a second-order integration over the fibers orientation probability density [37] and previous anatomical information provided by the image segmentation. We chose to stop this process after the reconstruction of six million streamlines. We then filtered reconstructed axonal beams, as described in [31], to remove potential biases and reduce the number of streamlines to about four million. Finally, the weight of an edge linking two ROIs (or vertices) was defined as the number of axonal beams shared by these two ROIs, as illustrated in Figure 2.4. The final result is thus a connected, weighted, undirected network, as depicted in Figure 2.5.

2.2 Results

Connectomes are complex networks characterized by several topological properties reported in the literature (see for example the review of some classical connectome measures and their interpretation by Rubinov and Sporns [6]). In this section, we want to assess that the methodological pipeline used for the construction of our dataset gives rise to networks significantly more structured (from a topological point of view) than random networks. First, we introduce the concept of a null model that we use in the following parts. Then, we present some global topological measures that we deemed representative when describing connectomes, and we compare their value in connectomes to their value in null models.

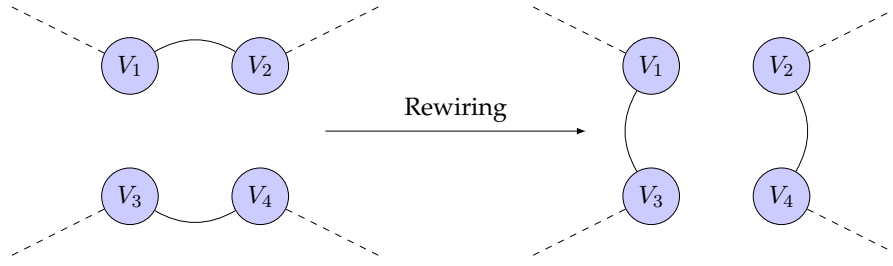


FIGURE 2.6: Illustration of one iteration of the Maslov-Sneppen rewiring algorithm.

2.2.1 Null models

To assess that a given topological measure is a significant property of connectomes, we need to compare this measure for both connectomes and networks whose organization has been randomized. Such random networks play the role of benchmarks and are referred to as null models. Ideally, when studying one particular measure, the connectome and the null model should share as many properties as possible, except for the measure of interest. In such a scenario, the potential contrast emerging between the connectome and the null model would only come from this particular measure. In practice, only a few of the main features of the networks (e.g. number of nodes and edges, degree and strength distribution) are maintained.

Two main categories of null models exist. The first one encompasses synthetic networks randomly generated according to some particular rules. The most famous models falling in this category come from the work of Erdős and Rényi [38], Watts and Strogatz [2], and Barabási and Albert [39]. The second category includes null models based on an iterative random rewiring of an original network (in our case, a connectome). One of the most popular rewiring algorithms was introduced by Malsov and Sneppen [40]. Fundamentally, this algorithm chooses at each iteration two non-adjacent edges (and thus, four vertices) at random. Then, these two edges are moved between the same four vertices in order to create two new edges. If at least one of the new edges already exists in the rewired network, the move is ignored, and a new iteration begins. This procedure, depicted in Figure 2.6, ensures that the rewired network has the same number of nodes and edges and the same degree distribution as the original network.

Hereafter, we use null models generated by a generalized version of the Maslov-Sneppen algorithm, appropriate for weighted networks, and introduced by Rubinov and Sporns in [41]. All the MATLAB functions we used come from the Brain Connectivity Toolbox⁴.

2.2.2 Characteristic path length

Let us consider a weighted undirected graph $G(\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$. The characteristic path length of G is defined as

$$L = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ i \neq j}} l_{ij} \quad (2.1)$$

⁴<https://sites.google.com/site/bctnet/>

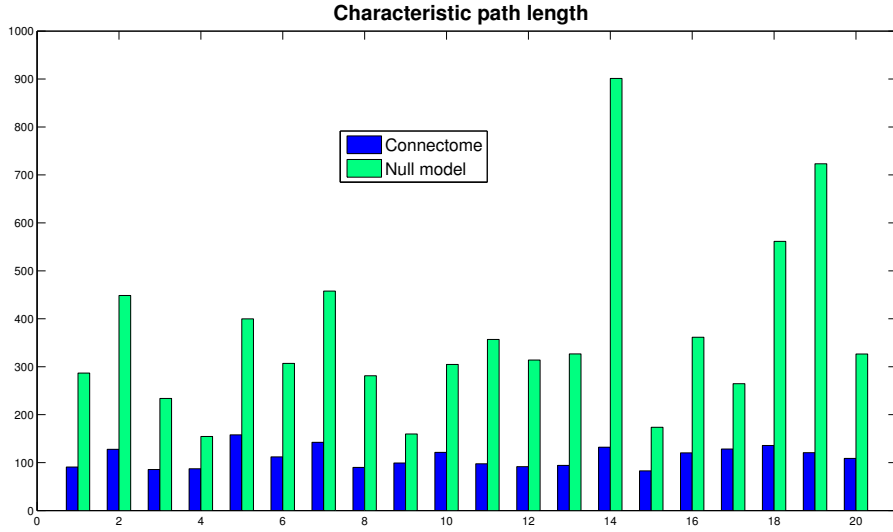


FIGURE 2.7: Characteristic path length computed for 20 connectomes of the dataset (blue) and their respective rewired null model (green).

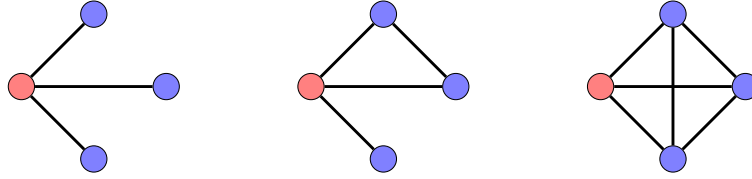


FIGURE 2.8: Illustration of the clustering coefficient of a node. The degree of the red node is 3. Left: No triangle is attached to the red node; $C_{red} = 0$. Center: One triangle is attached to the red node; $C_{red} = 1/3$. Right: Three triangles are attached to the red node; $C_{red} = 1$.

where $N = |\mathcal{V}|$ is the number of nodes and l_{ij} is the length of the shortest path between node i and node j .

When dealing with brain networks, the characteristic path length is often defined as a measure of integration. In other words, networks showing a short characteristic path length tend to promote, on average, an efficient routing of information between any pair of vertices. Figure 2.7 clearly shows that connectomes have a short characteristic path length compared to their null model. A global view of the distribution of the characteristic path length in connectomes and their respective null models is also available in Appendix A.

2.2.3 Clustering coefficient

Denoting k_i as the degree of node i and t_i as the number of closed triangles attached to this node, the clustering coefficient of a network is defined as

$$C = \frac{1}{N} \sum_{i=1}^N \frac{2t_i}{k_i(k_i - 1)} \quad (2.2)$$

Intuitively, as illustrated in Figure 2.8, the term of the sum in (2.2) can be understood as the proportion of pairs of i 's neighbors that are connected. A weighted variant of the

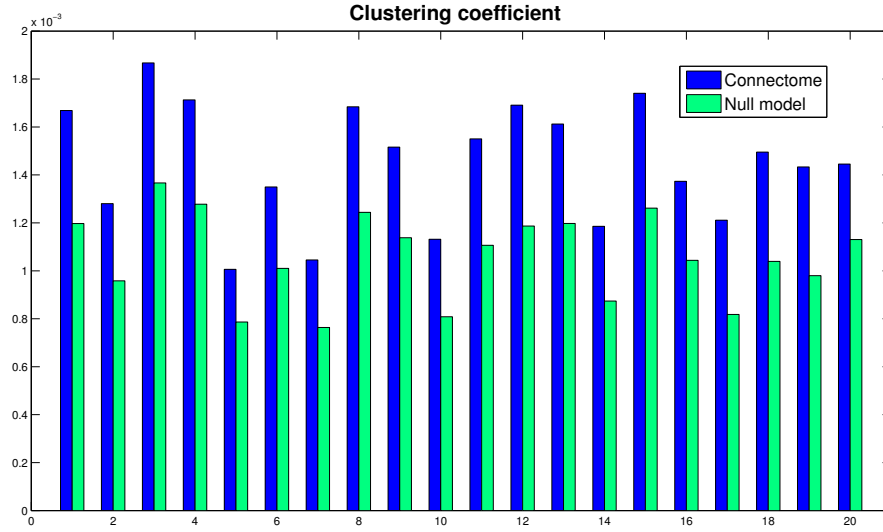


FIGURE 2.9: Weighted clustering coefficient computed for 20 connectomes of the dataset (blue) and their respective rewired null model (green).

clustering coefficient, based on the concept of subgraph intensities, has been proposed by Onnela et al. in [42] and is defined as

$$C_w = \frac{1}{N} \sum_{i=1}^N \frac{2}{k_i(k_i - 1)} \sum_{j,h} (\hat{w}_{ij} \hat{w}_{jh} \hat{w}_{hi})^{1/3} \quad (2.3)$$

where j and h are neighbors of i and $\hat{w}_{ij}$ denotes the weight of the edge linking i and j , scaled by the maximum weight of the network.

As opposed to the characteristic path length, the clustering coefficient is a measure of segregation in a network. This means that networks showing a high clustering coefficient are, in principle, able to locally specialize some processes within densely interconnected groups of vertices. However, the clustering coefficient is not directly interpretable in this way since it is a purely topological measure that does not involve any dynamics on the network. In Figure 2.9, we see that connectomes are characterized by a relatively high clustering coefficient. A global view of the distribution of the clustering coefficient in connectomes and their respective null models is also available in Appendix A.

2.2.4 Small-world organization

A short characteristic path length combined with a high clustering coefficient corresponds to a very particular network topology known as "small-world". This kind of networks tends to promote both segregation and integration, and thus is able to support complex dynamics. In [9], Bassett and Bullmore also argue that small-world brain networks are economical in terms of wiring cost and efficient in terms of information processing. Furthermore, these authors suggest that small-worldness could be the origin of resilience to certain pathological attacks. For all these reasons, the fact that all the connectomes constituting our dataset own the small-world property is highly meaningful when assessing that our generation process does not give rise to random networks.

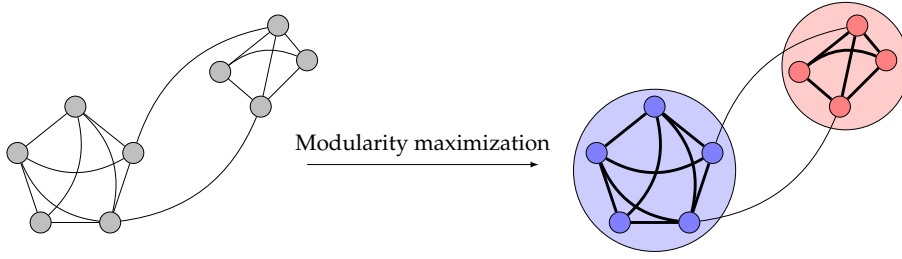


FIGURE 2.10: Illustration of community (or module) detection by modularity maximization.

2.2.5 Modularity

Modularity is a global measure whose concept is similar to that of the clustering coefficient. Given a partition of the network into non-overlapping groups of densely interconnected nodes (also known as modules or communities), this segregation index compares the density of connections within a module to that of connections lying between modules. Let m be the sum of all edge weights in the network, w_{ij} the weight of the edge linking nodes i and j , k_i the strength of node i (i.e. the sum of the weights of all edges linked to node i), c_i the index of the module of node i and δ the classical Kronecker delta. The weighted version of the modularity Q is

$$Q = \frac{1}{2m} \sum_{i,j=1}^N \left[w_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2.4)$$

Let us note that the leading factor $\frac{1}{2m}$ may vary depending on the definitions. The maximization of Q , illustrated in Figure 2.10, is the object of a lot of research, and many algorithms were developed to this purpose. Indeed, unravelling a community structure is of crucial interest when studying a complex network since these communities can play an important role in the dynamics supported by the network.

In Figure 2.11, we see that connectomes in our dataset exhibit a high modularity when optimized by the method introduced by Newman in [43], which confirms some results already reported in the literature. For instance, Mattar et al. show in [44] that communities inside functional connectomes correspond to known cognitive systems, while Meunier et al. demonstrate in [45] that the modular organization of both structural and functional connectomes is even more complex, namely hierarchically modular. A global view of the distribution of modularity in connectomes and their respective null models is also available in Appendix A.

2.3 Discussion

Origin of the data Motivated by the generation of a biologically consistent dataset of acceptable size, the choice to use data coming from the Human Connectome Project database seemed relevant. Indeed, this database provides multimodal neuroimaging data for more than 1000 subjects. These data, along with behavioral and demographic measures, were acquired following strict protocols, which ensures the stability of our dataset.

The size of our dataset was mainly constrained by the time required by the generation process. In future work, growing this dataset would only be beneficial for the relevance of the study.

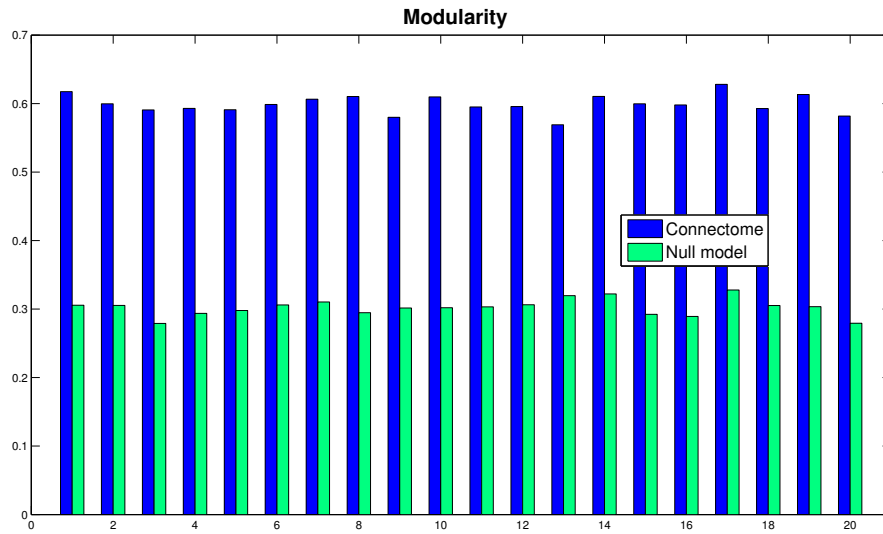


FIGURE 2.11: Modularity optimized following the method introduced by Newman in [43], for 20 connectomes of the dataset (blue) and their respective rewired null model (green).

Structural connectomes Two main classes of connectomes can be distinguished in the literature: structural (or anatomical) networks and their functional equivalent. While the former provide information about the physical links between each pair of ROIs, the latter define the weight of an edge as a measure of the co-activity of both corresponding ROIs, either at rest or during the execution of cognitive/motor tasks. The imaging modalities used in either case also drastically differ. As described in section 2.1, the combination of structural and diffusion MRI is the most common choice to build a structural connectome. For functional networks, functional MRI (fMRI) has become highly popular although neuroscientists are increasingly interested in other modalities such as ElectroEncephaloGraphy (EEG) or MagnetoEncephaloGraphy (MEG).

Because they directly (or indirectly in the case of fMRI) highlight the electrical activity of the brain, functional connectomes are attractive for studying brain dynamics. However, most functional studies only consider static connectomes, while time-varying networks could potentially be more relevant. Moreover, how to define the measure of co-activity between two neural regions remains arbitrary: many studies consider the correlation between time series (which makes the strong assumption of a linear relationship between the activity of two ROIs), whereas others consider the mutual information. Even if many studies have shown promising results, the artificial way functional connectomes are built could influence the consistency of the final network.

Furthermore, structural connectivity can be understood as the substrate from which co-activity patterns emerge [5, 11, 12], and somehow dominates functional connectivity. Hence, though the analysis of structural connectomes may turn out to be more subtle, it can potentially reveal deep features of the human brain.

Macroscale connectomics Considering the study of the human connectome at the macroscale has several advantages. First, neuroimaging modalities used at the macroscale are mostly noninvasive, safe and flexible, which allows the study of large groups of subjects, possibly over their lifetimes. Moreover, studying the human brain as a whole allows us to study the potential interactions between remote brain regions.

However, when studying macroscale connectomes, it is important to keep in mind that many neural mechanisms are deeply hidden in the structure of the brain. Microscale connectomics precisely addresses the study of these mechanisms at the scale of single neurons. The main disadvantages of this approach are the cost of invasive data acquisition which prevents the construction of large datasets, and the hard scalability to large neural systems.

In the future, the combination of results emanating from both microscale and macroscale connectomics should be of great interest, and opens up new exciting perspectives.

Choice of the brain atlas It is worth mentioning that several different brain atlases exist. Actually, this constitutes an important weakness of the overall connectomics field of research. Indeed, no consensus within the neuroscientific community exists about the strategy to adopt in order to define a suitable cortical parcellation. Of course, depending on the scale of interest, the atlases and their resolution will be different, but even at the macroscale there exists a wide range of atlases.

The atlas proposed by Desikan et al. [28] has the advantage to automatically recognize cortical areas in unseen MRI images, thus making it essential in a pipeline whose goal is to generate a large number of connectomes. Nevertheless, the main limitation of this kind of automated atlas is that the size of the regions resulting from the parcellation may highly vary from one another. This could potentially bias further results. Indeed, ROIs with large surface area are more likely to have a high degree than others, while this does not necessarily reflect the biological truth.

To address this problem, some authors such as Hagmann et al. in [46] suggest to randomly parcellate the cortical surface in regions of equal area. Unfortunately, following this approach, the defined ROIs have no biological sense, and the comparison between connectomes becomes meaningless. To address this problem, an interesting idea suggested by Hagmann et al. in [47] is to perform a random parcellation inside each anatomical ROI previously defined by an atlas. This would allow us to keep a biological correspondence between nodes and ROIs (several nodes being associated to the same ROI), while avoiding large differences in size between regions associated to individual nodes.

Furthermore, the atlas resolution (i.e., the number of ROIs that are defined) drastically varies across the studies, from less than 100 nodes to several thousands. As reported by Zalesky et al. in [48], this resolution can influence some topological measures of the connectome such as the small-worldness. Hence, it is important to always relate results from connectomics studies to the atlas used for the connectomes generation.

Finally, let us mention that we could have used the automated atlas introduced by Fischl et al. in [49] and updated in [50], known as the "Destrieux" atlas. This parcellation scheme defines 74 cortical ROIs per hemisphere and is equivalent in terms of quality to the Desikan atlas. However, we chose to use the latter, which defines fewer nodes, to avoid huge computational cost in the processes described in Chapter 3.

Limitations of tractography Tractography algorithms suffer from several drawbacks. First, the start and end points of reconstructed streamlines are poorly defined. Ideally, seed voxels should be randomly chosen in the gray matter, where the axons take their origin. However, water diffusion in gray matter is less constrained than in white matter, and information from diffusion imaging in these areas is thus poor. Moreover, tractography algorithms have difficulties in managing long streamlines. Actually, many algorithms require a fixed maximum streamline length. In our generation process, the use of anatomically-constrained tractography introduced by Smith et al. in [36] combined with a filtering post-step presented by Smith et al. in [31] allows us to greatly reduce the effect of these drawbacks.

In addition, diffusion MRI and tractography do not provide any information about the direction of reconstructed nerve fibers. This yields undirected connectomes. Furthermore, the number of streamlines to reconstruct is often chosen arbitrarily, and can span several orders of magnitude. For instance, in [25], Gu et al. consider the reconstruction of 10^5 streamlines, while the instructions we followed in the tutorial of the MRtrix3 toolbox recommend the reconstruction of 10^8 streamlines. In our case, we made a trade-off between accuracy of reconstruction and available time by choosing to reconstruct 4×10^6 streamlines. We are aware that this parameter can influence further results since the weights of the edges directly depend on the number of streamlines reconstructed. A simple solution would be to normalize the resulting adjacency matrix by the maximum edge weight for example.

Why extract a network representation from neuroimaging data? This question could be reworded as "Why do we not directly infer on neuroimaging data?". Indeed, at the time of deep learning and "big data", the direct analysis of raw data would appear not only feasible but also exciting. Nevertheless, we strongly believe that the advantages of the network representation are multiple.

First of all, let us reiterate that we are here interested in one particular aspect of the human brain: its structural connectivity. Hence, by focusing on a network representation, we somehow reduce the dimension of the data to study. Even if the connectome generation process is long, it results in a concise object not too expensive to further analyse.

In addition, graph theory is affordable in terms of knowledge: many notions are intuitive and easily understood by non-mathematicians. The accessibility and the convenience of graph theory strengthen the potential of connectomics and open the door to many scientists.

Finally, graph theoretical tools have the advantage of being general and robust. Studying the human brain through graph theory opens the field of neuroscience to a vast community of mathematicians and data scientists and leads to new perspectives.

Chapter 3

Classification of structural brain networks

«Do we all have the same brain structure? »

This question, as innocent as it seems, is fundamental when conducting studies about the human brain. Indeed, structural brain differences across individuals may potentially bias some experimental results. Furthermore, comparing the brains of subjects suffering from certain cerebral pathologies to those of healthy individuals could help us to understand these neurological disorders, and perhaps lead to early diagnosis procedures.

In this chapter, we address this question from the viewpoint of supervised learning. The motivation behind this is the following: Machine learning tools are becoming unavoidable in many scientific fields, due to the explosion of the quantity of data to be processed. Neuroscience is not an exception: the introduction of new neuroimaging techniques with impressive resolution has caused the resulting data to become nearly impossible to analyse without the help of powerful algorithms. In particular, we want to classify the connectomes composing our dataset (see Chapter 2), according to some label of interest. It turns out that the classification of structured data such as networks is a field of research that has recently attracted much interest within the machine learning community.

We begin by introducing the main tool we use for classification, namely Support Vector Machines (SVMs). Afterwards, we propose several approaches for classifying connectomes and we compare their respective performance for different labels. We show that, although results may vary according to the chosen label, the proposed approaches are competitive with state-of-the-art methods. Eventually, we discuss the validity of these results and propose future research ideas.

3.1 Methods

3.1.1 Classification and Support Vector Machines (from [51])

Let us consider a set of data points. Each of them is described by a vector of features. Some of the points belong to a class C_1 and the others belong to a class C_2 . For the sake of simplicity, we will only consider two classes (binary classification) in this document. The classification process aims at determining the class of new, unclassified data points. Among the multitude of classification methods available in the literature, we choose a type of algorithm known as Support Vector Machines (SVMs).

Since the early 2000s, the popularity of SVMs has continued to grow, due to their flexibility, simplicity and competitive performances compared to other classifiers. We introduce

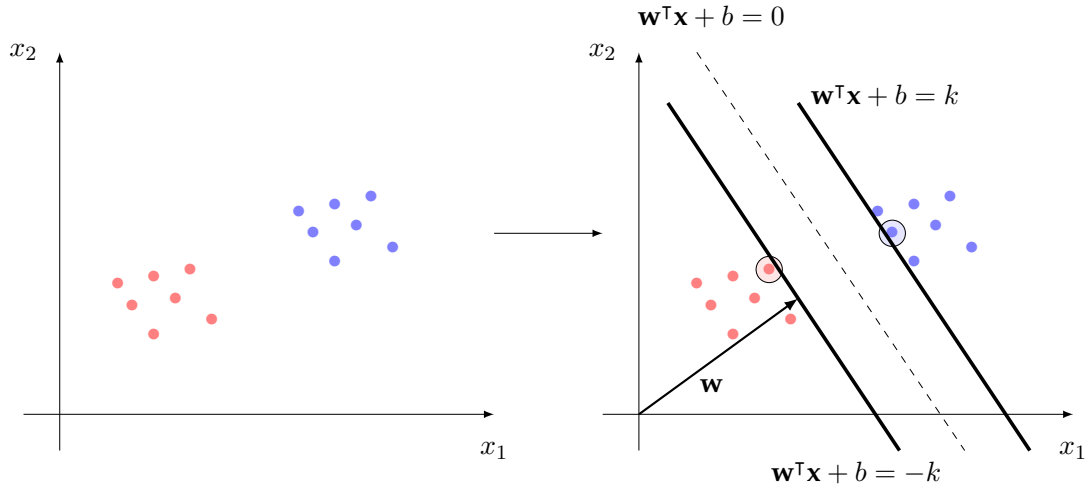


FIGURE 3.1: Illustration of the margin maximization, when classifying blue and red points. Highlighted points are called "support vectors".

here the fundamentals composing the theory underlying SVMs.

Given a set of N data points whose class is known, classification algorithms will try to define a separation between the two classes. Then, new data points are classified according to this separation. In the particular case of SVMs, we seek the maximal margin linear separation in the space spanned by the features representing the data points. In Figure 3.1, each data point $\mathbf{x}$ is described by two features x_1 and x_2 , and either belongs to class C_1 (red) or C_2 (blue). The dashed line represents the maximal margin separation, and the two solid lines represent the borders of the margin. Data points belonging to these borders are called support vectors.

Following the notations of Figure 3.1, we define the width of the margin as

$$\text{width} = \frac{2|k|}{\|\mathbf{w}\|_2} \quad (3.1)$$

Thus, the question of finding the maximal margin linear separation perfectly separating the two classes can be rewritten as an optimization problem:

$$\begin{aligned} \max_{\mathbf{w}, b} \quad & \frac{2|k|}{\|\mathbf{w}\|_2} \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{x} + b \leq -k, \quad \forall \mathbf{x} \in C_1 \\ & \mathbf{w}^T \mathbf{x} + b \geq k, \quad \forall \mathbf{x} \in C_2 \end{aligned} \quad (3.2)$$

The two constraints ensure that data points lie on the side corresponding to their respective class. Let us notice that there exists a scale for which we can choose $k = 1$. To simplify the notations, we index the data points with the superscript $i \in \{1, 2, \dots, N\}$, and attach to each of them a new variable y defined as follows:

$$y^i = -1, \quad \forall \mathbf{x}^i \in C_1 \quad (3.3)$$

$$y^i = 1, \quad \forall \mathbf{x}^i \in C_2 \quad (3.4)$$

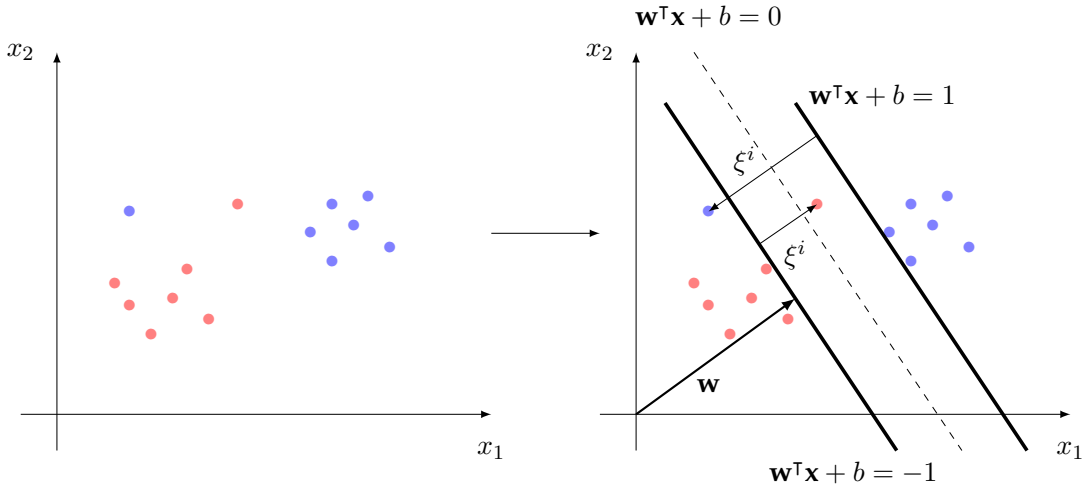


FIGURE 3.2: Illustration of the soft margin principle for two non-linearly separable classes, with slack variable ξ^i .

Thanks to these notations, we can rewrite our problem in the following manner:

$$\begin{aligned} \max_{\mathbf{w}, b} \quad & \frac{2}{\|\mathbf{w}\|_2} \\ \text{subject to} \quad & y^i(\mathbf{w}^T \mathbf{x}^i + b) \geq 1, \quad \forall i \in \{1, 2, \dots, N\} \end{aligned} \quad (3.5)$$

Conveniently, we can reformulate this as a minimization problem:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \|\mathbf{w}\|_2^2 \\ \text{subject to} \quad & y^i(\mathbf{w}^T \mathbf{x}^i + b) \geq 1, \quad \forall i \in \{1, 2, \dots, N\} \end{aligned} \quad (3.6)$$

When inspecting this last expression, let us notice that the optimization problem is convex. Hence, there exists a unique global minimum (when feasible) and the minimizers $\mathbf{w}$ and b are also unique. Furthermore, we can use efficient quadratic programming techniques to solve this problem. All these reasons make SVMs robust and attractive.

Hard and soft margins

Nevertheless, it could happen that the two classes are not linearly separable, resulting in a non-feasible problem. To address this issue, we introduce the concept of slack variables. As illustrated in Figure 3.2, a slack variable ξ^i attached to a data point $\mathbf{x}^i$ represents the distance from this point to the border of its respective class. For correctly classified points, we have $\xi^i = 0$. Thanks to these variables, we will allow the algorithm to misclassify some data points, while penalizing these misclassifications. The optimization problem becomes:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \|\mathbf{w}\|_2^2 + C \sum_{i=1}^N \xi^i \\ \text{subject to} \quad & y^i(\mathbf{w}^T \mathbf{x}^i + b) \geq 1 - \xi^i, \quad \forall i \in \{1, 2, \dots, N\} \\ & \xi^i \geq 0, \quad \forall i \in \{1, 2, \dots, N\} \end{aligned} \quad (3.7)$$

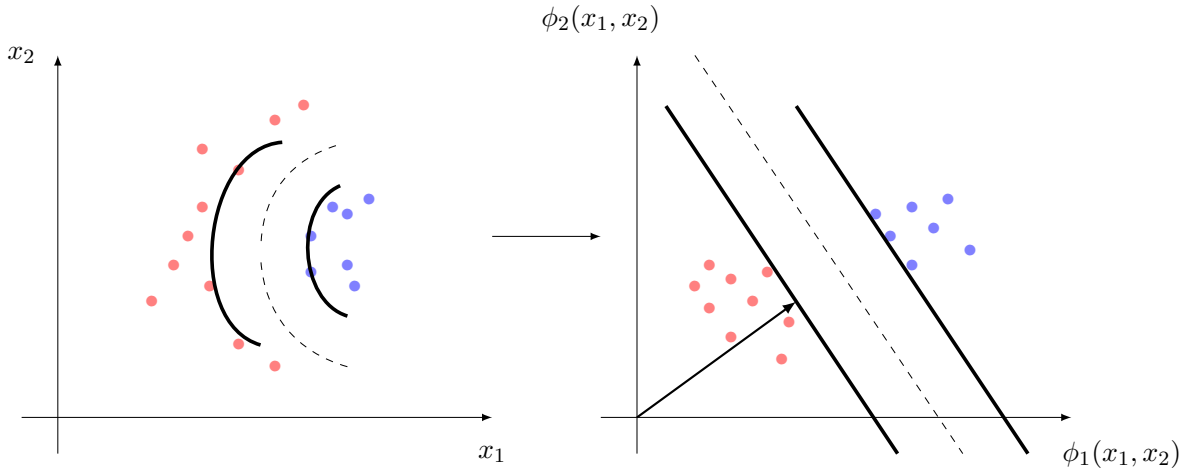


FIGURE 3.3: Illustration of the construction of an induced feature space in which classes are linearly separable.

In problem (3.7), the constant C tunes the trade-off between the margin width and the misclassifications. This problem is known as the soft-margin problem, meaning that the margin will allow misclassifications, as opposed to the hard-margin problem. Let us notice that soft-margin SVMs always have a solution and are more robust to outliers, while hard-margin SVMs have the advantage of being non-parametric. From now on and for the sake of generality, we will only consider soft-margin SVMs.

The kernel trick and Mercer's condition

To go further, it is possible to slightly modify the problem to allow nonlinear separation in the feature space spanned by the components of the data points $\mathbf{x}$. Indeed, let us consider the situation depicted in Figure 3.3. In the original feature space, classes are not linearly separable. Let us suppose that we can find a nonlinear mapping

$$\Phi(\mathbf{x}) = \begin{bmatrix} \phi_1(x_1, x_2) \\ \phi_2(x_1, x_2) \end{bmatrix} \quad (3.8)$$

that induces a new feature space in which classes become linearly separable.

The soft-margin SVM optimization problem then becomes:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & ||\mathbf{w}\|_2^2 + C \sum_{i=1}^N \xi^i \\ \text{subject to} \quad & y^i(\mathbf{w}^\top \Phi(\mathbf{x}^i) + b) \geq 1 - \xi^i, \quad \forall i \in \{1, 2, \dots, N\} \\ & \xi^i \geq 0, \quad \forall i \in \{1, 2, \dots, N\} \end{aligned} \quad (3.9)$$

In general, the induced feature space can have any number of dimensions (possibly an infinity). Intuitively, the higher the dimension of the induced space, the more chances we have to find a linear separation in this space. Unfortunately, finding a suitable (possibly high-dimensional) mapping Φ may turn out to be non trivial and expensive to compute. To address this issue, let us consider the dual formulation of problem (3.9):

$$\begin{aligned}
\min_{\alpha^i} \quad & \frac{1}{2} \sum_{i,j=1}^N \alpha^i \alpha^j y^i y^j (\Phi^\top(\mathbf{x}^i) \Phi(\mathbf{x}^j)) - \sum_{i=1}^N \alpha^i \\
\text{subject to} \quad & 0 \leq \alpha^i \leq C, \forall i \in \{1, 2, \dots, N\} \\
& \sum_{i=1}^N \alpha^i y^i = 0
\end{aligned} \tag{3.10}$$

The important thing to notice in problem (3.10) is that the mapping Φ only appears implicitly through a scalar product:

$$K(\mathbf{x}^i, \mathbf{x}^j) = \Phi^\top(\mathbf{x}^i) \Phi(\mathbf{x}^j) \tag{3.11}$$

In equation (3.11), we introduced the notation $K(\mathbf{x}^i, \mathbf{x}^j)$ which defines a function K called a kernel and which can be understood as a symmetrical measure of similarity between its arguments. This function computes the inner product of the representation of two data points in the induced feature space. Hence, since solving problem (3.10) is equivalent to solving problem (3.9), if we know the value of $K(\mathbf{x}^i, \mathbf{x}^j)$, we do not need to know explicitly the mapping Φ . This is referred to as the "kernel trick."

It turns out that many kernels exist in the literature. However, not every function K cannot be used as a kernel. A simplified version of the famous theorem proposed by James Mercer in [52] states that if the $N \times N$ symmetric matrix $\mathbf{K}$ defined by

$$[\mathbf{K}_{ij}]_{i,j=1}^N = [K(\mathbf{x}^i, \mathbf{x}^j)]_{i,j=1}^N \tag{3.12}$$

is positive semidefinite (i.e., all its eigenvalues are nonnegative), then there exists a corresponding mapping Φ such that equation (3.11) holds. This condition is essential to ensure the convexity of problem (3.10).

Dealing with unbalanced classes

In many real-world datasets, it is common for one class to be predominantly represented. For example, in classification tasks aiming at separating healthy people and people suffering from a particular disease, it is common that the dataset contains more healthy than sick subjects. If we consider the case of 90% of the individuals in the dataset being healthy, then a naive binary classifier whose outcome is always "healthy" would be right in 90% of cases. This is known as the class imbalance problem.

To address this problem, a simple idea is to penalize more severely misclassifications related to the minority class. Given a dataset of N points, let $\mathcal{N}$ and $\mathcal{P}$ denote the sets of negative and positive data points respectively. Thus, we have $|\mathcal{N}| + |\mathcal{P}| = N$. The soft-margin SVM formulation (3.9) becomes

$$\begin{aligned}
\min_{\mathbf{w}, b} \quad & \|\mathbf{w}\|_2^2 + C \left(\beta^- \sum_{i \in \mathcal{N}} \xi^i + \beta^+ \sum_{i \in \mathcal{P}} \xi^i \right) \\
\text{subject to} \quad & y^i (\mathbf{w}^\top \Phi(\mathbf{x}^i) + b) \geq 1 - \xi^i, \quad \forall i \in \{1, 2, \dots, N\} \\
& \xi^i \geq 0, \quad \forall i \in \{1, 2, \dots, N\}
\end{aligned} \tag{3.13}$$

By setting the parameters β^- and β^+ , we can grant more importance to the misclassification of points lying in the minority class.

In the context of connectomes classification, the data points are networks and the classes correspond to labels of interest that will be described in section 3.2. In the rest of this section, we propose several strategies either based on classical kernels built from a vector of topological/spectral features describing the connectomes, or based on specialized kernels that were developed particularly for graph classification.

3.1.2 Classical kernels

The linear kernel is probably the most simple and intuitive kernel since it does not involve any mapping into an induced feature space. It is defined by

$$K(\mathbf{x}^i, \mathbf{x}^j) = (\mathbf{x}^i)^\top \mathbf{x}^j \quad (3.14)$$

Actually, it corresponds to the case of Φ being the identity function, and results in a measure of similarity defined as the scalar product of the feature-based vector representation of two data points.

The main advantage of this kernel is that the components of vector $\mathbf{w}$ in problem (3.9) can be understood here as weights granted to each feature describing the data points. Hence, this allows us to highlight the most discriminative features, and to explain the results of the classification task.

Another famous and widely used kernel is the Gaussian kernel (also known as the radial basis function kernel) defined as

$$K(\mathbf{x}^i, \mathbf{x}^j) = \exp\left(-\frac{\|\mathbf{x}^i - \mathbf{x}^j\|_2^2}{2\sigma^2}\right) \quad (3.15)$$

where σ is a metaparameter to be tuned. In theory, it can be shown¹ that this function defines a mapping into an infinite-dimensional space, which makes it particularly interesting when considering SVMs. Unfortunately, it is no longer possible to identify the most discriminative features after using this kernel, since the optimization is made in a complex induced feature space.

Let us mention one last popular kernel, namely the degree- d polynomial kernel:

$$K(\mathbf{x}^i, \mathbf{x}^j) = ((\mathbf{x}^i)^\top \mathbf{x}^j + c)^d \quad (3.16)$$

In the expression (3.16), c is a metaparameter to be tuned and d is also a metaparameter that is often chosen *a priori*. Polynomial kernels are particularly popular in the natural language processing community (see for example the reference [53] from Goldberg et al.). However, as Gaussian kernels, they do not allow the identification of discriminative features, and thus they do not provide any direct explanation about the output of the classification process.

Topological and spectral feature vector

The first strategy we want to investigate to classify connectomes involves describing a connectome by a feature vector. Hence, the critical step is to choose the topological coefficients

¹By considering the expansion of the exponential function into an infinite series

that are the best suited to describe a connectome. In other words, these coefficients will constitute the fingerprint of a connectome. This approach was investigated by, among others, Li et al. in [54].

There exists a plethora of topological and spectral coefficients that can describe a network. We selected 15 of them that were already reported in [54] and [6], and whose implementation was either available in the Brain Connectivity Toolbox or straightforward to code. Here is the list of these coefficients (their mathematical definitions are available in Appendix B.1):

Topological and spectral features for graph classification:

- | | |
|-----------------------------------|--------------------------|
| 1. Mean degree | 9. Diameter |
| 2. Mean strength | 10. Modularity |
| 3. Density of edges | 11. Assortativity |
| 4. Average clustering coefficient | 12. Diffusion efficiency |
| 5. Transitivity coefficient | 13. Spectral radius |
| 6. Characteristic path length | 14. Frobenius norm |
| 7. Global efficiency | 15. Energy |
| 8. Radius | |

Then, we suggest using a linear kernel to allow us highlighting the most discriminative coefficients of this feature vector. The first strategy is summarized in Box 3.1.

Box 3.1: Strategy 1 - Topological/spectral feature vector

- Compute the topological/spectral coefficients described in Appendix B.1 to build a feature vector
- Use this vector to feed an SVM with linear kernel

Bag-of-edges and the curse of dimensionality

It is known that we can completely describe a graph using its weighted adjacency matrix. From this statement, it would seem obvious to vectorize the adjacency matrix and to use the result as a feature vector, called a bag-of-edges, whose components are the weights of the edges. Following this idea, the dimension of the resulting feature vector associated with a connectome of our dataset would be $\frac{84(84-1)}{2} = 3486$, since our connectomes have 84 nodes and are undirected (symmetric adjacency matrix). This would result in a feature vector whose dimension is much larger than the number of samples in the dataset (204 in our case). As a consequence, there is a high risk of overfitting, a well-known phenomenon in the machine learning community.

Overfitting occurs when a machine learning algorithm fits the training samples too precisely. Then, once exposed to unseen samples in a test set, the algorithm fails to generalize to these unseen samples, resulting in very poor performances. Overfitting particularly occurs when the dimension of the feature vectors is very high, compared to the size of the dataset. Hence, several methods were developed to reduce the dimension of the feature vector by

selecting the best features. This procedure is referred to as the feature selection. Here, we suggest the use of a specific feature selection method that is particularly convenient when using SVMs.

As discussed before, SVMs with linear kernel grant weights to features, according to their strength in terms of discrimination between classes. A procedure called Recursive Features Elimination (RFE), initially developed by Guyon et al. in [55], uses this property of linear SVMs to select the best features. Recursively, a linear SVM is trained with a given training set. After the training phase, we select a given percentage of the best features, according to the granted weights, and we ignore the others. Then, the process restarts on the same training set, with the remaining features. We stop the RFE when a given number of features is reached.

This leads us to our second strategy, summarized in Box 3.2. We propose the use of the weights of the edges as features in combination with an RFE scheme to prevent overfitting. Again, as the SVM uses a linear kernel, this strategy allows us to detect the most discriminative features; in our case, the most discriminative edges.

Box 3.2: Strategy 2 - Bag-of-edges + Recursive Features Elimination

- Build a feature vector from the weighted adjacency matrix
- Apply Recursive Features Elimination to get the 100 most discriminative edges
- Feed a linear SVM with the reduced feature vector

3.1.3 Graph kernels

Graphs are very particular objects. Their structure makes them difficult to describe precisely using a feature vector of reasonable size. From this statement, many researchers have begun to develop kernels especially designed for graph classification. Overall, the main idea is common to all these kernels: graphs that share many topological *patterns* are defined as similar. The key here is to select the patterns of interest.

For instance, in [56], Gärtner et al. propose to compare the presence of random walks of different length between graphs. In [57], Shervashidze et al. choose to count the number of graphlets of size 3, 4 or 5. A graphlet of size k is any subgraph composed of k nodes. In addition, Shervashidze et al. suggest in [58] to consider subtrees of height h . A (rooted) subtree is a subgraph containing a "root" node and no cycle. The height of a subtree is the maximum path length between its root and any of its other nodes, plus one.

We briefly present here the theoretical principles underlying two of the most common graph kernels.

Random walk kernel

Random walk kernels actually gather several different kernels. To measure the similarity between two graphs, the idea is to perform random walks on both graphs and to count the number of matching walks. We introduce here a particular method to compute such a kernel. This method was initiated by Gärtner et al. in [56] and taken over by Vishwanathan et

al. in [59].

First, we must introduce the notion of a direct product graph. Let us consider two graphs $G_1(\mathcal{V}_1, \mathcal{E}_1)$ and $G_2(\mathcal{V}_2, \mathcal{E}_2)$ with respective vertex sets $\mathcal{V}_1$ and $\mathcal{V}_2$, and edge sets $\mathcal{E}_1$ and $\mathcal{E}_2$. The vertex set of the direct product graph $G_\times$ associated to G_1 and G_2 is given by

$$\mathcal{V}_\times = \{(v_i, v_k) | v_i \in \mathcal{V}_1, v_k \in \mathcal{V}_2\} \quad (3.17)$$

Similarly, the edge set of $G_\times$ is given by

$$\mathcal{E}_\times = \{((v_i, v_k), (v_j, v_l)) | (v_i, v_j) \in \mathcal{E}_1 \wedge (v_k, v_l) \in \mathcal{E}_2\} \quad (3.18)$$

In other words, each node in $G_\times$ corresponds to a pair of nodes from G_1 and G_2 , and there exists an edge between two such nodes if the corresponding edges respectively exist in G_1 and G_2 . If $\mathbf{A}_1$ and $\mathbf{A}_2$ are the respective (binary) adjacency matrices of G_1 and G_2 , then we have

$$\mathbf{A}_\times = \mathbf{A}_1 \otimes \mathbf{A}_2 \quad (3.19)$$

where $\mathbf{A}_\times$ is the adjacency matrix of $G_\times$ and $\otimes$ denotes the Kronecker product. We can show that performing a random walk on $G_\times$ is equivalent to performing a random walk on both G_1 and G_2 simultaneously. This is why it is convenient to use the direct product graph of G_1 and G_2 . Similarly, we denote the weight matrix of the direct product graph $\mathbf{W}_\times$. Then, the random walk kernel between G_1 and G_2 is given by

$$K(G_1, G_2) = \sum_{k=0}^{\infty} \mu(k) \mathbf{q}_\times^T \mathbf{W}_\times^k \mathbf{p}_\times \quad (3.20)$$

where k denotes the length of the performed random walk, the function $\mu(k)$ gives the "weight" (i.e. the importance) granted to a random walk of length k and $\mathbf{p}_\times$ and $\mathbf{q}_\times$ are vectors whose components give, for each node, the probability that a walk respectively starts or stops from this node. In general, these probability distributions are chosen uniform. It is worth noticing that the successive powers of $\mathbf{W}_\times$ encode walks of different length on the direct product graph $G_\times$. If we consider the special case of $\mu(k) = \lambda^k$, $\lambda > 0$, we can rewrite equation (3.20) as

$$K_{RW}(G_1, G_2) = \mathbf{q}_\times^T (\mathbf{I} - \lambda \mathbf{W}_\times)^{-1} \mathbf{p}_\times \quad (3.21)$$

As explained in [59], we can compute the similarity defined in equation (3.21) by first solving the linear system

$$(\mathbf{I} - \lambda \mathbf{W}_\times) x = \mathbf{p}_\times \quad (3.22)$$

with a conjugate gradient method, and then computing the quantity $\mathbf{q}_\times^T x$.

Weisfeiler-Lehman kernel

The Weisfeiler-Lehman kernel is one of the most popular graph kernels. Initially developed by Shervashidze et al. in [60], it is based on the Weisfeiler-Lehman test of isomorphism between two graphs whose nodes are labelled, and considers the local substructure of each node. If the nodes are unlabelled, then we can simply consider that their label corresponds to their degree. The idea is to iteratively build a feature vector describing a graph. The components of this vector correspond to node labels. At each iteration, a new label is attributed

to define the nodes, they share the identity of their vertices. We will refer to this as the perfect nodes correspondence property. Comparing connectomes then amounts to comparing their (weighted) edges. We now propose two new kernels taking advantage of the perfect nodes correspondence property.

DeltaCon kernel

This kernel uses the similarity measure between graphs having perfect nodes correspondence introduced by Koutra et al. in [61]. Given two graphs G_1 and G_2 to compare, we can summarize the concept behind this similarity measure in three key steps:

DeltaCon [61] similarity measure between two graphs G_1 and G_2 :

1. Compute a node-to-node affinity matrix for both G_1 and G_2
2. Compute the distance d_{12} between the affinity matrices of G_1 and G_2
3. To get a similarity measure between 0 and 1, compute the quantity $\text{sim}(G_1, G_2) = \frac{1}{1+d_{12}}$

Node-to-node affinity measure In [61], the authors suggest the use of the Fast Belief Propagation (FaBP) affinity measure introduced by Koutra et al. in [62], although other such measures could be used without loss of generality. This measure of affinity between the nodes of a single graph relies on the assumption that nodes can influence each other. To understand this concept, let us consider a social network whose individuals are either "guilty" or not. If one individual has many guilty friends, it is likely to assume that this individual is also guilty. This is also referred to as the "guilt-by-association" principle.

In [62], the authors show that the original Belief Propagation problem, as described by Yedidia et al. in [63], can be approximated² by the following linear system:

$$[\mathbf{I} + a\mathbf{D} - c\mathbf{A}] \mathbf{b}_h = \phi_h \quad (3.27)$$

$$a = \frac{4h_h^2}{1 - 4h_h^2} \quad (3.28)$$

$$c = \frac{2h_h}{1 - 4h_h^2} \quad (3.29)$$

In equation (3.27), $\mathbf{I}$ denotes the identity matrix, $\mathbf{D}$ is the diagonal matrix of the degrees of the nodes, $\mathbf{A}$ is the (binary) adjacency matrix of the graph under study, ϕ_h is the vector of "prior beliefs" of the nodes and $\mathbf{b}_h$ is the vector of "final beliefs" of the nodes. In equations (3.28) and (3.29), h_h is a small constant that tunes the influence of the neighborhood of a node. When solving equation (3.27) for $\mathbf{b}_h$, we see that the matrix

$$\mathbf{S} = [\mathbf{I} + a\mathbf{D} - c\mathbf{A}]^{-1} \quad (3.30)$$

defines a pairwise measure of affinity between nodes of a graph.

²The proof mainly relies on the MacLaurin expansion of some terms of the equations governing the problem. The details, available in Appendices A and B of [62], are left to the reader.

Distance between matrices Among the different distance measures between matrices, the authors of [61] suggest to use the Jeffries-Matusita distance, applied to matrices³:

$$d_{JM}(\mathbf{S}_1, \mathbf{S}_2) = \left(\sum_i \sum_j \left((s_{1,ij})^{1/p} - (s_{2,ij})^{1/p} \right)^p \right)^{1/p} \quad (3.31)$$

Without loss of generality, we can choose $p = 2$ and get a distance measure closely related to the Euclidean distance:

$$d_{JM}(\mathbf{S}_1, \mathbf{S}_2) = \sqrt{\sum_i \sum_j (\sqrt{s_{1,ij}} - \sqrt{s_{2,ij}})^2} \quad (3.32)$$

The nodes' affinities $s_{.,ij}$ being between 0 and 1, the inner square roots in equation (3.32) are used to somehow "boost" them.

Finally, we define the DeltaCon kernel between two graphs G_1 and G_2 as

$$K_{DC}(G_1, G_2) = (1 + d_{JM}(\mathbf{S}_1, \mathbf{S}_2))^{-1} \quad (3.33)$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ are computed according to equation (3.30). Our third classification strategy is summarized in Box 3.3.

Box 3.3: Strategy 3 - DeltaCon kernel

- Build the Fast Belief Propagation affinity matrix for each connectome (equation (3.30))
- Build a kernel matrix by computing the quantity defined in equation (3.33) for each pair of connectomes
- Feed an SVM with this precomputed kernel matrix

Properties of the DeltaCon kernel As discussed by the authors of [61] (see the reference for illustrated examples), the DeltaCon similarity measure exhibits several important properties in accordance with intuition. Let us consider a graph G and a slightly modified version of G , namely $\tilde{G}$. We assume that G and $\tilde{G}$ are differentiated only by a few changes. First, it turns out that changes creating disconnected components in $\tilde{G}$ are more severely penalized by the DeltaCon measure than others. This is a direct consequence of the definition of the node-to-node affinity score.

Additionally, a specific change between G and $\tilde{G}$ will have a stronger influence on the DeltaCon similarity score if G and $\tilde{G}$ are sparse (i.e., they contain a small number of edges) than otherwise.

Finally, let us notice that when looking at equation (3.30), the generalization to a weighted adjacency matrix is straightforward. It turns out that a change of connectivity between G and $\tilde{G}$ will have a greater impact on the DeltaCon measure if the weight of the edges responsible for the change is large. For the rest of this section, we will always include the weights in our computation. Moreover, we smooth them with the logarithm function as recommended by Koutra et al. in the updated version of their article [64].

³This quantity was originally designed to measure the distance between two probability distributions.

Kernel based on the Regularized Laplacian

Motivated by the previous approach, we want to investigate the influence of the node-to-node affinity measure on the quality of the kernel. In fact, the measure of pairwise node affinity can be viewed as a kernel *on* a graph, as opposed to a kernel *between* graphs. In the literature, there exists a wide variety of kernels on a graph, especially in the semi-supervised learning field. Among them, we choose the Regularized Laplacian.

Given a graph G , its adjacency matrix $\mathbf{A}$ and its diagonal matrix of node degrees $\mathbf{D}$, the Laplacian matrix $\mathbf{L}$ of G is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad (3.34)$$

This matrix has several important properties. In particular, its spectrum can bring useful information about the structure of the graph. For convenience, we often use a variant of the Laplacian matrix called the normalized Laplacian matrix:

$$\mathcal{L} = \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} \quad (3.35)$$

We then introduce the Regularized Laplacian matrix:

$$\mathbf{S}_{RL} = (\mathbf{I} + \alpha \mathcal{L})^{-1} \quad (3.36)$$

In equation (3.36), $\mathbf{I}$ denotes the identity matrix and α is a parameter of the method. There exist many derivations and interpretations of the Regularized Laplacian. For example, Smola et al. present in [65] an elegant and general regularization framework leading to the Regularized Laplacian, and showing the links existing between the latter and the diffusion process on a graph.

Hence, by analogy to equation (3.33), we define the kernel based on the Regularized Laplacian between two graphs G_1 and G_2 as

$$K_{RL}(G_1, G_2) = (1 + d_{JM}(\mathbf{S}_{RL,1}, \mathbf{S}_{RL,2}))^{-1} \quad (3.37)$$

Our last classification strategy is summarized in Box 3.4.

Box 3.4: Strategy 4 - Kernel based on the Regularized Laplacian

- Build the Regularized Laplacian matrix for each connectome (equation (3.36))
- Build a kernel matrix by computing the quantity defined in equation (3.37) for each pair of connectomes
- Feed an SVM with this precomputed kernel matrix

3.2 Results

In the previous section, we presented the type of kernel-based algorithms we use for classification, namely Support Vector Machines. We also introduced four strategies for the classification of connectomes. Boxes 3.1, 3.2, 3.3 and 3.4 summarize them. In addition, we explained the principles underlying two state-of-the-art graph kernels: the random walk

TABLE 3.1: Spectral range of the proposed kernel matrices, on our dataset composed of 204 connectomes.

Method	$[\lambda_{min}, \lambda_{max}]$
DeltaCon kernel	[0.42, 65.23]
Kernel based on the Regularized Laplacian, $\alpha = 5$	[0.20, 134.3]
Kernel based on the Regularized Laplacian, $\alpha = 10$	[0.17, 139.97]
Kernel based on the Regularized Laplacian, $\alpha = 15$	[0.16, 141.6]

and the Weisfeiler-Lehman kernels.

In this section, we test these classification strategies with respect to four binary labels of interest: gender (male, female), handedness (left-handed, right-handed), alcohol abuse (positive, negative) and smoking (smoker, non-smoker). The aim is to train an SVM and to test its ability to classify connectomes with respect to these labels.

The metadata related to handedness, alcohol abuse and smoking are part of the restricted data from the Human Connectome Project [34]. For privacy reasons, we do not report here results about isolated subjects. We used the Scikit-learn PYTHON toolbox⁴ for the aspects related to Support Vector Machines, Recursive Feature Elimination and the performance evaluation procedure. The MATLAB implementation of both random walk and Weisfeiler-Lehman kernels are publicly available on the internet⁵. The feature vector of Strategy 1 (Box 3.1) was computed thanks to the Brain Connectivity Toolbox⁶. The DeltaCon kernel was computed using the code publicly available on the website⁷ of the first author of [61]. The kernel based on the Regularized Laplacian was coded in MATLAB.

In the rest of this section, we begin by presenting a preliminary analysis of our methods, our testing procedure and the performance evaluation metrics. Then, we present our results for each selected label of interest.

3.2.1 Preliminary analysis, testing procedure and performance evaluation

Kernel positive semidefiniteness

Before evaluating our methods, we need to assess their consistency. In particular, as mentioned in Section 3.1, we must check the positive semidefiniteness of our kernels in order to ensure the convexity of problem (3.10). As a reminder, a kernel matrix is positive semidefinite if and only if its eigenvalues are positive. We report the range of the (real) spectrum of the kernel matrices associated with our proposed methods in Table 3.1. We observe that, for our dataset, the proposed methods give rise to positive semidefinite kernel matrices.

Label redundancy and class imbalance

In the analysis of our results, we must take care of label redundancy. Let us consider a dataset composed of alcoholic people and abstinent people (i.e., that do not drink any alcohol). Let us imagine that we train a classifier with respect to two binary labels: 'Alcoholic

⁴<http://scikit-learn.org/>

⁵<http://www.mit.edu/~pinary/kdd/>

⁶<https://sites.google.com/site/bctnet/>

⁷<http://web.eecs.umich.edu/~dkoutra/#pub>

TABLE 3.2: Labels redundancy LR , with respect to our dataset composed of 204 connectomes. (Equation (3.38))

	Gender	Handedness	Alcohol abuse	Smoking
Gender	1	0.52	0.39	0.62
Handedness	-	1	0.23	0.67
Alcohol abuse	-	-	1	0.25
Smoking	-	-	-	1

TABLE 3.3: Class imbalance for each tested label, with respect to our dataset composed of 204 connectomes.

Label	Proportion	Class imbalance
Gender	91 - 113	Slightly unbalanced
Handedness	19 - 185	Strongly unbalanced
Alcohol abuse	42 - 162	Unbalanced
Smoking	54 - 150	Unbalanced

- Abstinent' and 'Regular alcohol consumer - Abstinent'. If the classifier succeeds to discriminate alcoholic and abstinent individuals, it will also succeed to classify regular alcohol consumer and abstinent subjects. Indeed the following implication holds: 'Alcoholic $\Rightarrow$ Regular alcohol consumer'. However, the contraposition is false. Hence, the statement "The classifier is able to detect regular alcohol consumers" will be wrong in this case. As illustrated by this example, label redundancy is conditioned by the dataset.

To quantify the redundancy between two labels l_1 and l_2 , we propose the following metric:

$$LR(l_1, l_2) = \frac{\text{\# of instances where } l_1 \text{ and } l_2 \text{ are in agreement}}{\text{Total \# of instances}} \quad (3.38)$$

Let us notice that $RL(l_1, l_2) = 1$ means that l_1 and l_2 are totally redundant and $RL(l_1, l_2) = 0$ means that l_1 and l_2 are totally anti-redundant. In these two cases, a given classifier will in fact solve exactly the same problem for both l_1 and l_2 . Table 3.2 reports the labels redundancy for our dataset, with respect to each label we will test.

As mentioned in Section 3.1, we also need to be aware of the class imbalance problem when analyzing our results. Table 3.3 reports, for each label, the repartition of the instances between the two classes.

Training and test phases

To evaluate the performance of our methods, we must follow a procedure in two phases. First, in the case where a method is parametric, we must select the best parameters. For each method, the values we tested for its parameters are given in Table 3.4. Then, once the parameters are fixed, we can test the performance of a method on unseen data.

We begin by splitting our dataset into a training set and a test set, as illustrated in Figure 3.5. In our case, we choose a test set representing 20% of the dataset. During the training

TABLE 3.4: Parameters of the methods.

Method	Parameter in equation (ref)	Value(s)
All methods	C in (3.9)	$C \in \{10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$
All methods	(β^-, β^+) in (3.13)	$(\beta^-, \beta^+) \in \left\{ \begin{array}{l} (3, 1), (3, 0.5), (2, 1), \\ (2, 0.5), (1, 1), (1, 2), \\ (0.5, 2), (1, 3), (0.5, 3) \end{array} \right\}$
Strategy 1 (Box 3.1)	/	/
Strategy 2 (Box 3.2)	Number of features to keep	100
Strategy 3 (Box 3.3)	h_h in (3.28, 3.29)	Upper bound derived in [62]
Strategy 4 (Box 3.4)	α in (3.36)	$\alpha \in \{5, 10, 15\}$
Random Walk kernel	λ in (3.21)	$\lambda \in \{10^{-5}, 10^{-3}, 10^{-1}\}$
Weisfeiler-Lehman kernel	k in (3.26)	$k \in \{0, 1, 2, 3, 4, 5\}$

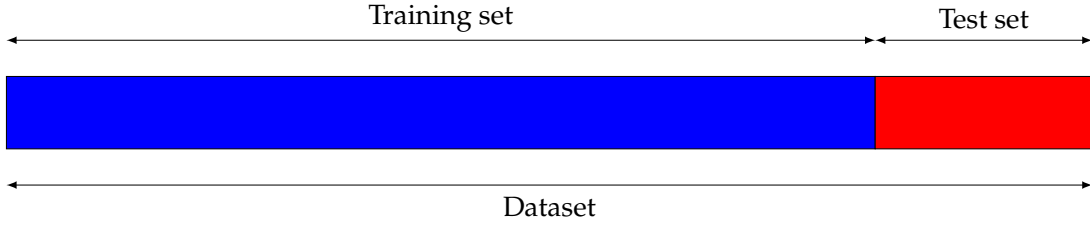


FIGURE 3.5: Splitting of the dataset into a training set and a test set.

phase, we split the training set into five parts of equal size. For each value of a parameter, we alternately test an SVM on one of these five parts (the validation set), after having trained it on the rest of the training set. This is called a 5-fold cross-validation (Figure 3.6). The parameter value giving the best score, averaged on the five independent tests, is kept for the test phase.

During the test phase, we feed the previously trained SVM with the test set and get a classification score. It is important to clearly separate the training set from the test set in order to check whether a method is able to generalize to unseen data or not. Furthermore, it is often recommended to repeat both training and test phases several times and to average the classification scores, in order to get more realistic results. We choose to repeat the procedure 30 times, with a random shuffling of the entire dataset before each execution.

Performance metrics

There exist many metrics allowing us to evaluate the performances of a classifier. To understand them, let us introduce the concept of a confusion matrix, illustrated in Figure 3.7. We consider a binary classification task where data points are either positive or negative. For a given trained classifier, the confusion matrix gives the repartition of the classifier predictions with respect to the actual (true) classes of the data points lying in the test set. When looking at this matrix, we can identify the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN).



FIGURE 3.6: Illustration of the 5-fold cross-validation procedure. Each line represents an independent test.

The classification accuracy is defined as follows:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (3.39)$$

$$\triangleq p_0 \quad (3.40)$$

In other words, it corresponds to the proportion of correct predictions made by the classifier. This metric is probably the most intuitive. However, it may not be the best suited for certain classification tasks, especially when classes are unbalanced. For instance, in a dataset containing 90% of positive and 10% of negative labels, a naive classifier whose outcome is always "positive" would achieve an accuracy of 90%. The confusion matrix corresponding to this example is depicted in Figure 3.8. In fact, the accuracy neglects the influence of the off-diagonal terms of the confusion matrix. The class imbalance problem was already addressed in the SVM formulation (3.13), thus for the training phase. We now need to take it into account in our performance evaluation during the test phase.

By denoting $n = TP + FP + FN + TN$, the number of data points used to evaluate the performances of the classifier (i.e., in the test set), let us introduce the following quantities:

$$p_c = \left(\frac{TP + FP}{n} \right) \left(\frac{TP + FN}{n} \right) + \left(\frac{FN + TN}{n} \right) \left(\frac{FP + TN}{n} \right) \quad (3.41)$$

$$\kappa = \frac{p_0 - p_c}{1 - p_c} \quad (3.42)$$

The latter quantity, κ , is known as Cohen's kappa, initially introduced by Jacob Cohen in [66], and belonging to $[-1, 1]$. The quantity p_c can be understood as the probability of random agreement between actual and predicted labels. An interpretation of Cohen's kappa provided by Landis and Koch in [67] is reported in Table 3.5. But why should we use Cohen's kappa instead of classification accuracy? Let us come back to our example of a naive classifier, whose confusion matrix is given in Figure 3.8. For this example, the classification accuracy is $p_0 = 0.9$, whereas Cohen's kappa is $\kappa = 0$. Hence, we see that Cohen's kappa can handle the class imbalance problem.

There is no consensus about the performance metric to use for classification tasks, as discussed by David Powers in [68]. In our results, we report both classification accuracy and Cohen's kappa.

	Actual: positive	Actual: negative
Predicted: positive	True Positive (TP)	False Positive (FP)
Predicted: negative	False Negative (FN)	True Negative (TN)

FIGURE 3.7: Confusion matrix.

TABLE 3.5: One interpretation of Cohen's kappa. (From [67])

κ	Strength of agreement
< 0.00	Poor
0.00 – 0.20	Slight
0.20 – 0.40	Fair
0.40 – 0.60	Moderate
0.60 – 0.80	Substantial
0.80 – 1.00	Almost perfect

3.2.2 Discrimination with respect to gender

The classification results for this label are reported in Table 3.6. We can observe that Strategy 2 is the best for both performance metrics, and that Strategies 3 and 4 are very competitive with respect to state-of-the-art graph kernels. When looking at Cohen's kappa scores, we see that the class imbalance is correctly handled by the SVM.

3.2.3 Discrimination with respect to handedness

The classification results for this label are reported in Table 3.7. This label, due to its strongly unbalanced classes in the dataset, is the perfect illustration of why we need to consider other performance metrics than accuracy. We see that this latter is very high, while Cohen's kappa is almost always zero. The SVM in this case chooses to always declare an individual as being right-handed, and totally fails to detect differences between right and left-handed. When the classes are strongly unbalanced, the modified SVM formulation (3.13) does not counter the imbalance⁸. Fortunately, Cohen's kappa allows us to detect the misbehaviour of our method.

3.2.4 Discrimination with respect to alcohol abuse

This label was defined according to the definition of alcohol abuse provided by the Diagnostic and Statistical Manual of Mental Disorders (4th edition), also known as "DSM-IV-TR" and published by the American Psychiatric Association. A subject was considered "positive"

⁸At least, with the parameters we tested (Table 3.4).

	Actual: positive	Actual: negative
Predicted: positive	90	10
Predicted: negative	0	0

FIGURE 3.8: Example of a confusion matrix for a naive classifier.

TABLE 3.6: Classification results for the label **Gender**, and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.

	Strategy 1	Strategy 2	Strategy 3	Strategy 4	RW	WL
Accuracy [%]	66.99 $\pm$ 7.4	83.09 $\pm$ 5.5	75.37 $\pm$ 5.7	77.32 $\pm$ 6.9	55.80 $\pm$ 7.7	62.20 $\pm$ 5.7
100 $\times$ κ [–]	35.45 $\pm$ 12.6	66.92 $\pm$ 11.2	47.95 $\pm$ 10.4	52.80 $\pm$ 12.7	19.41 $\pm$ 13.5	19.89 $\pm$ 13.9

TABLE 3.7: Classification results for the label **Handedness**, and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.

	Strategy 1	Strategy 2	Strategy 3	Strategy 4	RW	WL
Accuracy [%]	91.63 $\pm$ 3.7	90.08 $\pm$ 3.3	91.63 $\pm$ 3.7	91.63 $\pm$ 3.7	91.54 $\pm$ 3.8	91.63 $\pm$ 3.7
100 $\times$ κ [–]	0.00 $\pm$ 0.0	–1.65 $\pm$ 11.2	0.00 $\pm$ 0.0	0.00 $\pm$ 0.0	0.00 $\pm$ 0.0	0.00 $\pm$ 0.0

if he showed one or more symptom(s) reported in this manual.

The classification results for this label are reported in Table 3.8. We can see that Strategy 1 is the best, although its Cohen’s kappa score is low. The definition of the label may be too strict to define two classes of individuals. The most discriminative features according to Strategy 1 are reported in Appendix B.2.

3.2.5 Discrimination with respect to smoking

This label was defined according to the smoking history of the subjects (we only consider tobacco consumption). An individual was labeled "negative" if he had never smoked and "positive" if he tried at least once. Let us notice that this is relatively restrictive, and this definition can impact the results.

The classification results for this label are reported in Table 3.9. We observe that Strategies 1 and 3 are the best in terms of accuracy and Cohen’s kappa score respectively. The most discriminative features according to Strategy 1 are reported in Appendix B.2.

TABLE 3.8: Classification results for the label **Alcohol abuse**, and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.

	Strategy 1	Strategy 2	Strategy 3	Strategy 4	RW	WL
Accuracy [%]	79.51 $\pm$ 5.8	73.33 $\pm$ 5.7	79.51 $\pm$ 5.8	79.35 $\pm$ 5.9	79.51 $\pm$ 5.8	79.43 $\pm$ 5.7
100 $\times$ κ [–]	3.67 $\pm$ 12.3	2.46 $\pm$ 15.4	2.96 $\pm$ 9.8	–1.34 $\pm$ 3.3	–0.17 $\pm$ 4.4	–0.15 $\pm$ 0.8

TABLE 3.9: Classification results for the label **Smoking**, and for the four classification strategies (Boxes 3.1, 3.2, 3.3 and 3.4), the random walk kernel (RW) and the Weisfeiler-Lehman kernel (WL). Mean $\pm$ standard deviation, computed over 30 independent tests.

	Strategy 1	Strategy 2	Strategy 3	Strategy 4	RW	WL
Accuracy [%]	74.23 $\pm$ 5.0	65.53 $\pm$ 6.1	72.60 $\pm$ 6.0	73.98 $\pm$ 5.0	72.36 $\pm$ 5.4	73.98 $\pm$ 5.0
$100 \times \kappa$ [–]	4.66 $\pm$ 11.5	7.61 $\pm$ 15.4	16.56 $\pm$ 13.4	3.88 $\pm$ 11.1	6.31 $\pm$ 11.7	13.77 $\pm$ 18.1

3.3 Discussion

3.3.1 Limitations of the dataset

The main drawback of our dataset is probably its size. Indeed, 204 instances is a relatively small number of instances compared to other state-of-the-art datasets, especially when we want to test several labels. Although the classification with respect to gender and smoking provided encouraging results, growing our dataset would be the first thing to do in further studies. This would reduce the effect of class imbalance and result in a better-trained classifier.

3.3.2 Brain differences related to gender

When looking at Table 3.6, we observe that the SVM succeeds in finding a linear separation between male and female connectomes, in the space spanned by the weights of the edges. Intrigued by this observation, we wanted to investigate the origin of this separation. In Figure 3.9, we depicted the 400 most discriminative edges obtained by Recursive Features Elimination. These edges were traced in the connectome where their normalized weight was the highest. An interesting observation in this Figure is that inter-hemispheric connections are more present in female brain.

Let us consider a connectome $G(\mathcal{V}, \mathcal{E})$. We denote by $\mathcal{E}_{inter}$ the set of inter-hemispheric edges (i.e., linking two nodes lying in opposed hemispheres). The function $w : \mathcal{E} \rightarrow \mathbb{R} : e \rightarrow w(e)$ gives the weight of an edge e . We define the inter-hemispheric ratio as

$$\zeta = \frac{\sum_{e' \in \mathcal{E}_{inter}} w(e')}{\sum_{e \in \mathcal{E}} w(e)} \quad (3.43)$$

Figure 3.10 depicts the inter-hemispheric ratio for every connectome in our dataset. If we consider the red line as a threshold for a classification between male and female connectomes, 75% of the connectomes constituting our dataset are correctly classified. This is not as much as the accuracy reached by Strategy 2 in Table 3.6, but we emphasize here the fact that one single feature, the inter-hemispheric ratio, is able to discriminate male and female connectomes more efficiently than state-of-the-art graph kernels. Similar results about the inter-hemispheric ratio are discussed by Ingalhalikar et al. in [69].

3.3.3 Choice of the classifier

We chose to use the SVM algorithm because of its convenience and its popularity within the machine learning community. Nevertheless, other classifiers exist and could have been used in our study. Among the variety of classifiers existing in the literature, let us begin by

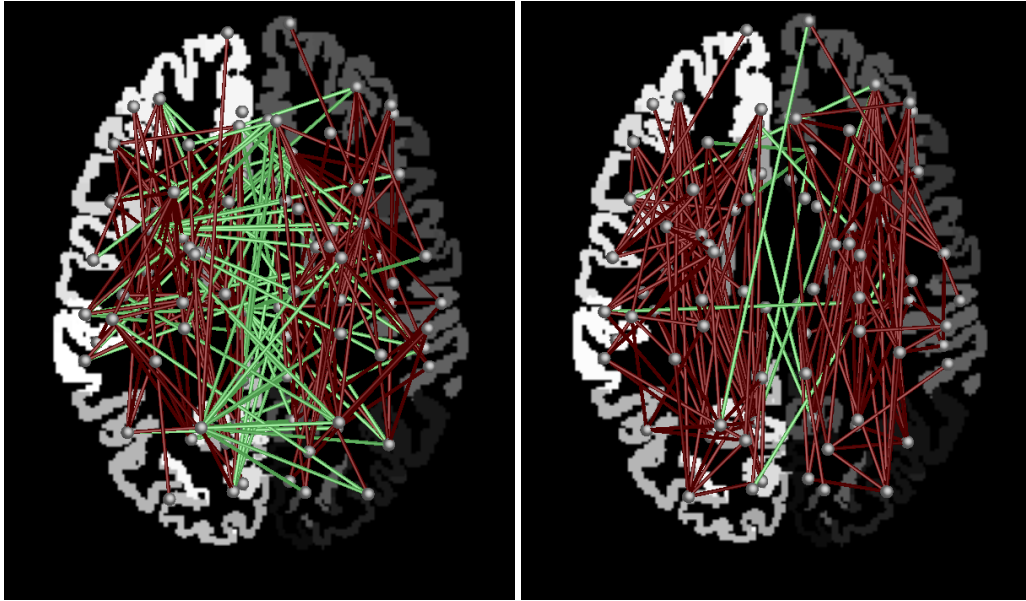


FIGURE 3.9: Left: average female connectome. Right: average male connectome. The 400 most discriminative edges obtained by RFE are depicted in the connectome where their weight is the highest. Inter-hemispheric edges are traced in green, intra-hemispheric edges are traced in red.

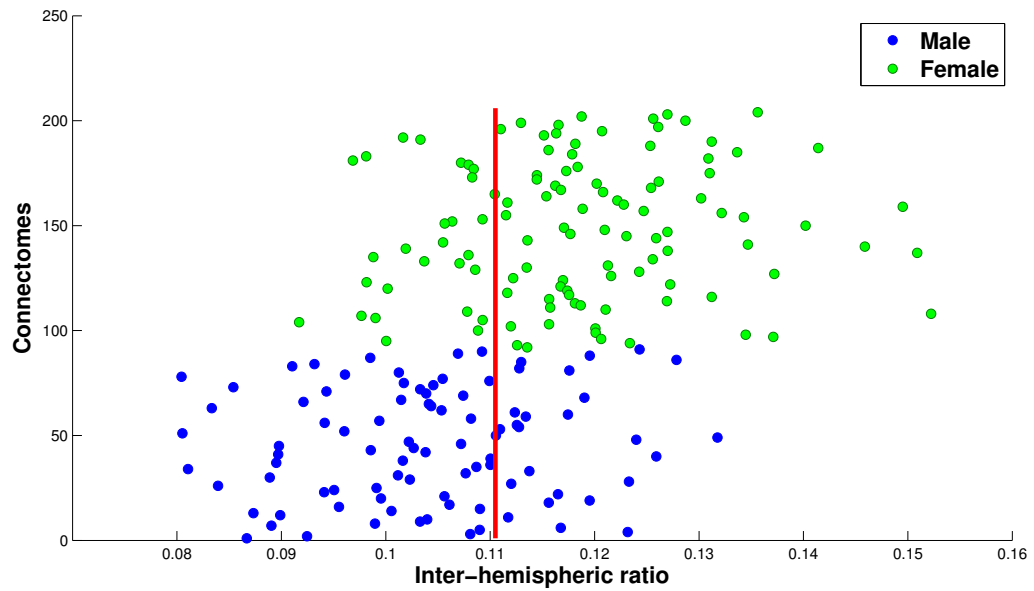


FIGURE 3.10: Scatter plot of the inter-hemispheric ratio for male and female connectomes. Threshold line: $\zeta^* = 0.111$. The expansion along the Y-axis is just for visualization.

mentioning the Random Forest approach, formally introduced by Breiman in [70]. Basically, Random Forests are based on the concept of Decision Trees. A Decision Tree consists of a rooted graph whose nodes correspond to features. Starting from the root, the algorithm makes a decision based on the value of the current feature: either it chooses to output a final decision on the class of a data point, or it chooses to go one step further in the tree to explore other features. The general idea behind Random Forests is to combine several decision trees (whose generation is seeded by a stochastic factor) and to average the results. Recently, Scornet showed in [71] that, by slightly modifying the mathematical definition of Random Forests, it is possible to "kernelize" the method. Hence, we could investigate this method with different graph kernels in future work.

Another popular classification method is the Linear Discriminant Analysis. Based on the work of the famous statistician Ronald Fisher (1890-1962), this method involves concepts from statistics and is closely related to the analysis of variance (ANOVA). Variants of this method include the Quadratic Discriminant Analysis and the kernel Fisher Discriminant Analysis introduced in [72]. As reported by Arbabshirani et al. in [73], this method has already been used in many articles about the prediction of brain disorders.

More intuitive methods based on a similarity measure between data points include the k Nearest Neighbors (k NN) classifier and the Sum-of-Similarities procedure⁹. The former method look for the k classified points closest to the point to classify, according to the similarity measure previously defined. Then, by means of a majority vote within the k nearest neighbors, the class of the new point is determined. The latter method computes the sum of the similarity scores of a point with all points belonging to a given class. This sum is computed for each class, and the class showing the highest sum of similarities is defined as the output of the classifier. Let us notice that these methods are straightforwardly generalizable to multiclass problems, and do not require the positive semidefiniteness condition on the pairwise similarity matrix.

Finally, motivated by the encouraging results of Strategy 2 (Box 3.2), let us mention two last approaches. First, the work by Zalesky et al. [74] aims at identifying brain differences across individuals by means of a test statistic performed on the adjacency matrix of connectomes. Compared to other methods, the so-called Network-Based Statistic looks for the most discriminative connected components in the graph. The second method we would like to mention comes from the PhD thesis of Daehan Won [75]. In his work, Won developed a combinatorial programming approach to detect discriminative edges in complex networks. Furthermore, his method was illustrated on human connectomes and showed encouraging results. These two methods, along with the Recursive Features Elimination procedure introduced in Section 3.1, seek to select the most discriminative components in a set of graphs, given a certain label of interest. The common goal is to sparsify the networks before classifying them. We strongly believe that this line of research has great potential in the field of graph classification.

3.3.4 Choice of the performance metric

The classification accuracy is probably the most often used metric to assess the performances of a classification technique. However, we wanted to show with our results that the accuracy, although intuitive, does not reflect the real behaviour of a classifier. In particular, a classifier designed to optimize the final accuracy could simply choose to ignore one class, if

⁹This method was introduced by Pr. Marco Saeuens in his course "LSINF2275 - Data Mining and Decision Making", taught at the Université Catholique de Louvain (UCL), Belgium.

the classes are unbalanced. Depending on the domain of application, this kind of behaviour could have a dramatic impact on the interpretation of the results.

To remedy this, we proposed to compare the classification accuracy with another performance metric, namely Cohen's kappa. As shown in our example, this metric is able to handle class imbalance. This allows us to get a more objective view of the behaviour of the classifiers. Nonetheless, Cohen's kappa does not achieve unanimity in the scientific community, as it has been denounced by Gwet in [76]. One particular flaw of this metric is that the threshold above which we can declare that a classifier is performant is not clearly defined. The interpretation of Cohen's kappa in Table 3.5 can help to get an idea about the quality of a classifier, but remains dataset dependent.

In the future, other metrics should be investigated. Among them, the F -score (or the F_1 -score) is a common alternative to classification accuracy. We can also define it from the confusion matrix (Figure 3.7):

$$\text{precision} = \frac{TP}{TP + FP} \quad (3.44)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3.45)$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (3.46)$$

Let us notice that this metric is defined with respect to the number of True Positives. When switching the classes to compute the F_1 -score with respect to the "negative" class, we could get a different score. Indeed, the F_1 -score is known to be asymmetric. Another widely used metric is the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) graph. This graph compares two quantities along several tests:

$$\text{True positive rate} = \frac{TP}{TP + FN} = \text{recall} \quad (3.47)$$

$$\text{False negative rate} = \frac{FP}{FP + TN} \quad (3.48)$$

The higher the AUC, the better the classifier. A very good introduction to ROC analysis is provided by Fawcett in [77].

3.3.5 Choice of the kernel

Due to the popularity of kernel-based methods, many kernels are available in the literature. In this chapter, we chose to introduce two strategies based on a linear kernel, two other strategies based on specialized graph kernels suited for the comparison of graphs showing the perfect nodes correspondence property, and two state-of-the-art graph kernels to compare our results. The use of a linear kernel was motivated by the *a posteriori* explanation of the classification outcome, provided by the weights granted to the features describing the connectomes (see Appendix B.2). The DeltaCon similarity measure has already shown encouraging results in the context of unsupervised learning, as reported in [61]. We thus wanted to investigate its potential in a supervised learning framework. Following this idea, we wanted to study the importance of the node-to-node affinity measure by proposing the kernel based on the Regularized Laplacian.

The choice of a kernel for graph classification is still an open question. Actually, we think that each kernel captures a particular facet of what a graph is. Based on this observation, the Multiple Kernel Learning (MKL) approach should be further studied. MKL suggests using a combination of several kernels in order to build a new kernel. Let $\mathcal{S} = \{\mathbf{K}_i\}_{i=1}^q$ be a set of q kernel matrices. A linear MKL approach consists of building a matrix $\tilde{\mathbf{K}}$ such that

$$\tilde{\mathbf{K}} = \sum_{i=1}^q \gamma_i \mathbf{K}_i \quad (3.49)$$

$$\sum_{i=1}^q \gamma_i = 1 \quad (3.50)$$

In equation (3.49), the coefficients $\gamma_i \geq 0$ must be determined by solving an optimization problem. Several general (not necessarily linear) formulations of this problem and its methods of resolution are provided by Sonnenburg et al. in [78] and by Gönen et al. in [79]. More particularly, in [80], Bach et al. are interested in how to include this optimization problem in the same framework as that of the SVM, by considering a conic combination of kernels. Multiple Kernel Learning was for example applied by Jie et al. [23] in the context of Mild Cognitive Impairment (MCI) diagnosis.

An important aspect that was not developed in our analysis is the computational cost related to kernel-based methods. Actually, this is one of the main flaws of graph kernels, and many studies try to improve the time complexity of existing kernels. For instance, the authors of [59] succeeded in drastically reducing the computational cost of the random walk kernel initially proposed by Gartner et al. [56]. During our experiments, the Weisfeiler-Lehman kernel appeared to be the most efficient in terms of time complexity. However, the authors of [61] also introduce a fast approximate version of the DeltaCon similarity measure. They argue that this version exhibits a linear time complexity, thereby making it particularly suited for comparing large networks. Future work should grant more importance to computational cost in order to compare graph kernels.

In Strategy 1 (Box 3.1), we used 15 topological and spectral features to describe our connectomes from the adjacency matrix. Of course, many other features can be used to describe a network. In [81], de Lange et al. analyze the Laplacian spectrum of brain networks from the macaque and the cat. Interestingly enough, it turns out that the Laplacian spectrum provides meaningful information about the structure of the graph. The authors also suggest a particular definition for the spectral distance between connectomes. This could be investigated in future work about brain networks classification.

Finally, let us reiterate that not all measures of similarity between graphs necessarily define a suitable kernel to feed an SVM. The positive semidefiniteness of the kernel matrix, as required by Mercer's theorem, should not be neglected when designing new kernels. Metzger et al. review in [82] the positive semidefiniteness of several graph kernels and particularly show that the random walk kernel meets Mercer's condition. The Weisfeiler-Lehman kernel, being based on a linear kernel, is also positive semidefinite. Although the DeltaCon kernel and the kernel based on the Regularized Laplacian give rise to positive semidefinite kernel matrices in the special case of our dataset (see Table 3.1), there is no proof (to the best of our knowledge) that this is always the case. To deal with indefinite kernels, Wu et al. introduce on the one hand in [83] several methods allowing us to slightly modify the kernel matrix in order to make it positive semidefinite. They noticed that these modifications can sometimes alter what they call the "semantic" of the kernel, i.e. the meaning (and the quality) of the

similarity measure. On the other hand, in the specific case of SVMs, Luss and d'Aspremont propose in [84] to treat the indefinite kernel matrix as a noisy version of the true positive semidefinite kernel matrix. Hence, in their method, they build a new kernel matrix, positive semidefinite, whose distance from the original kernel matrix is directly penalized in the SVM formulation. The optimization problem then remains convex. As a final remark, let us notice that the positive semidefiniteness of the similarity matrix is not required for some other classifiers such as the Sum-Of-Similarities method or the k -NN classifier.

3.3.6 Interpretation of a classifier decision

In Strategies 1 and 2 (Boxes 3.1, 3.2), we emphasized the fact that the linear kernel allows us to give an explanation about the classification result. Unfortunately, when using the kernel trick, the induced high-dimensional feature space is not explicitly available. Hence, it becomes difficult to understand the choice of a classifier. In the perspective of future clinical applications, classification methods should include an explanation of their decisions or at least provide a coefficient of certainty to inform the user on the reliability of the result.

In a binary classification framework, SVMs provide one single output: the class of the data point under study. No probability of belonging to one class or the other is directly available. SVMs are sometimes referred to as *discrete* classifiers. In [85], Niculescu-Mizil and Caruana examine some techniques allowing us to compute posterior probabilities, based on multiple outputs of a given classifier. Among these techniques, Platt scaling is often privileged in the case of SVMs. Basically, if we consider that the two classes correspond to $y = -1$ or $y = 1$ and if we denote the output of a classifier for a given data point x by $f(x)$, Platt scaling amounts to passing $f(x)$ through a sigmoid function:

$$P(y = 1|f(x)) = \frac{1}{1 + \exp(Af(x) + B)} \quad (3.51)$$

In equation (3.51), A and B are two parameters to be trained via cross-validation, or to be determined by maximizing a likelihood function. Despite the simplicity of this procedure, the authors of [85] explain that, in the case of SVMs, the prediction probabilities suffer from distortion.

To get explanations about classifier decisions, a recent article by Ribeiro, Singh and Guestrin [86] has particularly attracted our attention. In situations where the decision surface of a classifier is complex, as depicted in Figure 3.11, the authors propose to locally explain the decision of the classifier by training an interpretable model (such as the linear SVM) with neighboring data points, in an understandable feature space. The method is generalizable to any classifier and shows encouraging results when applied to image classification via deep learning. The authors also discuss at what extent we can trust a classifier. This question is fundamental at a time when machine learning and some of its "black-box" algorithms are being used to help the medical domain, and will probably be the object of future work.

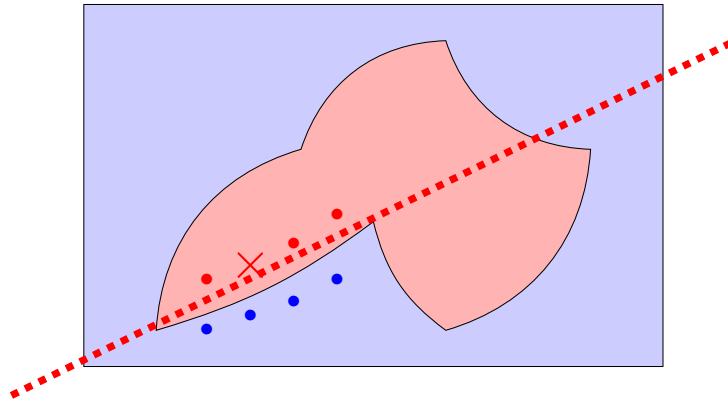


FIGURE 3.11: Local explanation of a classification decision using an interpretable classifier. To explain the classifier decision for the red cross, we can train a linear SVM locally, considering neighboring points in an understandable feature space.

Chapter 4

Controllability of structural brain networks

Given an abstract mathematical representation of the human brain as a complex network, many subfields related to graph theory can contribute to the analysis of the connectome. In particular, we are interested in this chapter in how control theory can help us to understand the relation between brain structure and function. Recent advances in control theory for networked systems open the door to new perspectives in the understanding of cognitive systems and their interactions. In the context of connectomics, controllability refers to the ability for the brain to be driven towards specific cognitive states corresponding to the execution of specific tasks.

Applications of control theoretic results to the human connectome are potentially numerous. Actually, influencing brain activity in order to steer it away from altered states is of great interest in the treatment of some neurological disorders. For instance, Muldoon et al. cite in [87] several examples of successful applications of Transcranial Magnetic Stimulation (TMS) and Deep Brain Stimulation (DBS) to pathologies such as hyperactivity, schizophrenia or Parkinson's disease. In this context, control theory is expected to improve existing methods by unravelling new pathways leading to target brain regions, and by optimizing the amount of energy transmitted to these areas.

In this chapter, we introduce basic notions of controllability for the structural connectome to provide an overview of the field. In particular, we follow the work of Gu et al. presented in [25] and compare their results to that obtained with our dataset.

4.1 Methods

We begin this section by introducing basic concepts of linear control. Then, we show how we can connect this theory to connectomics. Finally, we present different notions of controllability and interpret them in the case of brain networks.

4.1.1 The connectome as a linear system

Formally, a linear system is a system whose dynamics can be described by a set of linear equations. The general formulation of a continuous-time linear system composed of N interacting agents is given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_{\mathcal{K}}\mathbf{u}_{\mathcal{K}}(t), \quad t \geq 0 \quad (4.1)$$

In equation (4.1), $\mathbf{x}$ is the state vector. Each component $x_i(t)$ of $\mathbf{x}(t)$, $i \in \{1, 2, \dots, N\}$, represents the state of agent i at time t . We use the "dot" notation to denote the time derivative: $\dot{\mathbf{x}}(t) = \frac{d\mathbf{x}(t)}{dt}$. The $N \times N$ matrix $\mathbf{A}$ encodes the interactions between agents. We will not consider the case of time-varying or self¹ interactions in this document. Let us assume that we can control m agents. The respective indices of these agents are contained in set $\mathcal{K} = \{k_1, k_2, \dots, k_m\}$. The $N \times m$ matrix $\mathbf{B}_{\mathcal{K}}$ is then defined as

$$\mathbf{B}_{\mathcal{K}} = \begin{bmatrix} \mathbf{e}_{k_1} & \mathbf{e}_{k_2} & \dots & \mathbf{e}_{k_m} \end{bmatrix} \quad (4.2)$$

where $\mathbf{e}_i$ stands for the i -th canonical column vector of dimension N . Finally, the component u_{k_j} , $j \in \{1, 2, \dots, m\}$, of vector $\mathbf{u}_{\mathcal{K}}$ corresponds to the input signal applied to agent k_j . The discrete-time version of equation (4.1) is given by

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_{\mathcal{K}}\mathbf{u}_{\mathcal{K}}(t), \quad t = 1, 2, \dots \quad (4.3)$$

For the sake of simplicity, we will restrict our analysis to the discrete-time case.

In a structural connectome, an edge with weight w_{ij} links two ROIs i and j if they share w_{ij} reconstructed streamlines. To build the bridge between connectomics and linear control, we can reasonably assume that two anatomically connected ROIs influence each other functionally. Then, it is natural to replace matrix $\mathbf{A}$ in equation (4.3) by the adjacency matrix of a connectome in order to get a state representation of the anatomically constrained dynamics of the brain. Hence, the agents of the system correspond to the ROIs (i.e., the vertices) of the connectome and their state represent the neurophysiological activity of the associated neural region.

4.1.2 Global controllability

To define the notions related to the controllability of complex networks, we refer to the work of Pasqualetti et al. [88]. A discrete-time linear system represented by equation (4.3) is controllable in T steps if for every state vector $\mathbf{x}_f \in \mathbb{R}^N$ there exists an input vector $\mathbf{u}_{\mathcal{K}} \in \mathbb{R}^m$ such that $\mathbf{x}(T) = \mathbf{x}_f$. A classic result in linear control states that the system defined by equation (4.3) is controllable in T steps via the set of agents $\mathcal{K}$ if and only if the matrix

$$\mathcal{C}_{\mathcal{K},T} \triangleq \begin{bmatrix} \mathbf{B}_{\mathcal{K}} & \mathbf{A}\mathbf{B}_{\mathcal{K}} & \mathbf{A}^2\mathbf{B}_{\mathcal{K}} & \dots & \mathbf{A}^{T-1}\mathbf{B}_{\mathcal{K}} \end{bmatrix} \quad (4.4)$$

has full row rank.

Unfortunately, the previous condition does not provide any insight about the difficulty or the required energy to steer the system to a given target state. To address this issue, let us introduce the T -steps controllability Gramian $\mathcal{W}_{\mathcal{K},T}$:

$$\mathcal{W}_{\mathcal{K},T} \triangleq \sum_{\tau=0}^{T-1} \mathbf{A}^{\tau} \mathbf{B}_{\mathcal{K}} \mathbf{B}_{\mathcal{K}}^{\top} (\mathbf{A}^{\top})^{\tau} \quad (4.5)$$

$$= \mathcal{C}_{\mathcal{K},T} \mathcal{C}_{\mathcal{K},T}^{\top} \quad (4.6)$$

For the rest of this section, we consider an infinite time-horizon (i.e. $T \rightarrow \infty$) and omit the corresponding subscript to simplify the notations. Interestingly, it turns out that the spectrum of $\mathcal{W}_{\mathcal{K}}$ provides quantitative information about the controllability of system (4.3).

¹No self-interaction corresponds to matrix $\mathbf{A}$ satisfying $A_{ii} = 0, \forall i \in \{1, 2, \dots, N\}$.

Indeed, the authors of [88] show that the minimum energy² E^* required to steer a system to any target state satisfies

$$0 \leq E^* \leq \lambda_{\min}^{-1}(\mathcal{W}_{\mathcal{K}}) \quad (4.7)$$

where $\lambda_{\min}(M)$ denotes the smallest eigenvalue of a matrix M . In equation (4.7), we see that if $\lambda_{\min}(\mathcal{W}_{\mathcal{K}})$ is small, the required energy for the controllability of the system is potentially large. In particular, if $\lambda_{\min}(\mathcal{W}_{\mathcal{K}}) \leq 0$, the system is declared globally non controllable from a worst-case energy perspective (i.e. when considering the most difficult-to-reach state).

4.1.3 Regional controllability

It is possible to refine this notion of controllability to detect agents that are most influential in the transition between states. Especially, we distinguish agents allowing to drive the system to many "easy-to-reach" states and agents more involved in the transition towards "hard-to-reach" states. It is worth noticing that these notions are local. In this sense, the condition $\lambda_{\min}(\mathcal{W}_{\mathcal{K}}) \leq 0$ is in general not necessary to define the following concepts. In the following, we consider a graph $G(\mathcal{V}, \mathcal{E})$ composed of N vertices, and its weighted adjacency matrix $\mathbf{A}$.

Average controllability As explained in [88], average controllability with respect to a set of control nodes $\mathcal{K}$ is an inverse measure of the average amount of energy required to steer the system to many random target states. In other words, regions showing a high average controllability are able, on average, to steer the system to different states with little energy. In the context of connectomics, such "easy-to-reach" states include passive rest or mind-wandering for example.

Mathematically, we can show that the average input energy E^{av} associated to a set of control nodes $\mathcal{K}$ is proportional to $\text{Trace}(\mathcal{W}_{\mathcal{K}}^{-1})$. In particular, we can choose control nodes one at a time in order to measure their respective average controllability. Hence, we can write

$$E_i^{av} \sim \text{Trace}(\mathcal{W}_{\mathcal{K}}^{-1}) \quad (4.8)$$

where $\mathcal{K}$ only contains node i . Unfortunately, it turns out that the controllability Gramian is often ill-conditioned, resulting in a (nearly) singular matrix. Therefore, motivated by the inverse proportionality relation

$$\frac{\text{Trace}(\mathcal{W}_{\mathcal{K}}^{-1})}{n} \geq \frac{n}{\text{Trace}(\mathcal{W}_{\mathcal{K}})} \quad (4.9)$$

where n is the dimension of $\mathcal{W}_{\mathcal{K}}$, the authors of [25] and [88] choose to consider an alternative measure η_i for the average controllability of node i :

$$\eta_i \triangleq \text{Trace}(\mathcal{W}_{\mathcal{K}}), \quad \mathcal{K} = \{i\} \subset \mathcal{V} \quad (4.10)$$

This quantity can be computed even when $\mathcal{W}_{\mathcal{K}}$ is singular and has already been used in several papers (see the References of [88] for example).

Modal controllability As opposed to average controllability, modal controllability refers to the ability of some control nodes to drive the system towards "difficult-to-reach" states.

²In this context, the energy associated to a control input $\mathbf{u}_{\mathcal{K}}$ over a time-horizon T is defined as $E(\mathbf{u}_{\mathcal{K}}, T) = \sum_{\tau=0}^{T-1} \|\mathbf{u}_{\mathcal{K}}(\tau)\|_2^2$

It mainly relies on the spectral decomposition of the network adjacency matrix $\mathbf{A}$. In the context of connectomics, neural regions showing a high modal controllability are expected to be involved in cognitive or motor tasks for example.

In linear control theory, a pair eigenvalue-eigenvector of matrix $\mathbf{A}$ in system (4.3) is called a dynamic mode. As explained by Zhou et al. in [89], dynamic modes govern the overall dynamics of the system. Let us denote by $\mathbf{V} = [v_{ij}]$ the matrix whose columns are the eigenvectors of $\mathbf{A}$ and by $\lambda_j(\mathbf{A})$ the j -th eigenvalue of $\mathbf{A}$. By applying the Popov-Belovich-Hautus (PBH) test for controllability to symmetric matrices, we can show³ that the smaller an entry v_{ij} , the weaker the controllability of mode j from node i . From this statement, authors of [88] propose one definition for the modal controllability ϕ of node i :

$$\phi_i \triangleq \sum_{j=1}^N (1 - \lambda_j^2(\mathbf{A})) v_{ij}^2 \quad (4.11)$$

This quantity can be understood as a weighted measure of the controllability of all N modes from node i . Let us also observe that it does not depend on the controllability Gramian $\mathcal{W}_K$ and is therefore still defined when the system is not globally controllable from a worst-case energy perspective.

4.2 Results

In this section, we apply the same analysis as in [25] to our dataset. The main difference between our connectomes and that used in the original article is the resolution of the brain parcellation scheme (see Chapter 2). Gu et al. defined 234 ROIs (i.e. vertices) whereas we only consider 84. Codes for average and modal controllability are publicly available on Danielle Bassett's website⁴.

4.2.1 Preliminaries

First, to ensure the convergence of the series defined in equation (4.5) as $T \rightarrow \infty$, we require⁵ that $\rho(\mathbf{A}) < 1$, where $\rho(\mathbf{A})$ denotes the spectral radius of matrix $\mathbf{A}$: $\rho(\mathbf{A}) = \max\{|\lambda_i| : \lambda_i \in \text{spec}(\mathbf{A})\}$. To this end and for each connectome, we normalize the entries of its adjacency matrix by $\rho(\mathbf{A}) + 1$. Indeed, we have successively:

$$\frac{1}{\rho(\mathbf{A}) + 1} (\mathbf{A}x) = \frac{1}{\rho(\mathbf{A}) + 1} (\lambda x) \quad (4.12)$$

$$\underbrace{\left(\frac{\mathbf{A}}{\rho(\mathbf{A}) + 1} \right)}_{\tilde{\mathbf{A}}} x = \underbrace{\left(\frac{\lambda}{\rho(\mathbf{A}) + 1} \right)}_{<1} x \quad (4.13)$$

³The PBH test for controllability states that system (4.3) is controllable if and only if there exists no left eigenvector of $\mathbf{A}$ orthogonal to the columns of $\mathbf{B}$. If we consider a left eigenvector (and thus a mode) ω and a single input node i (thus, $\mathbf{B}$ is a column), then system (4.3) is non-controllable if $\omega^T \mathbf{B} = 0$. As all entries of $\mathbf{B}$ are zero except for the i -th, the i -th entry of ω will determine the controllability of the system. The PBH test for controllability is described in details in [89].

⁴<http://www.daniasbassett.com/>

⁵Although it is possible (and more convenient) to compute the controllability Gramian as a solution of the Lyapunov equation $\mathbf{A}\mathcal{W}_K\mathbf{A}^T - \mathcal{W}_K + \mathbf{B}\mathbf{B}^T = 0$. Nevertheless, if $\rho(\mathbf{A}) \geq 1$, the solution to this equation is not unique. This observation is also related to the stability of the system.

where x and λ are respectively some eigenvector and eigenvalue of A .

Secondly, as in [25], we consider in the following nine subsets of ROIs from the brain atlas we used in Chapter 2. These subsets will be referred to as "cognitive systems" and were derived by Power et al. in [90] on the basis of clustering methods performed on brain functional connectivity networks. The cognitive systems listed below are thus expected to associate anatomical ROIs to brain functions.

Cognitive systems considered in our analysis

- | | |
|----------------------|----------------------|
| 1. Auditory | 6. Default mode |
| 2. Visual | 7. Frontoparietal |
| 3. Somatosensory | 8. Cingulo-opercular |
| 4. Ventral attention | 9. Other |
| 5. Dorsal attention | |

The correspondence between ROIs and cognitive systems is available in Appendix C. Let us notice that the authors of [25] neglected the role of subcortical structures in cognitive systems in some of their results, while they were taken into account in the dynamics of system (4.3). In the following results, we therefore decide to add one system named "Subcortical".

4.2.2 Global controllability of the structural connectome

In contrast to the results presented in [25], the eigenvalues of the controllability Gramian associated to our connectomes are not strictly positive, when considering nodes one at a time as single control input. In other words, our connectomes appear to be non controllable, from a worst-case energy perspective, when the control input acts on one single node. Nevertheless, the minimum eigenvalue of $\mathcal{W}_K$ is always close to zero, as in [25]. Across all connectomes and all choices of control node, we obtained a mean minimum eigenvalue of -5.26×10^{-14} with standard deviation equal to 1.28×10^{-13} . Although the authors of [25] found that their connectomes were globally controllable, they insist on the fact that this result is purely theoretical, while in practice the connectome is extremely hard to control through one single ROI.

In view of this result, we focus our analysis on local notions of controllability, namely average and modal controllability, that are defined for each node of the connectome.

4.2.3 Correlation between controllability and strength

Interestingly, average and modal controllability (equations (4.10) and (4.11)) seem to be respectively correlated and anti-correlated with the strength (i.e. the weighted degree) of the corresponding node. The results reported in Table 4.1 are similar to that obtained by Gu et al. in [25].

These results are consistent with the intuition. Indeed, densely connected brain regions are expected to be more involved in the transition to many different "easy-to-reach" states, whereas transitions to "hard-to-reach" states will more likely require the intervention of sparsely connected areas.

TABLE 4.1: Pearson correlation coefficient between node strength and corresponding average and modal controllability measures, averaged across all 204 connectomes of our dataset. (Mean $\pm$ standard deviation)

Correlation	Average controllability of node i	Modal controllability of node i
Strength of node i	0.9103 ± 0.03	-0.9043 ± 0.02

4.2.4 Controllability-by-system interaction

We now want to investigate the relationship between cognitive systems and controllability. First, we compute both average and modal controllability vectors (i.e. one component by node in equations (4.10) and (4.11)) for each connectome of our dataset. Then, in order to get a qualitative overview of the relationship and to simplify the analysis, we average both controllability vectors over the 204 connectomes. Given these two "mean" controllability vectors, we choose to keep for each vector the 25 highest values, consistently with Gu et al. [25]. ROIs corresponding to these values are either named "average controllability hubs" or "modal controllability hubs", depending on the vector in which they appear. Finally, given the correspondence between ROIs and cognitive systems (Appendix C), we associate controllability hubs to cognitive systems and look for which of the latter is the most influential in either average or modal controllability.

Results related to average controllability are depicted in Figure 4.1. As expected, average controllability is the highest in the default mode system. This confirms the idea that average controllability measures the ability to switch between many "easy-to-reach" states. We also observe that the somatosensory system shows the second highest average controllability. Qualitatively, these results are in accordance with the work of Gu et al. [25].

Results related to modal controllability are depicted in Figure 4.2. We observe that sub-cortical structures are predominantly represented among modal controllability hubs, which was not reported in [25]. In addition, the frontoparietal system and brain regions gathered under the name "Other" show high and similar modal controllabilities, which may indicate that they are involved in functions requiring significant cognitive effort.

These results suggest a "controllability-by-system" paradigm: depending on the function of a system, its controllability diagnostic will be different. Furthermore, the controllability metrics we considered were computed on the basis of the network structure. Hence, this shows how structure somehow constrains function in the human connectome.

4.3 Discussion

Minimum set of control nodes As we have seen, none of our connectomes was controllable through a single control node, in contrast to the results of Gu et al. [25]. This could be the consequence of the coarser resolution of the brain atlas we used in Chapter 2. Nonetheless, the authors of [25] admit that, despite their theoretical results, global controllability (from a worst-case energy perspective) through a single control node is extremely hard to achieve in practice. One interesting question to be addressed is: 'What is the minimum set of nodes we should select to make the system controllable?'

As explained by Cortesi et al. in [91], many approaches to answer this question have been proposed, but most of them result in computationally hard combinatorial problems. Heuristic methods have also been developed to optimize particular controllability metrics (see [88]

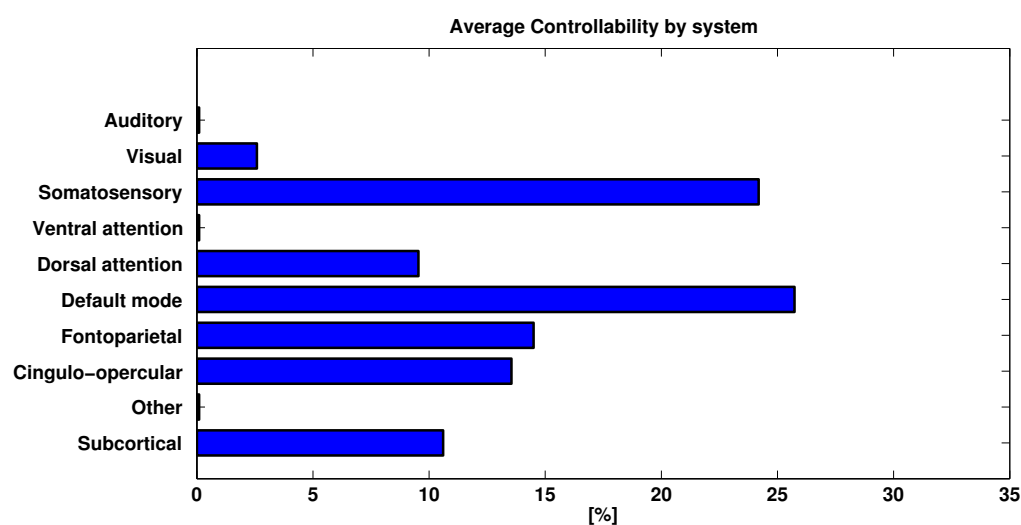


FIGURE 4.1: Percentage of average controllability by cognitive system, when considering the 25 strongest average controllability hubs and the correspondence in Appendix C. The proportions were normalized by the number of ROIs by system.

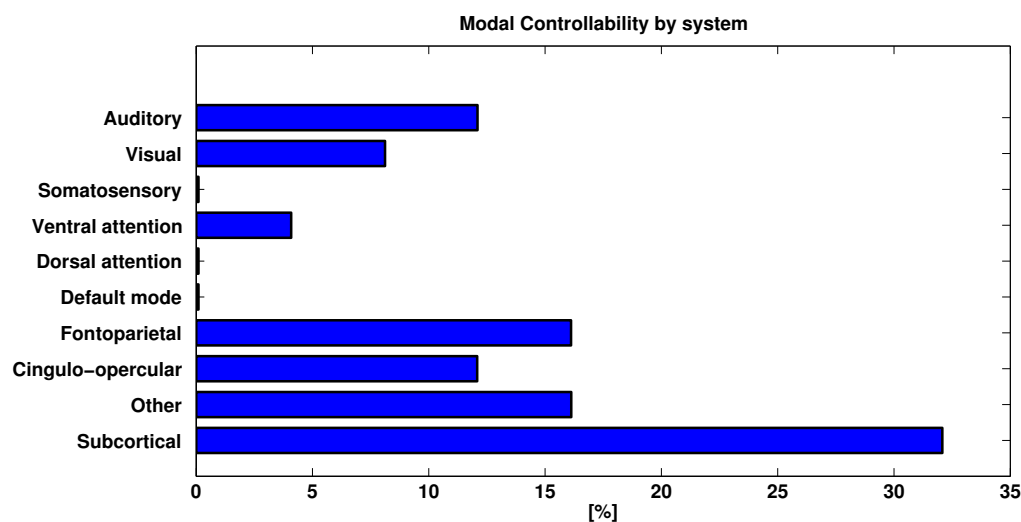


FIGURE 4.2: Percentage of modal controllability by cognitive system, when considering the 25 strongest modal controllability hubs and the correspondence in Appendix C. The proportions were normalized by the number of ROIs by system.

for instance). We suggest that applying such algorithms to the human connectome should be of great interest in future work.

Boundary controllability In [25], the authors study not only average and modal controllability of the human structural connectome, but also the so-called boundary controllability. This concept identifies nodes of the system that are most influential in decoupling trajectories (i.e. temporal sequences of states) between distinct "modules". Reworded in the context of connectomics, ROIs showing high boundary controllability are involved in transitions between states where disjoint cognitive systems are either coupled or decoupled. The results obtained by Gu et al. for this controllability metric are not as easy to interpret as that obtained for average and modal controllability, and the authors do not provide any reference on the application of this metric to other kinds of graphs to assess its consistency. To summarize their results, the repartition of "boundary controllability hubs" between cognitive systems is almost homogeneous, suggesting that each system owns its hubs in order to be coupled with other systems. The authors also argue that systems related to attention sustaining are more involved in the coupling/decoupling between systems, resulting in a (slightly) higher boundary controllability.

Definition of cognitive systems First of all, a finer anatomical parcellation scheme to define the ROIs should only be better in future work. This is also true for the definition of cognitive systems. Although the cognitive map we considered [90] is standard and was also considered in other articles (see the recent work of Mattar et al. [44] for example), we think that more attention should be paid to the definition of cognitive systems in this kind of study. For instance, restricting the membership of one ROI to a single cognitive system may limit the interpretation of the results. Several approaches designed to map cognitive systems involve clustering or community detection methods that detect non-overlapping groups of nodes in functional connectivity networks. Other techniques allowing the detection of overlapping groups (such as the clique percolation algorithm introduced by Palla et al. in [92]) could provide new interesting maps of cognitive systems.

Validity of the linear model Modelling the complex neural dynamics as a noise-free linear system may seem to be a coarse approximation. However, this first approach is doubly motivated. First, we did not study or simulate any dynamics on the network. We were only interested in how the structure of the connectome could *in fine* influence its dynamics. In this direction, the linear model for connectome dynamics is commonly accepted (see the recent work of Betzel et al. [93], Medaglia et al. [94] and Gu et al. [95]). As a first approach, linear models are well understood and can already provide meaningful information resulting from complex (though linear) interactions between brain regions. Secondly, encouraging results from the work of Honey et al. [96] and Muldoon et al. [87] have respectively shown that linear approximations allow to predict functional connectivity from structural connectomes, and that complex nonlinear models involving hundreds of parameters provide results about average and modal controllability that are consistent with that presented by Gu et al. [25]. Of course, although linear dynamics seems acceptable (at least in a first approach) at the macroscale, microscale connectomics should more likely consider the nonlinear dynamics of individual neurons.

Future prospects We strongly believe in the potential of controllability studies applied to connectomics. This emergent field has recently motivated many researchers, but the results mainly remain theoretical. Applied studies could bring new perspectives and challenges in future years.

Recent advances in establishing the role of network topology in controllability studies of the connectome have been made by Betzel et al. in [93]. Besides, Medaglia et al. investigated in [94] the correlation between on the one hand individual differences in terms of connectomic controllability, and on the other hand cognitive performances of these individuals. This kind of study could potentially enrich the field of connectome classification (see Chapter 3). Finally, all this research paves the way to a better understanding of the relationship between structure and function inside the brain, and future work combining both structural and functional connectivity data should be carried out in the context of controllability studies.

Chapter 5

Conclusion

Throughout this work, we presented the emergent field of connectomics, a discipline located at the intersection between complex network theory and neurosciences. Connectomics studies the brain as a complex network and allows the use of existing methods and algorithms in order to strengthen our knowledge of one of our most essential and fascinating organ. In particular, the use of machine learning algorithms applied to networked data offers exciting perspectives.

The objectives of the present work were threefold. First, starting from neuroimaging data, we built our own dataset composed of 204 structural connectomes (or brain graphs), using specialized algorithms based on widespread methods such as tractography.

Afterwards, we addressed the problem of classifying connectomes from our dataset with respect to some label of interest. The aim was to develop a machine learning framework able to identify differences between connectomes related to gender, handedness, alcohol consumption or smoking of the corresponding individuals. Along with two state-of-the-art graph classification techniques, we proposed four custom strategies involving methods of the literature that were not yet applied in this context. Our results show on the one hand that the proposed strategies provide in general better results than that of the considered state-of-the-art methods, and on the other hand that the results highly depend on the balance of the dataset. In particular, the classification with respect to gender provided evidence that male and female brains are intrinsically different. The classification with respect to smoking produced encouraging results but they do not lead to the conclusion of a marked difference between brains of smoker and non-smoker. The classification with respect to handedness and alcohol consumption highlighted the class imbalance problem and was therefore not conclusive.

Eventually, we introduced a promising field of connectomics related to the controllability of connectomes. We followed the work of a reference article and applied their analysis to our dataset. We showed that different mathematical notions of controllability are associated with different cognitive systems, and we motivated the interest of undertaking future studies in this direction.

In our research, we only considered structural connectomes for reasons we discussed, although many studies consider their functional counterpart. Nevertheless, the methods developed in Chapter 3 are valid for any kind of connectomes, and we would encourage future work testing them on functional connectomes. Besides, the main limitation encountered during our work was the time required to generate our dataset. We reached a reasonable dataset size within the allocated time, but further studies should start by growing this dataset.

Perspective and future prospects

The present work is the fruit of many thoughts and discussions, during which some ideas were formulated and then dismissed, although full of potential.

In particular, we believe that attention should be paid to the evolution of the work related to dynamic functional connectomes. The functions of our brain make it inherently dynamic, and new imaging modalities with impressive time resolution should help analyzing the dynamical reconfiguration of the brain, an aspect that was neglected until now.

Moreover, we would encourage future research aiming at quantifying the robustness of the connectome. Indeed, some work related to percolation theory already defined robustness metrics for complex networks such as the internet or power grids. Our feeling is that these metrics are not suited for connectomes since they do not reflect the biological reality underlying the abstract representation of the brain as a network. Hence, we think that new robustness metrics should be designed especially for connectomes, taking into account physiological factors related to our brain health.

To conclude, we strongly believe in the power of the union between graph theory, machine learning and neurosciences. Many studies have shown encouraging results, and a growing number of scientists turn their interest towards this field. It is a safe bet that connectomics will play an important role in the future of neurosciences.

Appendix A

Topological properties of connectomes

A.1 Z-scores histograms

To get a global overview of the topological differences existing between connectomes and null models, we depict in Figures A.1, A.2 and A.3 the Z-score histograms for the characteristic path length, the clustering coefficient and the modularity respectively.

To understand Z-score histograms, let us consider a particular topological coefficient c . In our dataset, we denote by x_i^c the value of coefficient c for connectome i . In the same way, we denote by $x_{i,0}^c$ the value of coefficient c for the null model associated to connectome i . Then, we denote by μ_0^c the mean of value of coefficient c in the population of null models, and by σ_0^c its standard deviation. For each connectome and null model respectively, the Z-score is given by

$$Z_i^c = \frac{x_i^c - \mu_0^c}{\sigma_0^c} \quad (\text{A.1})$$

$$Z_{i,0}^c = \frac{x_{i,0}^c - \mu_0^c}{\sigma_0^c} \quad (\text{A.2})$$

$$(\text{A.3})$$

Z-score histograms depict the distribution of a particular topological coefficient, "standardized" with respect to the population of null models, and provide an overall idea of how different is the value of a topological coefficient in connectomes and null models. These histograms must be understood carefully, since each null model is associated to a particular connectome, which explains the overlapping histograms in Figures A.1 and A.2.

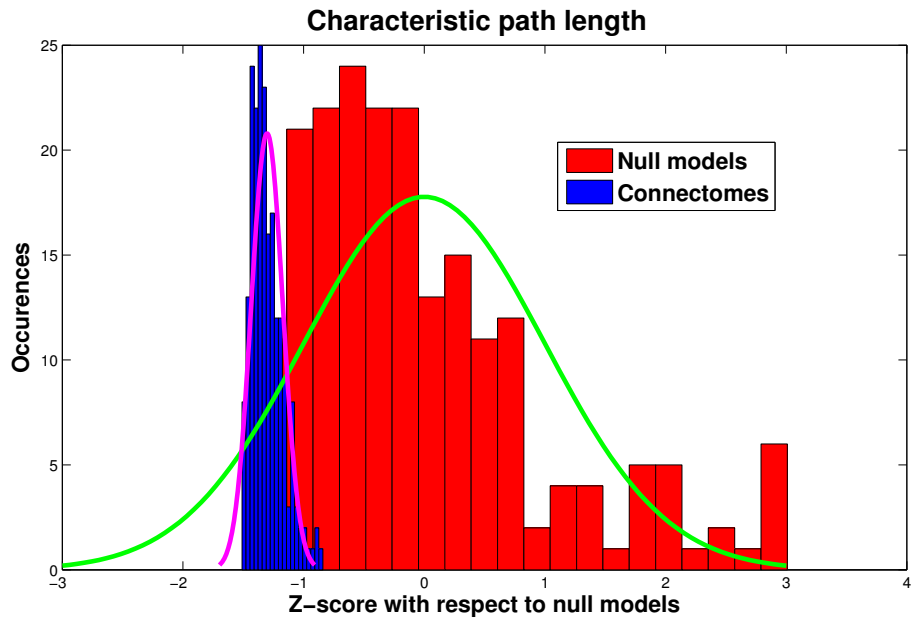


FIGURE A.1: Z-score histogram for the characteristic path length, on the whole dataset. (Z-score computed with respect to null models.)

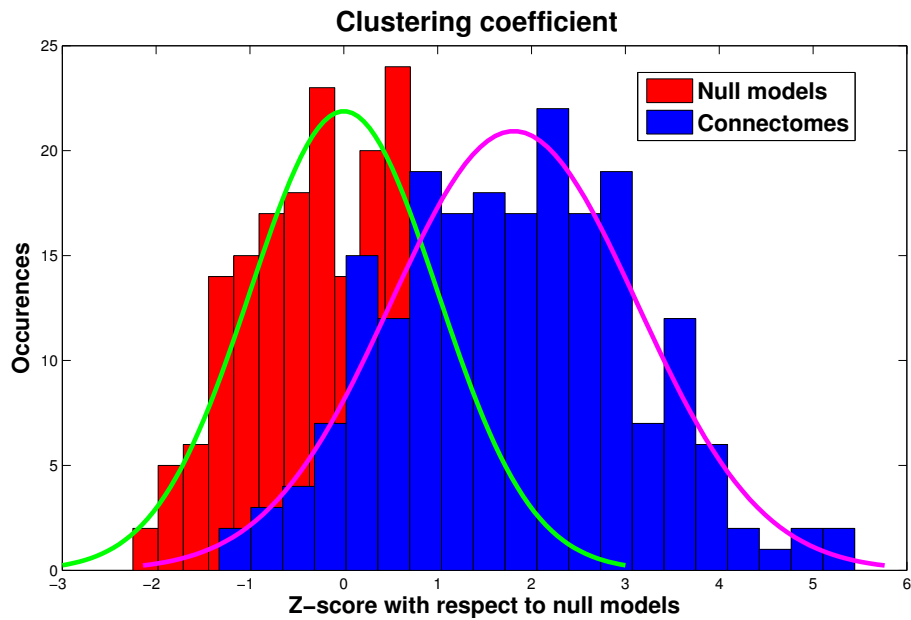


FIGURE A.2: Z-score histogram for the clustering coefficient, on the whole dataset. (Z-score computed with respect to null models.)

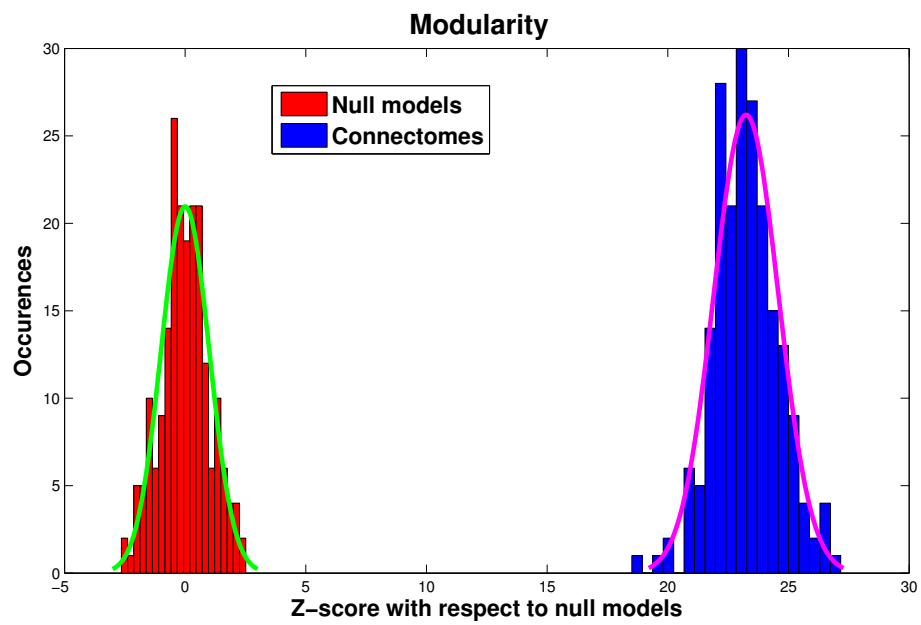


FIGURE A.3: Z-score histogram for the modularity on the whole dataset. (Z-score computed with respect to null models.)

Appendix B

Topological and spectral features for graph classification

B.1 Mathematical definition of the coefficients

Hereunder, we consider the mathematical definitions of the coefficients used to build the feature vector of our first classification strategy (see Box 3.1 in the main manuscript).

Let us consider an undirected graph $G(\mathcal{V}, \mathcal{E})$, with $N = |\mathcal{V}|$ nodes, and with adjacency matrix $\mathbf{A}$.

Mean degree

The degree k_i of a node i is the number of edges connecting i to its neighbors:

$$k_i = \sum_{j \neq i} \mathbf{A}_{ij}$$

The mean degree $\langle k \rangle$ is the algebraic average of the previous quantity over all nodes:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i \quad (\text{B.1})$$

Mean strength

The strength s_i of a node i is the sum of the weights of the edges connecting i to its neighbors:

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

where w_{ij} denotes the weight of the edge linking nodes i and j . The mean strength $\langle s \rangle$ is the algebraic average of the previous quantity over all nodes:

$$\langle s \rangle = \frac{1}{N} \sum_{i=1}^N s_i \quad (\text{B.2})$$

Density of edges

The density of edges χ in a graph with N nodes is the ratio between the actual number of edges composing the graph and the maximum number of edges that a graph with N nodes

can support (we consider one edge maximum between two distinct nodes):

$$\chi = \frac{2|\mathcal{E}|}{N(N-1)} \quad (\text{B.3})$$

Average clustering coefficient

The clustering coefficient of node i is the proportion of pairs of i 's neighbors that are connected. We consider here the average over all nodes of the weighted version of the clustering coefficient:

$$C_w = \frac{1}{N} \sum_{i=1}^N \frac{2}{k_i(k_i-1)} \sum_{j,h} (\hat{w}_{ij}\hat{w}_{jh}\hat{w}_{hi})^{1/3} \quad (\text{B.4})$$

where $\hat{w}_{ij}$ stands for the weight of the edge connecting nodes i and j , normalized by the maximum edge weight appearing in $\mathcal{E}$.

Transitivity coefficient

The transitivity coefficient T is an alternative to the clustering coefficient, estimating the probability that any two nodes connected to a third are also connected to each other:

$$T = \frac{2 \sum_{i=1}^N \sum_{j,h} (\hat{w}_{ij}\hat{w}_{jh}\hat{w}_{hi})^{1/3}}{\sum_{i=1}^N k_i(k_i-1)} \quad (\text{B.5})$$

Characteristic path length

A path is a sequence of edges connecting distinct nodes. The path length is the sum of the lengths of the edges composing the path. The shortest path between nodes i and j is denoted l_{ij} and can be computed using Dijkstra's algorithm. Let us specify that, in our context, the length of an edge is the inverse of its weight, meaning that stronger connections correspond to closer nodes. The characteristic path length L is then given by:

$$L = \frac{1}{N} \sum_{i=1}^N \frac{1}{N-1} \sum_{j=1, j \neq i}^N l_{ij} \quad (\text{B.6})$$

Global efficiency coefficient

The global efficiency coefficient E is closely related to the characteristic path length and has the advantage to be defined even when $l_{ij} = \infty$, as in the case of non-connected graphs:

$$E = \frac{1}{N} \sum_{i=1}^N \frac{1}{N-1} \sum_{j=1, j \neq i}^N l_{ij}^{-1} \quad (\text{B.7})$$

Radius

The eccentricity $\epsilon(i)$ of a node i is the maximum value among the shortest path lengths between i and any other node:

$$\epsilon(i) = \max_{j \in \mathcal{V}, j \neq i} l_{ij}$$

The radius R is the minimal eccentricity among the nodes of the graph:

$$R = \min_{i \in \mathcal{V}} \epsilon(i) \quad (\text{B.8})$$

Diameter

Reciprocally, the diameter D is the maximal eccentricity among the nodes of the graph:

$$D = \max_{i \in \mathcal{V}} \epsilon(i) \quad (\text{B.9})$$

Modularity

Modularity is a measure of how strong is the division of the nodes into distinct modules inside a graph. In other words, networks showing high modularity are composed of several sparsely connected groups of densely interconnected nodes. More details are available in Chapter 2. Given a partition, let us denote by c_i the group (or community) of node i and by m the sum of all edge weights in the graph. The modularity Q is given by:

$$Q = \frac{1}{2m} \sum_{i,j=1}^N \left[w_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (\text{B.10})$$

where δ is the classical Kronecker delta.

Several algorithms exist to find a partition that maximises Q . Among them, we chose the method introduced by Newman in [43] due to the modest size of our graphs.

Assortativity

Assortativity reflects the tendency of nodes in a graph to be connected to similar nodes. In our context, two nodes are considered similar if they share similar degree. Formally, the (degree) assortativity of an undirected network is computed as the Pearson correlation coefficient of the node degrees of connected pairs of nodes:

$$a = \frac{|\mathcal{E}|^{-1} \sum_{(i,j) \in \mathcal{E}} w_{ij} s_i s_j - \left[|\mathcal{E}|^{-1} \sum_{(i,j) \in \mathcal{E}} \frac{1}{2} w_{ij} (s_i + s_j) \right]^2}{\left[|\mathcal{E}|^{-1} \sum_{(i,j) \in \mathcal{E}} \frac{1}{2} w_{ij} (s_i^2 + s_j^2) - \left[|\mathcal{E}|^{-1} \sum_{(i,j) \in \mathcal{E}} \frac{1}{2} w_{ij} (s_i + s_j) \right]^2 \right)} \quad (\text{B.11})$$

where s_i is the strength of node i and w_{ij} is the weight of the edge linking nodes i and j .

Diffusion efficiency

Illustratively, let us consider sending messages from a starting node to a destination node in a network. Messages follow random walks through the edges of the network. Diffusion efficiency measures the number of messages that need to be sent to ensure at least one of them reaches the destination node.

First, denoting by $\mathbf{A}$ the weighted adjacency matrix of the graph and by $\mathbf{S}$ the diagonal matrix whose elements are the nodes strength, we define the following matrix:

$$\mathbf{U} = \mathbf{A}\mathbf{S}^{-1}$$

We can understand this matrix as the matrix of transition probabilities of a Markov chain whose states are network nodes. We next define the matrix $\mathbf{U}_j$ corresponding to matrix $\mathbf{U}$ whose j -th is set to zero. Then, the entries of matrix $\mathbf{U}_j^t$, where t denotes an integer exponent, provides the probabilities of a random walker to go from one node to another in exactly t steps, without the possibility to travel through node j .

Besides, we denote by X_{ij} the number of steps required by a random walker to go from node i to node j . We have

$$\begin{aligned}\langle X_{ij} \rangle &= \sum_{t=0}^{\infty} tP(X_{ij} = t) \\ &= \sum_{t=0}^{\infty} P(X_{ij} > t)\end{aligned}$$

It is important to notice that

$$P(X_{ij} > t) = \sum_{n=1}^N [\mathbf{U}_j^t]_{ni}$$

Thus, we get

$$\begin{aligned}\langle X_{ij} \rangle &= \sum_{t=0}^{\infty} \sum_{n=1}^N [\mathbf{U}_j^t]_{ni} \\ &= \sum_{n=1}^N [(\mathbf{I} - \mathbf{U}_j)^{-1}]_{ni}\end{aligned}$$

where $\mathbf{I}$ denotes the identity matrix. This quantity is also called the "mean first passage time" in the literature. Finally, Goñi et al. defined in [97] the diffusion efficiency coefficient as following:

$$E_{\text{diff}} = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle X_{ij} \rangle^{-1} \quad (\text{B.12})$$

Spectral radius

The spectral radius $\rho(\mathbf{A})$ is the largest eigenvalue of the adjacency matrix $\mathbf{A}$, in absolute value:

$$\rho(\mathbf{A}) = \max_i |\lambda_i|, \quad \lambda_i \in \text{spec}(\mathbf{A}) \quad (\text{B.13})$$

Frobenius norm

The Frobenius norm of a (real) adjacency matrix $\mathbf{A}$ is

$$\|\mathbf{A}\|_F = \sqrt{\text{Trace}(\mathbf{A}^\top \mathbf{A})} \quad (\text{B.14})$$

It is interesting to notice that in the case of $\mathbf{A}$ being symmetric, this expression reduces to

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^N \lambda_i^2}, \quad \lambda_i \in \text{spec}(\mathbf{A})$$

Energy

The energy of a graph, as defined by Gutman in [98], is given by

$$\Xi = \sum_{i=1}^N |\lambda_i|, \quad \lambda_i \in \text{spec}(\mathbf{A}) \quad (\text{B.15})$$

However, other authors consider an alternative definition (see for example [54]):

$$\Xi = \sum_{i=1}^N \lambda_i^2, \quad \lambda_i \in \text{spec}(\mathbf{A})$$

which is clearly related to the Frobenius norm of symmetric matrices.

B.2 Most discriminative features

We report here the most discriminative features of Strategy 1 (Box 3.1). For each label of interest, we rank topological and spectral features according to the weight granted by the linear kernel, in absolute value. As a reminder, the larger the weight granted to a feature, the higher its discriminative nature. Here, we keep only the features that appeared more than 20 times (over the 30 independent tests, using accuracy as performance metric) in the 5 most discriminative features. We do not report these features for the label "handedness" since the classifier totally failed to discriminate classes, due to the strong class imbalance.

- Gender:
 1. Global efficiency
 2. Diameter
 3. Frobenius norm
- Alcohol abuse:
 1. Mean strength
 2. Average clustering coefficient
 3. Transitivity coefficient
 4. Frobenius norm
 5. Energy
- Smoking:
 1. Average clustering coefficient
 2. Transitivity coefficient
 3. Global efficiency

Appendix C

Controllability-by-system

C.1 Correspondence between ROIs and cognitive systems

In the table hereunder, we give the correspondence between each ROI of our parcellation scheme and the cognitive systems introduced in Chapter 4. The column "ID" gives the index of the ROI in the adjacency matrices of the connectomes composing our dataset.

ID	Region of interest	Cognitive system
1	Left - Bank of the Superior Temporal Sulcus	Auditory
2	Left - Caudal Anterior Cingulate	Cingulo-Opercular
3	Left - Caudal Middle Frontal	Fronto-Parietal
4	Left - Cuneus	Visual
5	Left - Entorhinal	Visual
6	Left - Fusiform	Visual
7	Left - Inferior Parietal	Fronto-Parietal
8	Left - Inferior Temporal	Visual
9	Left - Isthmus Cingulate	Default Mode
10	Left - Lateral Occipital	Visual
11	Left - Lateral Orbitofrontal	Ventral Attention
12	Left - Lingual	Visual
13	Left - Medial Orbitofrontal	Fronto-Parietal
14	Left - Middle Temporal	Other
15	Left - Parahippocampal	Other
16	Left - Paracentral	Somatosensory
17	Left - Pars Opercularis	Cingulo-Opercular
18	Left - Pars Orbitalis	Cingulo-Opercular
19	Left - Pars Triangularis	Fronto-Parietal
20	Left - Pericalcarine	Visual
21	Left - Post Central	Somatosensory
22	Left - Posterior Cingulate	Default Mode
23	Left - Precentral	Somatosensory
24	Left - Precuneus	Default Mode
25	Left - Rostral Anterior Cingulate	Cingulo-Opercular

26	Left - Rostral Middle Frontal	Cingulo-Opercular
27	Left - Superior Frontal	Default Mode
28	Left - Superior Parietal	Dorsal Attention
29	Left - Superior Temporal	Auditory
30	Left - Supramarginal	Cingulo-Opercular
31	Left - Frontal Pole	Fronto-Parietal
32	Left - Temporal Pole	Other
33	Left - Transverse Temporal	Auditory
34	Left - Insula	Fronto-Parietal
35	Left - Cerebellum	/
36	Left - Thalamus	Subcortical
37	Left - Caudate	Subcortical
38	Left - Putamen	Subcortical
39	Left - Pallidum	Subcortical
40	Left - Hippocampus	Subcortical
41	Left - Amygdala	Subcortical
42	Left - Accumbens	Subcortical
43	Right - Thalamus	Subcortical
44	Right - Caudate	Subcortical
45	Right - Putamen	Subcortical
46	Right - Pallidum	Subcortical
47	Right - Hippocampus	Subcortical
48	Right - Amygdala	Subcortical
49	Right - Accumbens	Subcortical
50	Right - Bank of the Superior Temporal Sulcus	Auditory
51	Right - Caudal Anterior Cingulate	Cingulo-Opercular
52	Right - Caudal Middle Frontal	Fronto-Parietal
53	Right - Cuneus	Visual
54	Right - Entorhinal	Visual
55	Right - Fusiform	Visual
56	Right - Inferior Parietal	Fronto-Parietal
57	Right - Inferior Temporal	Visual
58	Right - Isthmus Cingulate	Default Mode
59	Right - Lateral Occipital	Visual
60	Right - Lateral Orbitofrontal	Ventral Attention
61	Right - Lingual	Visual
62	Right - Medial Orbitofrontal	Fronto-Parietal
63	Right - Middle Temporal	Other
64	Right - Parahippocampal	Other

65	Right - Paracentral	Somatosensory
66	Right - Pars Opercularis	Cingulo-Opercular
67	Right - Pars Orbitalis	Cingulo-Opercular
68	Right - Pars Triangularis	Fronto-Parietal
69	Right - Pericalcarine	Visual
70	Right - Post Central	Somatosensory
71	Right - Posterior Cingulate	Default Mode
72	Right - Precentral	Somatosensory
73	Right - Precuneus	Default Mode
74	Right - Rostral Anterior Cingulate	Cingulo-Opercular
75	Right - Rostral Middle Frontal	Cingulo-Opercular
76	Right - Superior Frontal	Default Mode
77	Right - Superior Parietal	Dorsal Attention
78	Right - Superior Temporal	Auditory
79	Right - Supramarginal	Cingulo-Opercular
80	Right - Frontal Pole	Fronto-Parietal
81	Right - Temporal Pole	Other
82	Right - Transverse Temporal	Auditory
83	Right - Insula	Fronto-Parietal
84	Right - Cerebellum	/

Bibliography

- [1] John G White et al. "The structure of the nervous system of the nematode *Caenorhabditis elegans*: the mind of a worm". In: *Phil. Trans. R. Soc. Lond* 314 (1986), pp. 1–340.
- [2] Duncan J Watts and Steven H Strogatz. "Collective dynamics of 'small-world' networks". In: *nature* 393.6684 (1998), pp. 440–442.
- [3] Alex Fornito, Andrew Zalesky, and Edward Bullmore. *Fundamentals of brain network analysis*. Academic Press, 2016.
- [4] Alex Fornito, Andrew Zalesky, and Michael Breakspear. "Graph analysis of the human connectome: promise, progress, and pitfalls". In: *Neuroimage* 80 (2013), pp. 426–444.
- [5] Ed Bullmore and Olaf Sporns. "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature Reviews Neuroscience* 10.3 (2009), pp. 186–198.
- [6] Mikail Rubinov and Olaf Sporns. "Complex network measures of brain connectivity: uses and interpretations". In: *Neuroimage* 52.3 (2010), pp. 1059–1069.
- [7] Olaf Sporns. "Structure and function of complex brain networks". In: *Dialogues Clin Neurosci* 15.3 (2013), pp. 247–262.
- [8] Cornelis J Stam and Jaap C Reijneveld. "Graph theoretical analysis of complex networks in the brain". In: *Nonlinear biomedical physics* 1.1 (2007), p. 3.
- [9] Danielle Smith Bassett and ED Bullmore. "Small-world brain networks". In: *The neuroscientist* 12.6 (2006), pp. 512–523.
- [10] Martijn P Van Den Heuvel and Olaf Sporns. "Rich-club organization of the human connectome". In: *Journal of Neuroscience* 31.44 (2011), pp. 15775–15786.
- [11] Olaf Sporns, Giulio Tononi, and Gerald M Edelman. "Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices". In: *Cerebral cortex* 10.2 (2000), pp. 127–141.
- [12] Martijn P Van Den Heuvel et al. "Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain". In: *Human brain mapping* 30.10 (2009), pp. 3127–3141.
- [13] Xiangnan Kong and Philip S Yu. "Brain network analysis: a data mining perspective". In: *ACM SIGKDD Explorations Newsletter* 15.2 (2014), pp. 30–38.
- [14] Stephen Smith¹ Diego Vidaurre² Matthew Glasser et al. "HCP beta-release of the Functional Connectivity MegaTrawl". In: ().
- [15] Urs Braun, Sarah F Muldoon, and Danielle S Bassett. "On human brain networks in health and disease". In: *eLS* (2015).
- [16] Cornelis J Stam. "Modern network science of neurological disorders". In: *Nature Reviews Neuroscience* 15.10 (2014), pp. 683–695.
- [17] David Mears and Harvey B Pollard. "Network science and the human brain: using graph theory to understand the brain and one of its hubs, the amygdala, in health and disease". In: *Journal of neuroscience research* (2016).

- [18] Wasifa Jamal et al. "Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates". In: *Journal of neural engineering* 11.4 (2014), p. 046019.
- [19] Mary-Ellen Lynall et al. "Functional connectivity and brain networks in schizophrenia". In: *Journal of Neuroscience* 30.28 (2010), pp. 9477–9487.
- [20] Martijn P van den Heuvel et al. "Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis". In: *Journal of Neuroscience* 30.47 (2010), pp. 15915–15926.
- [21] Mario Chavez et al. "Functional modularity of background activities in normal and epileptic brain networks". In: *Physical review letters* 104.11 (2010), p. 118701.
- [22] Chong-Yaw Wee et al. "Enriched white matter connectivity networks for accurate identification of MCI patients". In: *Neuroimage* 54.3 (2011), pp. 1812–1822.
- [23] Biao Jie et al. "Structural feature selection for connectivity network-based MCI diagnosis". In: *International Workshop on Multimodal Brain Image Analysis*. Springer. 2012, pp. 175–184.
- [24] Sara Llufríu et al. "Structural networks involved in attention and executive functions in multiple sclerosis". In: *NeuroImage: Clinical* 13 (2017), pp. 288–296.
- [25] Shi Gu et al. "Controllability of structural brain networks". In: *Nature communications* 6 (2015).
- [26] Patric Hagmann. "From diffusion MRI to brain connectomics". PhD thesis. Institut de traitement des signaux PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES PAR Docteur en médecine, Université de Lausanne, 2005.
- [27] Patric Hagmann et al. "White matter maturation reshapes structural connectivity in the late developing human brain". In: *Proceedings of the National Academy of Sciences* 107.44 (2010), pp. 19067–19072.
- [28] Rahul S Desikan et al. "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest". In: *Neuroimage* 31.3 (2006), pp. 968–980.
- [29] Brian Patenaude et al. "A Bayesian model of shape and appearance for subcortical brain segmentation". In: *Neuroimage* 56.3 (2011), pp. 907–922.
- [30] Fatima Tensaouti. "Tractographie par IRM de diffusion: algorithmes, validation, reproductibilité et applications". PhD thesis. Université de Toulouse, Université Toulouse III-Paul Sabatier, 2010.
- [31] Robert E Smith et al. "SIFT: spherical-deconvolution informed filtering of tractograms". In: *Neuroimage* 67 (2013), pp. 298–312.
- [32] Matthew F Glasser et al. "The minimal preprocessing pipelines for the Human Connectome Project". In: *Neuroimage* 80 (2013), pp. 105–124.
- [33] Bruce Fischl. "FreeSurfer". In: *Neuroimage* 62.2 (2012), pp. 774–781.
- [34] David C Van Essen et al. "The WU-Minn human connectome project: an overview". In: *Neuroimage* 80 (2013), pp. 62–79.
- [35] J Tournier, Fernando Calamante, Alan Connelly, et al. "MRtrix: diffusion tractography in crossing fiber regions". In: *International Journal of Imaging Systems and Technology* 22.1 (2012), pp. 53–66.

- [36] Robert E Smith et al. "Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information". In: *Neuroimage* 62.3 (2012), pp. 1924–1938.
- [37] JD Tournier, F Calamante, and A Connelly. "Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions". In: *Proc. 18th Annual Meeting of the Intl. Soc. Mag. Reson. Med.(ISMRM)*. 2010, p. 1670.
- [38] Paul Erdos and Alfréd Rényi. "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.
- [39] Réka Albert and Albert-László Barabási. "Statistical mechanics of complex networks". In: *Reviews of modern physics* 74.1 (2002), p. 47.
- [40] Sergei Maslov and Kim Sneppen. "Specificity and stability in topology of protein networks". In: *Science* 296.5569 (2002), pp. 910–913.
- [41] Mikail Rubinov and Olaf Sporns. "Weight-conserving characterization of complex functional brain networks". In: *Neuroimage* 56.4 (2011), pp. 2068–2079.
- [42] Jukka-Pekka Onnela et al. "Intensity and coherence of motifs in weighted complex networks". In: *Physical Review E* 71.6 (2005), p. 065103.
- [43] Mark EJ Newman. "Modularity and community structure in networks". In: *Proceedings of the national academy of sciences* 103.23 (2006), pp. 8577–8582.
- [44] Marcelo G Mattar et al. "A functional cartography of cognitive systems". In: *PLoS Comput Biol* 11.12 (2015), e1004533.
- [45] David Meunier, Renaud Lambiotte, and Edward T Bullmore. "Modular and hierarchically modular organization of brain networks". In: *Frontiers in neuroscience* 4 (2010), p. 200.
- [46] Patric Hagmann et al. "Mapping human whole-brain structural networks with diffusion MRI". In: *PloS one* 2.7 (2007), e597.
- [47] Patric Hagmann et al. "MR connectomics: principles and challenges". In: *Journal of neuroscience methods* 194.1 (2010), pp. 34–45.
- [48] Andrew Zalesky et al. "Whole-brain anatomical networks: does the choice of nodes matter?" In: *Neuroimage* 50.3 (2010), pp. 970–983.
- [49] Bruce Fischl et al. "Automatically parcellating the human cerebral cortex". In: *Cerebral cortex* 14.1 (2004), pp. 11–22.
- [50] Christophe Destrieux et al. "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature". In: *Neuroimage* 53.1 (2010), pp. 1–15.
- [51] Michel Verleysen and John Lee. *Lecture notes of the course LELEC2870 - Machine Learning : regression, dimensionality reduction and data visualization*. 2016.
- [52] James Mercer. "Functions of positive and negative type, and their connection with the theory of integral equations". In: *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character* 209 (1909), pp. 415–446.
- [53] Yoav Goldberg and Michael Elhadad. "splitSVM: fast, space-efficient, non-heuristic, polynomial kernel computation for NLP applications". In: *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*. Association for Computational Linguistics. 2008, pp. 237–240.
- [54] Geng Li et al. "Effective graph classification based on topological and label attributes". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5.4 (2012), pp. 265–283.

- [55] Isabelle Guyon et al. "Gene selection for cancer classification using support vector machines". In: *Machine learning* 46.1 (2002), pp. 389–422.
- [56] Thomas Gärtner, Peter Flach, and Stefan Wrobel. "On graph kernels: Hardness results and efficient alternatives". In: *Learning Theory and Kernel Machines*. Springer, 2003, pp. 129–143.
- [57] Nino Shervashidze et al. "Efficient graphlet kernels for large graph comparison". In: *Artificial Intelligence and Statistics*. 2009, pp. 488–495.
- [58] Nino Shervashidze and Karsten M Borgwardt. "Fast subtree kernels on graphs." In: *NIPS*. 2009, pp. 1660–1668.
- [59] S Vichy N Vishwanathan et al. "Graph kernels". In: *Journal of Machine Learning Research* 11.Apr (2010), pp. 1201–1242.
- [60] Nino Shervashidze et al. "Weisfeiler-lehman graph kernels". In: *Journal of Machine Learning Research* 12.Sep (2011), pp. 2539–2561.
- [61] Danai Koutra, Joshua T Vogelstein, and Christos Faloutsos. "Deltacon: A principled massive-graph similarity function". In: *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM. 2013, pp. 162–170.
- [62] Danai Koutra et al. "Unifying guilt-by-association approaches: Theorems and fast algorithms". In: *Machine Learning and Knowledge Discovery in Databases* (2011), pp. 245–260.
- [63] Jonathan S Yedidia, William T Freeman, and Yair Weiss. "Understanding belief propagation and its generalizations". In: *Exploring artificial intelligence in the new millennium* 8 (2003), pp. 236–239.
- [64] Danai Koutra et al. "Deltacon: Principled Massive-Graph Similarity Function with Attribution". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 10.3 (2016), p. 28.
- [65] Alexander J Smola and Risi Kondor. "Kernels and regularization on graphs". In: *Learning theory and kernel machines*. Springer, 2003, pp. 144–158.
- [66] Jacob Cohen. "A coefficient of agreement for nominal scales". In: *Educational and psychological measurement* 20.1 (1960), pp. 37–46.
- [67] J Richard Landis and Gary G Koch. "The measurement of observer agreement for categorical data". In: *biometrics* (1977), pp. 159–174.
- [68] David Martin Powers. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation". In: (2011).
- [69] Madhura Ingalhalikar et al. "Sex differences in the structural connectome of the human brain". In: *Proceedings of the National Academy of Sciences* 111.2 (2014), pp. 823–828.
- [70] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.
- [71] Erwan Scornet. "Random forests and kernel methods". In: *IEEE Transactions on Information Theory* 62.3 (2016), pp. 1485–1500.
- [72] Sebastian Mika et al. "Fisher discriminant analysis with kernels". In: *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*. IEEE. 1999, pp. 41–48.
- [73] Mohammad R Arbabshirani et al. "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls". In: *NeuroImage* 145 (2017), pp. 137–165.

- [74] Andrew Zalesky, Alex Fornito, and Edward T Bullmore. "Network-based statistic: identifying differences in brain networks". In: *Neuroimage* 53.4 (2010), pp. 1197–1207.
- [75] Daehan Won. "Optimization and Machine Learning Frameworks for Complex Network Analysis". PhD thesis. 2016.
- [76] Kilem Gwet. "Kappa statistic is not satisfactory for assessing the extent of agreement between raters". In: *Statistical methods for inter-rater reliability assessment* 1.6 (2002), pp. 1–6.
- [77] Tom Fawcett. "An introduction to ROC analysis". In: *Pattern recognition letters* 27.8 (2006), pp. 861–874.
- [78] Sören Sonnenburg et al. "Large scale multiple kernel learning". In: *Journal of Machine Learning Research* 7.Jul (2006), pp. 1531–1565.
- [79] Mehmet Gönen and Ethem Alpaydın. "Multiple kernel learning algorithms". In: *Journal of Machine Learning Research* 12.Jul (2011), pp. 2211–2268.
- [80] Francis R Bach, Gert RG Lanckriet, and Michael I Jordan. "Multiple kernel learning, conic duality, and the SMO algorithm". In: *Proceedings of the twenty-first international conference on Machine learning*. ACM. 2004, p. 6.
- [81] Siemon C de Lange, Marcel A de Reus, and Martijn P van den Heuvel. "The Laplacian spectrum of neural networks". In: *Frontiers in computational neuroscience* 7 (2013).
- [82] Cornelia Metzger et al. "Graph Kernels—a Synthesis Note on Positive Definiteness". In: *Actes de la conférence d'apprentissage CAp*. 2012, pp. 223–237.
- [83] Gang Wu, Edward Y Chang, and Zhihua Zhang. "An analysis of transformation on non-positive semidefinite similarity matrix for kernel machines". In: *Proceedings of the 22nd International Conference on Machine Learning*. Vol. 8. 2005.
- [84] Ronny Luss and Alexandre d'Aspremont. "Support vector machine classification with indefinite kernels". In: *Advances in Neural Information Processing Systems*. 2008, pp. 953–960.
- [85] Alexandru Niculescu-Mizil and Rich Caruana. "Predicting good probabilities with supervised learning". In: *Proceedings of the 22nd international conference on Machine learning*. ACM. 2005, pp. 625–632.
- [86] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?: Explaining the Predictions of Any Classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2016, pp. 1135–1144.
- [87] Sarah Feldt Muldoon et al. "Stimulation-based control of dynamic brain networks". In: *PLoS Comput Biol* 12.9 (2016), e1005076.
- [88] Fabio Pasqualetti, Sandro Zampieri, and Francesco Bullo. "Controllability metrics, limitations and algorithms for complex networks". In: *IEEE Transactions on Control of Network Systems* 1.1 (2014), pp. 40–52.
- [89] Kemin Zhou, John Comstock Doyle, Keith Glover, et al. *Robust and optimal control*. Vol. 40. Prentice hall New Jersey, 1996.
- [90] Jonathan D Power et al. "Functional network organization of the human brain". In: *Neuron* 72.4 (2011), pp. 665–678.
- [91] Fabrizio L Cortesi, Tyler H Summers, and John Lygeros. "Submodularity of energy related controllability metrics". In: *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE. 2014, pp. 2883–2888.

- [92] Gergely Palla et al. "Uncovering the overlapping community structure of complex networks in nature and society". In: *Nature* 435.7043 (2005), pp. 814–818.
- [93] Richard F Betzel et al. "Optimally controlling the human connectome: the role of network topology". In: *Scientific Reports* 6 (2016).
- [94] John D Medaglia et al. "Cognitive Control in the Controllable Connectome". In: *arXiv preprint arXiv:1606.09185* (2016).
- [95] Shi Gu et al. "Optimal Trajectories of Brain State Transitions". In: *NeuroImage* (2017).
- [96] CJ Honey et al. "Predicting human resting-state functional connectivity from structural connectivity". In: *Proceedings of the National Academy of Sciences* 106.6 (2009), pp. 2035–2040.
- [97] Joaquín Goñi et al. "Exploring the morphospace of communication efficiency in complex networks". In: *PLoS One* 8.3 (2013), e58070.
- [98] Ivan Gutman. "The energy of a graph: old and new results". In: *Algebraic combinatorics and applications* (2001), pp. 196–211.

