Università degli Studi di Milano - Bicocca

Dipartimento di Economia, Metodi Quantitativi e Strategie di Impresa

Corso di Dottorato di Ricerca in Statistica - XXVIII Ciclo

Statistical Network Analysis: a Multiple Testing Approach

Candidate:
Agnese Maria Di Brisco

Supervisor:
Prof. Piero Quatto

December 1, 2015

“ I sogni durano poco, si sa, ed anche il più bello cede il passo alla mattina.
Immersa nel buio ho sognato il mio sogno più bello;
ora sono sveglia e tremo, ma mi circonda la pura luce del mattino. ”

Acknowledgements

Grazie Prof. Piero Quatto, per la tua profonda cultura,
Grazie Prof.ssa Manuela Berlingeri, per avermi introdotta alla ricerca,
Grazie Prof.ssa Fulvia Mecatti, per le importanti opportunità che il dottorato mi ha offerto,
Grazie Prof.ssa Paola Chiodini, per la tua disarmante sincerità,
Grazie Prof.ssa Patrizia Farina, per ogni tuo consiglio,
Grazie Gaia, sembriamo così diverse, siamo così simili,
Grazie Igor, per ogni abbraccio, per ogni caffè,
Grazie Emanuela, per il tuo brio,
Grazie Fernando, per aver distrutto ogni mia illusione,
Grazie Papà, perchè mi ricordi quanto sono forte,
Grazie Mamma, perchè mi hai resa bella come il sole,
Grazie a tutti coloro che ho incontrato e conosciuto fino ad ora, a chi ha creduto in me, a chi mi ha insegnato qualcosa ed a chi, magari anche solo per un istante, mi ha sorriso,
E brava amor mio, perchè stai imparando a respirare.

Contents

List of Figures

List of Tables

Alla mia Nonna, che mi aspetta nel nostro posto felice

Introduction

The problem of multiple testing has aroused new interest in the last decades, especially since 1950s when Tukey and Scheffé formalized it first. There exist many applicative problems that require a proper statistical method for controlling the probability of false rejections and the test power while dealing with a finite set of multiple comparisons. Because of the increasing number of applications that require a multiple testing approach, because of the increasing power of computers that allows to perform intensive algorithms and, as a counterpart, because of the limitations of the historical multiple testing approaches, the methodological research in this specific field is of great interest and in rapid evolution.
The methodological interest that motivated first this Thesis, hence, concerns the analysis of the multiple testing problem in a Bayesian perspective and the evaluation of the robustness of our proposal when dealing with dependent test statistics.

Furthermore, during our research, it has been launched a fruitful collaboration with the Department of Psychology and with the new Milan Center for Neuroscience (Neuro-mi). One of the most intriguing challenge in neuroscience concerns, indeed, the construction of brain networks that can be interpreted as maps of the interactions between brain regions. The multiple testing methodology is a well-established approach for the construction of brain networks, but the praxis, in neurofunctional applicative studies, is to adopt traditional methods for multiple comparisons that are known to have many limitations.
The analysis of neurofunctional data is a complex process that requires multiple skills and that can not disregard from a valid and proper statistical method. Hence, the practical aims that motivated this Thesis concerns the analysis of neurofunctional magnetic resonance images (fMRI) data in order to describe and synthesize the pattern of functional connectivity of human brains in resting state conditions.

The human brain is a proper example of a complex system, i. e. a system

characterized by a set of elementary units and their interactions, that can be represented by means of a network.
In Chapter 1 we discuss the network paradigm and we argument the reasons of its success in modern science and in the specific field of neuroscience. The main aim of this Chapter is to review the mathematical theory of graphs since it provides a useful formalization of the concept of network. This Chapter aims also to formalize the most relevant definitions about graphs in matrix form and, at last, to give special attention to the main network measures available for the analysis of networks.
Since the MRI scans are, nowadays, very accurate, researchers can define small brain regions of interest (ROIs): the higher the amount of ROIs, the greater the problem of dimensionality in the multiple comparisons. The problem of multiple testing in brain network construction is an open issue in applied statistical research.
In Chapter 2 we illustrate, from a methodological point of view, many of the estimators that have been historically proposed with respect to the probability of first kind and of second kind, given a set of multiple comparisons. Many traditional methods have good properties under the assumption of independence of test statistics; furthermore, the traditional approaches adopted in fMRI data analysis usually allow to control only for the false rejections, without providing any information on the error of second kind or on the test power.
In this regard, in Section 2.3 of Chapter 2 we propose, in an innovative way, a definition of the false non-discovery rate and of the power (BP) in a Bayesian perspective (Berlingeri, 2015), in analogy with Efron (2010) which introduced the Bayes false discovery rate (FDR). In addition, we explore the main finite properties of the estimates of FDR and BP; we investigate also the asymptotic properties of those estimates, both with theoretical considerations and with simulation studies, both under the hypothesis of independence of the p-values and under some assumptions of dependence.
In Chapter 3, we adopt a combination of the most recent techniques and tools to analyze resting state fMRI data, with the innovative Bayesian solution to control for multiple testing problems and with graph theory. This is done to explicitly testing two alternative theories about age-related changes in resting state functional connectivity, namely the de-differentiation hypothesis and the localization hypothesis. The de-differentiation hypothesis would be due to a less efficient neuromodulation of neuronal noise that would cause more

dense and less sparse neuronal networks; the localization hypothesis, conversely, would be due to an increment of neurofunctional connections between brain regions that belong to the same anatomical macro-structures.
Thirty-five young and thirty-five elderly participants underwent a resting state fMRI study. For each participant we extracted 116 time-series that represented the spontaneous low frequency fluctuations of the 116 regions Automatic Anatomical Labelling Atlas and we computed, by means of an empirical Bayesian approach, subject-specific balanced network, i.e. a network in which the false discovery rate and the power were balanced. Each single-subject network was then described by means of network and distance measures. These measures were used to test between-groups differences.

Chapter 1

Network Analysis

1.1 Reasons for a successful paradigm

The network paradigm is increasingly common in many fields of research, such as neuroscience, information technology or sociology. Consequently, mathematicians and statisticians have spent many effort in improving the theoretical knowledge of this paradigm, in developing efficient algorithms of analysis and in investigating the properties and the characteristics of networks.
But why, precisely, this paradigm has been so successful?
Before introducing any mathematical concept or any formal definition about networks, the main aim of this Section, indeed, is to answer this question.
In this manner, we hope not only to be able to explain the reasons of the success of network models in science, but also, implicitly, to argue why we too have adopted this specific paradigm for our applied research on neurofunctional data.

In the *Oxford Dictionary*, a *network* is generically defined as a *collection of interconnected things*. If we analyze in depth this definition, we find out that it implies the knowledge and the specification of two different sets: the set of elements considered, namely the "things", and the set of interconnections between those elements.
The development and diffusion of the network paradigm reflects a "tendency towards a system-level perspective in science" (Kolaczyk and Csárdi, 2014). In fact, as Eric D. Kolaczyk effectively pointed out in his book "Statistical analysis of network data" (2014), in modern science there is an increasing interest in evaluating how the composing elements of a complex system cooperate and are interconnected: as a consequence the focus shifts from the analysis of each single element in its singularity, to the analysis of the whole system in its complexity.

In this regard, a network can be described as a mathematical tool for modeling and synthesizing complex systems.
In order to better understand the network paradigm we aim to explain what we precisely mean by complex system and, as a second step, we aim to formalize in mathematical language, by means of the theory of graphs, some basic concepts about networks.
The best way to clarify what a complex system is, seems to make a distinction between complex systems and complicated systems.
The adjective *complicated* derives from the Greek word *plèkein* and means *with folds*: as a consequence, the better way to understand a complicated system is to *unfold* it with a linear analysis. On the other hand, the adjective *complex* derives from the Greek word *plèko* and means *with connections*: network models, indeed, offer mathematical tools to properly analyze those connections.
Hence, a complex system is formally defined as an *open* system, namely a system that interacts with the environment, characterized by a set of elementary units and a set of connections between pairs of units.
As we have already underlined, in the last decades networks became an essential statistical method for the description and the analysis of complex systems: from an epistemological point of view, network models are a successful paradigm because they allow to hive off a complex system into simpler units and to study their significant connections.
A statistical network analysis can be effectively performed in many areas of application: there are numerous examples of studies carried out on different classes of complex systems such as neurological, social, economic, physical or computer science.
The mathematical theory of graphs provides the most effective "language" for the formalization of a network, as well as for the analysis of its features. The statistical analysis of networks, therefore, can not disregard from graph theory whose distinctive rationale will be extensively detailed in Section 1.2.

1.1.1 Brain networks as a special case of networks models

The human brain is a proper example of a complex system: as a matter of fact, it is composed by brain regions that are anatomically connected and that cooperate both in performing specific tasks or in resting state conditions.
In neuroscience, Functional Magnetic resonance imaging (fMRI) analysis have become more and more widespread, both due to the increasing availability of

the fMRI scanners and due to the increasing precision of the acquired brain images. Nowadays, a huge amount of brain imaging data is available[1] and researchers are looking for adequate methods of analysis.

In this regard, graph theory provides an effective way to represent a complex system: in the particular contest of neuroscience and neuroimaging, these representations, and the methodological and statistical procedures underlying graphs generation, are referred to as "brain network analyses".
The human brain, as a complex system, can be modeled as a network and can be graphically represented: a node of the brain network model corresponds to a specific brain region, while a link of the network corresponds to an interaction between two nodes.
With regard to the set of interactions between pairs of units or nodes, at least three different kind of connections (namely the anatomical, the functional and the effective connection) can be taken into consideration.
The *anatomical* connectivity refers to the physical and synaptic connections between neuronal elements.
The *functional* connectivity is defined as the temporal covariance (or correlation), or, more generally, as the deviation from statistical independence, that exists between distributed and often spatially remote neuronal units (Friston et al., 1993; Friston, 1994). The deviations from statistical independence between neuronal elements are commonly captured in a covariance matrix (or a correlation matrix), or equivalently in an adjacency matrix, to represent the functional connectivity of the explored system (Sporns, 2002).
At last, the *effective* connectivity refers to the impact of a brain region of a neuronal element on another.
Since the correlation, or covariation, concerns the intensity of the signal but not its direction, the functional connectivity is symmetric; on the contrary, the effective connectivity measures the asymmetric effect of a brain region on each other brain region and can be interpreted as a causal connection.

The main objective of a brain network analysis concerns the construction of a brain network, that is a map of the brain containing only the significant connections among brain regions of interest (Gold et al., 2010).

[1] The Human Connectome Project, for example, is a five years project sponsored by the Harvard Consortium, that aims to map the complete structural and functional neural connections across individuals and that provides lots of freely available fMRI data.

In the application on real fMRI data, whose results are illustrated in Chapter 3, we have restricted our attention on the analysis of the pattern of functional connectivity. We excluded the anatomical connectivity as we did not have any interest in investigating the synaptic connections between neurons; on the other hand we excluded the effective connectivity because we were not interested in the analysis of causal connections.
In fact, the practical interest that motivated first this study concerns the construction of a network capable of describing and synthesizing the pattern of functional connectivity of a human brain.

1.2 Graph Theory

The mathematical theory of graphs provides a useful formalization of the concept of network, defined as set of elements and their interconnections, and also a useful way of representation.
Historically, Euler is recognized as the first mathematician to have solved a problem, the famous problem of the seven bridges of Königsberg, with a argumentation based on graphs.
In order to determine if there exists a close walk that crosses exactly once each of the seven bridges, Euler formalized the problem by drawing an undirected graph such that each node corresponds to a shore while each link corresponds to a bridge (see Alexanderson, 2006; Gribkovskaia, Halskau, and Laporte, 2007, for an historical review). Although this paper by Euler was his only contribution to graph theory, this solution was a precursor to a new way of thinking a system as a set of interconnected objects, namely as a complex system.

1.2.1 Main Definitions

A *graph*, denoted by $G = (V, E)$ (in accordance to the notation in Drton, 2009), is defined as a mathematical object consisting of a finite set of nodes (or vertices equivalently) V and a finite set of links (or edges equivalently) E, such that $E \subseteq (V \times V)/\{(v, v) : v \in V\}$.
The number of nodes of a graph G, $N_v = |V|$, represents its *order* while the number of links, $N_e = |E|$, represents its *size*.

If we restrict our attention to brain networks analysis, each anatomical brain region corresponds to a node of the brain network, while the connection between two anatomical regions is represented by means of a link.
A graph can be classified, with respect to the characteristics of its links, as a *directed* graph or *digraph*, whenever all its links are directed, or as an *undirected* graph otherwise. A link $(v, w) \in E$ is said to be *directed* if $(w, v) \notin E$, and is denoted by $v \to w$; on the contrary a link is said to be *undirected* if $(w, v) \in E$, and is denoted by $v - w$. If $v \to w$ then the node v is a *parent* of w that is said *children*, while if $v - w$ then v and w are said *neighbors*.

Usually in brain network analysis, and also in the application on real fMRI data illustrated in Chapter 3, the main objective concerns the construction of an interpretable brain map made up by functional connections between couples of brain regions. The functional connectivity can be quantified by means of covariation's or correlation's measures that are symmetric definite. As a consequence, a link has not a direction and the resulting brain network is undirected. For this reason, in what follows, we will provide definitions and hints only with regards to undirected graphs (for a discussion of directed graphs see Drton, 2009; Kolaczyk and Csárdi, 2014).

Given a set of nodes, the pattern of connectivity among nodes depends on the set of links: in order to fully describe the characteristics of a graph, hence, it is of interest to formalize such a pattern of connectivity by means of specific definitions.
Two nodes in V are said to be *adjacent* if they are connected by a link in E; in an analogous way, two links are said to be *adjacent* if they are connected by a common node referred to as endpoint.
A node is said to be *incident* on a link if it is an endpoint of that link. Given those primary notions of adjacency and incidence, it follows the intuitive concept of degree of a node. Formally, the *degree* d_v of a node $v \in V$ is equal to the number of links incident on v. Intuitively, the node degree can be interpreted as a measure of connectivity of a graph: the higher the degree of a node, the more interconnected the node with the graph.

A graph is said to be *regular* if any node has the same degree (if the common degree is equal to d the graph is said *d-regular*). The attribute 'regular' refers to the fact that, in that particular case, the same level of connectivity is

shared by all the nodes in the graph.

Having introduced some fundamental concepts about connectivity in a graph, as a next step it is interesting to formalize the concept of paths between nodes. In this regard, a *walk* in the graph G is a sequence of nodes connected by a sequence of links; the length of a walk is denoted by the lowercase letter l.
A *path* is a particular case of walk characterized by a set of distinct nodes $v_0, \ldots, v_k$ such that v_i and v_{i+1} are adjacent for all $i = 1, \ldots, k$; differently a *trails* is a walk in which the links are adjacent and distinct.
Furthermore, a *cycle* is a path that begins and ends in the same node: if a graph has no cycles then is said to be *acyclic*. A directed graph with no cycles is synthetically referred to as DAG (*Directed Acyclic Graph*).
Whenever exists a walk from a node $u \in V$ to another node $v \in V$, then the node v is said to be *reachable* starting from u. If the property of reachability is owned by any node in the graph, then the graph G is said to be *connected*: specifically, starting from any node in the set V, any other node of the graph is reachable.

Of sure interest is the quantification of the *distance* between two nodes $u, v \in V$ of a graph by means of the total amount of links that connect them. Usually, however, there exists more than one path between u and v: the praxis, in this case, is to define the distance as the length of the shortest path between the nodes. Conversely, if in a graph does not exist a path between v and u, then the distance between this couple of nodes is conventionally set equal to infinity. The *diameter* of a graph is equal to the length of its longest path.

Furthermore, we aim to introduce an extension of undirected graphs by defining the *weighted* graphs. Formally, a graph is said to be weighted if each link has an associated weight. A link between two nodes denotes the presence of a connection, while the corresponding weight quantifies the intensity of this connection. By way of example, in brain networks the presence of a link between two brain regions reveals the existence of a functional connection, an the intensity of the signal can be quantified by means of the correlation or covariation index. All the definitions and notions previously introduced can be easily extended to weighed graphs. As an example, the degree of a node becomes equal to the sum of weights of the links incident on that node.

Finally, a *subgraph* $H = (V_H, E_H)$ of a graph $G = (V, E)$ is also a graph in which $V_H \subseteq V$ and $E_H \subseteq E$. A particular case of subgraph is the *induced subgraph*, that is a subgraph $G' = (V', E')$ of a graph $G = (V, E)$ such that $V' \subseteq V$ is defined *a priori* and $E' \subseteq E$ is the set of links in G among the set of nodes V'.
Let us consider a not connected graph G, if there exists a subgraph of G that is connected then this latter is defined as a *component* of G.

A graph is said to be *complete* if every node is connected to every other node by a link: in an analogous way a *clique* of a graph G is a complete subgraph of G. A clique of G is maximal if it is not contained in any other complete subgraph.

1.2.2 Matrix Algebra

Having defined the graph $G = (V, E)$ and its main characteristics (Section 1.2.1), as a second step one may aim to rigorously attend some analysis on that graph. In this regard, it may be useful to formalize the information contained in the graph by converting it in matrix form.
A graph, as we have introduced yet, is a set V of nodes connected by a set E of links. The information about the pattern of connectivity of a graph can be effectively stored in a symmetric matrix $\mathbf{A}$, called *adjacency matrix*:

$$A_{i,j} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \textit{otherwise} \end{cases}$$

The adjacency matrix has dimensions $N_v \times N_v$, where N_v is the number of nodes in the graph. If two nodes are connected by a link in the graph, then the corresponding entry of $\mathbf{A}$ is set equal to 1, otherwise it is set equal to 0.

The adjacency matrix, as a storage element, contains all the informations about the connectivity on the original graph; in addition, the matrix algebra enables to attend many operations and analysis on the network in an effective and simple way, in particular in an IT perspective.

By way of example, many of the definitions introduced in the previous Section (1.2.1), can be converted in algebraic operations on the adjacency matrix. A simple example concerns the concept of degree d_i associated to node i. In

graph theory d_i is formally defined as the number of links that connect node i to the remaining nodes of graph G; such a definition is clear and easy to understand but it does not provide any operational rule to compute d_i. Given a graph $G = (V, E)$, one would have to count, for each node, how many nodes are connected with it by a link. Having converted the original graph in an adjacency matrix, the computation of d_i comes down to the summary $A_{i\cdot} = \sum_j A_{ij}$; such a summary can be performed equivalently by row or column because of the symmetry of $\mathbf{A}$.

Another interesting example of graphs analysis by means of matrix algebra concerns the quantification of the number of walks of same length in a graph. In fact, recalling that a walk is a sequence of nodes connected by a sequence of links, one may be interested in evaluating how much walks of length r connect the nodes i and j in the graph. Similarly to the previous example on the degree, with the sole graph theory this operation can be very demanding. On the contrary, given an adjacency matrix associated to the original graph, the number of walks of length r is simply equal to A^r_{ij}, where $\mathbf{A}^r$ is the r-th power of the adjacency matrix.
It appears clear, hence, that the matrix algebra applied on graphs provide an effective mean of analysis; furthermore, the operational rules on an adjacency matrix can be converted easily in computer language.

In the following Section 1.3 on network measures, we will introduce some relevant indicators that synthesize and summarize the topology of a graph; because of the effectiveness and because of the simplicity of the matrix algebra applied to graph theory, we will define all the measures of interest in terms of the adjacency matrix of a graph, by adopting directly the algebraic notation rather than the graphs notation.

1.2.3 Graph representation

A graph is a mathematical object containing many informations on the pattern of connectivity of a set of elements. In the adjacency matrix the whole pattern of connectivity is translated in algebraic language, in order to facilitate the computation of synthetic indicators about the characteristics of the graph. Moreover, despite the adjacency matrix provides a full description of the graph, it appears necessary also to represent it by means of a visual image. A visual representation of a graph guarantees an intuitive understanding; this

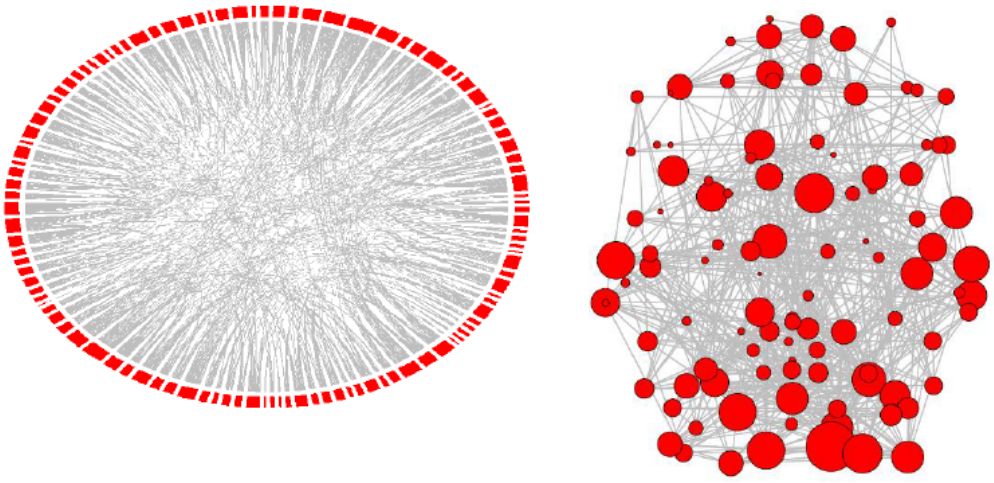
is particularly relevant as the network analysis is extensively adopted in multidisciplinary research.

A well-established representation of a graph provides that:

- each node is represented as a full-fill dot, at times labeled with the node's name or ID;
- each link is represented as a line (for weighted graphs, usually the thickness of the line is proportional to the weight of the link).

In the particular contest of brain network representations, it is quite common to represent the graph in stereotactic coordinates. As an alternative, brain networks can be represented by a chord diagram, that is a particular kind or graphical representation such that the nodes are displayed radially around a circle while the links are drawn as arcs rather than straight lines.

FIGURE 1.1: On the left a brain network is represented with a chord diagram, on the right the same brain network is represented in stereotactic coordinates



1.3 Network Measures

Once a complex system has been represented by means of a network, its graph topology can be quantitatively described by a wide variety of measures. As a consequence, the purpose of this Section is to introduce some relevant network measures that have been extensively described in the wide literature on Networks.

The network measures are synthetic indicators of specific network characteristics and provide a description of the their main properties.

In the specific contest of neuroimaging, for instance, when constructing a brain network, researchers have usually to deal with a huge number of nodes and links. The brain area scanned with a magnetic resonance is, in fact, partitioned into at least 90 brain regions; as a consequence, the total amount of possible links is equal to 4005. It appears clear, hence, that the sole graphical representation is not sufficient to understand the characteristics of the network: as a result, the analysis of the characteristics of a brain network requires inevitably the selection of a set of proper indicators.

1.3.1 Nodes and links characteristics

In Subsection 1.2.1 we have already introduced some synthetic indicators that provide a description of the characteristics of nodes and links. Let us recall, for example, the definitions of *order* an *size* of a graph; the order of a graph is equal to the number of rows, or columns by symmetry, of the adjacency matrix, while the size of an unweighted graph corresponds to the number of "1" values that are displayed above the main diagonal of the adjacency matrix (Kolaczyk and Csárdi, 2014). While the order is a nodes characteristic, the size is a links characteristic: those two measures, taken together, provide a first evidence of the 'dimension' of the graph.

A second indicator of nodes characteristics is represented by the *degree* of a node $v \in V$, that is equal to the number of links that connect v to the remaining nodes of the graph. The degree of a node quantifies properly the level of connectivity of a specific node in a graph. Each node is associated with its degree and, as a consequence, the definition of *degree distribution* results naturally. Let us define, indeed, f_d the fraction of nodes with degree equal to d

$$f_d = \frac{\#\{v \in V | d_v = d\}}{|V|}$$

the collection of f_d for all possible $d \geq 0$ characterizes the degree distribution of a graph. Because of the Glivenko-Cantelli theorem, the empirical degree distribution of a graph is a good estimate of the unknown degree distribution. Given the degree distribution, there exist many standard statistical approaches to synthesize it. By way of example, an histogram or a box plot are proper

graphical tools that allow to individuate the modal class or asymmetries of the distribution. The arithmetic mean $\bar{d} = \sum_{v \in V} d_v$ or the median can be calculated, furthermore, as measures of central tendency of the degree distribution, while the standard deviation quantifies the amount of dispersion among nodes degrees.

Finally, another measure that is related to the degree distribution of a graph is the *number of hubs*. An hub is generally defined as a fulcrum of a network, as its pivot node: formally, in the literature about network measures, an hub is defined as a node whose degree is greater than the mean network degree (Wijk, Stam, and Daffertshofer, 2010). Obviously, in a graph we can count more than one hub; the total number of hubs, in fact, represents a measure of network efficiency; the higher the number of hubs, the more interconnected the nodes of the network (Heuvel and Sporns, 2013).

1.3.2 Measures of centrality

The concept of *centrality* is referred to the relevance that a node has in the graph structure. Intuitively, it is easy to understand what the concept of relevance of a node in a graph means; however, if we were asked to formalize this concept, it would be clear that there is not a unique definition of *relevance*. Freeman (1979), while investigating the concept of centrality in social networks, pointed out that 'there is certainly no unanimity on exactly what centrality is or on its conceptual foundations, and there is little agreement on the proper procedure for its measurement.' At each formal definition of *relevance*, therefore, corresponds a suitable network measure and in what follows we will explore some well-established measures of centrality.

The measure of *closeness centrality* enables to quantify how much a node is close to other nodes in the graph. Formally, the closeness centrality (Freeman, 1979) of node v is equal to

$$L_v = \frac{N_v - 1}{\sum_{u \in V} d_{u,v}}$$

where $d_{u,v}$ is the shortest path between nodes u and v. With respect to this measure of centrality, the concept of relevance is interpreted in terms of distance of a node to the remaining nodes of the graph.

The measure of *betweenness centrality* is also related to the paths structure in a graph but, in addition, takes into account the position of a node with respect to the other nodes. The formal definition of this measure, for node v, is given by:

$$B_v = \frac{1}{(N_v - 1)(N_v - 2)} \sum_{s \neq t \neq v \in V} \frac{\sigma_{s,t|v}}{\sigma_{s,t}}$$

where $\sigma_{s,t}$ is equal to the number of shortest paths between nodes s and t while $\sigma_{s,t|v}$ is equal to the number of shortest paths, constrained to pass through v, between nodes s and t.

1.3.3 Measures of local density

A relevant property of graphs, that is particularly interesting when dealing with brain networks, is the property of *density*, namely the tendency of a graph to be almost fully connected. Usually such a tendency can be observed only for a subset of nodes of a graph and, for this reason, it is preferable to consider the property of *local density*.

In graph theory, a *clique* is a complete subgraph of the original graph G. In order to identify the pattern of localized density, it may seems appropriate to identify all the cliques of G. However, in real networks, is uncommon to observe cliques made up by more than three nodes, except for highly globally dense networks: as a consequence, it appears appropriate to weaken the definition of clique.
In this regard, it is interesting to define the concept of *core* of a graph G that is a subset of G composed by nodes with similar degree: by way of example a *k-core* is a subset of G such that its nodes have degree at least equal to k.

An alternative measure of local density quantifies, with a measure between 0 and 1, how a subgraph H is close to be a clique.

$$den(H) = \frac{|E_H|}{|V_H|(|V_H| - 1)/2}$$

where $|V_H|$ and $|E_H|$ are respectively the order and the size of H.

1.3.4 Measures of segregation

An interesting operation on the nodes of a graph consists in clustering them into groups or modules. Many of the common clustering algorithms that exist in the statistical literature may be applied to this specific contest of analysis (Rokach and Maimon, 2005). Consider, by way of example, the hierarchical clustering that classified the objects of interest, the nodes of a graph in our application, into a set of finite modules: usually the clustering analysis can be performed with an agglomerative algorithm, namely an algorithm that merges into the same cluster the most similar elements, or with a divisive algorithm, namely an algorithm that separates into different clusters the most dissimilar elements. An extension of the classical hierarchical algorithm has been proposed by Cattinelli et al. (2013), and is invariant to data ordering.

Given a set of modules, one may be interested in evaluating how much such a partition captures a nontrivial module structure.
In this regard the Newman's modularity measure Q (Newman, 2006) is defined as:

$$Q = \frac{1}{N_e} \sum_{i,j \in V} \left(a_{ij} - \frac{a_{i.} a_{.j}}{N_e} \right) \delta_{m_i, m_j} \tag{1.1}$$

where m_i is the module containing node i and δ_{m_i, m_j} is an indicator function that assumes value 1 if $m_i = m_j$ and value 0 otherwise.
Such a measure quantifies the modular structure beyond the expected at random one and ranges between -1/2 and 1 (Brandes et al., 2008).

Chapter 2

Multiple Hypotheses Testing

2.1 Focusing the problem

Traditionally, when testing a null hypothesis against a single alternative, the objective is to maximize the power of the test by checking the error of first kind, that is the probability of rejecting the null hypothesis when it is true. Therefore, similarly to the well-established paradigm for single hypothesis, when dealing with multiple comparisons we are interested in evaluating the rate of false discoveries and the power.

The problem of multiple comparisons has been historically introduced by Tukey and Scheffé in the 1950s (Benjamini and Braun, 2002) and has aroused new interest with the diffusion of genomic studies that allow to compare the expression levels of thousands of genes simultaneously (Dudoit, Shaffer, and Boldrick, 2003).
In many applications may be needed to simultaneously test a finite number of hypotheses. In particular, neuropsychologists have to deal with a huge amount of multiple comparisons, by way of example in the analysis of functional Magnetic Resonance Imaging (fMRI), while constructing brain maps or brain networks, that describe and summarize the pattern of brain activations (Lindquist, 2008).
The practical interest that motivated first this research project concerns, in fact, the construction of a network capable of describing and synthesizing the pattern of functional connectivity of a human brain. As already pointed out, in neuroimaging literature the human brain connectivity is commonly classified into anatomical, functional and effective connectivity (Friston, 1994; Sporns, 2002).
With regard to this specific study, we focused our attention on the analysis of the pattern of functional connectivity, that is measured as the correlation

among couples of brain regions (Sporns, 2002); there exist many statistical measures that allow to quantify such a correlation as the Pearson's correlation coefficient or the Spearman's rank coefficient. The pattern of functional connectivity is usually summarized into a brain network, that can be interpreted as a map of the brain composed only by the significant signals among brain regions of interest (Gold et al., 2010). In particular, as we mean to construct an interpretable brain map made up only by effective connections between couples of ROIs, only signals higher than a proper threshold are kept. The resulting brain network is, hence, unweighted and, since the correlation concerns the intensity of the signal but not its direction, also undirected.
From a statistical point of view, the determination of which signals are statistically significant with respect to a proper threshold results in a multiple testing problem. In fact, each region is compared with the remaining brain regions and all the comparisons have to be performed conjointly: given k brain regions, the correlation coefficients to be tested are $\frac{k(k-1)}{2}$. As a consequence, the more the total amount of brain areas, the more the total amount of simultaneous comparisons, the more the level of detail of the resulting brain network.
More formally, let us consider $r_1, \dots, r_j, \dots, r_m$ distinct correlation coefficients, where m is the number of conjoint tests; we are interested in the simultaneous comparison of the following hypotheses for the correlation parameter ρ_j corresponding to r_j,

$$H_0 : \rho_j = 0 \qquad vs \qquad H_1 = \bar{H}_0 : \rho_j \neq 0$$

The statistic test t_j is, hence, given by,

$$t_j = \frac{r_j}{\sqrt{1 - r_j^2}} \sqrt{n-2} \qquad for\ j = 1, \dots, m \tag{2.1}$$

(Gibbons and Chakraborti, 2011). Under H_0, the test statistic t_j for each comparison is assumed approximately distributed as a Student's t with $n-2$ degrees of freedom (Zimmerman et al., 2006), where n is the length of the originating time series for each brain region.
We usually define the same rejection region for each simultaneous test, equivalently with respect to the statistic test or to its associated p-value,

$$\Gamma = \{|r_j| \geq \tau\} = \{p_j \leq \gamma\}$$

We point out that τ is the rejection threshold on the basis of which we

constructed the brain network. The more τ is high, close to 1, the more the resulting brain network is conservative; in fact as τ increases, the number of links included in the brain network decreases. In the literature concerning brain networks construction with a multiple comparison perspective does not yet exist a common approach to identify a suitable threshold. In Section 2.2 we review the well-established literature on multiple comparisons, while in Section 2.3 we propose a method of threshold's selection that controls the probability of false discoveries and the power of the comparisons in a bayesian perspective.

2.2 Traditional Methods

Let us consider the general problem of simultaneously testing $H_1, \ldots, H_m$ hypothesis.
The easiest solution to this multiple testing problem could consist in ignoring its multiplicity by testing, individually, each hypothesis at a fixed level α. However, we can observe that with such a procedure, as m increases also the probability of false rejections rapidly increases: as a consequence the tests are not indeed controlled at level α.

TABLE 2.1: Probability of at least one false rejection, P, as m increases.

m	1	2	5	10	15	20	30	40	50	100
P	0.05	0.10	0.23	0.40	0.54	0.64	0.79	0.88	0.92	0.99

It is evident, hence, that the approach of individually testing each hypothesis is not effective: an appropriate approach, as a consequence, has necessarily to take into account the multiplicity of the problem, for example by controlling right the probability of at least one false rejection.

2.2.1 Family-wise error rate

In the light of the results reported in Table 2.1, the probability of at least one false rejection seems the probability that deserves to be controlled in order to manage the family of multiple comparisons.
Since, for a single hypothesis the probability of false rejection is set less or

equal to α, in an analogous way, for m simultaneous hypotheses, the probability of one or more false rejection, referred to as *FWER*, can be naturally set not greater then a fix level.
The acronym FWER stands for *family-wise error rate* where the term "family", in particular, refers to the finite set of hypotheses considered.

In this Subsection we desire to recall several methods based on FWER that have been proposed in the last century (Lehmann, 2006) and that still are adopted in numerous applications studies, in particular in the functional neuroimaging literature (Nichols and Hayasaka, 2003).

A statistical method for multiple comparisons that controls FWER is a method such that there exists a constant α for which:

$$FWER \leq \alpha$$

When dealing with multiple comparisons, it is usually preferable to define the statistical approach in terms of p-values of the individual tests, rather then in terms of test statistics; in this regard, the following Lemma assumes a particular relevance.

Lemma 1. *Let us define t_i as the test statistic associated to the $i-th$ null hypothesis and S_α as the rejection region at level α.*
For any $\alpha < \alpha'$, we assume that $S_\alpha \subset S_{\alpha'}$ and also that $P(t_i \in S_\alpha) \leq \alpha$ for all $0 < \alpha < 1$. The p-value is, hence, defined as:

$$p_i = inf\{\alpha \ : \ t_i \in S_\alpha\}$$

Hence, for the $i-th$ null hypothesis, the following properties hold:

1.
$$P(p_i \leq u) \leq u$$

2.
$$P(p_i \leq u) \geq P(t_i \in S_u)$$

and p_i is uniformly distributed on $(0,1)$.

Proof. Under the assumption of the i$-th$ null hypothesis to be true, the event $\{p_i \leq u\}$ implies $\{t_i \leq S_{u+\epsilon}\}$ for any small $\epsilon > 0$. Then, by assumption:

$$P(p_i \leq u) \leq P(t_i \leq S_{u+\epsilon}) \leq u + \epsilon \rightarrow u$$

as $\epsilon \rightarrow 0$ and so (*1.*) is given.
In order to demonstrate (2.) it is sufficient to point out that the event $\{t_i \leq S_u\}$ implies $\{p_i \leq u\}$ and so, by monotonicity:

$$P\left(t_i \in S_u\right) \leq P(p_i \leq u)$$

Finally, given S_α such that $P(t_i \in S_\alpha) \leq \alpha$ for all $0 < \alpha < 1$, p_i is uniformly distributed on $(0,1)$ under the null hypothesis. □

If a p-value verify the property (*1.*) of Lemma 1, then such a p-value is referred to as a *valid* p-value (Casella and Berger, 2002).

The first method we aim to define and explore is the Bonferroni method (Lehmann, 2006), a well-established approach that controls FWER.

Let us assume the $H_1, \ldots, H_m$ conjoint hypotheses to be independent and let us define the corresponding p-values $p_1, \ldots, p_m$.
The Bonferroni method provides that the hypothesis H_i is rejected if $p_i \leq \alpha/m$, where m is the total amount of multiple comparisons.
The Bonferroni procedure at level α implies that $FWER \leq \alpha$ for the family of m hypotheses simultaneously tested; such an implication can be easily demonstrated:

Proof. Let us define I as the subset of true hypotheses with $|I|$ denoting its cardinality. Then:

$$FWER = P(reject\ H_i, \forall i \in I) \leq \sum_{i \in I} P(reject\ H_i) =$$

$$= \sum_{i \in I} P\left(p_i \leq \frac{\alpha}{m}\right) \leq \sum_{i \in I} \frac{\alpha}{m} \leq \frac{|I|\alpha}{m} \leq \alpha$$

We recall that the first inequality is true because of the Boole inequality, that, for the generic event E_i, states that

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i) \tag{2.2}$$

□

The Bonferroni procedure is very simple but it does not represent a proper approach in case the number of simultaneously tested hypotheses, m, is high. In fact, each hypothesis H_i is tested at level α/m, and this quantity is as smaller as m increases, since α is fixed: therefore, given m very high, there would be a tendency to accept nearly all the hypotheses and, as a consequence, to accept also the false hypotheses. Precisely for this reason, the Bonferroni procedure is usually criticized as a too much conservative procedure.

In order to overcome the characteristic of the Bonferroni procedure to be excessively conservative, there have been proposed, in the statistical literature, many methodologies for multiple comparisons based on a *step down* iterative procedure: the Holm procedure described below is a relevant example of this kind (Holm, 1979).

Given m conjoint hypotheses, the notation $H_{(1)}, \ldots, H_{(m)}$ is referred to the ordered hypotheses and, similarly, the notation $p_{(1)} \leq \cdots \leq p_{(m)}$ is referred to the corresponding ordered p-values.
The Hold procedure is a step-down iterative procedure that can be outlined step by step as follows:

(i) The hypothesis H_1, that corresponds to the hypothesis with the smallest p-value, is tested at level α/m. If $p_{(1)} \geq \alpha/m$, then all the hypotheses are accepted and the procedure ends; otherwise if $p_{(1)} < \alpha/m$, the hypothesis $H_{(1)}$ is rejected and the procedure continues to step (ii);

(ii) The hypothesis H_2 is tested at level $\alpha/(m-1)$. If $p_{(1)} < \alpha/m$ and $p_{(2)} \geq \alpha/(m-1)$, then the hypotheses $H_{(2)}, \ldots, H_{(m)}$ are accepted and the procedure ends; otherwise, if $p_{(1)} < \alpha/m$ and $p_{(2)} < \alpha/(m-1)$ the hypothesis $H_{(2)}$is rejected and the procedure continues to step (iii);

(iii) The procedure is iterated until, at step k, it occurs that $p_{(k)} \geq \alpha/(m-k+1)$: the first $k-1$ ordered hypotheses are rejected, while the remaining are accepted.

In what follows it is demonstrated that the Holm procedure satisfies the inequality $FWER \leq \alpha$:

Proof. Let us define I as the subset of true hypotheses, and j as the smallest random index such that:

$$p_{(j)} = \min_{i \in I} p_i$$

Clearly $j \leq m - |I| + 1$. If the following holds:

$$p_{(1)} \leq \alpha/m, p_{(2)} \leq \alpha/(m-1), \ldots, p_{(j)} \leq \alpha/(m-j+1)$$

Then the procedure commits a false rejection and, as a consequence, it holds that:

$$\min_{i \in I} p_i = p_{(j)} \leq \alpha/(m-j+1) \leq \alpha/|I|$$

In conclusion, by the Boole inequality, the probability of erroneously rejecting $p_{(j)}$ is less of equal than α:

$$P\left(\min_{i \in I} \hat{p}_i \leq \alpha/|I|\right) \leq \sum_{i \in I} P\left(\hat{p}_i \leq \alpha/|I|\right) \leq \alpha$$

And hence the Thesis is given. □

The Holm procedure can be generalized in a general step-down procedure that consists in determining if the most significant test has to be rejected.
In particular, given the $p_{(1)} \leq \cdots \leq p_{(m)}$ ordered p-values, let us consider also a set of ordered constant values $\alpha_1 \leq \alpha_2 \leq \ldots \alpha_m$.
The first step of this general step-down procedure consists in verifying if $p_{(1)} \geq \alpha_1$: if the inequality is satisfied, then all the hypotheses are accepted; otherwise if

$$p_{(1)} < \alpha_1, \ldots, p_{(r)} < \alpha_r$$

then the first $r \leq m$ hypotheses are rejected.
In case the generic α_i is set equal to $\frac{\alpha}{m-i+1}$, than the Holm procedure is obtained again.

A different iterative procedure for multiple comparisons, is the *step-up* procedure that, as opposed to the step-down procedure, starts by determining if

the least significant p-value has to be accepted. More formally, the entire set of hypotheses is rejected if $p_{(m)} < \alpha_m$, otherwise if:

$$p_{(m)} \geq \alpha_m, \ldots, p_{(r+1)} \geq \alpha_{r+1}, p_{(r)} \geq \alpha_r$$

the first $r-1 \leq m$ hypotheses are rejected.

The procedures based on FWER control for at least one false rejection; this approach can be extended by admitting more then one false rejection, expecially if m is large, as commonly occurs in many applications. Such a procedure is referred to as the $k-$FWER procedure because it controls for k or more false rejections (Lehmann and Romano, 2012).
In analogy with the definition of FWER, the $k-$FWER quantity is defined as:

$$k - FWER = P(reject\, at\, least\, k\, hypotheses\, H_i\, ,\ \ \forall i \in I)$$

Having defined the probability of at least k false rejections, Lehmann and Romano (2012) also introduced a step-down procedure that controls $k-$FWER at level α, namely a procedure such that $k - FWER \leq \alpha$.
Interestingly, the procedure based on $k-$FWER holds without any assumptions on the dependence structure of the p-values, while all the procedures that control FWER require the assumption of independence among p-values.
As it requires less restrictive assumptions, the procedure based on $k-$FWER can be interpreted as a less severe procedure with respect to procedures based on FWER and is preferable whenever such an independence hypothesis on the p-values is violated.

The first procedure proposed by Lehmann and Romano (2012) can be interpreted as an extension of the Bonferroni method in case one aims to control $k-$FWER rather than FWER.
Let us consider a general multiple testing problem, and let us adopt the same notation as before; such a procedure rejects any H_i if, for its corresponding p-value, it occurs that:

$$p_i \leq \frac{k\alpha}{m}$$

It is easy to prove that this procedure controls $k-$FWER at level α, because of the Markov inequality:

Proof. Let us define I the subset of true hypotheses and N the number of false rejections. Then by the Markov inequality:

$$P(N \geq k) \leq \frac{\mathbb{E}(N)}{k} = \frac{\mathbb{E}\left(\sum_{i \in I} \mathbf{1}_{p_i \leq \frac{k\alpha}{m}}\right)}{K} = \sum_{i \in I} \frac{P\left(p_i \leq \frac{k\alpha}{m}\right)}{k}$$
$$\leq \sum_{i \in I} \frac{k\alpha/m}{k} = |I| \frac{\alpha}{m} \leq \alpha$$

□

With respect to FWER, we have introduced the step-down procedure, and in particular the procedure proposed by Holm, as an extension of the simpler Bonferroni method. In an analogous way, with respect to $k-$FWER, the procedure we have just described can be easily extended into a generic step-down approach. Given the constant value α_i such that:

$$\alpha_i = \begin{cases} \frac{k\alpha}{m} & i \leq k \\ \frac{k\alpha}{m+k-i} & i > k \end{cases}$$

The step-down procedure that controls $k-$ FWER at level α follows the same steps of the Holm procedure and, for this reason, can be interpreted as its counterpart. The following argumentation demonstrates that the procedure strongly controls $k-$ FWER.

Proof. Let us consider I as the subset of true hypotheses and $|I| \geq k$ (otherwise no proof is needed). Consider the ordered p-values corresponding to the true hypotheses and denoted as:

$$q_1 \leq \cdots \leq q_{|I|}$$

Consider, furthermore, j to be the smallest random index such that $p_{(j)} = q_{(k)}$. Th largest random index j is observed if the smallest p-values correspond to the $m - |I|$ false null hypotheses while the other p-values correspond to the true null hypotheses and hence, the following is verified:

$$k \leq j \leq m - |I| + k$$

In addition, if it holds that:

$$p_{(1)} \leq \alpha_1, p_{(2)} \leq \alpha_2, \ldots, p_{(j)} \leq \alpha_j$$

Hence, at least k false rejections are generated from the generalized step-down procedure; as a consequence:

$$q_{(k)} = p_{(j)} \leq \alpha_j = \frac{k\alpha}{m+k-j} \leq \frac{k\alpha}{|I|}$$

Finally, to sum up and conclude:

$$P\left(q_{(k)} \leq \frac{k\alpha}{|I|}\right) \leq \frac{k\alpha}{|I|} \leq \alpha$$

□

2.2.2 False discovery rate

The possible outcomes of a problem of m multiple comparisons can be effectively summarize as shown in Table 2.2:

TABLE 2.2: Summary of the possible outcomes in case of m hypotheses simultaneously tested.

Hypotesis	Accepted	Rejected	Total
Null true	U	V	m_0
Alternative true	T	S	m_1
	W	R	m

It is relevant to notice that R, the number of rejected hypotheses, is an observed random variable, while U, V, S e T are not observable random variables because, in real application problems, is unknown which null hypotheses are true. Furthermore, V represents the number of true null hypotheses erroneously rejected, hereafter referred to as false discoveries: the proportion of false discoveries (FDP) is, hence, defined as V/R.
According to the proposed notation in Table 2.2, the FWER quantity can be expressed as $P(V \geq 1)$ while the $k-$FWER can be expressed as $P(V \geq k)$.
Lehmann and Romano (2012) proposed an approach that controls the FDP at a given probability; in particular, given γ and $\alpha \in (0,1)$, a procedure such that:

$$P(FDP > \gamma) \leq \alpha$$

Benjamini and Hochberg (1995), in a pioneering paper in the field of multiple testing, have defined a new error measurement, called *false discovery rate* (FDR), that is equal to the expected value of the FDP:

$$FDR = \mathbb{E}\left(\frac{V}{R}\right)$$

In order to better understand the properties of the FDR, it is interesting to recall two important properties of the FDP:

1. If all null hypotheses are true, namely if $m_0 = m$, then $S = 0$ and $V = R$. If $V = 0$ no hypothesis is rejected and therefore the rate of false discoveries is equal to 0. If $V > 0$ then $\frac{V}{R} = 1$ and $FDR = P(V \geq 1)$ and thus is equivalent to control FDR or FWER .

2. If $m_0 < m$ and $V > 0$ then the proportion $V/R \leq 1$ and $P(V \geq 1) \geq FDR$: controlling FWER is the same as controlling FDR, but the vice versa does not hold. For this reason, any procedure that controls FDR can yield a power's increase as it is less conservative than procedures that control FWER. In particular, the greater is the number of null hypotheses that are not true the greater is S and, then, FDR is smaller if compared to FWER.

Having verified that controlling FDR guarantees an increase in the test power than controlling FWER, it is of sure interest, hence, to examine the controlling procedure for the FDR measurement error.

In accordance with the notation previously introduced on the set of m hypotheses and their correnspondingp-values and in accordance with the Bonferroni method, we aim to define a procedure such that the first k ordered hypotheses $H_{(1)}, \ldots, H_{(k)}$ are rejected if:

$$k = \max\{i : p_{(i)} \leq \frac{i}{m} q^*\} \tag{2.3}$$

Benjamini and Hochberg (1995) have demonstrated the following:

Theorem 2. *For each possible configuration of the null hypotheses, and under the assumption of independence of the test statistics, the procedure defined in 2.3 controls FDR at level q^*, namely it guarantees that $FDR \leq q^*$.*

The proof of Theorem 2 arises immediately from the following Lemma.

Lemma 3. *For all the first $0 \leq m_0 \leq m$ independent p-values that correspond to the true null hypotheses and for all the subsequent $m_1 = m - m_0$ p-values that correspond*

to the false null hypotheses, the multiple testing procedure defined in 2.3 satisfies the inequality:

$$\mathbb{E}\left[\frac{V}{R}\middle| P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}\right] \leq \frac{m_0}{m} q^*$$

where P_i refers to the p-value random variable, while p_i refers to its observed value and $p_1 \leq \cdots \leq p_{m_1}$ are ordered without loss of generality.

Proof. The Lemma is trivially true if $m = 1$.
By induction, let us assume that the Lemma is true for every $m' \leq m$; then the Lemma can be proven for $m + 1$.

If $m_0 = 0$, then all the null hypotheses are false and therefore the proportion V/R is identically equal to zero; as a consequence:

$$\mathbb{E}\left[\frac{V}{R}|P_1 = p_1, \ldots, P_m = p_m\right] = 0 \leq \frac{m_0}{m+1} q^*$$

On the other hand, if $m_0 > 0$, let us define P'_i, for $i = 1, \ldots, m_0$, the p-values corresponding to true null hypotheses and $P'_{(m_0)}$ the largest one; we recall that those p-values are independent random variables uniformly distributed between 0 and 1 (see 1).
Finally, let j_0 to be the largest j between 0 and m_1 such that:

$$p_j \leq \frac{m_0 + j}{m+1} q^*$$

and let $p'' = \frac{m_0+j_0}{m+1} q^*$.
By conditioning the expected value also on $P'_{(m_0)} = p$ it follows that:

$$\mathbb{E}\left[\frac{V}{R}|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m1}\right] =$$

$$= \int_0^1 \mathbb{E}\left[\frac{V}{R}|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m1}\right] m_0 p^{(m_0-1)}\, dp =$$

$$= \int_0^{p''} \mathbb{E}\left[\frac{V}{R}|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m1}\right] m_0 p^{(m_0-1)}\, dp$$

$$+ \int_{p''}^1 \mathbb{E}\left[\frac{V}{R}|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m1}\right] m_0 p^{(m_0-1)}\, dp$$

where $m_0 p^{(m_0-1)}$ is the density of the random variable $P'_{(m_0)}$.

Note that the integral has been split into two addends to highlight the integration point p''.

In the first integral, the integration interval is equal to $[0, p'']$ (i.e. $p \leq p''$): all the $m_0 + j_0$ hypotheses, then, are rejected and $\frac{V}{R} = \frac{m_0}{m_0+j_0}$. Furthermore, the expected value is constant and moves out of the integral while the primitive of the remaining density is trivially equal to p^{m_0}. In conclusion, the first integral is equal to:

$$\frac{m_0}{m_0 + j_0}(p'')^{m_0} = \frac{m_0}{m_0 + j_0}\left(\frac{m_0 + j_0}{m+1}\right)^{m_0}(q^*)^{m_0} = \frac{m_0}{m+1}q^*(p'')^{m_0-1}$$

In the second integral, the integration interval is equal to $[p'', 1]$ (i.e.$p > p''$): by analyzing together the m null hypotheses, both true and false, the i-th ordered hypothesis is rejected if it exists a k such that $p_{(k)} \leq [k/(m+1)]q^*$, with $i \leq k \leq m_0 + j - 1$, or, equivalently, it is rejected if:

$$\frac{p_{(k)}}{p} \leq \frac{k}{p(m+1)}q^*$$

By induction, given that the thesis is true for m, then for $m+1$ it holds that:

$$\mathbb{E}\left[\frac{V}{R}\middle| P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}\right] \leq \frac{m_0 - 1}{(m+1)p}q^*$$

Then, the second integral is less than or equal to:

$$\frac{m_0 - 1}{m+1}q^* \int_{p''}^{1} m_0 p^{m_0-2}) = \frac{m_0}{m+1}q^*[1 - (p'')^{m_0-1}]$$

In order to conclude, it is worth to notice that the thesis is true for $m+1$, under the assumption that it was true for m, because of the following inequality:

$$\mathbb{E}\left[\frac{V}{R}|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m1}\right]$$

$$\leq \frac{m_0}{m+1}q^*(p'')^{m_0-1} + \frac{m_0}{m+1}q^*[1 - (p'')^{m_0-1}] = \frac{m_0}{m+1}q^*$$

We want to remark that the hypothesis of independence among the test statistics has not been exploited in the proof of this Lemma.

□

Because of Lemma 3, Theorem 2 can be immediately demonstrated by applying the property of conditional expected values: hence, the procedure in

2.3 guarantees that FDR is controlled at level q^*.

The *acceptance/rejection* procedure defined in 2.3, with FDR controlled at level $q^* = \alpha$, is a step-down procedure. In fact, given the p-values in ascending order, the first step of such a procedure consists of comparing $p_{(m)}$ and α: if the m-th p-value is smaller than the type I error α, then all the null hypotheses are rejected; otherwise only the hypothesis $H_{(m)}$ is accepted and the procedure keeps on by comparing the $(m-1)-$th p-value and $\alpha(1-\frac{1}{m})$. The algorithm ends at step $i-$th if $p_{(i)} \leq \alpha(1-\frac{i-1}{m})$: in this case the hypotheses from the first to the i-th are rejected.

The proposed FDR controlling procedure can also be viewed as a posterior maximization, as proved by the following Theorem by Benjamini and Hochberg (1995).

Theorem 4. *The solution to the problem of maximum under constraints which leads to choose α such that $r(\alpha)$, the number of rejection at that level, is maximized under the constraint that $\alpha m / r(\alpha) \leq q^*$ is equivalent to the controlling procedure 2.3.*

Proof. In order to prove this Theorem it is sufficient to observe that, chosen α as the control level, if $p_{(i)} \leq \alpha \leq p_{(i+1)}$ then $r(\alpha) = i$.
As a consequence, by replacing the value i in the constraint, we obtain $\frac{\alpha m}{i} \leq q^*$ and, hence $\alpha \leq \frac{i}{m} q^*$. By recalling that procedure 2.3 leads to finding the $p_{(k)}$ ordered p-value such that $k = \max\{i : p_{(i)} \leq \frac{i}{m} q^*\}$, we deduce that $\alpha = p_{(k)}$ is the solution of the maximum under constraint problem. □

2.2.3 Positive False Discovery Rate

In their pioneering paper, Benjamini and Hochberg (1995) investigated some alternative formulations of the false discovery rate such as:

$$\mathbb{E}\left(\frac{V}{R}\middle| R > 0\right) \qquad (a)$$

$$\frac{\mathbb{E}[V]}{\mathbb{E}[R]} \qquad (b)$$

The authors motivated their propensity for FDR, rather than for the alternatives (a) and (b), by observing that if all the null hypotheses are true, i.e. $m_0 = m$, then $V/R = 1$ and so both (a) and (b) can not been controlled as they are identically equal to 1. On the contrary FDR is controllable in this case and, as remarked in the previous Section, controlling FDR

is equivalent to controlling FWER: in fact, FDR can be equivalently written as $FDR = \mathbb{E}\left(\frac{V}{R}\middle| R > 0\right) P(R > 0)$, and clearly it depends on $P(R > 0)$.

Differently from the methodological proposal of Benjamini and Hochberg (1995), that has been widely welcomed in the statistical literature, Storey (2002) and Storey and Tibshirani (2003) has further revolutionized the methodology concerning multiple comparisons by proposing the quantity (a) as the appropriate quantity to be controlled.
His methodological proposal originates, firstly, by noting that it is true that if $m = m_0$ then the rate of false discoveries is equal to 1, as observed by Benjamini and Hochberg (1995), but this seems an unlikely eventuality and a not so relevant case of study.
Furthermore, when at least one rejection has occurred, i.e. $R > 0$, and the significance level of control is equal to α, it is fundamental to observe that FDR is actually controlled at level $\frac{\alpha}{P(R>0)}$ and not at level α, as one might mistakenly suppose; on the contrary, when controlling the quantity (a), this misunderstanding is completely avoided.

Formally the *positive False Discovery Rate* (pFDR) is defined as:

$$pFDR = \mathbb{E}\left(\frac{V}{R}\middle| R > 0\right)$$

The adjective *positive* points out that the expected value of false discoveries is conditioned to positive values of the variable R, that represents the number of hypotheses rejected.

In case of m conjoint hypotheses $H_1, \ldots, H_m$, the corresponding independent test statistics are indicated with $T_1, \ldots, T_m$ and Γ represents a given rejection region.
The subscript 0 denotes the $i-$th null hypothesis H_{i0}, while the notation $\overline{H}_{i0}$ denotes the $i-$th alternative hypothesis.
Under the assumption of identical distribution between $T_i|H_{i0}$ and $T_i|\overline{H}_{i0}$, the hypotheses are independent Bernoulli random variables with $P(H_{i0}) = \pi_0$ and $P(\overline{H}_{i0}) = \pi_1$.
pFDR can be written equivalently as a function of the rejection region Γ:

$$pFDR(\Gamma) = \mathbb{E}\left(\frac{V(\Gamma)}{R(\Gamma)}\middle| R(\Gamma) > 0\right)$$

where $V(\Gamma) = \#\{T_i : T_i \in \Gamma | H_{i0}\}$ and $R(\Gamma) = \#\{T_i : T_i \in \Gamma\}$. As the conjoint hypotheses are Bernoulli random variables with parameters π_0 and π_1, then the following theorem holds (Storey, 2002):

Theorem 5. *Let us consider m identical hypotheses with independent test statistics $T_1, \ldots, T_m$ and a rejection region Γ. If the null hypotheses are true with an a priori probability equal to π_0, then pFDR can be written equivalently as:*

$$pFDR = \frac{\pi_0 P(T \in \Gamma | H_0)}{P(T \in \Gamma)} = P(H_0 | T \in \Gamma)$$

It is clear, hence, that pFDR has a natural Bayesian interpretation, under the assumption of independence.

Proof. At first, it is worth to observe what follows:

$$pFDR(\Gamma) = \mathbb{E}\left(\frac{V(\Gamma)}{R(\Gamma)} \middle| R(\Gamma) > 0 \right) =$$

$$= \sum_{k=1}^{m} \mathbb{E}\left(\frac{V(\Gamma)}{R(\Gamma)} \middle| R(\Gamma) = k \right) P(R(\Gamma) = k | R(\Gamma) > 0) =$$

$$= \sum_{k=1}^{m} \mathbb{E}\left(\frac{V(\Gamma)}{k} \middle| R(\Gamma) = k \right) P[R(\Gamma) = k | R(\Gamma) > 0]$$

Under the assumption of independence and identical distribution of the the test statistics, it holds that:

$$\mathbb{E}[V(\Gamma) | R(\Gamma) = k] = \mathbb{E}\left[\sum_{i=1}^{m} \mathbb{1}(T_i \in \Gamma) \mathbb{1}(H_{i0}) \middle| (T_1, \ldots, T_k) \in \Gamma, (T_{k+1}, \ldots, T_m) \notin \Gamma \right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{k} \mathbb{1}(H_{i0}) \middle| (T_1, \ldots, T_k) \in \Gamma, (T_{k+1}, \ldots, T_m) \notin \Gamma \right] =$$

$$= \mathbb{E}\left[\sum_{i=1}^{k} \mathbb{1}(H_{i0}) \middle| T_i \in \Gamma \right] = kP(H_0 | T \in \Gamma)$$

Hence, in conclusion:

$$pFDR(\Gamma) = \sum_{k=1}^{m} \frac{kP(H_0 | T \in \Gamma)}{k} P[R(\Gamma) = k | R(\Gamma) > 0] = P(H_0 | T \in \Gamma)$$

□

Under the assumptions of independence and identical distribution of the test statistics, pFDR is equivalent to $\frac{\mathbb{E}[V]}{\mathbb{E}[R]}$, as demonstrated in the following:

Theorem 6. *Under the same assumptions of Theorem 5, it holds the following:*

$$\mathbb{E}\left(\frac{V(\Gamma)}{R(\Gamma)}\middle| R(\Gamma) > 0\right) = \frac{\mathbb{E}[V]}{\mathbb{E}[R]}$$

Proof. Because of the independence and identical distribution of the m test statistics, it holds what follows:

$$\mathbb{E}[V(\Gamma)] = m\pi_0 P(T \in \Gamma | H_0)$$

and

$$\mathbb{E}[R(\Gamma)] = mP(T \in \Gamma)$$

Consequently:

$$\frac{\mathbb{E}[V]}{\mathbb{E}[R]} = \frac{\not{m}\pi_0 P(T \in \Gamma | H_0)}{\not{m} P(T \in \Gamma)} = pFDR(\Gamma)$$

□

Accordingly to the notation in Storey (2002), the measurement error pFDR has been expressed as a function of the test statistic T; in an equivalent way, it can be easily written as a function of p-values, a widespread custom in multiple testing methods.
The rejection region Γ is referred to the test statistics, under the assumption of identical distribution; in an equivalent way, the rejection region can be defined with respect to the p-values and is denoted by the interval $[0, \gamma]$, with $0 \leq \gamma \leq 1$.

The result of Theorem 5 can be written with respect to the parameter γ as follows:

$$pFDR(\gamma) = \frac{\pi_0 P(p_j \leq \gamma | H_0)}{P(p_j \leq \gamma)} = \frac{\pi_0 \gamma}{P(p_j \leq \gamma)}$$

where p_j represents the generic p-value random variable, under the assumption of identical distribution among p-values.

pFDR depends on π_0 that corresponds to the unknown a priori probability of a null hypothesis to be true. A conservative estimate of π_0, defined with

respect to a tuning parameter λ, is given by the following:

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_j > \lambda\}}{(1-\lambda)m} = \frac{W(\lambda)}{(1-\lambda)m} \tag{2.4}$$

On the other hand, a natural estimate of $P(p_j \leq \gamma)$ is given by the following plug-in estimator:

$$\hat{P}(p_j \leq \gamma) = \frac{\#\{p_j \leq \gamma\}}{m} = \frac{R(\gamma)}{m} \tag{2.5}$$

In conclusion, a proper estimate for pFDR, given the rejection parameter γ and the tuning parameter λ, is equal to:

$$\widehat{pFDR}_\lambda(\gamma) = \frac{\gamma W(\lambda)}{(1-\lambda)R(\gamma)} \tag{2.6}$$

We have already pointed out that for $R(\gamma) = 0$, the pFDR estimate can not be defined: in order to overcome this limitation, in such a specific case it may be convenient to replace $R(\gamma)$ with the constant 1, and the adopted notation is $R(\gamma) \vee 1$.

In addition, Storey (2002) proposed to conveniently adjust the estimate of pFDR for small samples.
It is worth to notice, in fact, that $1 - (1-\gamma)^m$ represents a lower bound of the probability $P(R(\gamma) > 0)$, because $\gamma = P(p_j \leq \gamma | H_0)$ and, consequently, $(1-\gamma)^m$ is equal to the probability of not rejecting any null hypothesis, under the assumption of independence among p-values.
Therefore, a good conservative estimate of pFDR for finite samples is:

$$\widehat{pFDR}_\lambda(\gamma) = \frac{\gamma W(\lambda)}{(1-\lambda)\{R(\gamma) \vee 1\}(1-(1-\gamma)^m)}$$

A proper estimate for FDR can be define in a similar way, just by recalling that FDR is not conditioned on the quantity $R(\gamma)$ to be strictly positive.

$$\widehat{FDR}_\lambda(\gamma) = \frac{\gamma W(\lambda)}{(1-\lambda)\{R(\gamma) \vee 1\}}$$

The estimation algorithm for $pFDR(\gamma)$ and $FDR(\gamma)$, proposed by Storey (2002), is characterized by the following steps:

1. First, compute the $p_1, \ldots, p_m$ p-values corresponding to the m multiple comparisons;

2. As a second step, compute the estimates of the a priori probabilities π_0 and of $P(p_j \leq \gamma)$, accordingly respectively to 2.4 and 2.5;

3. Then, for each rejection region $[0, \gamma]$, depending on γ, compute the estimate of $pFDR(\gamma)$ for a chosen λ, accordingly to 2.6;

4. Furthermore, adopt a bootstrap approach on the m p-values, by extracting B bootstrap samples of $p_1, \ldots, p_m$, and compute the B bootstrap estimates of pFDR, denoted as $\widehat{pFDR}^b_\lambda(\gamma)$;

5. Define a confidence interval of level $1 - \alpha$ for $pFDR(\gamma)$ by adopting the quantile of level $1 - \alpha$ of the estimates $\widehat{pFDR}^b_\lambda(\gamma)$ as the upper bound.

By means of an analogous step by step procedure, also $FDR(\gamma)$ can be estimated, just by recalling that:

$$\widehat{FDR}_\lambda(\gamma) = \frac{\gamma \hat{\pi}_0(\lambda)}{\hat{P}(p_j \leq \gamma)} \tag{2.7}$$

The estimation algorithm for pFDR guarantees an increase in the test power when compared to the algorithm proposed by Benjamini and Hochberg (1995). In order to demonstrate this statement, let us consider at first the m ordered p-values $p_{(1)}, \ldots, p_{(m)}$: the method proposed by Benjamini and Hochberg (1995) aims to determine the value k such that

$$k = \max\{i : p_{(i)} \leq \frac{i}{m} q^*\}$$

With such an approach, indeed, the first k ordered hypotheses are rejected and the FDR is controlled at level q^*.
On the contrary, the method proposed by Storey (2002) aims to estimate FDR accordingly to 2.7. Hypothetically, suppose to choose the most conservative estimate for the a priori probability of a null hypothesis to be true, namely $\hat{\pi}_0 = 1$. As a consequence, FDR is controlled at level q^* and the first l ordered hypotheses are rejected when:

$$l = \max\{i : \widehat{FDR}_{p(i)} \leq q^*\} = \max\{i : p_{(i)} \leq \frac{i}{m} q^*\}$$

We have shown, hence, that by adopting the most conservative estimate for π_0, then $l = k$ and so the two estimation methods lead to reject the same number of hypotheses at level q^*.

On the contrary, if the a priori probability is estimated with 2.4, that is a conservative estimate but not the most conservative one, then $l > k$ at level q^*. It seems evident, hence, that by adopting a *good* estimate for π_0, the estimation approach for FDR proposed by Storey (2002) leads to refuse a greater number of hypotheses than the estimation approach by Benjamini and Hochberg (1995) and, as a consequence, leads to a greater test power.

We have already introduced the parameter λ as a tuning parameter from which the estimate of the a priori probability π_0 depends. Rather than simply fix a constant value for λ, Storey (2002) proposed to select the optimal λ, within a finite set of possible values, by adopting an algorithm that minimizes the bootstrap mean square error of the pFDR estimates:

$$\lambda_{best} = \min_{\lambda \in [0,1]} \widehat{MSE}(\lambda)$$

If $pFDR(\gamma)$ were known, it would be possible to calculate directly the bootstrap MSE, depending only on λ, as:

$$\frac{1}{B} \sum_{b=1}^{B} [(\widehat{pFDR}_{\lambda}^{b}(\gamma) - pFDR(\gamma)]^2$$

However, the true $pFDR(\gamma)$ is not known and has to be estimated: because of Theorem 7, then $\forall \lambda \in [0,1]$

$$\mathbb{E}[\widehat{pFDR}_{\lambda}(\gamma)] \geq \min_{\lambda} \mathbb{E}[\widehat{pFDR}_{\lambda}(\gamma)] \geq pFDR(\gamma)$$

Therefore the chosen plug-in estimate for $pFDR(\gamma)$ is $\min_{\lambda} \mathbb{E}[\widehat{pFDR}_{\lambda}(\gamma)]$ and, in conclusion, the mean square error estimate is :

$$\widehat{MSE}(\lambda) = \frac{1}{B} \sum_{b=1}^{B} [(\widehat{pFDR}_{\lambda}^{b}(\gamma) - \min_{\lambda} \mathbb{E}[\widehat{pFDR}_{\lambda}(\gamma)]]^2$$

2.2.4 Properties of FDR and pFDR

Finite Properties

The aim of this Section is to show that the estimates 2.6 and 2.7 referred respectively to pFDR and FDR, are conservative and ensure a strong control of the quantities of interest ((Storey, 2002)). In this regard, it holds the following Theorem:

Theorem 7. $\forall\gamma$ *and* π_0*:*

$$\mathbb{E}(\widehat{pFDR}_\lambda(\gamma)) \geq pFDR(\gamma)$$

$$\mathbb{E}(\widehat{FDR}_\lambda(\gamma)) \geq FDR(\gamma)$$

Proof. The inequality for pFDR is demonstrated by recalling that:

$$\widehat{pFDR}_\lambda(\gamma) = \frac{\gamma W(\lambda)}{(1-\lambda)\{R(\gamma)\vee 1\}(1-(1-\gamma)^m)}$$

and by observing that pFDR can be equivalently written with the following formulation

$$pFDR(\gamma) = \mathbb{E}\left[\frac{V(\gamma)}{R(\gamma)\vee 1}\right]\frac{1}{P(R(\gamma)>0)}$$

Moreover, it has already been noted that $P[R(\gamma) > 0] \leq 1-(1-\gamma)^m$ under the assumption of independence of the test statistics.
For this reason, it holds that:

$$\mathbb{E}[pFD\hat{R}_\lambda(\gamma)] - pFDR(\gamma)$$

$$= \mathbb{E}\left[\frac{\gamma W(\lambda)}{(1-\lambda)\{R(\gamma)\vee 1\}(1-(1-\gamma)^m)}\right] - \mathbb{E}\left[\frac{V(\gamma)}{R(\gamma)\vee 1}\right]\frac{1}{P(R(\gamma)>0)}$$

$$\geq \mathbb{E}\left[\frac{\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)}{\{R(\gamma)\vee 1\}(P(R(\gamma)>0)}\right]$$

By conditioning the expected value on $R(\gamma)$

$$\mathbb{E}\left[\frac{\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)}{\{R(\gamma)\vee 1\}(P(R(\gamma)>0)}\middle| R(\gamma)\right] = \frac{\frac{\gamma\mathbb{E}[W(\lambda)|R(\gamma)]}{1-\lambda} - \mathbb{E}[V(\gamma)|R(\gamma)]}{\{R(\gamma)\vee 1\}P(R(\gamma)>0)}$$

Under the hypothesis of independence, recalling the definitions of the quantities of interest, we observe that: $\mathbb{E}[W(\lambda)|R(\gamma)]$ is a non-increasing function of $R(\gamma)$, because for increasing number of $\{p_i \leq \gamma\}$ does not increase the number of $\{p_i > \lambda\}$, while $\mathbb{E}[V(\gamma)|R(\gamma)]$ is a non-decreasing function of $R(\gamma)$, because for increasing number of $\{p_i \leq \gamma\}$ does not decrease the number of $\{p_i \leq \gamma\}$, given that the null hypothesis is true.

Accordingly with these considerations on the monotony, we can apply the Jensen's inequality, by conditioning on the positive values of $R(\gamma)$, and we get:

$$\mathbb{E}\left[\frac{\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)}{R(\gamma)P(R(\gamma)>0)}\middle| R(\gamma)>0\right] \geq \frac{\mathbb{E}\left[\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)|R(\gamma)>0\right]}{\mathbb{E}[R(\gamma)|R(\gamma)>0]P(R(\gamma)>0)}$$
$$= \frac{\mathbb{E}\left[\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)|R(\gamma)>0\right]}{\mathbb{E}[R(\gamma)]}$$

Furthermore, by noting that $\mathbb{E}[R(\gamma)] \geq P(R(\gamma)>0)$, the following inequality holds, by conditioning on $R(\gamma)=0$:

$$\mathbb{E}\left[\frac{\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)}{\{R(\gamma)\vee 1\}(P(R(\gamma)>0)}\middle| R(\gamma)=0\right] = \mathbb{E}\left[\frac{W(\lambda)\gamma}{(1-\lambda)P(R(\gamma)>0)}\middle| R(\gamma)=0\right]$$
$$\geq \mathbb{E}\left[\frac{W(\lambda)\gamma}{(1-\lambda)\mathbb{E}[R(\gamma)]}\middle| R(\gamma)=0\right] = \frac{\mathbb{E}\left[\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)|R(\gamma)=0\right]}{R(\gamma)>0}$$

In conclusion:

$$\mathbb{E}[pFD\hat{R}_\lambda(\gamma)] - pFDR(\gamma) \geq \mathbb{E}\left[\frac{\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)}{\{R(\gamma)\vee 1\}(P(R(\gamma)>0)}\right]$$
$$\geq \frac{\mathbb{E}\left[\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)|R(\gamma)>0\right]}{\mathbb{E}[R(\gamma)]}P(R(\gamma)>0) + \frac{\mathbb{E}\left[\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)|R(\gamma)=0\right]}{R(\gamma)>0}P(R(\gamma)=0)$$
$$= \frac{\mathbb{E}\left[\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)\right]}{\mathbb{E}[R(\gamma)]} \geq 0$$

The numerator of the last fraction is certainly positive because:

$$E\left[\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)\right] \geq \frac{m\pi_0(1-\lambda)\gamma}{1-\lambda} - m\pi_0\gamma = 0$$

As a consequence the thesis for pFDR is demonstrated.

Having verified the thesis for pFDR, is straightforward to demonstrate the thesis also for FDR, just by observing that:

$$\frac{1}{P(R(\gamma)>0)}[\mathbb{E}(FD\hat{R}_\lambda(\gamma)) \geq FDR(\gamma)] = \mathbb{E}\left[\frac{\frac{\gamma W(\lambda)}{1-\lambda} - V(\gamma)}{\{R(\gamma)\vee 1\}(P(R(\gamma)>0)}\right] \geq 0$$

□

In practical applications, it is necessary to truncate the estimates at 1 as, by definition, $FDR \leq pFDR \leq 1$: the results of Theorem 7 are demonstrated under the assumption that this truncation is not performed.
The following Theorem provides further informations on the truncated estimates by showing how such a truncation procedure reduces the variability of the estimates.

Theorem 8.

$$\mathbb{E}\{[\widehat{pFDR}_\lambda(\gamma) - pFDR(\gamma)]^2\} > \mathbb{E}\{[\widehat{pFDR}_\lambda(\gamma) \wedge 1 - pFDR(\gamma)]^2\}$$

and

$$\mathbb{E}\{[\widehat{FDR}_\lambda(\gamma) - FDR(\gamma)]^2\} > \mathbb{E}\{[\widehat{FDR}_\lambda(\gamma) \wedge 1 - FDR(\gamma)]^2\}$$

where notation $\widehat{FDR}_\lambda(\gamma) \wedge 1$ *indicates the estimates truncated at 1.*

Proof. The proof follows immediately from the property of the iterated expected value:

$$\mathbb{E}\{[\widehat{pFDR}_\lambda(\gamma) - pFDR(\gamma)]^2 | \widehat{pFDR}_\lambda(\gamma) > 1\}$$

$$> \mathbb{E}\{[\widehat{pFDR}_\lambda(\gamma) \wedge 1 - pFDR(\gamma)]^2 | \widehat{pFDR}_{\lambda(\gamma)>1}\}$$

□

In order to conclude the analysis of the main properties of pFDR, we aim to recall the following Theorem that shows how the estimate 2.6 of pFDR, not adjusted for finite samples, is a maximum likelihood estimate for a transformation of pFDR.

Theorem 9. *Under the assumptions of independence and identical distribution among the p-values, the estimate 2.6, referred to as* $\hat{Q}_\lambda(\gamma)$*, is a maximum likelihood estimate for*

$$\frac{\pi_0 + \pi_1[1 - g(\lambda)]/(1 - \lambda)}{\pi_0} pFDR(\gamma)$$

where $g(\lambda) = P(p_j \leq \lambda | \overline{H}_0)$ *is the test power, depending on* λ*.*

Proof. The likelihood can be written as:

$$[\pi_0\gamma + \pi_1 g(\gamma)]^{R(\gamma)}[1 - \pi_0\gamma - \pi_1 g(\gamma)]^{m - R(\gamma)}$$

It is straightforward to observe that it is the likelihood function of a binomial random variable with parameter $p = \pi_0\lambda + \pi_1 g(\lambda)$. Hence, the maximum likelihood estimate for the parameter p is equal to $p_{ML} = R(\gamma)/m$. Because of the invariance property of the maximum likelihood estimates and since the maximum likelihood estimate of $m(1-p)$ is equal to $W(\lambda)$, because the hypotheses are distributed as independent Bernoulli random variables, hence the thesis is given:

$$\frac{\pi_0 + \pi_1[1-g(\lambda)]/(1-\lambda)}{\pi_0} pFDR_{ML}(\gamma) = \frac{\pi_0 + \pi_1[1-g(\lambda)]/(1-\lambda)}{\pi_0}\frac{\pi_0\gamma}{p_{ML}}$$

$$= \frac{\pi_0(1-\lambda) + \pi_1(1-g(\lambda)}{1-\lambda}\frac{m\gamma}{R(\gamma)} = \frac{\gamma[m(1-p)]}{R(\gamma)(1-\lambda)} = \frac{\gamma W(\lambda)}{(1-\lambda)R(\gamma)} = \hat{Q}_\lambda$$

□

Asymptotic properties

Under the null hypotheses, we recall that the p-values are assumed to be identically distributed as Bernoulli random variables with parameter equal to π_0. Under the alternative hypotheses, the p-values are still assumed to be identically distributed with an unknown cumulative distribution that is usually denoted by F, while the corresponding density distribution is denoted by f.

The following Theorem illustrates the asymptotic behavior of $k = \max\{i : p_{(i)} \leq \frac{i\alpha}{m}\}$, that is the thresholding point of the procedure that controls FDR:

Theorem 10. *Under the following assumptions:*

1. *F is strictly concave;*
2. *$F'(0) > \beta$, where*
$$\beta = \frac{1-\alpha A_0}{\alpha A_1};$$
3. *$A_1 > 0$, namely at least one null hypothesis is false.*

where $A_0 = m_0/m$ and $A_1 = m_1/m$ are, respectively, the proportions of null and alternative hypotheses.
Let us denote with u^* the unique solution of the equation $F(u) = \beta u$.
Then, for $m \to \infty$, $k/m \xrightarrow{p} u^*/\alpha$.

Proof. Because of the assumptions on the cumulative distribution F under the alternative hypotheses, the solution u^* of the system $F(u) = \beta u$ exists and is

unique.
Furthermore $\beta > 1$ because it can be written equivalently as $\beta = 1 + \frac{1-\alpha(A_1+A_0)}{\alpha A_1}$ and $\frac{1-\alpha(A_1+A_0)}{\alpha A_1} > 0$.
These conditions imply that the line βu is above the bisector of the first quadrant and crosses F exactly once. Let us define:

$$a_m = \frac{mu^*(1-\epsilon_m)}{\alpha}$$

$$b_m = \frac{mu^*(1+\epsilon_m)}{\alpha}$$

where $\epsilon_m \to 0$ and $\epsilon_m \sqrt(m) \to m$. By recalling the equation $F(u) = \beta u$, we get:

$$u^* = \frac{F(u^*)}{\beta} = \frac{\alpha A_1 F(u^*)}{1-\alpha A_0}$$

We can write a_m and b_m in an equivalent way as follows:

$$a_m = \frac{mF(u^*)A_1(1-\epsilon_m)}{1-\alpha A_0}$$

$$b_m = \frac{mF(u^*)A_1(1+\epsilon_m)}{1-\alpha A_0}$$

Since $A_1 \leq 1 - \alpha A_0$, then $\frac{1}{\beta} \leq \alpha$ and so $u^* \leq \alpha F(u^*) \leq \alpha$.
As a consequence $F(u^*) \leq F(\alpha)$ and, because of the concavity of F, $f(u^*) < \beta$.
Given

$$N_i = \#\left\{j : \ p_j \leq \frac{i\alpha}{m}\right\} = \sum_{j=1}^{m} \mathbf{1}\left(p_j \leq \frac{i\alpha}{m}\right)$$

then:

$$\{k > b_m\} = \cup_{d>b_m}\{p_{(d)} \leq \frac{d\alpha}{m}\} = \cup_{d>b_m}\{N_{(d)} \geq d\}$$

Hence for the property of sub-additivity of the probability measures:

$$P(k > b_m) \leq \sum_{d>b_m} P(N_d \geq d)$$

Let us define:

$$\mu(t) = A_0\alpha t + A_1 m F\left(\frac{\alpha t}{m}\right)$$

If d is an integer, then

$$\mu(d) = \mathbb{E}(N_d) = \sum_i P\left(p_i \leq \frac{d\alpha}{m}\right)$$

□

At last, Storey (2002) has demonstrated that the estimate of pFDR is conservative also for large samples. In this regard, it holds the following:

Theorem 11. *The following limit holds with probability 1:*

$$lim_{m\to\infty}\left[\widehat{pFDR}_\lambda(\gamma)\right] = \frac{\pi_0 + \pi_1[1-g(\lambda)]/(1-\lambda)}{\pi_0} pFDR(\gamma) \geq pFDR(\gamma)$$

Proof. Because of the law of large numbers:

1. $$\hat{P}(p_j \leq \gamma) \xrightarrow{ae} P(p_j \leq \gamma);$$

2. $$\frac{W(\lambda)}{m} \xrightarrow{ae} (1-\lambda)\pi_0 + (1-g(\lambda)\pi_1$$

   Since $P(p_j \geq \lambda | H_0) = 1-\lambda$ and $P(p_j \geq \lambda | H_1) = 1-g(\lambda)$.

Hence the thesis is given. □

A similar result can be demonstrated also for FDR, just by observing that $\widehat{FDR}_\lambda(\gamma) \sim \widehat{pFDR}_\lambda(\gamma)$.

pFDR under dependence

With respect to the pFDR, all its relevant properties have been demonstrated under the assumption of independence and identical distribution among the p-values. In particular, the hypothesis of independence is commonly not adequate in many application studies. For this reason, the weakening of the hypothesis of independence is one of the most exciting and studied directions of research in the field of multiple comparisons.

Given m multiple comparisons and $T_1, \ldots, T_m$ test statistics, Storey (2003) has proposed the following Theorem that generalize the definition of pFDR in case of dependency among the test statistics, under the assumption of a common rejection region Γ for each test.

Theorem 12.

$$pFDR(\Gamma) = \sum_{k=1}^{m} \sum_{i_1,\ldots,i_k} \frac{1}{k} \sum_{j=1}^{k} P(H_{i_j} = 0, (T_{i_1}, \ldots, T_{i_k}) \in \Gamma, (T_{i_{k+1}}, \ldots, T_{i_m}) \notin \Gamma | R(\Gamma) > 0)$$

Proof.

$$pFDR(\Gamma) = \sum_{k=1}^{m} \sum_{i_1,\dots,i_k} \mathbb{E}\left(\frac{V(\Gamma)}{R(\Gamma)} \middle| (T_{i_1},\dots,T_{i_k}) \in \Gamma, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma \right) \times P((T_{i_1},\dots,T_{i_k}) \in \Gamma, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma | R(\Gamma) > 0)$$

$$= \sum_{k=1}^{m} \sum_{i_1,\dots,i_k} \mathbb{E}\left(\frac{\sum_{j=1}^{k}(1-H_{i_j})}{k} \middle| (T_{i_1},\dots,T_{i_k}) \in \Gamma, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma \right) \times P((T_{i_1},\dots,T_{i_k}) \in \Gamma, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma | R(\Gamma) > 0)$$

$$= \sum_{k=1}^{m} \sum_{i_1,\dots,i_k} \frac{1}{k} \sum_{j=1}^{k} P(H_{i_j} = 0 | (T_{i_1},\dots,T_{i_k}) \in \Gamma, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma) \times P((T_{i_1},\dots,T_{i_k}) \in \Gamma, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma | R(\Gamma) > 0)$$

$$= \sum_{k=1}^{m} \sum_{i_1,\dots,i_k} \frac{1}{k} \sum_{j=1}^{k} P(H_{i_j} = 0, (T_{i_1},\dots,T_{i_k}) \in \Gamma, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma | R(\Gamma) > 0)$$

□

Furthermore, if the conjoint tests are dependent but exchangeable the thesis of Theorem 12 can be simplified as follows:

Corollary 1. Let us consider $(H_1, T_1), \dots, (H_m, T_m)$ as exchangeable random variables, then:

$$pFDR(\Gamma_\alpha) = \sum_{k=1}^{m} P(H_1 = 0 | (T_{i_1},\dots,T_{i_k}) \in \Gamma_\alpha, (T_{i_{k+1}},\dots,T_{i_m}) \notin \Gamma_\alpha) P(R = k | R > 0)$$

It is clear, hence, that the fundamental result stated in Theorem 5 does not hold in case of dependent test statistics; nevertheless, under weaken assumption than the independence, such a result still holds asymptotically (Storey, 2003; Storey and Tibshirani, 2001).

Theorem 13. *Suppose with probability 1 that:*

$$\sum_{i=1}^{m} \frac{(1-H_i)}{m} \to \pi_0$$

In addition, let us define for $m \to \infty$ and for each $\alpha > 0$:

$$\frac{V_m(\Gamma_\alpha)}{\sum_{i=1}^{m}(1-H_i)} = \frac{\sum_{i=1}^{m}(1-H_i)\mathbb{1}(T_i \in \Gamma_\alpha)}{\sum_{i=1}^{m}(1-H_i)} \to G_0(\alpha)$$

$$\frac{S_m(\Gamma_\alpha)}{\sum i = 1^m H_i} = \frac{\sum_{i=1}^m H_i \mathbb{1}(T_i \in \Gamma_\alpha)}{\sum_{i=1}^m H_i} \to G_1(\alpha)$$

Hence, the continuous functions G_0 and G_1 are the asymptotic type I error and power, respectively.
Then, for each $\delta > 0$ such that $\pi_0 G_0(\delta) + (1-\pi_0)G_1(\delta) > 0$, the followings statements hold:

i.
$$\limsup_{m\to\infty} \left| \frac{V_m(\Gamma_\alpha)}{R_m(\Gamma_\alpha) \vee 1} - P_\infty(H_0|X \in \Gamma_\alpha) \right| \overset{ae}{=} 0$$

ii.
$$\limsup_{m\to\infty} |FDR_m(\Gamma_\alpha) - P_\infty(H_0|X \in \Gamma_\alpha)| = 0$$

iii.
$$\limsup_{m\to\infty} |pFDR_m(\Gamma_\alpha) - P_\infty(H_0|X \in \Gamma_\alpha)| = 0$$

where
$$P_\infty(H_0|X \in \Gamma_\alpha) = \frac{\pi_0 G_0(\alpha)}{\pi_0 G_0(\alpha) + (1-\pi_0)G_1(\alpha)}$$

Proof. It is worth recalling that $V_m(\Gamma_\alpha)/m$ is the empirical distribution of the type I error, since it is defined as the frequency of true hypotheses rejected.
Then its corresponding theoretical distribution is equal to $\pi_0 G_0(\alpha)$. By the Glivenko-Cantelli theorem it follows that:

$$\lim_{m\to\infty} \sup_{\alpha\geq\delta} \left| \frac{V_m(\Gamma_\alpha)}{m} - \pi_0 G_0(\alpha) \right| \overset{ae}{=} 0$$

With an analogous argumentation:

$$\lim_{m\to\infty} \sup_{\alpha\geq\delta} \left| \frac{[V_m(\Gamma_\alpha) + S_m(\Gamma_\alpha)] \vee 1}{m} - [\pi_0 G_0(\alpha) + (1-\pi_0)G_1(\alpha)] \right| \overset{ae}{=} 0$$

Since $[\pi_0 G_0(\alpha) + (1-\pi_0)G_1(\alpha)]$ is a positive non decreasing function, then:

$$\lim_{m\to\infty} \sup_{\alpha\geq\delta} \left| \frac{m}{[V_m(\Gamma_\alpha) + S_m(\Gamma_\alpha)] \vee 1} - \frac{1}{[\pi_0 G_0(\alpha) + (1-\pi_0)G_1(\alpha)]} \right| \overset{ae}{=} 0$$

Let us define
$$Q_m(\alpha) = \frac{V_m(\Gamma_\alpha)}{[V_m(\Gamma_\alpha) + S_m(\Gamma_\alpha)] \vee 1}$$

it may be observed that:

$$|Q_m(\alpha) - P_\infty(H_0|X \in \Gamma_\alpha)|$$

$$\leq \left| \frac{V_m(\Gamma_\alpha) - m\pi_0 G_0(\alpha)}{[V_m(\Gamma_\alpha) + S_m(\Gamma_\alpha)] \vee 1} \right| + \left| \frac{m\pi_0 G_0(\alpha)}{[V_m(\Gamma_\alpha) + S_m(\Gamma_\alpha)] \vee 1} - \frac{\pi_0 G_0(\alpha)}{[\pi_0 G_0(\alpha) + (1-\pi_0) G_1(\alpha)]} \right|$$

Hence the first statement of the Theorem is demonstrated.

Furthermore, in order to demonstrate the asymptotic convergence of FDR, it is worth observing that

$$|Q_m(\alpha) - P_\infty(H_0|X \in \Gamma_\alpha)| \overset{ae}{\leq} 2$$

As a consequence:

$$0 = \mathbb{E}[\lim_{m\to\infty} \sup_{\alpha \geq \delta} |Q_m(\alpha) - P_\infty(H_0|X \in \Gamma_\alpha)|] =$$

$$\lim_{m\to\infty} \mathbb{E}[\sup_{\alpha \geq \delta} |Q_m(\alpha) - P_\infty(H_0|X \in \Gamma_\alpha)|]$$

$$\lim_{m\to\infty} \sup_{\alpha \geq \delta} \mathbb{E}[|Q_m(\alpha) - P_\infty(H_0|X \in \Gamma_\alpha)|] \geq 0$$

Where $\mathbb{E}[Q_m(\alpha)] = FDR_m(\Gamma_\alpha)$. In conclusion:

$$\lim_{m\to\infty} \sup_{\alpha \geq \delta} |pFDR_m(\Gamma_\alpha) - FDR_m(\Gamma_\alpha)|] \leq \lim_{m\to\infty} \left| \frac{1}{P(R_m(\delta) > 0)} - 1 \right| = 0$$

From the second statement, it follows immediately the asymptotic convergence of pFDR. □

2.2.5 False Non Discovery Rate

Until now, while analyzing the problem of multiple comparisons, we have focused only on the type I error, that is related to the number of false rejections; now, we aim to introduce some adequate measures for the type II error, that is related to the number of false acceptances.

By recalling Table 2.2, and in accordance to that notation, Craiu and Sun (2008) proposed a new quantity, referred to as *Non-Discovery Rate*, as a possible measure of type II error since it is defined as the expected proportion of

acceptances among false null hypotheses:

$$NDR = \frac{\mathbb{E}(T)}{m_1}$$

Clearly, if $m_1 = 0$ the NDR is not yet well-defined. The main advantage of this quantity is that its complement to one can be interpreted as the average power (Dudoit, Shaffer, and Boldrick, 2003):

$$1 - NDR = 1 - \frac{\mathbb{E}(T)}{m_1} = \frac{\mathbb{E}(S)}{m_1}$$

Genovese and Wasserman (2002) proposed, instead, a counterpart of FDR, referred to as the *False Nondiscovery Rate*, that is defined as the expected proportion of false negatives among all the hypotheses that are not rejected:

$$FNR = \mathbb{E}\left(\frac{T}{W}\right)$$

Differently from NDR, the complement to one of FNR is not equal to the average power of the multiple comparisons, as one may expect in a frequentist perspective.
At last, Storey (2003) proposed a natural counterpart of pFDR, referred to as the *positive False Nondiscovery Rate*:

$$pFNR = \mathbb{E}\left(\frac{T}{W}\middle| W > 0\right)$$

In analogy with Theorem 5 about pFDR, it is easy to prove the following:

Theorem 14. *Consider m hypotheses with independent test statistics $T_1, \ldots, T_m$ and a common rejection region Γ. If the null hypotheses are true with an a priori probabilities of π_0 then:*

$$pFNR = P(\overline{H}_0 | T \notin \Gamma)$$

Proof. At first, it is worth noting that:

$$pFNR(\Gamma) = \mathbb{E}\left(\frac{T(\Gamma)}{W(\Gamma)}\middle| W(\Gamma) > 0\right) =$$

$$= \sum_{k=1}^{m} \mathbb{E}\left(\frac{T(\Gamma)}{W(\Gamma)}\middle| W(\Gamma) = k\right) P(W(\Gamma) = k | W(\Gamma) > 0) =$$

$$= \sum_{k=1}^{m} \mathbb{E}\left(\frac{T(\Gamma)}{k}\middle| W(\Gamma) = k\right) P[W(\Gamma) = k | W(\Gamma) > 0]$$

Under the strong assumption that the statistics are independent and identically distributed, it holds the following:

$$\begin{aligned}
&\mathbb{E}[T(\Gamma)|W(\Gamma) = k] \\
&= \mathbb{E}\left[\sum_{i=1}^{m} \mathbb{1}(T_i \notin \Gamma)\mathbb{1}(H_i = 1) \,\middle|\, (T_1, \dots, T_k) \notin \Gamma, (T_{k+1}, \dots, T_m) \in \Gamma\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{k} \mathbb{1}(H_i = 1) \,\middle|\, (T_1, \dots, T_k) \notin \Gamma, (T_{k+1}, \dots, T_m) \in \Gamma\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{k} \mathbb{1}(H_i = 1) \,\middle|\, T_i \notin \Gamma\right] = kP(\overline{H}_0|T \notin \Gamma)
\end{aligned}$$

And, hence, the thesis is given:

$$pFDR(\Gamma) = \sum_{k=1}^{m} \frac{kP(\overline{H}_0|T \notin \Gamma)}{k} P[W(\Gamma) = k|W(\Gamma) > 0] = P(\overline{H}_0|T \notin \Gamma)$$

□

Sala (2014) investigated the trade-off between pFDR and pFNR, and proposed a method for balancing those two types of error, that has been successfully applied on fMRI data.

2.3 Bayes False Discovery Rate and Bayes Power in Multiple Testing

Let us recall the general setting of multiple comparisons we are interested in and let us introduce some fundamental definitions. Consider, in fact, m simultaneous hypotheses such that H_0 denotes the null hypothesis with an a priori probabilities of $\pi_0 = P(H_0)$, while $H_1 = \bar{H}_0$ denotes the alternative hypothesis with an a priori probability of $\pi_1 = P(H_1) = 1 - \pi_0$.

In this study, and in the perspective of a brain network construction, we aim to control, as a function of the rejection region γ, the probability of false discoveries by means of the Bayes FDR (Efron, 2010):

$$FDR(\gamma) = P(H_0|p_j \leq \gamma) = \frac{P(p_j \leq \gamma|H_0)P(H_0)}{P(p_j \leq \gamma)} = \frac{\gamma\pi_0}{F(\gamma)}, \tag{2.8}$$

the probability of false non discoveries by means of the Bayes False Non-discovery Rate (FNR):

$$FNR(\gamma) = P(\overline{H}_0|p_j > \gamma) = 1 - P(H_0|p_j > \gamma) = 1 - \frac{(1-\gamma)\pi_0}{1-F(\gamma)}, \tag{2.9}$$

and the Bayes power (BP) (Berlingeri, 2015):

$$BP(\gamma) = P(p_j \leq \gamma|\overline{H}_0) = \frac{P(\overline{H}_0|p_j \leq \gamma)P(p_j \leq \gamma)}{P(\overline{H}_0)} = \frac{F(\gamma) - \gamma\pi_0}{1-\pi_0}, \tag{2.10}$$

where F represents the p-values cdf.

The Bayes FDR, FNR and BP have been defined under the assumption that the p-values are uniformly distributed under H_0: in particular, it holds that $P(p_j \leq \gamma|H_0) = \gamma$ and all the results in the following Sections will be demonstrated under this specific setting.
In case one assumes, however, the p-values to be valid, accordingly to 1 of Lemma 1, it holds that $P(p_j \leq \gamma|H_0) \leq \gamma$. As a consequence, the Bayes quantities of interest can be written in this more general case as:

$$FDR(\gamma) \leq \frac{\gamma\pi_0}{F(\gamma)}, \tag{2.11}$$

$$FNR(\gamma) \leq 1 - \frac{(1-\gamma)\pi_0}{1-F(\gamma)}, \tag{2.12}$$

$$BP(\gamma) \geq \frac{F(\gamma) - \gamma\pi_0}{1-\pi_0}, \tag{2.13}$$

It is clear that the direction of the inequalities is conservative in all the three cases: the two errors measurement, in fact, will be overestimated while the Bayes power will be underestimated.
Hence, the theoretical results demonstrated under the assumption that $P(p_j \leq \gamma|H_0) = \gamma$, can be generalized in case of valid p-values.

2.3.1 Not empirical Bayesian estimates

Given the plug-in estimator of the unknown cdf of the p-values:

$$\hat{F}(\gamma) = \frac{\#\{p_j \leq \gamma\}}{m} \tag{2.14}$$

By substituting 2.14 into 2.8, 2.9 and 2.10, we get the estimates of the Bayes FDR, of the Bayes FNR and of the BP, respectively, denoted as $E[\hat{FDR}(\gamma)]$, $E[\hat{FNR}(\gamma)]$ and $E[\hat{BP}(\gamma)]$; those estimates are said to be not empirical because π_0 is assumed to be known. Due to:

$$E[\hat{F}(\gamma)] = F(\gamma) = \gamma\pi_0 + BP(\gamma)\pi_1 \tag{2.15}$$

and thanks to the Jensen's inequality, we obtain the following relations on the expected values of the estimates:

$$\mathbb{E}[\widehat{FDR}(\gamma)] \geq \frac{\gamma\pi_0}{\mathbb{E}[\hat{F}(\gamma)]} = FDR(\gamma) \tag{2.16}$$

$$\mathbb{E}[\widehat{FNR}(\gamma)] \leq 1 - \frac{(1-\gamma)\pi_0}{1-\mathbb{E}[\hat{F}(\gamma)]} = FNR(\gamma) \tag{2.17}$$

$$\mathbb{E}[\widehat{BP}(\gamma)] = \frac{\mathbb{E}[\hat{F}(\gamma)] - \gamma\pi_0}{1-\pi_0} = BP(\gamma) \tag{2.18}$$

Indeed, the estimate of the Bayes FDR is conservative (Storey, 2002) and the estimate of the BP is correct; conversely the direction of the inequality for the estimate of the Bayes FNR suggests that, on average, the probability of false non-discoveries is underestimated. Because of this lack of conservativeness we choose not to study further the Bayes FNR.

2.3.2 Empirical Bayesian estimates

Let us consider the case in which π_0 is unknown and hence it has to be estimated empirically from the data. By applying the conservative estimate of the a priori probability π_0 (Storey, 2002):

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_j > \lambda\}}{m(1-\lambda)} = \frac{1-\hat{F}(\lambda)}{1-\lambda}, \tag{2.19}$$

which has expected value:

$$E[\hat{\pi}_0(\lambda)] = \frac{1-E[\hat{F}(\lambda)]}{1-\lambda} = \frac{1-F(\lambda)}{1-\lambda} = \pi_0 + \frac{1-BP(\lambda)}{1-\lambda}\pi_1 \geq \pi_0. \tag{2.20}$$

And by substituting 2.14 and 2.19 in 2.8 and 2.10, we obtain the following empirical Bayes estimates:

$$\widehat{FDR}_{\lambda_1}(\gamma) = \frac{\gamma\hat{\pi}_0(\lambda_1)}{\hat{F}(\gamma)}, \tag{2.21}$$

and [1]

$$\widehat{BP}_{\lambda_2}(\gamma) = \frac{\hat{F}(\gamma) - \gamma\hat{\pi}_0(\lambda_2)}{1 - \hat{\pi}_0(\lambda_2)} \tag{2.22}$$

.

In order to estimate the optimal values λ_i $(i = 1, 2)$ of the tuning parameters, we resample the m p-values with replacement B times, we calculate the Bootstrap versions of 2.21 and 2.22 over a range of λ values (e.g. from 0 to 1 with step 0.1) and we minimize the Bootstrap estimates of the corresponding mean square errors (Storey, 2002). Such a bootstrap procedure allows us also to construct the one-sided $(1-\alpha)$-confidence intervals for the two Bayes parameters by taking, respectively, the $(1-\alpha)$-quantiles of the $\hat{FDR}_{\lambda_1}(\gamma)$ Bootstrap distributions as the corresponding upper confidence bound of the probability of false discoveries, and the α-quantile of the $\hat{BP}_{\lambda_2}(\gamma)$ Bootstrap distribution as the lower confidence bound of the power. Since it is not sufficient to control the Bayes FDR alone, we propose to balance these two types of error rates by choosing a suitable value of the threshold γ such that the Bayes FDR is low and the BP is reasonably high.

2.3.3 Large Sample Theory

Asymptotic results in case of independence

Let us assume the independence of the m p-values, then by the strong law of large numbers:

$$\hat{F}(\gamma) \overset{a.s.}{\to} E[\hat{F}(\gamma)] = F(\gamma),$$

by the continuous mapping theorem (Billingsley, 1986), that states that continuous functions are limit preserving, it follows that:

$$\hat{\pi}_0(\lambda) \overset{a.s.}{\to} E[\hat{\pi}_0(\lambda)] = \pi_0 \left[1 + \frac{F1 - BP(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right] \geq \pi_0,$$

[1] The estimate of the a priori probability π_0 is truncated at 1 (Storey, 2002). If $\hat{\pi}_0(\lambda) = 1$ then the estimate of the power is not yet well defined. In order to handle this problem, we suggest, in case $\hat{\pi}_0(\lambda) = 1$, to estimate the power with $\hat{BP}_{\lambda_2}(\gamma) = \hat{F}(\gamma)$, according to (Karr, 1993).

$$\widehat{FDR}_\lambda(\gamma) \overset{a.s.}{\to} \frac{\gamma E[\hat{\pi}_0(\lambda)]}{E[\hat{F}(\gamma)]} = \frac{\gamma\pi_0}{F(\gamma)}\left[1 + \frac{1 - BP(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right] = FDR(\gamma)\left[1 + \frac{1 - BP(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right] \geq FDR(\gamma), \tag{2.23}$$

and

$$\widehat{BP}_\lambda(\gamma) \overset{a.s.}{\to} \frac{E[\hat{F}(\gamma)] - \gamma E[\hat{\pi}_0(\lambda)]}{1 - E[\hat{\pi}_0(\lambda)]} = \frac{F(\gamma) - \gamma\pi_0\left[1 + \frac{1-BP(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right]}{1 - \pi_0\left[1 + \frac{1-BP(\lambda)}{1-\lambda}\frac{\pi_1}{\pi_0}\right]} = \frac{1-\lambda}{BP(\lambda)-\lambda}BP(\gamma) - \gamma\frac{1-BP(\lambda)}{BP(\lambda)-\lambda} \tag{2.24}$$

Therefore, when λ is chosen so that $BP(\lambda)$ is close to 1 (for example, if λ is near 1), the empirical Bayes estimates asymptotically approximate the actual values of the corresponding parameters.

Asymptotic results in case of dependence

Let us consider now the general case of not independent p-values: the pattern of dependency among p-values is unknown, a priori, and also hard to model from the data. As a consequence, in order to prove similar results as in Section 2.3.3, it is necessary to formalize some hypothesis on the conjoint probability of the p-values.

For instance, if

$$P(p_i \leq \gamma, p_j \leq \gamma) = F(\gamma)^2 + o(1) \quad for|i-j| \to \infty$$

then, by the Bernstein's law of large numbers (Shiryaev and Lyasoff, 2012),

$$\hat{F}(\gamma) \overset{p}{\to} E[\hat{F}(\gamma)] = F(\gamma),$$

from which the convergences 2.23 and 2.24 holds only in probability (by the continuous mapping theorem).
In order to guarantee the almost sure convergence for 2.23 and 2.24, also in case of dependent p-values, we have then to strengthen some hypotheses.
The Glivenko-Cantelli theorem determines the almost surely convergence of the empirical distribution function for a set of n independent and identically-distributed random variables, with $n \to \infty$. Let us extend this theorem for

stationary ergodic sequences $X_1, \dots, X_k, \dots$ defined on a probability space (Ω, F, P). For each $k \geq 0$ and $n \geq 1$, if the distribution functions of $(X_1, \dots, X_n)$ and $(X_{k+1}, \dots, X_{k+n})$ are the same, then the sequence is said to be stationary. More formally, consider a measureable function T on Ω and assume that T is also measure preserving, that is $P(T^{-1}A) = P(A)$ for all $A \in F$. The function T is also ergodic if $P(A) = 0$ or 1 for all $A \in J$, where $J \subset F$ is the sub-field induced by the invariant sets (a set $B \in F$ is invariant if $T^{-1}B = B$). Consequently, in case of stationary ergodic p-values,

$$\sup_{\gamma} \{\hat{F}(\gamma) - F(\gamma)\} \overset{a.s.}{\to} 0$$

(Dehling and Philipp, 2002).

2.3.4 Simulation Studies

In this Section we propose some numerical results that support and enrich our theoretical considerations.

Simulations in case of independence

In order to demonstrate numerically the properties of the not empirical Bayes estimates of FDR, FNR and BP in case of independent p-values we simulated $m = 1000$ independent and normally distributed random variables $Z_i \sim N(\mu, 1)$ where $\mu = 0$ under the null hypothesis while $\mu = 2$ under the alternative hypothesis. For each i-th multiple comparison, $i = 1, \dots, m$, we computed the associated p-value $p_i = 2 - 2\Phi(|z_i|)]$, where z_i is the observed value of the i-th normal random variable. We iterated our simulation $N = 10000$ times for $\pi_0 = 0.1, \dots, 0.9$ and for two rejection regions depending on $\gamma = 0.01$ and $\gamma = 0.001$. At last, we computed the not empirical Bayes estimates, with π_0 known and equal to the real simulated value; the results are reported in Table 2.3.

Accordingly to the theoretical results in Section 2.3.1, we observed that the expected value, over N iterations, of the Bayes FDR estimates is conservative, while the expected value of the Bayes FNR systematically underestimate or is equal to the true value of the probability of false non-discoveries. While this first simulation allowed us to verified the properties of the not empirical Bayes estimates, the following simulation studies enabled the analysis of the empirical Bayes estimates, both in case of independent p-values and under

TABLE 2.3: Simulation in case of independent p-values over a range of $\pi_0 = 0.1, \ldots, 0.9$ and $\gamma = 0.01, 0.001$; for FDR, FNR and Power we reported the true value, the Monte Carlo not empirical Bayes estimate (indicated with hat), the true mean square error and the 95% bootstrap confidence interval.

π_0	FDR	$\widehat{FDR}$	MSE	95% CI	FNR	$\widehat{FNR}$	MSE	95% CI	BP	$\widehat{BP}$	MSE	95% CI
$\gamma = 0.01$												
0.1	0.0039	0.0039	4.49e-08	[0;0.0043]	0.8671	0.8670	5.95e-06	[0;0.8709]	0.2824	0.2824	2.0e-04	[0.2578;1]
0.2	0.0088	0.0088	2.49e-07	[0;0.0097]	0.7436	0.7435	1.81e-05	[0;0.7504]	0.2824	0.2823	2.55e-04	[0.2556;1]
0.3	0.0150	0.0150	8.19e-07	[0;0.0167]	0.6284	0.6284	3.11e-05	[0;0.6376]	0.2824	0.2824	2.93e-04	[0.2534;1]
0.4	0.0231	0.0232	2.29e-06	[0;0.0261]	0.5209	0.5208	4.22e-05	[0;0.5317]	0.2824	0.2824	3.48e-04	[0.2504;1]
0.5	0.0342	0.0344	5.95e-06	[0;0.0392]	0.4203	0.4203	4.79e-05	[0;0.4322]	0.2824	0.2821	4.15e-04	[0.2463;1]
0.6	0.0504	0.0508	1.62e-05	[0;0.0590]	0.3258	0.3258	5.03e-05	[0;0.3381]	0.2824	0.2822	5.36e-04	[0.2414;1]
0.7	0.0763	0.0770	4.98e-05	[0;0.0917]	0.2370	0.2370	4.76e-05	[0;0.2490]	0.2824	0.2823	7.48e-04	[0.2338;1]
0.8	0.1241	0.1256	1.93e-04	[0;0.1557]	0.1534	0.1534	3.96e-05	[0;0.1644]	0.2824	0.2823	1.21e-03	[0.2209;1]
0.9	0.2417	0.2466	1.42e-03	[0;0.3328]	0.0745	0.0745	2.74e-05	[0;0.0834]	0.2824	0.2830	2.96e-03	[0.1889;1]
$\gamma = 0.001$												
0.1	0.0011	0.0011	1.40e-08	[0;0.0014]	0.8904	0.8904	1.17e-06	[0;0.8921]	0.0984	0.0984	9.98e-05	[0.0825;1]
0.2	0.0025	0.0026	7.87e-08	[0;0.0031]	0.7831	0.7831	3.91e-06	[0;0.7862]	0.0984	0.0984	1.10e-04	[0.0815;1]
0.3	0.0043	0.0044	2.63e-07	[0;0.0054]	0.6780	0.6780	7.36e-06	[0;0.6823]	0.0984	0.0985	1.25e-04	[0.0803;1]
0.4	0.0067	0.0068	7.55e-07	[0;0.0086]	0.5751	0.5751	1.09e-05	[0;0.5804]	0.0984	0.0985	1.48e-04	[0.0788;1]
0.5	0.0101	0.0103	2.17e-06	[0;0.0132]	0.4744	0.4744	1.41e-05	[0;0.4803]	0.0984	0.0982	184e-04	[0.0765;1]
0.6	0.0150	0.0154	6.04e-06	[0;0.0205]	0.3756	0.3756	1.53e-05	[0;0.3819]	0.0984	0.0983	2.26e-04	[0.0740;1]
0.7	0.0232	0.0239	1.96e-05	[0;0.0336]	0.2789	0.2789	1.49e-05	[0;0.2851]	0.0984	0.0985	2.99e-04	[0.0703;1]
0.8	0.0390	0.0411	9.69e-05	[0;0.0643]	0.1841	0.1841	1.29e-05	[0;0.1899]	0.0984	0.0981	4.63e-04	[0.0635;1]
0.9	0.0838	0.0927	1.51e-03	[0;0.1938]	0.0911	0.0911	8.45e-06	[0;0.0956]	0.0984	0.0987	9.99e-04	[0.0495;1]

specific patterns of dependency. As a consequence, the main aim of the following simulation studies was the comparison between the behavior of the Bayes FDR and of the BP, in case of independent p-values and under some specific patterns of dependency. Because of the not conservative behavior shown by the Bayes FNR, according to 2.17, we investigated further only the Bayes FDR and the BP. All the following simulation studies, both under independence and under dependence of p-values, were performed for $m = 1000$ multiple comparisons and for the parameters $\pi_0 = 0.1, \ldots, 0.9$ and $\gamma = 0.01, 0.001$. As already observed in Section 2.3.2, the empirical Bayesian estimates of FDR and BP depend not only on γ but, also, on the tuning parameters λ_1 and λ_2, respectively. In order to estimate the optimal values of those tuning parameters, hence, we performed our simulations for $\lambda = 0.1, \ldots, 0.9$. According to Storey (2002) we selected the optimal λ as the value that minimizes the bootstrap mean square error (MSE).
With respect to the bootstrap MSE of the Bayes FDR and BP, we considered the standard plug-in (Shao and Tu, 1995) estimates $\widehat{FDR}_\lambda(\gamma)$ and $\widehat{BP}_\lambda(\gamma)$, respectively, and hence:

$$\lambda_1^{opt} = \text{argmin}\widehat{MSE}_{FDR}(\lambda) = \text{argmin}\frac{1}{B}\sum_{b=1}^{B}\{\widehat{FDR}_\lambda^b(\gamma) - \widehat{FDR}_\lambda(\gamma)\}^2 \quad (2.25)$$

$$\lambda_2^{opt} = \text{argmin}\widehat{MSE}_{BP}(\lambda) = \text{argmin}\frac{1}{B}\sum_{b=1}^{B}\{\widehat{BP}_\lambda^b(\gamma) - \widehat{BP}_\lambda(\gamma)\}^2 \quad (2.26)$$

In case of independent p-values, we simulated $m = 1000$ independent and normally distributed random variables, each one with variance equal to 1 and mean equal to 0 under the null hypothesis while equal to 2 under the alternative hypothesis. The simulations results are shown in Table 2.4.

If we compare the MSE of FDR and power, in case of not empirical Bayes estimates and in case of empirical Bayes estimates, we observe that the a priori information on π_0 induces smaller MSE for both the Bayes FDR and the BP, for all the parameters considered.

TABLE 2.4: Simulation in case of independent p-values over a range of $\pi_0 = 0.1, \ldots, 0.9$ and $\gamma = 0.01, 0.001$; for FDR and Power we reported the true value, the optimal λ that minimizes the bootstrap mean square error, the Monte Carlo empirical bayes estimate (indicated with hat), the true mean square error and the 95% bootstrap confidence interval.

π_0	FDR	λ_1^{opt}	$\widehat{FDR}$	MSE	95% CI	BP	λ_1^{opt}	$\widehat{BP}$	MSE	95% CI
$\gamma = 0.01$										
0.1	0.0039	0.4	0.0111	5.24e-05	[0;0.0130]	0.2824	0.5	0.3418	3.88E-03	[0.3115;1]
0.2	0.0088	0.4	0.0159	5.28e-05	[0;0.0186]	0.2824	0.4	0.3506	5.07E-03	[0.3175;1]
0.3	0.0150	0.4	0.0221	5.35e-05	[0;0.0257]	0.2824	0.4	0.3519	5.33E-03	[0.3155;1]
0.4	0.0231	0.4	0.0303	5.83e-05	[0;0.0354]	0.2824	0.3	0.3676	7.92E-03	[0.3270;1]
0.5	0.0342	0.4	0.0413	6.26e-05	[0;0.0487]	0.2824	0.3	0.3675	8.09E-03	[0.3218;1]
0.6	0.0504	0.4	0.0575	7.71e-05	[0;0.0687]	0.2824	0.2	0.3968	1.43E-02	[0.3424;1]
0.7	0.0763	0.4	0.0834	1.21e-04	[0;0.1017]	0.2824	0.2	0.3983	1.53E-02	[0.3322;1]
0.8	0.1241	0.4	0.1326	3.26e-04	[0;0.1673]	0.2824	0.1	0.4649	3.70E-02	[0.3731;1]
0.9	0.2417	0.4	0.2545	1.80e-03	[0;0.3487]	0.2824	0.1	0.4668	4.53E-02	[0.3170;1]
$\gamma = 0.001$										
0.1	0.0011	0.6	0.0028	2.93e-06	[0;0.0036]	0.0984	0.6	0.1171	5.01E-04	[0.0981;1]
0.2	0.0025	0.5	0.0044	3.77e-06	[0;0.0055]	0.0984	0.6	0.1168	4.98E-04	[0.0962;1]
0.3	0.0043	0.5	0.0062	4.11e-06	[0;0.0079]	0.0984	0.5	0.1201	6.58E-04	[0.0976;1]
0.4	0.0067	0.5	0.0086	4.99e-06	[0;0.0111]	0.0984	0.5	0.1202	7.10E-04	[0.0958;1]
0.5	0.0101	0.5	0.0121	7.32e-06	[0;0.0159]	0.0984	0.4	0.1228	9.13E-04	[0.0956;1]
0.6	0.0150	0.5	0.0172	1.27e-05	[0;0.0234]	0.0984	0.4	0.1232	1.00E-03	[0.0920;1]
0.7	0.0232	0.5	0.0259	3.11e-05	[0;0.0371]	0.0984	0.3	0.1286	1.50E-03	[0.0912;1]
0.8	0.0390	0.6	0.0422	1.13e-04	[0;0.0664]	0.0984	0.2	0.1421	3.02E-03	[0.0922;1]
0.9	0.0838	0.9	0.0927	1.39e-03	[0;0.1981]	0.0984	0.1	0.1688	8.72E-03	[0.0854;1]

Simulations in case of dependence

In case of dependent p-values, otherwise, we assumed the set of random variables $Z_1, \ldots, Z_{1000}$ to have a multivariate normal distribution. Each marginal distribution $Z_i \sim N(\mu, 1)$, as well as before, has the same variance and mean equal to 0 under the null hypothesis while equal to 2 under the alternative hypothesis; in order to explore different pattern of dependency among z-values we defined different covariance matrices.
Specifically we explored three different pattern of dependency among variables (Myers et al., 2012): the autoregressive pattern, the constant pattern, and the unstructured pattern of dependency. With respect to the autoregressive pattern of dependency, we defined each correlation among couples of random variables equals to:

$$Corr(Z_i, Z_j) = \rho^{(|i-j|)}$$

And we fixed $\rho = 0.8$. Conversely, for the constant and the unstructured patterns, we considered an hub-structured matrix of covariances. In particular we split the z-values into 100 clusters, each one composed by 10 z-values, and we imposed a pattern of dependency among z-values only if belonging to the same cluster, while setting to 0 the correlations between z-values belonging to different clusters. The matrix of variances and covariances thus constructed results in a block matrix with 100 sub-matrices of size 10×10 on its diagonal. In case of constant pattern of dependency, we fixed each correlation within clusters equal to the others; in particular, we simulated this pattern of covariances with respect to three different constant values: $c = 0.1, 0.5, 0.9$. In case of unstructured pattern of dependency, differently, the correlations values in each sub-matrix were randomly assigned in the range (0;0.35). The simulations results are shown in Tables 2.5 and 2.6.

In Figures 2.1 and 2.2 the true Bayes FDR values and the FDR empirical Bayes estimates in case of constant pattern of dependency, for the three constant values simulated $(c = 0.1, 0.5, 0.9)$ are plotted, given $\gamma = 0.01$ and $\gamma = 0.001$ respectively,. This comparison reveals that the more high the constant c, the more conservative the estimate of the Bayes FDR. The empirical Bayes estimates of BP show a similar behavior (Figures 2.3 and 2.4).

Furthermore, we investigated the behavior of Bayes FDR and BP estimates in case of independence of the p-values versus the case of autoregressive and hub-unstructured patterns of dependence among p-values. Specifically, we

TABLE 2.5: Simulation in case of dependent p-values given an autoregressive matrix of covariances, over a range of $\pi_0 = 0.1, \ldots, 0.9$ and $\gamma = 0.01, 0.001$; for FDR and Power we reported the true value, the optimal λ that minimizes the bootstrap mean square error, the Monte Carlo empirical bayes estimate (indicated with hat), the true mean square error and the 95% bootstrap confidence interval.

π_0	FDR	λ_1^{opt}	$\widehat{FDR}$	MSE	95% CI	BP	λ_1^{opt}	$\widehat{BP}$	MSE	95% CI
$\gamma = 0.01$										
0.1	0.0039	0.5	0.0104	4.81e-05	[0;0.0123]	0.2824	0.5	0.3418	5.08E-03	[0.3117;1]
0.2	0.0088	0.4	0.0163	6.98e-05	[0;0.0190]	0.2824	0.4	0.3523	6.59E-03	[0.3193;1]
0.3	0.0150	0.4	0.0228	8.39e-05	[0;0.0266]	0.2824	0.4	0.3489	6.33E-03	[0.3129;1]
0.4	0.0231	0.4	0.0311	1.09e-04	[0;0.0365]	0.2824	0.3	0.3677	9.90E-03	[0.3269;1]
0.5	0.0342	0.4	0.0421	1.50e-04	[0;0.0498]	0.2824	0.3	0.3688	1.08E-02	[0.3227;1]
0.6	0.0504	0.4	0.0607	3.24e-04	[0;0.0728]	0.2824	0.2	0.3925	1.65E-02	[0.3381;1]
0.7	0.0763	0.4	0.0879	6.87e-04	[0;0.1079]	0.2824	0.2	0.3971	1.96E-02	[0.3305;1]
0.8	0.1241	0.4	0.1376	1.75e-03	[0;0.1756]	0.2824	0.1	0.4680	4.40E-02	[0.3747;1]
0.9	0.2417	0.9	0.2735	1.52e-02	[0;0.3960]	0.2824	0.1	0.4885	8.12E-02	[0.3097;1]
$\gamma = 0.001$										
0.1	0.0011	0.6	0.0029	4.08e-06	[0;0.0037]	0.0984	0.6	0.1175	9.98E-04	[0.0985;1]
0.2	0.0025	0.6	0.0044	5.62e-06	[0;0.0057]	0.0984	0.6	0.1176	1.09E-03	[0.0972;1]
0.3	0.0043	0.6	0.0064	8.41e-06	[0;0.0083]	0.0984	0.5	0.1190	1.18E-03	[0.0969;1]
0.4	0.0067	0.6	0.0092	1.76e-05	[0;0.0121]	0.0984	0.5	0.1200	1.52E-03	[0.0957;1]
0.5	0.0101	0.6	0.0127	3.15e-05	[0;0.0172]	0.0984	0.4	0.1246	2.04E-03	[0.0972;1]
0.6	0.0150	0.7	0.0194	1.26e-04	[0;0.0282]	0.0984	0.4	0.1226	2.40E-03	[0.0917;1]
0.7	0.0232	0.5	0.0298	3.08e-04	[0;0.0460]	0.0984	0.3	0.1310	3.71E-03	[0.0932;1]
0.8	0.0390	0.6	0.0536	2.21e-03	[0;0.0994]	0.0984	0.2	0.1395	5.90E-03	[0.0903;1]
0.9	0.0838	0.9	0.1498	3.83e-02	[0;0.3112]	0.0984	0.1	0.1798	2.42E-02	[0.0893;1]

TABLE 2.6: Simulation in case of dependent p-values given an hub-unstructured matrix of covariances, over a range of $\pi_0 = 0.1, \ldots, 0.9$ and $\gamma = 0.01, 0.001$; for FDR and Power we reported the true value, the optimal λ that minimizes the bootstrap mean square error, the Monte Carlo empirical bayes estimate (indicated with hat), the true mean square error and the 95% bootstrap confidence interval.

π_0	FDR	λ_1^{opt}	$\widehat{FDR}$	MSE	95% CI	BP	λ_1^{opt}	$\widehat{BP}$	MSE	95% CI
$\gamma = 0.01$										
0.1	0.0039	0.5	0.0103	4.28e-05	[0;0.0122]	0.2824	0.5	0.3407	3.95E-03	[0.3105;1]
0.2	0.0088	0.4	0.0160	5.63e-05	[0;0.0187]	0.2824	0.4	0.3518	5.42E-03	[0.3189;1]
0.3	0.0150	0.4	0.0220	5.65e-05	[0;0.0257]	0.2824	0.4	0.3535	5.74E-03	[0.3173;1]
0.4	0.0231	0.4	0.0303	6.58e-05	[0;0.0355]	0.2824	0.3	0.3685	8.45E-03	[0.3278;1]
0.5	0.0342	0.4	0.0414	7.58e-05	[0;0.0488]	0.2824	0.3	0.3677	8.49E-03	[0.3219;1]
0.6	0.0504	0.4	0.0577	1.07e-04	[0;0.0690]	0.2824	0.2	0.3960	1.44E-02	[0.3418;1]
0.7	0.0763	0.4	0.0848	2.01e-04	[0;0.1035]	0.2824	0.2	0.3935	1.47E-02	[0.3276;1]
0.8	0.1241	0.4	0.1326	5.02e-04	[0;0.1678]	0.2824	0.1	0.4689	3.95E-02	[0.3764;1]
0.9	0.2417	0.5	0.2569	3.37e-03	[0;0.3541]	0.2824	0.1	0.4751	5.25E-02	[0.3208;1]
$\gamma = 0.001$										
0.1	0.0011	0.6	0.0028	3.13e-06	[0;0.0036]	0.0984	0.7	0.1158	5.26E-04	[0.0968;1]
0.2	0.0025	0.5	0.0044	4.20e-06	[0;0.0056]	0.0984	0.6	0.1175	6.10E-04	[0.0970;1]
0.3	0.0043	0.5	0.0062	4.68e-06	[0;0.0079]	0.0984	0.5	0.1206	7.87E-04	[0.0982;1]
0.4	0.0067	0.5	0.0087	6.59e-06	[0;0.0113]	0.0984	0.5	0.1198	8.07E-04	[0.0955;1]
0.5	0.0101	0.5	0.0122	1.06e-05	[0;0.0161]	0.0984	0.5	0.1196	8.85E-04	[0.0925;1]
0.6	0.0150	0.5	0.0175	2.18e-05	[0;0.0239]	0.0984	0.4	0.1232	1.27E-03	[0.0923;1]
0.7	0.0232	0.5	0.0264	5.32e-05	[0;0.0382]	0.0984	0.3	0.1282	1.74E-03	[0.0908;1]
0.8	0.0390	0.5	0.0435	2.04e-04	[0;0.0698]	0.0984	0.2	0.1430	3.75E-03	[0.0926;1]
0.9	0.0838	0.9	0.0989	3.54e-03	[0;0.2185]	0.0984	0.1	0.1695	9.82E-03	[0.0848;1]

FIGURE 2.1: Comparison between the true FDR (black circle) and the FDR empirical Bayes estimates in case of constant pattern of dependency with constant c equal to 0.1 (red triangle), to 0.5 (green triangle) and to 0.9 (blue cross); values are plotted over a range of $\pi_0 = 0.1, \ldots, 0.9$ and given $\gamma = 0.01$.

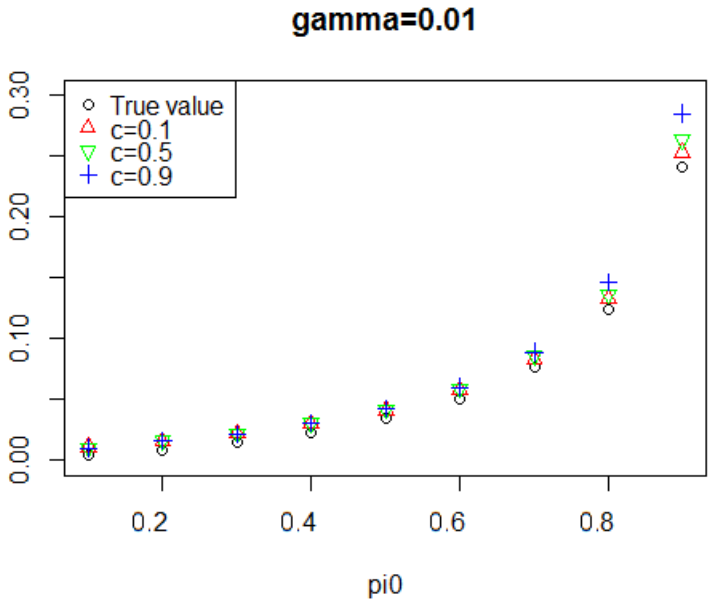


FIGURE 2.2: Comparison between the true FDR (black circle) and the FDR empirical Bayes estimates in case of constant pattern of dependency with constant c equal to 0.1 (red triangle), to 0.5 (green triangle) and to 0.9 (blue cross); values are plotted over a range of $\pi_0 = 0.1, \ldots, 0.9$ and given $\gamma = 0.001$.

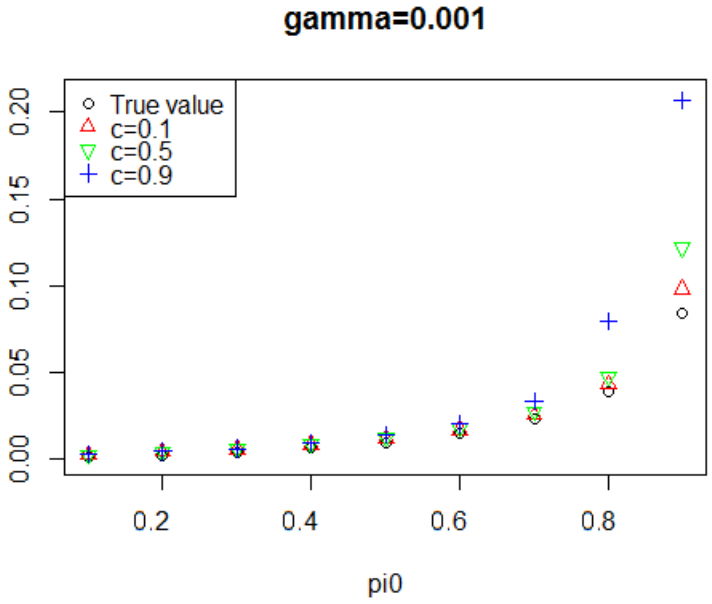


FIGURE 2.3: Comparison between the true Power (black circle) and the Power's empirical Bayes estimates in case of constant pattern of dependency with constant c equal to 0.1 (red triangle), to 0.5 (green triangle) and to 0.9 (blue cross); values are plotted over a range of $\pi_0 = 0.1, \ldots, 0.9$ and given $\gamma = 0.01$.

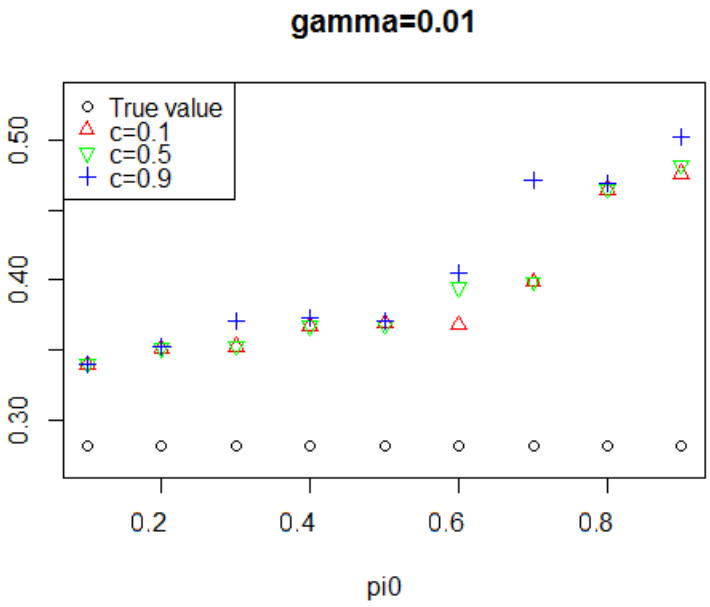
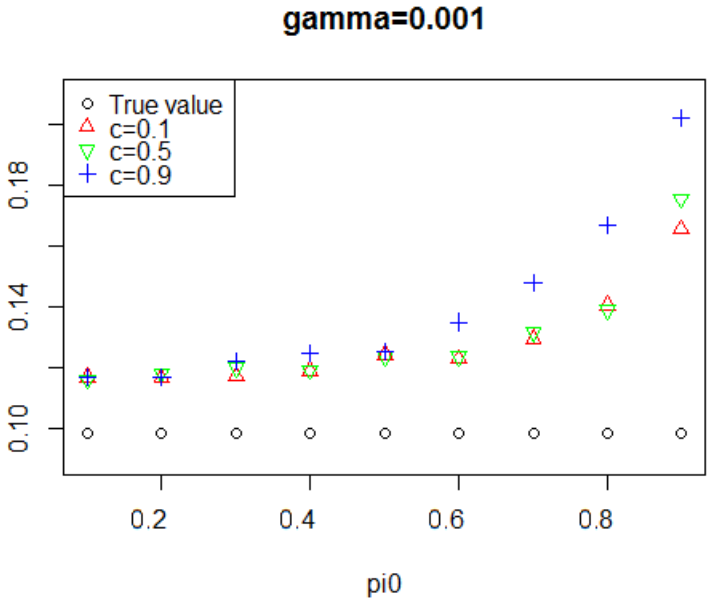
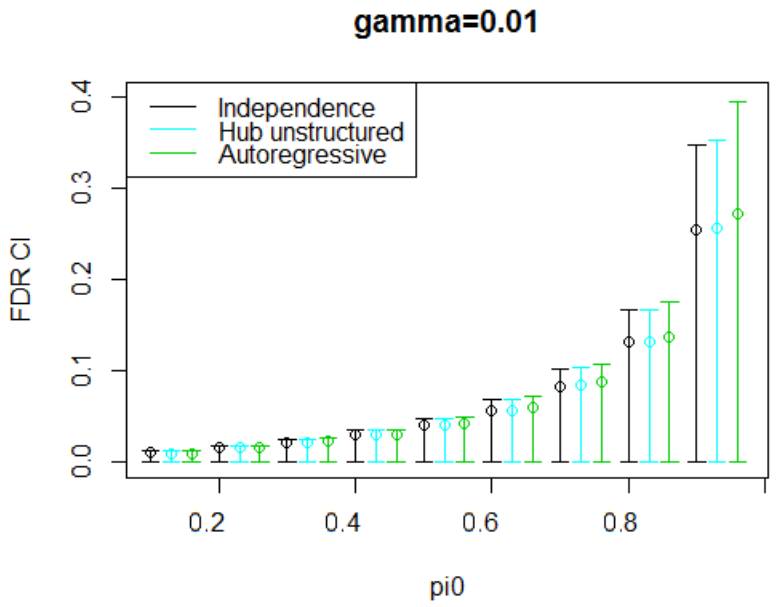


FIGURE 2.4: Comparison between the true Power (black circle) and the Power's empirical Bayes estimates in case of constant pattern of dependency with constant c equal to 0.1 (red triangle), to 0.5 (green triangle) and to 0.9 (blue cross); values are plotted over a range of $\pi_0 = 0.1, \ldots, 0.9$ and given $\gamma = 0.001$.



observe that, in case of independence, the MSE of the Bayes FDR and BP are smaller than the corresponding MSE in case of dependence. Moreover, as the hub-unstructured correlation matrix induces a weaker pattern of dependency than the autoregressive correlations matrix, the MSE in Table 2.6 are all smaller than the MSE in Table 2.5. In Figures 2.5 and 2.6 the Bootstrap confidence intervals of the Bayes FDR are plotted, for $\gamma = 0.01, 0.001$: the more the distance from a condition of independence, the wider the bootstrap confidence intervals.

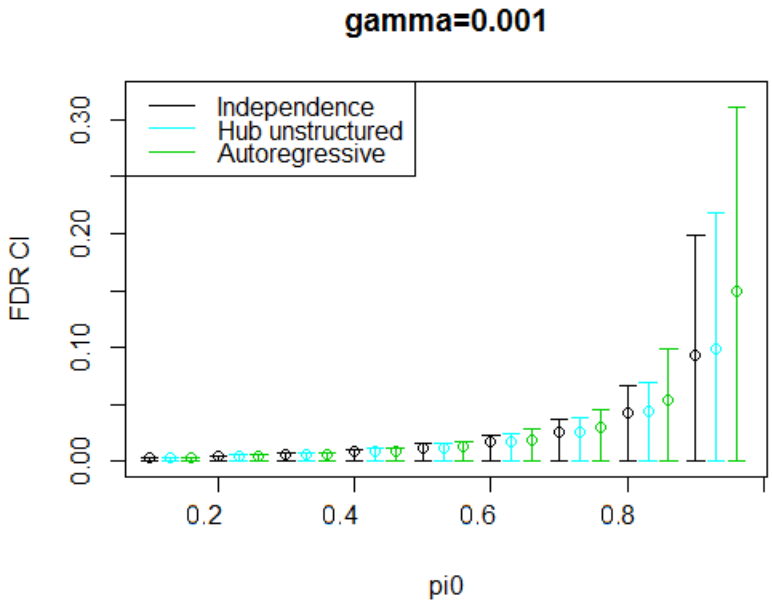
FIGURE 2.5: Comparison between the Bayes FDR confidence intervals in case of independence (black line), in case of hub-unstructured pattern of dependency (light-blue line), and in case of autoregressive pattern of dependency (green line); values are plotted over a range of $\pi_0 = 0.1, \ldots, 0.9$ and given $\gamma = 0.01$.



Simulation with a block bootstrap procedure

At last, we realized a simulation, in the spirit of the application on real fMRI data, in order to investigate the adequacy of the bootstrap procedure on the p-values, as proposed by Storey, versus a block bootstrap procedure on the original data series, in case of dependency among p-values. First of all, we have to point out that the standard bootstrap procedure implies the assumption of independence on the bootstrapped values. Whenever the p-values are dependent the standard bootstrap does not allow to handle the dependency

FIGURE 2.6: Comparison between the Bayes FDR confidence intervals in case of independence (black line), in case of hub-unstructured pattern of dependency (light-blue line), and in case of autoregressive pattern of dependency (green line); values are plotted over a range of $\pi_0 = 0.1, \ldots, 0.9$ and given $\gamma = 0.001$.



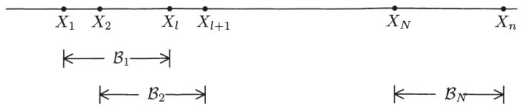
structure. A proper bootstrap procedure in case of dependency is the moving block bootstrap (MBB) procedure (Lahiri, 2013), in which the only assumption of independence is adopted among blocks of values.

In fact, in a MBB procedure, usually, one aims to fix the length, l, of each block long enough such that the observations which are more distant than l are nearly independent. As a consequence, it is worth to observe that, while the original data are dependent, the sampled blocks are independent by construction.

One of the main strength of the MBB is that it does not require any parametric model assumptions on the data; furthermore the MBB allows to capture at any lag the finite-dimensional joint distribution of the unknown underlying time series.

This simulation has two levels of complexity: we not only had to simulate a vector of p-values as in the previous simulations, but we had also to simulate the dataset from which the p-values are computed. Therefore, let us assume a network composed by 50 nodes; for each node we collect a time series of

FIGURE 2.7: The collection of N overlapping blocks under the MBB.



$n = 200$ measurements. In addition, recall that we are interested in the construction of an unweighted and undirected network such that a link between two nodes is placed if the correlation among those nodes is statistically significant, namely if the correlation measure is higher than a proper threshold. The network construction results in a problem of multiple testing; in particular we have $m = \frac{50(50-1)}{2} = 1225$ conjoint comparisons.

Analogously to the example on real data, we computed the Spearman's rank correlation coefficients, whose test statistic is defined in 2.1. We assumed the set of random variables $T_1, \ldots, T_{1225}$ to have a multivariate normal distribution; each variable, hence, is marginally distributed as a $N(0,1)$ under H_0 and as a $N(2,1)$ under H_1. The choice of the multivariate normal approximation guarantees a simplification in this simulation's step, but depends also on the numerous properties of the multivariate normal distribution, primarily the equivalence of the concepts of independence and incorrelation. Given the simulated z-values, we computed the corresponding p-values $p_j = 1 - \Phi(z_j)$ for $i = 1, \ldots, m$, where Φ is the cdf of a standard normal. In order to explore the case of dependent p-values, we designed an interpretable and meaningful pattern of dependency. With this regard, we simulated a hub-structured matrix of correlations among p-values such that the blocks of p-values were induced by an a priori set of clusters or modules of the original nodes. In particular, let us assume 5 modules each one composed by 10 nodes; in Table 2.7 is shown a portion of the matrix of the z-values:

We enumerated the p-values in this order: at first, we enumerated the p-values given by the comparisons between nodes belonging to the same module, and only later we enumerated the remaining p-values. The classification into modules of the nodes induced a partition of the p-values into 10 blocks, composed by 10 p-values each, and into 45 blocks composed by 25 p-values each. It seemed plausible to assume that the p-values belonging to the same block were dependent, while p-values belonging to different blocks were assumed to be independent.

TABLE 2.7: Portion of the p-values symmetric matrix. Each column and row corresponds to a ROI (or node equivalently). The p-values are enumerated accordingly to the following rule: the p-values associated to nodes belonging to the same module are enumerated first, starting from the top left of the matrix.

Nodes	I	II	III	IV	V	VI	VII	...
I	1	p_1	p_2	p_3	p_4	p_{101}	p_{102}	
II		1	p_5	p_6	p_7	p_{106}	p_{107}	
III			1	p_8	p_9	p_{111}	p_{112}	
IV				1	p_{10}	p_{116}	p_{117}	
V					1	p_{121}	p_{122}	
VI						1	p_{11}	
VII							1	
...								...

The pattern of dependency simulated, hence, has an hub-structure, as illustrated in Table 2.8.

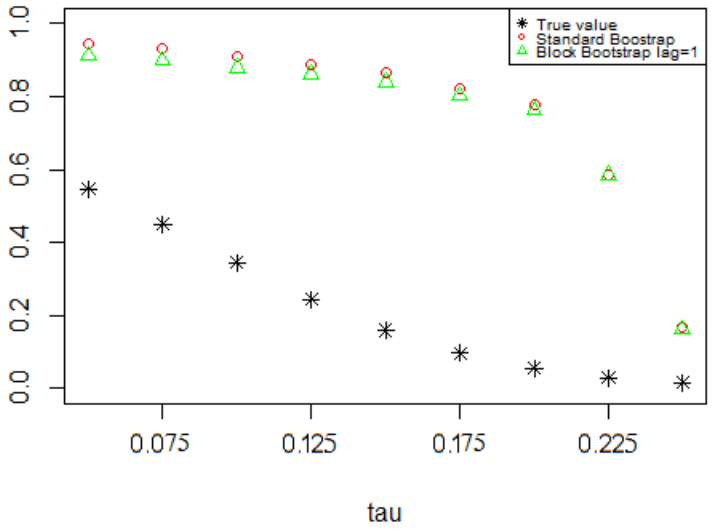
TABLE 2.8: Portion of the p-values symmetric matrix. The pattern of dependency has an hub-structure: correlations within blocks are set equal to c, while the remaining correlations are set equal to 0.

P-values	p_1	p_2	...	p_9	p_{10}	p_{11}	p_{12}	...
p_1	1	c		c	c	0	0	
p_2		1		c	c	0	0	
...			1	c	c	0	0	
p_9				1	c	0	0	
p_{10}					1	0	0	
p_{11}	0	0	0	0	0	1	c	
p_{12}	0	0	0	0	0		1	
...								...

We simulated a constant pattern of correlations among p-values, in each block, and we performed our simulation for c=0.1. We computed the empirical Bayes estimates for FDR and BP, with respect to the parameters $\tau = 0.05, \ldots, 0.25$ and $\lambda = 0.1, \ldots, 0.9$, and given $\pi_0 = 0.8$; hence we compared the results obtained with the bootstrap on the p-values with those obtained with a block-bootstrap on the original data.
In Figure 2.8 we show the true Bayes FDR value compared to the empirical

Bayes estimates of FDR, computed respectively with a standard bootstrap on the p-values and with a block bootstrap on the original data, at lag 1.

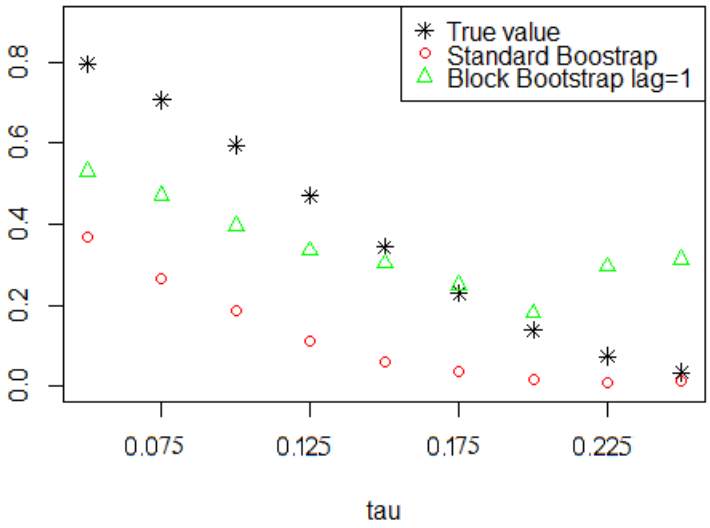
FIGURE 2.8: Comparison between the true Bayes FDR (black star), the Bayes FDR estimates in case of standard bootstrap on the p-values (red circle) and the Bayes FDR estimates in case of block bootstrap, with lag 1, on the original data (green triangle); values are plotted over a range of $\tau = 0.05, \ldots, 0.25$ and given $\pi_0 = 0.8$. The simulated pattern of dependency has a constant hub-structure, with c=0.1.



The Monte Carlo estimates and also the MSE values are quite similar. This simulation result suggests that, even in case of dependency among p-values, the estimation process of FDR, based on the standard bootstrap procedure, is robust.

Similarly to the previous figure, in Figure 2.9 we show the true power value compared to the empirical Bayes estimates of BP, computed respectively with a standard bootstrap on the p-values and with a block bootstrap on the original data, at lag 1. With the standard bootstrap procedure, for every τ, the power is underestimated; on the contrary, given a block bootstrap procedure, the BP is underestimated only for $\tau \leq 0.15$. In order to assure a fully conservative approach, hence, the standard bootstrap procedure is preferable.

FIGURE 2.9: Comparison between the true BP (black star), the BP estimates in case of standard bootstrap on the p-values (red circle) and BP estimates in case of block bootstrap, with lag 1, on the original data (green triangle); values are plotted over a range of $\tau = 0.05, \ldots, 0.25$ and given $\pi_0 = 0.8$. The simulated pattern of dependency has a constant hub-structure, with c=0.1.



Chapter 3

Application on Neuro-functional data

3.1 Aim of the study

One of the most significant changes in the demography of the last decades is the continuous increment of estimated-life; projection studies suggest that a large proportion of those born in western countries from year 2000 might have a life expectancy of about 100 years (Christensen et al., 2009). Although this perspective may be exciting, we have to admit that "old age" is a complicated, and still mysterious, period of human life. As we age we need to cope with a series of changes that can be more or less significant. This coping process can be influenced by a large variety of biological, cognitive and social factors that will inevitably make it more or less successful. In a society that is projected to grow older and older in the next fifty years, it is not surprising that the study of these factors and of their interactions became, in the last decade, one of the most investigated topics in life sciences. With this regard, one of the main efforts of psychologists and cognitive scientists was to explore the pattern of age-related behavioural, neurofunctional and morphometrical changes in different cognitive domains.

From 1992, when the first functional neuroimaging study (Tempel and Perlmutter, 1992) on aging appeared in the international literature, to October 2015, more than 220 papers had been published on this topic. This large number of neuroimaging findings, together with a large pool of behavioral studies, contributed to the development of a number of theories and models about the aging processes (see Barulli and Stern, 2013; Dennis and Cabeza, 2008, for a recent review), with the most of them referring to the concept of compensation.

Resting state age-related changes could be addressed in terms of either de-differentiation, or localization. These two theoretical frameworks make two opposite predictions about the manifestation of age-related changes in resting state functional connectivity. According to the de-differentiation hypothesis healthy aging would be characterized by a significant reduction of the signal-to-noise ratio (i.e. the ratio between the desired signal and the non-neural background noise) and, as a result, by a highly dense and less organized neurofunctional network. On the contrary, according to the localization hypothesis resting state age-related changes would be characterized by a reduction of the long-term neurofunctional connections and, as a consequence, by a detectable increment of intra-lobar neurofunctional connections. This means that as we age, the connections between distant brain regions would get lost in favour of the neurofunctional connections between brain regions that belong to the same anatomical macro-structures.
To explicitly test these hypotheses, we constructed subject-specific balanced unweighted and undirected networks, for the younger and the older participants, by adopting a thresholding method which, as already described in Section 2.3, is based on the identification of a statistical threshold τ capable of balancing the I error rates and the power, in a bayesian perspective.
Furthermore, to explicitly describe each subject-specific network, we identify a set of relevant network indexes, each one extracted from each network:

1. the global number of significant links between brain regions;
2. the degree centrality of each node (i.e. each brain region);
3. the mean degree centrality of each network;
4. the number of hubs in each network;
5. the modularity measure (i.e. a measure that quantifies the modular structure beyond the expected at random one);
6. we created a new index that measured the level of between-modules (BM) functional connectivity.

We adopted these indexes to explicitly test the de-differentiation and the localization hypotheses according to the scenarios reported in Table 3.1.

The two hypotheses were also tested by means of a distance measure (namely the length of the significant links) according to the scenarios reported in Figure 3.1.

FIGURE 3.1: Graphical representation of the de-differentiation and of the localization hypothesis in healthy aging.

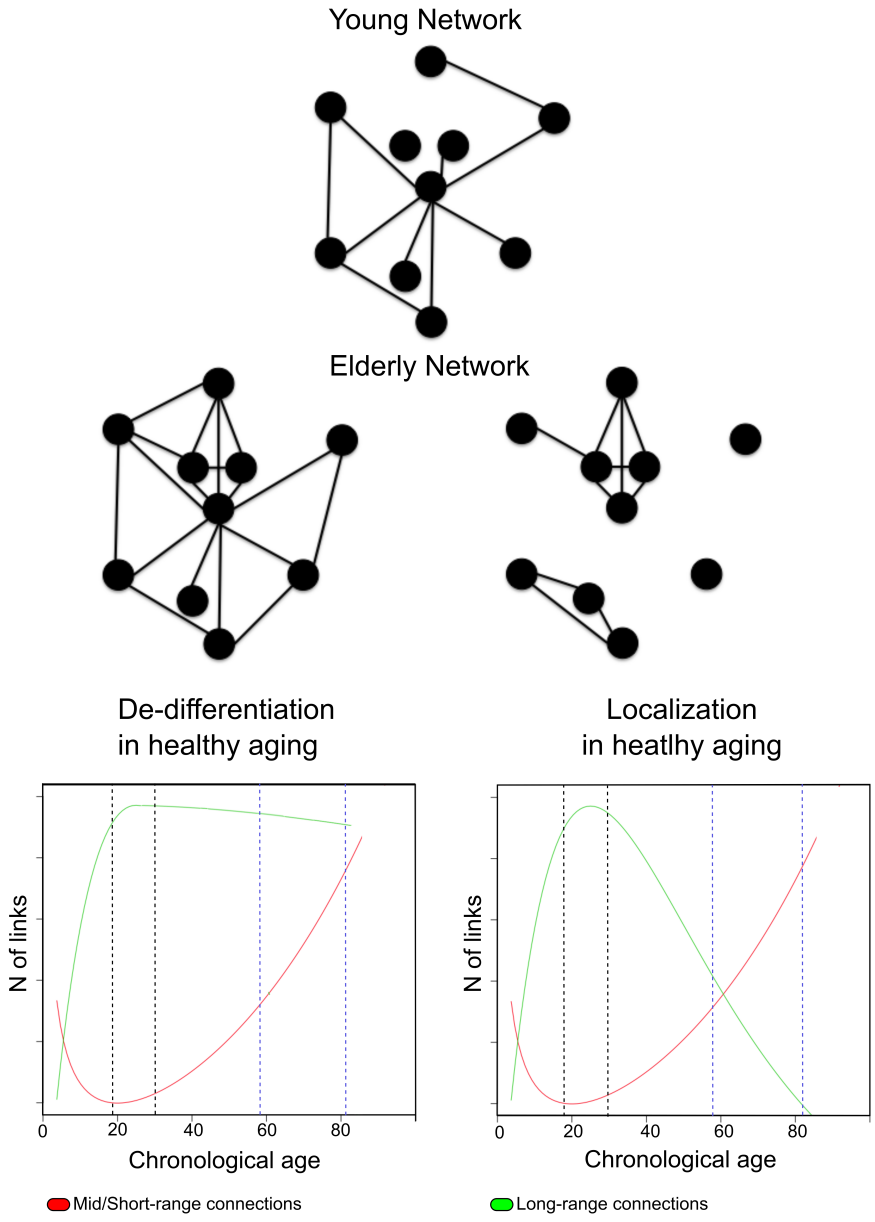


TABLE 3.1: Possible Scenarios. Here we report the expected results, for each singe network measure, in the context of the de-differentiation and of the localization hypothesis.

Measure	**It is a measure of**	**De-differentiation**	**Localization**
Number of links	Size of the network	Elderly>Young	n.e.
Mean network degree	Density of the network	Elderly>Young	n.e.
Number of hubs	Number of vertices characterized by especially high degree	Elderly>Young	n.e.
Modularity	Degree to which the network may be subdivided into such clearly delineated and nonoverlapping groups (modules)	Elderly<Young	Elderly>Young
Between-modules Index	Proportion of links that go out from one module among all the links that exist outside that module	Elderly>Young	Elderly<Young

In the upper panel of Figure 3.1 is represented an hypothetical young network together with the hypothetical elderly network in the case of de-differentiation (on the left) and in case of localization (on the right). According to the de-differentiation hypothesis, healthy elderlies would developed new short-range neurofunctional links (lower panel on the left, red curve), while maintaining the long-range neurofunctional links (green) developed during childhood. As a result, the elderly network would be either denser, or differently organized, than the youngers' one. On the contrary, according to the localization hypothesis, healthy elderlies would progressively loose long-range connections (lower panel on the right, green curve) in favour of short-range links (red). As a result, the healthy elderlies network would be as dense as the youngers' one, but differently organized. The black dotted lines isolate the young participants' age range, while the blue dotted lines isolate the elderlies' age-range considered for this study. The red and the green pattern in the range $0 < \text{age} < 20$ has been designed on the basis of the results reported in the study by Fair et al. (2009).

3.2 Data Acquisition

In order to verify the hypothesis of de-differentiation in healthy elderlies we acquired the MRI scans of 70 healthy participants, 35 elderlies (17 males; age = 66.5 $\pm$ 7.4 years, min = 52, max = 81) and 35 young subjects (17 males; age = 26.5 $\pm$ 4.3 years, min = 19, max = 34).
All the participants had a normal cognitive profile and none of them had history of major medical disorders or was under psychotropic medication. All the participants gave their written informed consent to the MRI scan. The data were collected in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

The MRI data were acquired using a 1.5 T General Electrics Scanner at the Neuroradiology Department of the Niguarda Ca' Granda Hospital, Milan, Italy and a 1.5 T Siemens Avanto scanner at the Department of Diagnostic Radiology and Bioimages of IRCCS Galeazzi, Milan, Italy .
The rsfMRI acquisition parameters for the General Electrics Scanner and for the Siemens Scanner were as follows:

TABLE 3.2: Summary of the rsfMRI acquisition parameters

	General Electrics Scanner	Siemens Scanner
Repetition time (TR)	3000 ms	3000 ms
Echo time (TE)	60 ms	60 ms
Flip angle	90°	90°
Field-of-view (FOV)	$240 \times 240\ mm^2$	$280 \times 210\ mm^2$

During the MRI exam, that lasted for 10 minutes, the subjects were instructed to lie down supine with their eyes closed and to try not to think anything in particular while being awake. Foam padding was positioned around the head to restrict participants' head motion.
Resting-state functional images were acquired using an Echo Planar Imaging (EPI) sequence sensitive to BOLD contrast, resulting in a series of 200 T2 weighted BOLD volumes.

A total of 35 contiguous transverse slices of 4 mm thickness were collected with no intersection gap for each volume in both cases.
After the fMRI session, a high-resolution T1 weighted volumetric scan was acquired by using a 3D Magnetization Prepared Rapid Gradient-Echo (MP-RAGE) sequence with the following parameters: TE = 60 ms, flip angle = 90°, FOV = $256 \times 192\ mm^2$, inversion time (TI) = 768 ms, thus obtaining 145 slices

(thickness = 1 mm) with no interslice gap, and acquired on oblique sections parallel to the AC-PC line in order to cover the entire brain volume. As a final result, we obtained a high-resolution T1 image with a voxels size of $1 \times 1 \times 1$ mm for each participant.

All the images were analyzed using Matlab R2014 (MathWorks, Natick, MA, USA). In particular, the image preprocessing was carried out using a Matlab toolbox DPARSF-A ('Data Processing Assistant for Resting-State fMRI Advanced') (Chao-Gan and Yu-Feng, 2010). DPARSF-A is based on Statistical Parametric Mapping software functions (SPM8) and REST software (Song et al., 2011).

After converting DICOM files to NIFTI images, the raw data were corrected for any movement by means of a least squares approach called the Levenberg-Marquardt method (Moré, 1978) and a six parameters (rigid body) spatial transformation. It is worth noting that the subject's motion parameters did not exceed 2.5 mm and the 2.5 degrees of rotation in any dimension through the resting-state run. Afterwords, each individual structural T1 weighted image was coregistered to the mean functional image T2 weighted scan, to obtain an overlap between functional and anatomical image.

These transformed structural images were then segmented into grey matter (GM), white matter (WhM) and cerebrospinal fluid (CSF) compartments, using the so called New Segmentation that is a unified segmentation algorithm (Ashburner and Friston, 2005).
Motion-corrected functional volumes were stereotactically normalized to the Montreal Neurological Institute (MNI) template (Collins et al., 1994) using the normalization parameters estimated during the unified segmentation on T1 weighted image (Ashburner and Friston, 2005). As a result, we obtained 200 normalized volumes with voxels of $3 \times 3 \times 3$ mm.

The normalized rsfMRI data were then smoothed with a Gaussian isotropic filter, with a Full-Width-Half-Maximum (FWHM) = $4 \times 4 \times 4$ mm. In the following step, data preprocessing included removal, by means of a regression analysis, of the global whole-brain signal (Fox et al., 2005; Fransson, 2005; Greicius et al., 2003), of the six parameters of head motion, of the CSF and the WhM signals. This was done to reduce the effects of non-neuronal BOLD

fluctuations (Fox et al., 2005; Kelly et al., 2008).

rsfMRI data were then submitted to a temporal band-pass filtering (0.01 to 0.08 Hz). The choice of this frequency interval was guided by the fact that prior works on resting-state fMRI functional connectivity have found that low frequency fluctuations (LFFs) are reported to be of physiological importance (Biswal et al., 1995) and are supposed to reflect spontaneous neuronal activity (Lu et al., 2007).

Finally, DPARSF-A allowed us to extract, for each participant, 116 time-series representing the mean LFFs within the 116 non-overlapping anatomical Regions of Interest (ROIs), defined by the automated anatomical labelling (AAL) atlas previously validated by Tzourio-Mazoyer et al. (2002). The MNI coordinates of the center of the 116 ROIs were taken from Liu et al. (2007) and are reported in Table 3.3. All regions are bilaterally symmetric; left-hand and right-hand sided homologues are distinguished by the prefix L and R, respectively.

TABLE 3.3: Cortical and subcortical regions defined in AAL template image in standard stereotaxic space.

ID	Brain region name	MNI coordinates		
		X	Y	Z
1	L precentral gyrus	-39	-6	51
2	R precentral gyrus	41	-8	52
3	L superior frontal gyrus. dorsolateral	-18	35	42
4	R superior frontal gyrus. dorsolateral	22	31	44
5	L superior frontal gyrus. orbital part	-17	47	-13
6	R superior frontal gyrus. orbital part	18	48	-14
7	L middle frontal gyrus. lateral part	-33	33	35
8	R middle frontal gyrus. lateral part	38	33	34
9	L middle frontal gyrus. orbital part	-31	50	-10
10	R middle frontal gyrus. orbital part	33	53	-11
11	L opercular part of inferior frontal gyrus	-48	13	19
12	R opercular part of inferior frontal gyrus	50	15	21
13	L area triangularis	-46	30	14
14	R area triangularis	50	30	14

Continue on next page ⟶

Continue from previous page ⟵

ID	Brain region name	MNI coordinates		
		X	Y	Z
15	L orbital part of inferior frontal gyrus	-36	31	-12
16	R orbital part of inferior frontal gyrus	41	32	-12
17	L rolandic operculum	-47	-8	14
18	R rolandic operculum	53	-6	15
19	L supplementary motor area	-5	5	61
20	R supplementary motor area	9	0	62
21	L olfactory cortex	-8	15	-11
22	R olfactory cortex	10	16	-11
23	L superior frontal gyrus. medial part	-5	49	31
24	R superior frontal gyrus. medial part	9	51	30
25	L superior frontal gyrus. medial orbital part	-5	54	-7
26	R superior frontal gyrus. medial orbital part	8	52	-7
27	L gyrus rectus	-5	37	-18
28	R gyrus rectus	8	36	-18
29	L insula	-35	7	3
30	R insula	39	6	2
31	L anterior cingulate gyrus	-4	35	14
32	R anterior cingulate gyrus	8	37	16
33	L middle cingulate	-5	-15	42
34	R middle cingulate	8	-9	40
35	L posterior cingulate gyrus	-5	-43	25
36	R posterior cingulate gyrus	7	-42	22
37	L hippocampus	-25	-21	-10
38	R hippocampus	29	-20	-10
39	L parahippocampal gyrus	-21	-16	-21
40	R parahippocampal gyrus	25	-15	-20
41	L amygdala	-23	-1	-17
42	R amygdala	27	1	-18
43	L calcarine sulcus	-7	-79	6
44	R calcarine sulcus	16	-73	9
45	L cuneus	-6	-80	27
46	R cuneus	14	-79	28
47	L lingual gyrus	-15	-68	-5

Continue on next page ⟶

Continue from previous page ⟵

ID	Brain region name	MNI coordinates		
		X	Y	Z
48	R lingual gyrus	16	-67	-4
49	L superior occipital	-17	-84	28
50	R superior occipital	24	-81	31
51	L middle occipital	-32	-81	16
52	R middle occipital	37	-80	19
53	L inferior occipital	-36	-78	-8
54	R inferior occipital	38	-82	-8
55	L fusiform gyrus	-31	-40	-20
56	R fusiform gyrus	34	-39	-20
57	L postcentral gyrus	-42	-23	49
58	R postcentral gyrus	41	-25	53
59	L superior parietal lobule	-23	-60	59
60	R superior parietal lobule	26	-59	62
61	L inferior parietal lobule	-43	-46	47
62	R inferior parietal lobule	46	-46	50
63	L supramarginal gyrus	-56	-34	30
64	R supramarginal gyrus	58	-32	34
65	L angular gyrus	-44	-61	36
66	R angular gyrus	46	-60	39
67	L precuneus	-8	-25	70
68	R precuneus	7	-32	68
69	L paracentral lobule	-7	-56	48
70	R paracentral lobule	10	-56	44
71	L caudate nucleus	-11	11	9
72	R caudate nucleus	15	12	9
73	L putamen	-24	4	2
74	R putamen	28	5	2
75	L globus pallidus	-18	0	0
76	R globus pallidus	21	0	0
77	L thalamus	-11	-18	8
78	R thalamus	13	-18	8
79	L transverse temporal gyri	-42	-19	10
80	R transverse temporal gyri	46	-17	10

Continue on next page ⟶

Continue from previous page ⟵

ID	Brain region name	MNI coordinates		
		X	Y	Z
81	L superior temporal gyrus	-53	-21	7
82	R superior temporal gyrus	58	-22	7
83	L superior temporal pole	-40	15	-20
84	R superior temporal pole	48	15	-17
85	L middle temporal gyrus	-56	-34	-2
86	R middle temporal gyrus	57	-37	-1
87	L middle temporal pole	-36	15	-34
88	R middle temporal pole	44	15	-32
89	L inferior temporal gyrus	-50	-28	-23
90	R inferior temporal gyrus	54	-31	-22
91	L Cerebelum-10	-22	-34	-42
92	R Cerebelum-10	27	-34	-41
93	L Cerebelum-3	-8	-37	-19
94	R Cerebelum-3	13	-34	-19
95	L Cerebelum-4-5	-14	-43	-17
96	R Cerebelum-4-5	18	-43	-18
97	L Cerebelum-6	-22	-59	-22
98	R Cerebelum-6	26	-58	-24
99	L Cerebelum-7b	-31	-60	-45
100	R Cerebelum-7b	34	-63	-48
101	L Cerebelum-8	-25	-55	-48
102	R Cerebelum-8	26	-56	-49
103	L Cerebelum-9	-10	-49	-46
104	R Cerebelum-9	10	-49	-46
105	L Cerebelum-Crus1	-35	-67	-29
106	R Cerebelum-Crus1	38	-67	-30
107	L Cerebelum-Crus2	-28	-73	-38
108	R Cerebelum-Crus2	33	-69	-40
109	Vermis-1-2	2	-39	-20
110	Vermis-10	1	-46	-32
111	Vermis-3	2	-40	-11
112	Vermis-4-5	2	-52	-6
113	Vermis-6	2	-67	-15

Continue on next page ⟶

Continue from previous page ⟵

ID	Brain region name	MNI coordinates		
		X	Y	Z
114	Vermis-7	2	-72	-25
115	Vermis-8	2	-64	-34
116	Vermis-9	2	-55	-35

End Table

To construct the ROIs we adopted the draw toolbox available with the software MRIcron; in particular, we designed 116 independent spherical volume of interest (VOI) centered in each stereotaxic coordinate with 4 mm of radius (as a consequence we obtained 116 files, one for each VOI). The 116 .voi files were then converted into the .nii format to make our ROIs compatible with the software DPARSF-A.

3.3 Brain Network Construction

For each participant, the first element of the analysis is a 200×116 matrix where 200 is the length of a time-series of the mean LFFs, while 116 are the non-overlapping anatomical ROIs. The functional connectivity (formally defined in Friston et al. (1993) and Friston (1994) as the temporal correlation, or, more generally, as the deviation from statistical independence, that exists between brain units) has been computed for each subject by means of the Spearman's rank correlation coefficients r, because such a procedure results in a completely non-parametric approach; the correlation values, obtained from the comparison between all possible pairs of ROIs, are stored in a correlation vector $(r_1, \ldots, r_m)$ of length $m = \frac{116(116-1)}{2} = 6670$.

We therefore tested, for each participant, m hypotheses H_{0j} of null correlation versus the alternative hypotheses of positive correlation:

$$H_0 : \rho_j \leq 0 \qquad vs \qquad H_1 = \bar{H}_0 : \rho_j > 0$$

Under $\rho_j = 0$, the statistic test t_j

$$t_j = \frac{r_j}{\sqrt{1 - r_j^2}} \sqrt{n-2} \qquad for\ j = 1, \ldots, m$$

is distributed approximately as Student's t distribution with n-2 degrees of freedom, where n = 200 is the length of each ROI's specific time-series. Given the approximated null distribution of the random variable t_j, the p-values are defined as

$$p_j = 1 - F_{n-2}(t_j)$$

where F_{n-2} represents the Student's distribution function, with $n-2$ degrees of freedom. The same rejection region is assumed for each test, and is thus defined as

$$\Gamma = \{r_j \geq \tau\} = \{p_j \leq \gamma\}$$

where τ is the rejection threshold while

$$\gamma = 1 - F_{n-2}(\sqrt{\frac{(n-2)\tau^2}{1-\tau^2}}) \tag{3.1}$$

γ represents the Per-Comparison Error Rate (PCER), that is the significance level of a single comparison. It is worthy to note that the threshold τ and γ are related, so that at the increasing of the threshold τ corresponds the decreasing of γ. At this stage, it is worthy to remember that by testing the correlation vector we are actually simultaneously testing 6670 single hypotheses. From a multiple testing perspective there are two different kinds of probabilities that deserve to be controlled: the probability of false discoveries (also called Bayes False Discovery Rate (Efron, 2010), i.e. $FDR(\gamma) = P(H_0\ true|p_j \leq \gamma)$, and the power, i.e. $BP(\gamma) = P(p_j \leq \gamma|H_0\ not\ true)$ (Berlingeri, 2015).
In order to estimate and balance those two probabilities, we adopted here the empirical Bayes approach illustrated in Section 2.3, that still holds under the assumption of the p-values to be valid.
With $\hat{F}(\gamma)$ we denote the empirical distribution of the observed p-values:

$$\hat{F}(\gamma) = \frac{\#\{p_j \leq \gamma\}}{m}$$

Let us introduce, now, a second parameter, say λ, such that all the non-null cases give p-values less then λ. The conservative estimate of the a priori probability of a null hypothesis to be true is defined by Storey (2002) as:

$$\pi_0 = \frac{\#\{p_j > \lambda\}}{m(1-\lambda)}$$

Such a procedure, nevertheless, imply the independence of the p-values, that

seems a too strong hypothesis in our framework. Although the hypothesis of independence of the p-values seems too strong in our framework, we decided however to bootstrap directly the p-values for estimating the optimal value of the tuning parameter λ, according to Storey (2002), in the light of the robustness shown by the estimators (Section 2.3.2).
As in Storey (2002), the optimal value of the tuning parameter λ, i.e. the value that minimized the bootstrap mean square error of FDR, was selected for each participant. The empirical estimates (Section 2.3.2) were then used to obtain the estimates of $FDR(\gamma)$ and $BP(\gamma)$:

$$\widehat{FDR}_\lambda(\gamma) = \frac{\gamma\hat{\pi}_0}{\#\{p_j \leq \gamma\}/m} = \frac{\gamma\#\{p_j > \lambda\}}{(1-\lambda)\#\{p_j \leq \gamma\}}$$

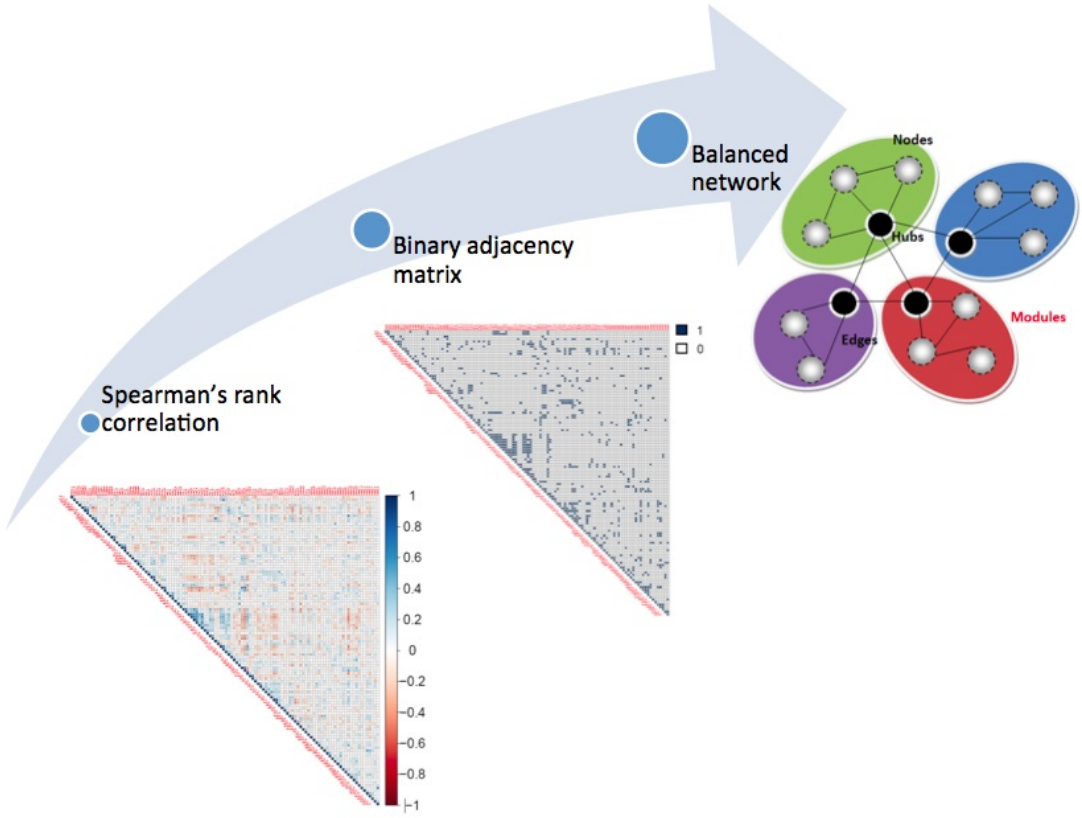
$$\widehat{BP}_\lambda(\gamma) = \frac{\#\{p_j \leq \gamma\}(1-\lambda) - \gamma\#\{p_j > \lambda\}}{m(1-\lambda) - \#\{p_j > \lambda\}}$$

To choose an *appropriate* threshold of the rejection region, it is important to control the probability of false rejections, but also to take into account the corresponding power; this means that we select the threshold that guarantees the best balance between these two forces.
The first step of our procedure consists in the selection of a suitable threshold γ that permits to control for a certain proportion of false positives; here we choose to keep the FDR below the 1% for all the participants:

$$FDR_\lambda(\gamma) \leq 0.01 \tag{3.2}$$

This choice allowed us to control, across participants, for the same overall amount of FDR. Inequality 3.2 is satisfied for a subset of γ values; hence as a second step the minimumγ value for which 3.2 is satisfied is chosen (for each participant). This approach allowed us to create adjacency matrix in which the choice of the γ threshold, that is related to τ according to 3.1, is completely data-driven and maximizes the power, given the maximum overall amount of possible, and acceptable for this specific field of research, false positives. Finally, by keeping the FDR under the 1% we were able to check, post-hoc, whether our participants had similar level of threshold τ; this in turn allowed us to suggest that the overall signal-to-noise ratio across participants was similar, at least at a descriptive level.

FIGURE 3.2: Rationale of the brain network construction.



In Figure 3.2 it is summarized the rationale adopted to design the subject-by-subject balanced brain networks. In order to fully illustrate our procedure of brain network construction first we concentrate on the analysis of a single participant.

By way of example, let us select the subject $60-th$, a 33 years old woman: the main objective of this application on real data concerned the construction of a brain map or network that allowed us to capture the structure of functional connectivity among brain regions.

As a first step, we extracted a 200×116 matrix (that includes the time-series representing the mean low frequency fluctuations for each ROI of interest) according to the procedure describe in 3.2.

Since we did not have any distributional information on the data and in order to make as few assumptions as possible, as a second step, we computed the correlation's coefficients by means of the Spearman's rank coefficient.

As we selected 116 anatomical non-overlapping ROIs, we performed 6670 multiple comparisons. In this regard, we conjointly controlled the probability of false discoveries and the power; specifically, we adopted the empirical estimates of FDR and BP, as we did not have any additional or a priori information on the probability π_0.
Furthermore, we carried out the analysis for τ values within the range $\{0.05, 0.1, \ldots, 0.4, 0.45\}$ and for λ values within the range $\{0.05, 0.1, \ldots, 0.4, 0.45\}$.
We selected the optimal values of λ_1 and λ_2 accordingly to 2.25 and 2.26 and we computed the bootstrap MSE by means of a standard bootstrap procedure on the p-values. The resulting values of the Bayes FDR and BP, depending only on γ, are reported in Table 3.4.

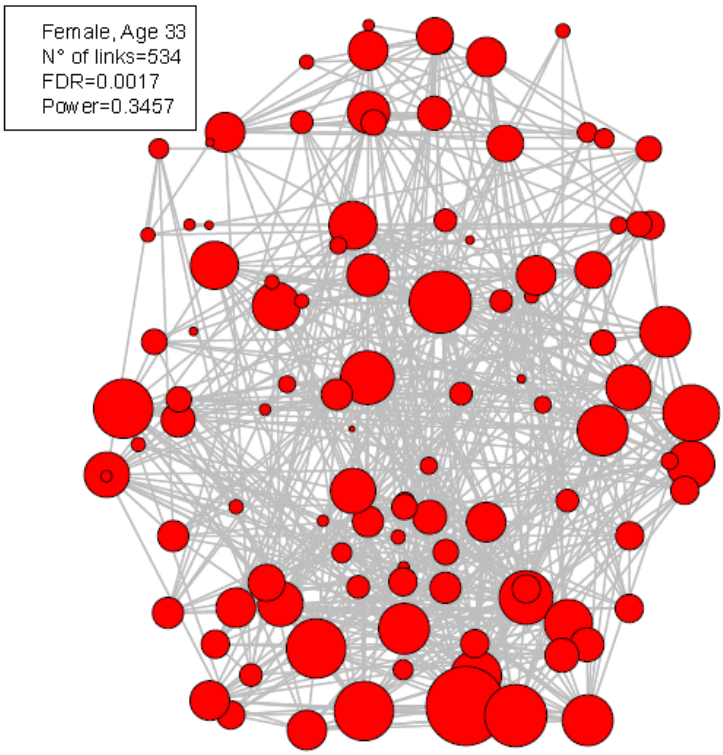
TABLE 3.4: $\widehat{FDR}(\gamma)$ and $\widehat{BP}(\gamma))$ for $\tau = 0.2, \ldots, 0.45$. In bold the τ that guarantees the maximum BP and that corresponds to the maximum value under the constraint $FDR \leq 0.01$.

τ	FDR	BP
0.250	0.0017124	0.3456607
0.275	0.0005063	0.2816666
0.300	0.0001195	0.2210831
0.325	0.0000262	0.1685835
0.350	0.0000047	0.1335729
0.375	0.0000007	0.1050432
0.400	0.0000001	0.0836455

As a result, the γ value that ensures the best balance between the Bayes FDR and the BP is $\gamma = 0.000178$ (or equivalently $\tau = 0.25$). In fact, the trade-off between the Bayes FDR and the BP implies that as the probability of false discoveries increases, the power also increases. The balanced γ is, hence, the maximum value of γ value such that the Bayes FDR is smaller than the maximum error considered to be acceptable. In our application, the maximum acceptable error was set equal to 0.01: if the maximum γ that guarantees $FDR \leq 0.01$ is chosen than the BP is maximum, under such a constraint. On the contrary, if, instead of choosing the maximum γ that guarantees $FDR \leq 0.01$, we select a smaller γ, still $FDR \leq 0.01$ but the BP would be smaller.

At last, the adjacency matrix can be visualized as a graph in which each node corresponds to one of the 116 ROIs, and each link corresponds to a significant connection between a pair of ROIs (i.e. to the 1 values reported in the i-adjacency matrix). The resulting brain network is reported in Figure 3.3.

FIGURE 3.3: Graphical representation in stereotactic coordinates of the brain network of the healthy participant (woman, 33 years old). Each red circle corresponds to a ROI: the wider the diameter of the circle, the higher the degree associated to the ROI (i.e. the number of links starting from that node); each gray line, otherwise, corresponds to a significant link between couple of ROIs. Given the 6670 possible links, only 534 were significant with a probability of false discoveries equal to 0.0017 and an overall power of 0.3457.



Finally, from the subject-by-subject balanced brain networks, i.e. the graph, we extracted a number of specific network measures that were used to explicitly test between-groups differences. This last step is extensively described in what follows.

3.4 Brain Network Analysis

Graph topology can be quantitatively described by a wide variety of measures; we selected four different known measures and a new ad-hoc created network index to investigate the differences between young and elderly participants. The first and simplest measure is represented by the number of links included in each subject-by-subject network; in the case of unweighted and undirected networks, the number of links corresponds to the number of '1' values that are displayed above the main diagonal of the adjacency matrix (Kolaczyk and Csárdi, 2014).
The second index is the mean network degree: a measure of density of the network. In general the degree k_i of a node i is equal to the number of connections that links the node it-self to the rest of the network (Rubinov and Sporns, 2010); the mean network degree is obtained by averaging the k_i calculated for the entire network according to:

$$\bar{k} = \frac{1}{n} \sum_{i=1}^{n} k_i$$

where n is the number of all the nodes of the network (in our specific case n=116).
The third measure considered is the number of hubs of a graph: an hub can be defined as a node whose specific degree is greater than the mean network degree (Wijk, Stam, and Daffertshofer, 2010). The number of hubs represents a measure of network efficiency; the higher the number of hubs, the more interconnected the nodes of the network (Heuvel and Sporns, 2013).
The fourth measure is the Newman's modularity Q (Newman, 2006) that quantifies the level of intra-regional connectivity and ranges between -1/2 and 1 (Brandes et al., 2008).

$$Q = \frac{1}{l} \sum_{i,j \in N} \left(a_{ij} - \frac{k_i k_j}{l} \right) \delta_{m_i, m_j}$$

where a_{ij} is the connection status between i and j, so that $a_{ij} = 1$ when link (i, j) exists and $a_{ij} = 0$ otherwise, l is the number of links, m_i is the module containing node i, δ_{m_i,m_j} is an indicator function that assumes value 1 if $m_i = m_j$ and 0 otherwise and finally N is the set of all nodes in the network. The prefixed regions, i.e. our modules, correspond to the following anatomical non-overlapping macro-areas: 1) left dorsolateral prefrontal cortex, 2) left ventromedial prefrontal cortex and diencephalon, 3) left parietal and temporal cortices, 4) left cerebellum 5) right dorsolateral prefrontal cortex, 6) right ventromedial prefrontal cortex, 7) right lateral parietal regions, 8), right temporal cortex 9) right cerebellum 10) middle-line structures, 11) retrosplenial region. The macro-areas were defined by adopting the modified hierarchical clustering algorithm described in Cattinelli et al. (2013) and by setting a spatial resolution of 17 mm: a spatial resolution that corresponded to the first percentile of the distribution of the Euclidean distances between the 116 ROIs (the matrix of the Euclidean distances is reported in Figure 3.4). A list of the AAL regions included in each macro-area is reported in Table 3.3 while the medium coordinates and standard deviations of the eleven macro-areas are reported in Table 3.5.

TABLE 3.5: Centroids and standard deviations of the eleven macro-areas (modules) as defined by the modified hierarchical algorithm described in the paper by Cattinelli et al. (2013).

ID	ClusterLabel	**X** (SD)	**Y** (SD)	**Z** (SD)
1	Left dorsolateral prefrontal cortex	**-36** (138.654)	**28** (100.457)	**28** (131.783)
2	Left ventromedial prefrontal cortex and diencephalon	**-23** (134.722)	**12** (202.811)	**-10** (112.086)
3	Left parietal and temporal cortices	**-47** (61.968)	**-28** (167.465)	**22** (246.957)
4	Left cerebellum	**-11** (146.944)	**-53** (122.432)	**-29** (129.849)
5	Right dorsolateral prefrontal cortex	**34** (11.225)	**14** (193.915)	**0** (21.074)
6	Right ventromedial prefrontal cortex	**4** (84.722)	**44** (7.874)	**3** (199.186)
7	Right lateral parietal regions	**43** (103.923)	**-38** (204.613)	**48** (101.719)
8	Right temporal cortex	**54** (47.223)	**-23** (120.955)	**2** (145.155)
9	Right cerebellum	**28** (81.394)	**-51** (140.633)	**-32** (125.842)
10	Middle-line structures	**-1** (105.763)	**-28** (216.777)	**43** (214.975)
11	Retrosplenial region	**3** (25.101)	**-78** (54.328)	**12** (152.283)

The last measure computed is the new ad-hoc created index (hence called

between-modules index; BM) that represents a measure of the between-regional connectivity. The BM is defined as the proportion of links that connects one specific module k to others modules on the total number of links that are not included in k. Thus, this index represents a measure of the neurofunctional localization, the higher the BM value, the less localized the neurofunctional connectivity of the macro-areas.

In particular, BM can be defined as

$$BM_k = \frac{\sum_{i \in I_k; j \notin I_k} a_{ij}}{n - \sum i, j \in I_k a_{ij}}$$

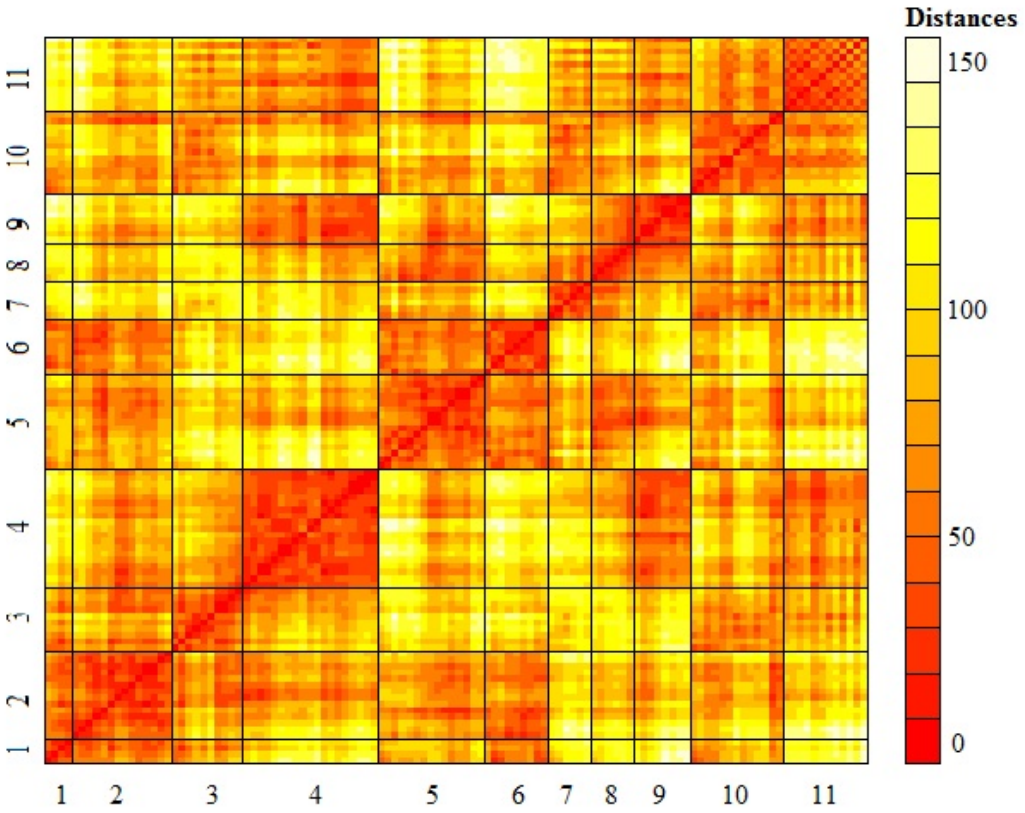
where n is the number of nodes and $I_k = \{i_1, \ldots, i_{nk}\}$ is the set of all nodes contained in the $k-th$ module, with n_k equal to the number of nodes belonging to the $k-th$ module, and $N = \cup_{k=1}^{K} I_k$ is the set of all nodes in the network, with K equal to the total number of modules.
Once computed, for each subject, the network measures described in the previous section, we tested the differences between elderly and young participants with respect to each measure.
Let consider, for example, the BM index: the aim is to test whether the mean measure for the elderlies, BM_e, is equal to the mean measure for the youngers, BM_y, versus a unilateral alternative hypothesis.

$$H_0 \quad : \quad BM_e = BM_y \quad against \quad H_1 : \quad BM_e > BM_y$$

It is known that the one-sided t-test is the uniformly most powerful unbiased test for this type of hypotheses if the group's measures originate from normal distributions. Such an assumption is not easily justifiable in this contest; for this reason we adopted a permutation test approach (Lehmann, 2006) that does not require any specific distributional assumption, and that is also adequate for small samples.

Finally, to specifically test the localization hypothesis, the length of the links included in our balanced network were classified into 4 categories according to the Euclidean Matrix (Figure 3.4) whose elements e_{ij} represent the Euclidean distance between ROI_i and ROI_j (defined by the stereotactic coordinates reported in Table 3.3). We then computed the deciles of the distribution in order to empirically determine the cut-points between the 4 categories: 1) short-range if $e_{ij} \leq e_{0.2}$, 2) as short.mid-range if $e_{0.2} < e_{ij} \leq e_{0.5}$, 3) mid.long-range if $e_{0.5} < e_{ij} \leq e_{0.8}$ and 4) long-range if $e_{ij} > e_{0.8}$.

FIGURE 3.4: The figure represents the matrix of the Euclidean distances between the 116 ROIs computed by the clustering algorithm in Cattinelli et al. (2013). On the basis of this matrix we identified 11 independent anatomical modules: 1) left dorsolateral prefrontal cortex, 2) left ventromedial prefrontal cortex and diencephalon, 3) left parietal and temporal cortices, 4) left cerebellum 5) right dorsolateral prefrontal cortex, 6) right ventromedial prefrontal cortex, 7) right lateral parietal regions, 8), right temporal cortex 9) right cerebellum 10) middle-line structures, 11) retrosplenial region.



3.5 Results

3.5.1 Young's networks

We have calculated the estimates of the Bayes FDR and the BP, and their corresponding 95% confidence intervals for each young subject for a range of τ values $T = \{0.15, 0.20, \ldots, 0.50\}$ and for a range of λ values $R = \{0.05, \ldots, 0.95\}$. The 95% CIs were constructed using $B = 500$ bootstrap replications. The results of this process of estimation are reported in Table 3.6.

TABLE 3.6: Point estimates and 95% confidence interval (CI) for the FDR and power. in the group of young subjects.

Sub. ID	τ	$\widehat{FDR}_\lambda(\gamma)$	CI 95% FDR	$\widehat{BP}_\lambda(\gamma)$	CI 95% BP
1	0.25	0.00213	[0;0.00233]	0.33130	[0.30461;0.35941]
2	0.25	0.00110	[0;0.00117]	0.52197	[0.49470;0.54891]
3	0.2	0.00999	[0;0.01048]	0.69285	[0.66897;0.71895]
4	0.25	0.00120	[0;0.00127]	0.47391	[0.44747;0.50581]
5	0.25	0.00142	[0;0.00153]	0.42853	[0.40293;0.45654]
6	0.25	0.00143	[0;0.00154]	0.42058	[0.39330;0.44925]
7	0.25	0.00215	[0;0.00233]	0.33134	[0.30369;0.35787]
8	0.25	0.00154	[0;0.00166]	0.42476	[0.38873;0.44410]
9	0.25	0.00168	[0;0.00181]	0.37910	[0.35311;0.40971]
10	0.25	0.00178	[0;0.00193]	0.40745	[0.37901;0.43612]
11	0.25	0.00118	[0;0.00126]	0.43865	[0.41339;0.46655]
12	0.25	0.00144	[0;0.00155]	0.41353	[0.38780;0.44186]
13	0.25	0.00174	[0;0.00188]	0.36346	[0.33683;0.39046]
14	0.25	0.00169	[0;0.00182]	0.36882	[0.34169;0.39566]
15	0.2	0.00871	[0;0.00914]	0.70069	[0.67565;0.72504]
16	0.25	0.00193	[0;0.00209]	0.39457	[0.36282;0.42338]
17	0.25	0.00151	[0;0.00163]	0.43095	[0.40488;0.45927]
18	0.25	0.00171	[0;0.00184]	0.34566	[0.31937;0.36937]
19	0.25	0.00121	[0;0.00129]	0.48324	[0.45693;0.50773]
20	0.2	0.00865	[0;0.00911]	0.72203	[0.69819;0.74617]
21	0.25	0.00220	[0;0.00241]	0.33860	[0.30714;0.36885]
22	0.25	0.00194	[0;0.00212]	0.34260	[0.31597;0.36951]
23	0.25	0.00214	[0;0.00234]	0.33359	[0.30689;0.36271]
24	0.25	0.00305	[0;0.00339]	0.25775	[0.23362;0.28537]
25	0.25	0.00126	[0;0.00135]	0.47695	[0.44939;0.50666]
26	0.25	0.00234	[0;0.00256]	0.32766	[0.29836;0.36070]
27	0.25	0.00153	[0;0.00165]	0.35983	[0.33472;0.38542]
28	0.25	0.00183	[0;0.00199]	0.36259	[0.33283;0.39052]
29	0.25	0.00223	[0;0.00242]	0.30757	[0.28183;0.33604]
30	0.25	0.00174	[0;0.00188]	0.36763	[0.34084;0.39702]
31	0.25	0.00321	[0;0.00354]	0.25417	[0.23198;0.28067]
32	0.25	0.00157	[0;0.00170]	0.41002	[0.38240;0.44147]
33	0.25	0.00171	[0;0.00185]	0.39178	[0.36467;0.42010]
34	0.25	0.00132	[0;0.00141]	0.45168	[0.42425;0.48017]
35	0.25	0.00166	[0;0.00179]	0.38069	[0.35481;0.40670]

TABLE 3.7: Point estimates and 95% confidence interval (CI) for the FDR and power, in the group of elderly subjects.

Sub. ID	τ	$\widehat{FDR}_\lambda(\gamma)$	CI 95% FDR	$\widehat{BP}_\lambda(\gamma)$	CI 95% BP
1	0,25	0,00148	[0;0,00160]	0,43717	[0,41026;0,46618]
2	0,25	0,00163	[0;0,00175]	0,38839	[0,35838;0,41643]
3	0,25	0,00153	[0;0,00165]	0,41161	[0,38324;0,43767]
4	0,25	0,00186	[0;0,00202]	0,34567	[0,31939;0,37211]
5	0,2	0,00964	[0;0,01015]	0,68993	[0,64838;0,70770]
6	0,25	0,00133	[0;0,00143]	0,45203	[0,42404;0,48047]
7	0,25	0,00226	[0;0,00244]	0,31504	[0,28808;0,34711]
8	0,2	0,00838	[0;0,00880]	0,72818	[0,70291;0,75082]
9	0,25	0,00136	[0;0,00146]	0,41295	[0,38669;0,43949]
10	0,25	0,00176	[0;0,00190]	0,39987	[0,38390;0,44090]
11	0,25	0,00161	[0;0,00173]	0,39388	[0,36725;0,42438]
12	0,25	0,00136	[0;0,00144]	0,43867	[0,40968;0,46917]
13	0,25	0,00171	[0;0,00184]	0,38870	[0,35949;0,41767]
14	0,25	0,00154	[0;0,00166]	0,42181	[0,39188;0,45163]
15	0,25	0,00141	[0;0,00151]	0,43798	[0,40978;0,46883]
16	0,25	0,00165	[0;0,00178]	0,38872	[0,36136;0,41561]
17	0,25	0,00147	[0;0,00157]	0,45710	[0,42947;0,48915]
18	0,25	0,00138	[0;0,00147]	0,48486	[0,45742;0,51010]
19	0,25	0,00188	[0;0,00203]	0,34459	[0,31875;0,37159]
20	0,25	0,00195	[0;0,00213]	0,37631	[0,34704;0,40750]
21	0,25	0,00183	[0;0,00198]	0,36259	[0,33473;0,39149]
22	0,25	0,00177	[0;0,00190]	0,39206	[0,36463;0,42149]
23	0,25	0,00182	[0;0,00196]	0,37288	[0,34501;0,40358]
24	0,25	0,00152	[0;0,00163]	0,40719	[0,38087;0,43544]
25	0,2	0,00950	[0;0,01002]	0,63935	[0,61252;0,66610]
26	0,25	0,00174	[0;0,00187]	0,33852	[0,31308;0,36107]
27	0,25	0,00156	[0;0,00168]	0,39214	[0,36596;0,41713]
28	0,2	0,00990	[0;0,01050]	0,73504	[0,70688;0,76067]
29	0,25	0,00244	[0;0,00269]	0,30106	[0,27173;0,33032]
30	0,25	0,00298	[0;0,00329]	0,26155	[0,23531;0,28792]
31	0,25	0,00178	[0;0,00192]	0,35394	[0,32603;0,37963]
32	0,2	0,00799	[0;0,00842]	0,76580	[0,74006;0,78972]
33	0,25	0,00127	[0;0,00134]	0,44199	[0,41645;0,46811]
34	0,25	0,00296	[0;0,00324]	0,27419	[0,24651;0,30680]
35	0,25	0,00123	[0;0,00131]	0,47506	[0,44688;0,50394]

The selected tuning parameter τ is similar for each subject: only in three cases τ is equal to 0.2, for the remaining subjects it is equal to 0.25. The mean FDR in this group is small equal to 0.0024 while the corresponding mean power is equal to 0.41.

3.5.2 Elderly's networks

The results of the point estimated of FDR and BP, and the corresponding the 95% bootstrap confidence intervals for the group of elderly subjects over a range T of values of τ and a range R of values of λ are reported in Table 3.7.

Even for the elderly subjects the tuning parameter τ is similar, equal to 0.2 for five subjects and 0.25 for the others. The mean FDR in this group is small equal to 0.0028 while the corresponding mean power is equal to 0.4351.

3.5.3 Between-groups Differences: Network Measures

In order to test the two alternative hypotheses, namely the de-differentiation hypothesis versus the localization hypothesis, we computed subject-by-subject balanced networks from which we extracted the subject-by-subject network measures described in the methods section. In Table 3.8, a summary of the between-groups differences is reported.

TABLE 3.8: Summary of the network measures. The significant between-groups differences are reported in bold.

	N. of links	Mean degree	Number of hubs
Elderly Mean	658,7714	11,3581	54,1429
Elderly SD	257,1495	4,4336	4,8881
Young Mean	617,3143	10,6433	53,9143
Young SD	222,6740	3,8392	4,3777
p-value	0,239	0,240	0,429
Bootstrap CI	(-68,72; 155)	(-1,18; 2,67)	(-1,89; 2,43)
H_1	Elderly>Young	Elderly>Young	Elderly>Young

	Modularity	BM index
Elderly Mean	0,0387	0,0712
Elderly SD	0,0308	0,0028
Young Mean	0,0535	0,0699
Young SD	0,0309	0,0028
p-value	**0.026**	**0,027**
Bootstrap CI	(-1,29e-04;-0,029)	(1,317e-06;0,0026)
H_1	Elderly<Young	Elderly>Young

TABLE 3.9: Relative contribution of each macro-area to the modularity. The significant between-groups differences are reported in bold.

Modul ID	Young	Elderly	p-value
1	0.01070	0.01607	0.059
2	0.05095	0.06763	0.145
3	0.01578	0.01951	0.302
4	0.07855	0.10062	0.201
5	0.02179	0.03568	**0.007**
6	0.05238	0.06585	0.287
7	0.03408	0.04311	0.126
8	0.17510	0.13640	0.085
9	0.10977	0.11237	0.864
10	0.40039	0.33487	**0.038**
11	0.04961	0.06789	0.176
Total	1	1	

TABLE 3.10: Relative contribution of each macro-area to the BM index. The significant between-groups differences are reported in bold.

Modul ID	Young	Elderly	p-value
1	0.07167	0.07612	0.355
2	0.16154	0.16549	0.571
3	0.03074	0.03011	0.825
4	0.17646	0.16777	0.173
5	0.05947	0.06110	0.510
6	0.12109	0.10686	**0.013**
7	0.04912	0.04560	0.480
8	0.03157	0.02887	0.344
9	0.10785	0.10603	0.756
10	0.10578	0.12029	0.153
11	0.08471	0.09086	0.224
Total	1	1	

Even though from the mere descriptive level, the number of links, the mean degree and the number of hubs may seem higher for the elderlies when compared with the youngers (providing the first hint towards the de-differentiation hypothesis), those differences were not statistically significant according to the permutation tests adopted here. On the contrary, significant between-groups differences emerged in the other two network measures: modularity and BM index. In particular, while the mean modularity was significantly lower, the mean BM index was significantly higher in elderlies than in young healthy participants.

In particular, the highest between-groups differences in the modularity index emerged from the right dorsolateral prefrontal cortex and the middle-line structures (Table 3.9); while the module that contributed the most to the between-group difference in the BM index was the right ventromedial prefrontal cortex (Table 3.10).

TABLE 3.11: Summary of the distance's distributions among groups. The significant differences are reported in bold.

	Short-range Proportion	Short.Mid-range Proportion
Elderly Mean	0.368	0.288
Young Mean	0.377	0.270
p-value	0.1861	**0.0027**
Bootstrap CI	(-0.029;0.011)	(0.006;0.030)
H1	Elderly>Young	Elderly>Young
	Mid.Long-range Proportion	Long-range Proportion
Elderly Mean	0.219	0.125
Young Mean	0.225	0.127
p-value	0.1866	0.3549
Bootstrap CI	(-0.021;0.008)	(-0.016;0.011)
H1	Elderly<Young	Elderly<Young

The de-differentiation and the localization hypotheses can also differently affect the length of functional connections between ROIs. In figure 3.1 we made clear the expected patterns in the two alternative hypotheses, for the different classes of links. The mean distribution of the links lengths between elderly and young participants was statistically different (p-value = 0.01889 of the two-sample Kolmogorov-Smirnov test). With respect to the proportion of links included in the 4 categories described previously, only the proportion of mid-short range connections was significantly different between the two groups (Table 3.11).

3.6 Discussion

We adopted here a combination of the most recent techniques and tools to analyse rsfMRI data, with innovative solutions to control for multiple testing problems and with graph theory. From the methodological point of view, the main contribution of this application is the description of a completely data-driven and robust method to create subject-specific balanced networks. Here, the attribute "balanced" refers to a network that has been obtained by controlling both for the type I error and the power.

When approaching the problem of describing a complex neural network, one has to consider that each participant, regardless of his/her age, is unique and, as a consequence, may present a specific pattern of functional connectivity. This inevitably implies that the subject's uniqueness has to be taken into account when we are interested in studying functional connectivity at the group level. In other words, we have to take into account within-group variability. While in massively univariate neuroimaging studies this problem has been classically approached by adopting random effects methods, when dealing with a functional connectivity problem, an alternative solution is (i) to obtain a balanced network for each single subject, (ii) to describe each balanced network by means of specific network measures that, in turn, can be used to make inferences at the group level. By doing this, the problem of subjects' uniqueness is solved at the first level of analysis, before any group-inferential process is done. In particular, to obtain subject-by-subject balanced network we developed a new methodological approach based on subject-by-subject threshold selection. As described previously in Section 1.2.1 to compute a graph that represents functional connectivity, a correlation matrix between the nodes of interest has to be computed, as a first step. Secondly, to "transform" a correlation matrix into a binary adjacency matrix we have to identify the significant correlation and this means that we have to conjointly test a great number of hypotheses (in our specific case: 6670). In the statistical literature, methods had been developed to address the problem of multiple testing, but in neuroimaging the praxis consists in typically focusing on methods that control for the type I error, while ignoring, in the most of the cases, the problem of statistical power. The empirical Bayes approach applied here first on neurofunctional data permits to bridge this gap. By controlling for the type I error rate we restricted the number of false discoveries and by computing the empirical power we provided information on the expected number of correct rejections: the

higher the power, the more reliable and sensible the test and as a consequence, the more reliable and sensible the resulting subject-specific neural network and the ensuing network measures. It is worthy to note that beyond the reliability and sensibility, our methodological approach to construct balanced network can also be considered robust from the statistical point of view. The robustness is guaranteed by the adoption of a non-parametric approach (Spearman's rank correlation coefficient) in the computation of the correlation matrix, i.e. by the adoption of a test that is free from the distribution assumptions typical of the parametric tests. Finally, our results, and as a consequence our conclusions, can be considered robust as we implemented, to test between-groups differences (and hence to test the two alternative theoretical frameworks described in the introduction), a distributional-free permutation test, rather than a classical one-sided t-test.

3.6.1 De-differentiation vs Localization: evidence from the Network Measures

Beyond the methodological innovation described Section 2.3, the aim of this applicative study was to explicitly test two alternative hypotheses about age-related changes in resting state functional connectivity: namely the de- differentiation hypothesis and the localization hypothesis (this latter hypothesis being formulated here for the first time in the context of healthy aging). In particular, the formulation of the localization hypothesis in healthy aging had been put forward on the basis of a review of the age-related changes across the entire life span. The two hypotheses were explicitly tested by means of a collection of network measures, according to the logic described in Table 3.1.
The first, and maybe more intuitive, index computed here was the number of links included in the balanced networks. This index was used to test the de-differentiation hypothesis on the basis of a simple consideration: if de-differentiation manifests it-self in terms of a reduction of the signal-to-noise ratio, then - once the statistical features of subject-by-subject networks had been balanced - one could easily expect to find a more dense network (i.e. a higher number of links) in the group of healthy elderly than in the group of young participants. In particular, according to the de-differentiation hypothesis the increment of the number of links in the elderlies would be due to the presence of a higher number of false positive (an inevitable consequence of the

reduction of the signal-to-noise ratio). Our empirical findings do not support this hypothesis as we found a relatively similar number, on average, of significant links in the young and in the elderly balanced network. A similar pattern of results emerged also for the mean degree centrality and for the number of hubs: empirical measures of network density and efficiency respectively. With this regard, it may be assumed that our negative findings depend by the choice of the statistical threshold applied to the subject-by-subject correlation matrices. As our method is based on the balancing between the FDR and the BP, and as de-differentiation may be conceptualized as an artefact due to a decrement of signal-to-noise ratio, one may argue that by balancing FDR and BP we are inevitably optimizing the signal-to-noise ratio at a subject-by-subject level. In other words, one may assume that the lack of empirical evidence in favor of the de-differentiation hypothesis may depend by the specific level of τ (i.e. by threshold level set to create the balanced adjacency matrix). However, this does not seem to be the case because the level of τ for the youngers and the elderlies was similar. A similar level of τ across subjects suggests that the groups were similarly balanced (see Table 3.6 and 3.7) and that, in turn, the level of signal-to-noise ratio in the raw data could be considered similar in these two groups.
The lack of between-group differences in the number of links, hubs and in the mean network degree may rather suggest that the healthy elderlies' resting-state networks globally maintain a "juvenile" level of efficiency. This is to say that, elderlies' networks may be de-differentiated not because of a general lack of network efficiency triggered by a reduction of the signal-to-noise ratio, but rather because the intrinsic organization of the resting state networks would change as age goes by. This assumption is supported by the results of the last two network measures: the modularity index and the BM index. Indeed, the healthy elderly participants showed, on average, a lower level of modularity and a higher number of between modules connections than young participants.

3.6.2 De-differentiation vs Localization: evidence from the Distance Measures

To further address the de-differentiation/localization dichotomy, we classified each link included in each subject-specific network according to 4 categories: 1) short-range connections, 2) short.mid-range connections, 3) mid.long-range

links and 4) long-range links. The results of this analysis are straightforward: young and elderly participants showed, on average, a different distribution of the links, in particular, the healthy elderlies' networks were characterized by a higher proportion of mid-short range connections. The increment of mid-short range connections, coupled with a stable proportion of long-range connection is, according to the scenarios depicted in Figure 3.1, a clear sign of de-differentiation. This result is in line with the finding from the network measures described in the previous section and with the results by Wang et al. (2010).

Conclusion

The idea of this Thesis arose during a doctoral lesson, when the multiple testing problem caught my attention; at a later stage, I started a collaboration with Neuro-MI on the analysis of neurofunctional data with a multiple testing approach. The Thesis, hence, has both a theoretical and practical aim and, as a consequence, has led to both theoretical and practical results.
We have attempted to describe the multiple testing problem and we have proposed to control for the Bayes FDR and BP, in a Bayesian perspective. We demonstrated, both with theoretical considerations and via simulation studies, that our approach is robust with respect to violation of the hypothesis of independence among p-values. Indeed, we demonstrated that the not empirical Bayes estimate of FDR is conservative (Section 2.3.1) while the not empirical Bayes estimate of BP is unbiased (Section 2.3.2). With respect to the empirical Bayes estimates of FDR and BP, moreover, we studied their asymptotic properties. In case of independent p-values, we demonstrated the almost sure convergence of those estimates (Section 2.3.3). In case of dependent p-values, on the contrary, we demonstrated the same almost sure convergence by assuming stationary ergodic p-values (Section 2.3.3). By means of some simulation studies, we enriched our analysis on the proposed estimators, by exploring some patterns of dependency, well-established in the statistical literature (Section 2.3.4). We provided evidence to support the robustness of our estimates and we also showed that the standard bootstrap approach, although not formally adequate, can be still adopted in case of moderate dependence.

Furthermore, through a case study, we proved how easily applicable our Bayesian approach is. In the neurofunctional literature there was not a systematic way to choose the threshold τ. A simple approach is to select the threshold only by controlling the probability of first kind (Achard et al., 2006; Cauda et al., 2011): clearly this approach does not take into account the power of the test and can yield to misleading results. Bullmore and Bassett (2011) proposed a completely empirical approach based on the exploration of network properties as a function of changing threshold. This method, however, provides only a heuristic in the threshold selection without quantifying for the false

positives and false negatives (Seo et al., 2013). Our approach satisfies both the need for a systematic and reproducible approach of threshold selection, and the need for a joint control of FDR and power of the tests. In the application on real data we focused our attention on the analysis of 70 participants and, hence, we constructed subject by subject brain networks. As a second step, we analyzed those networks by means of a set of network measures in order to assess if there exist some differences between elderlies and young subjects, and whether these differences are in accordance to the de-differentiation hypothesis or to the localization hypothesis.

Our Bayesian approach has the double advantage of enjoying good properties and of being easily applicable not only in neuroscience, but also in many different application field, such as economics, social science or computer science to mention a few.

Bibliography

Achard, S. et al. (2006). "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs". In: *J Neurosci* 26.1, pp. 63–72.

Alexanderson, G. (2006). "About the cover: Euler and Königsberg's Bridges: A historical view". In: *Bulletin of the american mathematical society* 43.4, pp. 567–573.

Ashburner, J. and K. J. Friston (2005). "Unified segmentation". In: *Neuroimage* 26.3, pp. 839–51.

Barulli, D. and Y. Stern (2013). "Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve". In: *Trends Cogn Sci* 17.10, pp. 502–9.

Benjamini, Y. and H. Braun (2002). "John W. Tukey's contributions to multiple comparisons". In: *Annals of Statistics*, pp. 1576–1594.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300.

Berlingeri M.; Di Brisco, A. M.; Zapparoli L.; Sberna M.; Gallucci M.; Banfi G.; Quatto P. (2015). "De-differentiation in healthy elderlies: evidence from resting-state fMRI and a new empirical-bayesian approach". *Submitted*.

Billingsley, P. (1986). *Measure and Probability*. John Willey.

Biswal, B. et al. (1995). "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI". In: *Magn Reson Med* 34.4, pp. 537–41.

Brandes, U. et al. (2008). "On Modularity Clustering". In: *Knowledge and Data Engineering, IEEE Transactions on* 20.2, pp. 172–188.

Bullmore, E. T. and D. S. Bassett (2011). "Brain graphs: graphical models of the human brain connectome". In: *Annu Rev Clin Psychol* 7, pp. 113–40.

Casella, G. and R. L. Berger (2002). *Statistical inference*. Vol. 2. Duxbury Pacific Grove, CA.

Cattinelli, I. et al. (2013). "A novel approach to the problem of non-uniqueness of the solution in hierarchical clustering". In: *IEEE Trans Neural Netw Learn Syst* 24.7, pp. 1166–73.

Cauda, F. et al. (2011). “Functional connectivity of the insula in the resting brain”. In: *Neuroimage* 55.1, pp. 8–23.

Chao-Gan, Y. and Z. Yu-Feng (2010). “DPARSF: A MATLAB Toolbox for "Pipeline" Data Analysis of Resting-State fMRI”. In: *Front Syst Neurosci* 4, p. 13.

Christensen, K. et al. (2009). “Ageing populations: the challenges ahead”. In: *Lancet* 374.9696, pp. 1196–208.

Collins, D. L. et al. (1994). “Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space”. In: *J Comput Assist Tomogr* 18.2, pp. 192–205.

Craiu, R. V. and L. Sun (2008). “Choosing the lesser evil: trade-off between false discovery rate and non-discovery rate”. In: *Statistica Sinica* 18, pp. 861–879.

Dehling, H. and W. Philipp (2002). *Empirical process techniques for dependent data*. Springer.

Dennis, N. A. and R. Cabeza (2008). “Neuroimaging of healthy cognitive aging”. In: *The handbook of aging and cognition* 3, pp. 1–54.

Drton, M. (2009). “Discrete chain graph models”. In: *Bernoulli*, pp. 736–753.

Dudoit, S., J. P. Shaffer, and J. C. Boldrick (2003). “Multiple hypothesis testing in microarray experiments”. In: *Statistical Science*, pp. 71–103.

Efron, B. (2010). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol. 1. Cambridge University Press.

Fair, D. A. et al. (2009). “Functional brain networks develop from a "local to distributed" organization”. In: *PLoS Comput Biol* 5.5, e1000381.

Fox, M. D. et al. (2005). “The human brain is intrinsically organized into dynamic, anticorrelated functional networks”. In: *Proc Natl Acad Sci U S A* 102.27, pp. 9673–8.

Fransson, P. (2005). “Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis”. In: *Hum Brain Mapp* 26.1, pp. 15–29.

Freeman, L. C. (1979). “Centrality in social networks conceptual clarification”. In: *Social networks* 1.3, pp. 215–239.

Friston, K. J. (1994). “Functional and effective connectivity in neuroimaging: A synthesis”. In: *Human Brain Mapping* 2.1-2, pp. 56–78.

Friston, K. J. et al. (1993). “Functional connectivity: the principal-component analysis of large (PET) data sets”. In: *J Cereb Blood Flow Metab* 13.1, pp. 5–14.

Genovese, C. and L. Wasserman (2002). "Operating characteristics and extensions of the false discovery rate procedure". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.3, pp. 499–517.

Gibbons, J. D. and S. Chakraborti (2011). *Nonparametric statistical inference*. Springer.

Gold, B. T. et al. (2010). "Age-related slowing of task switching is associated with decreased integrity of frontoparietal white matter". In: *Neurobiol Aging* 31.3, pp. 512–22.

Greicius, M. D. et al. (2003). "Functional connectivity in the resting brain: a network analysis of the default mode hypothesis". In: *Proc Natl Acad Sci U S A* 100.1, pp. 253–8.

Gribkovskaia, I., Ø. Halskau, and G. Laporte (2007). "The bridges of Königsberg—a historical perspective". In: *Networks* 49.3, pp. 199–203.

Heuvel, M. P. van den and O. Sporns (2013). "Network hubs in the human brain". In: *Trends Cogn Sci* 17.12, pp. 683–96.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure". In: *Scandinavian journal of statistics*, pp. 65–70.

Karr, A. F. (1993). *Probability*. Springer Texts in Statistics. Springer New York.

Kelly, A. M. et al. (2008). "Competition between functional brain networks mediates behavioral variability". In: *Neuroimage* 39.1, pp. 527–37.

Kolaczyk, E. D. and G. Csárdi (2014). *Statistical analysis of network data with R*. Vol. 65. Springer.

Lahiri, S. N. (2013). *Resampling methods for dependent data*. Springer Science and Business Media.

Lehmann, E. L. (2006). *Testing statistical hypotheses*. 3rd ed. Springer Texts in Statistics.

Lehmann, E. L. and J. P. Romano (2012). *Generalizations of the familywise error rate*. Springer.

Lindquist, M. A. (2008). "The statistical analysis of fMRI data". In: *Statistical Science* 23.4, pp. 439–464.

Liu, Y. et al. (2007). "Whole brain functional connectivity in the early blind". In: *Brain* 130.Pt 8, pp. 2085–96.

Lu, H. et al. (2007). "Synchronized delta oscillations correlate with the resting-state functional MRI signal". In: *Proc Natl Acad Sci U S A* 104.46, pp. 18265–9.

Moré, J. J. (1978). "The Levenberg-Marquardt algorithm: implementation and theory". In: *Numerical analysis*. Springer, pp. 105–116.

Myers, R. H. et al. (2012). *Generalized linear models: with applications in engineering and the sciences*. Vol. 791. John Wiley and Sons.

Newman, M. E. J. (2006). "Modularity and community structure in networks". In: *Proceedings of the National Academy of Sciences* 103.23, pp. 8577–8582.

Nichols, T. and S. Hayasaka (2003). "Controlling the familywise error rate in functional neuroimaging: a comparative review". In: *Statistical methods in medical research* 12.5, pp. 419–446.

Rokach, L. and O. Maimon (2005). "Clustering methods". In: *Data mining and knowledge discovery handbook*. Springer, pp. 321–352.

Rubinov, M. and O. Sporns (2010). "Complex network measures of brain connectivity: uses and interpretations". In: *Neuroimage* 52.3, pp. 1059–69.

Sala S.; Quatto, P.; Valsasina P.; Agosta F.; Filippi M. (2014). "pFDR and pFNR estimation for brain networks construction". In: *Stat Med* 33.1, pp. 158–69.

Seo, E. H. et al. (2013). "Whole-brain functional networks in cognitively normal, mild cognitive impairment, and Alzheimer's disease". In: *PLoS One* 8.1, e53922.

Shao, J. and D. Tu (1995). *The jackknife and bootstrap*. Springer-Verlag, New York.

Shiryaev, A. N. and A. Lyasoff (2012). *Problems in probability*. Springer Science and Business Media.

Song, X. W. et al. (2011). "REST: a toolkit for resting-state functional magnetic resonance imaging data processing". In: *PLoS One* 6.9, e25031.

Sporns, O. (2002). "Network analysis, complexity, and brain function". In: *Complexity* 8.1, pp. 56–60.

Storey, J. D. (2002). "A direct approach to false discovery rates". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.3, pp. 479–498.

— (2003). "The positive false discovery rate: a Bayesian interpretation and the q-value". In: *Annals of statistics*, pp. 2013–2035.

Storey, J. D. and R. Tibshirani (2001). "Estimating false discovery rates under dependence, with applications to DNA microarrays". In: *Technical of Report 2001* 28.

— (2003). "Statistical significance for genomewide studies". In: *Proc Natl Acad Sci U S A* 100.16, pp. 9440–5.

Tempel, L. W. and J. S. Perlmutter (1992). "Vibration-induced regional cerebral blood flow responses in normal aging". In: *J Cereb Blood Flow Metab* 12.4, pp. 554–61.

Tzourio-Mazoyer, N. et al. (2002). "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain". In: *Neuroimage* 15.1, pp. 273–89.

Wang, L. et al. (2010). "Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition". In: *Neuroimage* 50.3, pp. 862–72.

Wijk, B. C. van, C. J. Stam, and A. Daffertshofer (2010). "Comparing brain networks of different size and connectivity density using graph theory". In: *PLoS One* 5.10, e13701.

Zimmerman, M. E. et al. (2006). "The relationship between frontal gray matter volume and cognition varies across the healthy adult lifespan". In: *Am J Geriatr Psychiatry* 14.10, pp. 823–33.